

**De novo development of novel DM1 toxic ncRNA targeting small molecules
and its biological evaluation**

Raul Ondoño Molina

<http://hdl.handle.net/10803/687314>

Data de defensa: 13-12-2022

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. The access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

DOCTORAL THESIS

Title	De novo development of novel DM1 toxic ncRNA targeting small molecules and its biological evaluation
Presented by	Raul Ondoño Molina
Centre	IQS School of Engineering
Department	Química Orgànica i Farmacèutica
Directed by	Dr. Roger Estrada Prof. Dr. José I. Borrell

Pels pares, mon germà i la Iris

ACKNOWLEDGMENTS

En primer lloc m'agradaria agrair la possibilitat de haver dut a terme aquesta tesi a l'IQS que amb el suport econòmic de la beca de doctorat ha fet possible que em pogués dedicar plenament a la investigació durant 3 anys.

D'altra banda i també voldria mencionar als meus directors, el Roger i l'Iñaki, que han estat a les dures i a les madures, donant-me suport i consell sempre que l'he necessitat (i quan he necessitat una estirada d'orelles també).

Estaré eternament agraït a l'Alex Perálvarez i l'Elena, ells ens van treure del pou quan ja semblava que tot estava perdut i ens van oferir una casa on treballar i aprendre bioquímica en un entorn de cooperació fantàstic a la unitat de biofísica de la UAB.

Vorrei ringraziare di cuore Germana, Claudia e Beatrice. Ero un estraneo e loro mi hanno trattato come uno di famiglia, insegnandomi e aiutandomi con qualcosa di così lontano dalla mia formazione come la biologia cellulare. Grazie mille.

A part d'ells També m'agradaria agrair l'oportunitat i la paciència rebudes tant de els companys com els de els professors que m'han suportat tots aquests anys: Especialment als membres del GQF, amb qui hem compartit molts anys de feina i esbarjo. Especial record a aquells que ara tenen un futur brillant a IQS, com l'Ana Belén, el Raimon, l'Albert i també la Maia i la Gemma.

Un record per tots aquells doctorands de la meva generació amb qui vem aconseguir trencar moltes barreres i vem acabar sent família. L'Eli, el Joaquim, el Jaume, l'Alba, la Laura, la Mireia, el Mario i tots els membres dels grups de GQF, FotoQs, GEMAT i Bioch, recordarem sempre totes les aventures a la Fira.

A qui més agraït estic i estaré sempre es a la meva família, ells han aguantat tot de mí, que és molt i no podria haver tingut una de millor. Ells ho han fet tot possible, tesi inclosa, si soc algú és per vosaltres.

Finalment, dedico aquest treball a la Iris, la persona que més entén que em passa pel cap. Sempre agrairé que la tesi ens hagi fet trobar-nos, ara ja està i la següent ets tu. Compta amb mi pel que necessitis, sempre hi seré.

A tots una abraçada,

Raul

SUMMARY

Myotonic Dystrophy type 1 (DM1) is an incurable neuromuscular disorder caused by toxic DMPK transcripts that carry CUG repeat expansions in the 3' untranslated regions (3'UTR). The intrinsic complexity and lack of crystallographic data make noncoding RNA regions challenging targets to study in the field of drug discovery. In DM1, toxic transcripts tend to stall in the nuclei forming complex inclusion bodies called foci and sequestering many essential alternative splicing factors such as Muscleblind-like 1 (MBNL1). Most DM1 phenotypic features stem from the reduced availability of free MBNL1, and therefore many therapeutic efforts are focused on recovering its regular activity. For that purpose, in the present thesis, we decide to target CUG RNA to free MBNL1. The in-silico screening using structure-based drug design techniques of novel candidates based on two different approaches is described. Furthermore, synthetic pathways are developed for the selected candidates based on the click chemistry approach. Finally, to assess their biological activity, an already described biochemical test is tuned, and cellular models and patient-derived muscular cells are used to evaluate the most promising candidates. The obtained results may lead to subsequent generations of ligands, highlighting a new affordable treatment against DM1.

SUMARIO

La distrofia miotónica de tipo 1 (DM1) es un trastorno neuromuscular incurable causado por las transcripciones tóxicas del gen DMPK. Estos transcritos llevan expansiones de repeticiones CUG en las regiones no traducidas 3' (3'UTR). La complejidad intrínseca y la falta de datos cristalográficos hacen que las regiones de ARN no codificante sean objetivos difíciles de estudiar en el campo del desarrollo de nuevos fármacos. En la DM1, los transcritos tóxicos tienden a estancarse en el interior de los núcleos formando complejos cuerpos de inclusión llamados foci y secuestrando muchos factores de *splicing* alternativo esenciales como el Muscleblind-like 1 (MBNL1). La mayoría de las características fenotípicas de la DM1 se derivan de la reducida disponibilidad de MBNL1 libre, por lo que muchos esfuerzos terapéuticos se centran en recuperar su actividad regular. Para ello, en la presente tesis, decidimos utilizar como diana terapéutica el ARN CUG, con el fin de liberar MBNL1. Por lo que respecta al diseño de nuevas estructuras, se describe el cribado *in-silico* mediante técnicas de diseño de fármacos basadas en estructura usando dos premisas diferentes de abordar CUG. Además, se desarrollan vías sintéticas para los candidatos seleccionados basadas en química click. Por último, para evaluar su actividad biológica, se pone a punto un ensayo bioquímico ya descrito, y se utilizan modelos celulares y células musculares derivadas de pacientes para evaluar los candidatos más prometedores. Los resultados obtenidos pueden conducir a posteriores generaciones de ligandos, poniendo de manifiesto un nuevo tratamiento asequible contra la DM1.

SUMARI

La distròfia miotònica de tipus 1 (DM1) és un trastorn neuromuscular incurable causat per les transcripcions tòxiques del gen DMPK. Aquests transcrits porten expansions de repeticions CUG a les regions no traduïdes 3' (3'UTR). La complexitat intrínseca i la falta de dades cristal·logràfiques fan que les regions d'ARN no codificant siguin objectius difícils d'estudiar en el camp del desenvolupament de nous fàrmacs. En el cas de la DM1, els transcrits tòxics tendeixen a estancar-se a l'interior dels nuclis formant complexos cossos d'inclusió anomenats *foci* i segrestant molts factors de *splicing* alternatiu essencials com el Muscleblind-like 1 (MBNL1). La majoria de les característiques fenotípiques de la DM1 es deriven de la reduïda disponibilitat de MBNL1 lliure, per la qual cosa molts esforços terapèutics es centren en recuperar la seva activitat regular. Per a això, en la present tesi, decidim utilitzar com a diana terapèutica l'ARN CUG, amb la finalitat d'alliberar MBNL1. Pel que respecta al disseny de noves estructures, es descriu el cribratge *in-silico* mitjançant tècniques de disseny de fàrmacs basades en estructura usant dues premisses diferents d'abordar CUG. A més, es desenvolupen vies sintètiques per als candidats seleccionats basades en química clic. Finalment, per a avaluar la seva activitat biològica, es posa a punt un assaig bioquímic ja descrit, i s'utilitzen models cel·lulars i cèl·lules musculars derivades de pacients per a avaluar els candidats més prometedors. Els resultats obtinguts poden conduir a posteriors generacions de lligands, posant de manifest un nou tractament assequible contra la DM1.

ABBREVIATIONS

AEBSF	4-(2-Aminoethyl)benzene-1-sulfonyl fluoride
AID	PubChem Article ID
Amp^R	Ampicillin resistant
APA	Alternative Poly-Adenylation
API	Active Pharmaceutical Ingredient
ASO	Antisense oligonucleotides
ATP	Adenosine triphosphate
bp	base pair
Cas	CRISPR associated protein
CD	Cyclic Dichroism
CDI	Carbonyl diimidazole
CDK4	Cyclin-dependent kinase 4
CELF1	CUGBP Elav-Like Family Member 1
clc	Chloride channel
CNBP	Cellular nucleic acid-binding protein
COSY	Correlation spectroscopy
CRISPR	clustered regularly interspaced palindromic repeats
CXCR4	C-X-C chemokine receptor type 4
cy	Cyclohexane
DABCO	1,4-diazabicyclo[2.2.2]octane
DCC	N,N'-Dicyclohexylcarbodiimide
DCM	Dichloromethane
DCU	N,N'-Dicyclohexylurea
DEPC	Diethyl pyrocarbonate
DM1/2	Myotonic Dystrophy Type 1/2
DM200	DM1 mouse model
DM5	DM1 mouse model
DMAP	4-Dimethylaminopyridine
DMF	<i>N,N</i> -dimethylformamide
DMPK	Dystrophia myotonica protein kinase
DMSO	Dimethyl sulfoxide
DMSXL	Mice bearing over a thousand CTG repetitions
DMWD	Dystrophia myotonica WD repeat-containing protein gene
DNA	Deoxyribonucleic acid
dppf	diphenylphosphinoferrocene
dsDNA	Double-stranded DNA
EDC-HCl	1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide hydrochloride

EDTA	Ethylenediaminetetraacetic acid
eIF2α	Eukaryotic translation initiation factor 2 alpha
ER	Estrogen Receptor
ESI-TOF	Electrospray ionization - Time of Flight
FAM	Fluorescein amidite
FDA	The United States Food and Drug Administration (FDA or USFDA)
FISH	Fluorescence in-situ Hybridization
FRAX	X-Fragile syndrome
FRET	Förster Resonance Energy Transfer
GBVI/WSA	Generalized Born/Volume Integral Implicit Solvent Model
GSK3β	Glycogen synthase kinase-3 beta
GST	Glutathione S-transferase
HD	Huntington Disease
HDR	Homology-directed repair
HGNC	HUGO Gene Nomenclature Committee
HIV	Human Immunodeficiency Viruses
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HRMS	High Resolution Mass Spectrometry
HSA^{LR}	Mice expressing CUG Long Repetitions
HSQC	Heteronuclear single quantum coherence spectroscopy
hTERT	Human Telomerase Reverse Transcriptase
HTRF	Homogeneous Time Resolved Fluorescence
HTS	High Throughput Screening
INSR	Insulin receptor protein coding gene
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IR	Infrared Spectroscopy
ITC	Isothermal titration calorimetry
IUPAC	International Union of Pure and Applied Chemistry
J	Coupling constant
JW	Janus-Wedge model
KO	Knock out
LB	Lysogeny broth
LBDD	Ligand-Based Drug Design
lncRNA	Long non-coding RNA
mAb	monoclonal Antibody
MBNL	Muscleblind Like Splicing Regulator
MD	Molecular Dynamics
miRNA	micro-RNA
MMPBSA	Molecular Mechanics Poisson-Boltzmann Surface Area

MOE	Molecular Operative Environment
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide based assay
MW	Micro Waves
MYOD1	Myoblast determination protein 1
Myosin HC	Myosin Heavy Chain
NBS	<i>N</i> -bromosuccinimide
ncRNA	Non-coding RNA
NCS	<i>N</i> -chlorosuccinimide
NIS	<i>N</i> -iodosuccinimide
NMR	Nuclear Magnetic Resonance Spectroscopy
NPT	Virtual system with constant Number of atoms, Pressure and Temperature
OD₆₀₀	Optical density at 600 nm
OEA	Organic Elemental Analysis
PAGE	Polyacrylamide gel electrophoresis
PCA	Principal Component Analysis
PDB	Protein Data Bank
PDB	Protein Data Bank
PDM	Proximal Myotonic Dystrophy
PMSF	Phenylmethanesulfonyl fluoride
ppm	part per million
PROMM	Proximal Myotonic Myopathy
RAN	Repeat Associated Non-AUG translation
RBP	RNA-Binding Protein
RMSD	Root-Mean-Square Deviation
RNAi	Interfering RNA
rt	room temperature
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SBDD	Structure-Based Drug Design
SBMA	Spinal-Bulbar Muscular Atrophy
SCA	Spinocerebellar Ataxia
SDF	Structure Data File
SDS	Sodium Dodecyl Sulphate
SERCA1	Sarcoplasmic/endoplasmic Reticulum Calcium ATPase 1
sgRNA	Small guide RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SIX5	SIX Homeobox 5 gene
SMILES	Simplified Molecular-Input Line-Entry System
S_N2	Nucleophilic Substitution type 2

SPPS	Solid-Phase Peptide Synthesis
SSM	Slipped Strand Mismatching
TEA	Triethylamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
TR	Time-resolved
TRED	Trinucleotide Repeat Expansion Disease
Tris	Tris(hydroxymethyl)aminomethane
UTR	Un-Transcribed Region
UV	Ultraviolet
Y3H	Yeast 3 Hybrid
ZnF	Zinc Finger
ZNF9	Zinc Finger protein 9 gene

INDEX

1	INTRODUCTION	1
1.1	ORPHAN DISEASES	3
1.1.1	<i>Genetic disorders</i>	3
1.1.1.1	<i>Chromosomal disorders</i>	3
1.1.1.2	<i>Complex disorders</i>	3
1.1.1.3	<i>Monogenic disorders</i>	3
1.1.1.3.1	<i>TRED's</i>	4
1.2	MYOTONIC DYSTROPHIES	5
1.2.1	<i>Myotonic dystrophy type 1 (DM1)</i>	7
1.2.1.1	<i>Symptoms, prognosis & prevalence</i>	7
1.2.1.2	<i>Molecular basis</i>	9
1.2.1.3	<i>The direct implication of genetic alteration and cis effect</i>	10
1.2.1.4	<i>Alternative polyadenylation (APA)</i>	11
1.2.1.5	<i>Nuclear foci</i>	11
1.2.1.6	<i>RAN translation</i>	12
1.2.1.7	<i>Protein implications</i>	12
1.2.2	<i>Muscleblind-like proteins</i>	13
1.2.3	<i>CELF1 Overexpression</i>	15
1.2.4	<i>miRNA dysregulation</i>	15
1.2.5	<i>Splicing pattern imbalance</i>	16
1.3	THERAPEUTIC STATE OF THE ART	17
1.3.1	<i>DNA targeting</i>	17
1.3.1.1	<i>CRISPR genomic engineering</i>	17
1.3.2	<i>Protein targeting</i>	18
1.3.3	<i>RNA targeting</i>	18
1.3.3.1	<i>Oligonucleotides</i>	19
1.3.3.2	<i>Oligopeptides</i>	19
1.3.3.3	<i>Small molecular entities</i>	20
1.3.4	<i>The study of Myotonic Dystrophy at IQS School of Engineering</i>	21
1.4	REFERENCES	23

2	ADAPTING AID 2675	33
2.1	INTRODUCTION	35
2.1.1	<i>Need for a test</i>	35
2.1.2	<i>Biological evaluation assays</i>	35
2.1.2.1	<i>Animal models</i>	35
2.1.2.2	<i>Cell models</i>	36
2.1.2.3	<i>In vitro screening</i>	36
2.1.3	<i>AID 2675 Bioassay</i>	36
2.1.3.1	<i>Time-resolved fluorescence energy transfer (HTRF)</i>	38
2.2	OBJECTIVES	39
2.3	RESULTS	40
2.3.1	<i>Plasmid design</i>	40
2.3.2	<i>Bacteriological culture studies</i>	42
2.3.3	<i>BL21(DE3)pLysS transformation</i>	42
2.3.3.1	<i>Bacterial strains transformation</i>	43
2.3.3.2	<i>Verifying transformation</i>	44
2.3.3.3	<i>Induction studies</i>	46
2.3.4	<i>Recombinant protein production</i>	48
2.3.4.1	<i>SDS-PAGE check</i>	49
2.3.4.2	<i>MBNL1 quantification</i>	50
2.3.5	<i>Troubleshooting</i>	51
2.3.5.1	<i>Ultracentrifugation.</i>	51
2.3.5.2	<i>Cobalt affinity resin</i>	52
2.3.5.3	<i>Avoiding bacteriophages</i>	53
2.3.5.4	<i>RNase</i>	53
2.3.6	<i>Tuning of test conditions</i>	54
2.4	CHAPTER DISCUSSION	55
2.4.1	<i>Final standard procedure</i>	56
2.5	REFERENCES	58

3	GROOVE BINDERS	61
3.1	INTRODUCTION	63
3.1.1	<i>Overview</i>	63
3.1.2	<i>The groove bendir approach</i>	64
3.1.2.1	<i>The grooves</i>	64
3.1.3	<i>Candidate selection</i>	65
3.1.3.1	<i>Hoechst</i>	66
3.1.3.2	<i>Pentamidine</i>	67
3.1.3.3	<i>This work</i>	68
3.2	OBJECTIVES	69
3.3	RATIONAL DESIGN OF POTENTIAL GROOVE BINDERS	70
3.3.1	<i>Assembly of the chemical library</i>	70
3.3.1.1	<i>Scaffold definition</i>	71
3.3.1.2	<i>Linker selection</i>	72
3.3.1.3	<i>Binding groups selection</i>	73
3.3.2	<i>Combinatorial library enumeration</i>	74
3.3.3	<i>Molecular conformation calculation</i>	74
3.3.4	<i>Molecular descriptor calculation</i>	75
3.3.5	<i>Diversity selection</i>	76
3.3.6	<i>Prediction of the binding mechanism of selected compounds</i>	76
3.4	SYNTHESIS	82
3.4.1	<i>Structures 3.6: methylenebisphenyl spacer</i>	83
3.4.1.1	<i>Retrosynthetic analysis and synthesis</i>	84
3.4.1.1.1	<i>Synthesis of 3.15: acid chloride formation</i>	85
3.4.1.1.2	<i>Synthesis of 3.6{4.1}: Schotten-Baumann double amidation</i>	86
3.4.2	<i>Structures 3.13: oxybisphenyl spacer</i>	87
3.4.2.1	<i>Retrosynthetic analysis</i>	88
3.4.2.2	<i>Synthesis of binding moieties</i>	89
3.4.2.2.1	<i>Synthesis of 3.18</i>	90
3.4.2.2.2	<i>Synthesis of 3.20</i>	91
3.4.2.3	<i>Amidation couplings</i>	92
3.4.2.3.1	<i>Acylic substitution</i>	92
3.4.2.3.2	<i>Schotten-Baumann Couplings</i>	94
3.4.2.4	<i>Post coupling transformations</i>	95

3.4.2.4.1 4,4'-oxybis(N-(3-guanidinophenyl)benzamide) (3.13{1.1}).	95
3.4.3 Structures 3.14: urea bisphenyl spacer	96
3.4.3.1 Retrosynthetic análisis	96
3.4.3.2 Spacer synthesis	98
3.4.3.2.1 Synthesis of 3.27	99
3.4.3.2.2 Synthesis of 3.25	100
3.4.3.2.3 Synthesis of 3.27 (II)	100
3.4.3.2.4 Synthesis of 3.28	101
3.4.3.2.5 Synthesis of 3.14	101
3.4.3.3 Obtaining the binding groups	101
3.4.3.3.1 Synthesis of 3.28	102
3.4.3.4 Couplings	102
3.4.3.4.1 Synthesis of 3.14{7.1}	103
3.4.3.4.2 Synthesis of 3.14{2.1}	104
3.4.3.4.3 Synthesis of 3.32	104
3.4.3.5 Post coupling transformations	105
3.5 BIOLOGICAL EVALUATION	106
3.5.1 AID 2675	106
3.5.2 Polarization fluorescence assays	107
3.5.3 Cell biology evaluation	108
3.6 FURTHER STUDIES	111
3.7 CHAPTER DISCUSSION AND PERSPECTIVES	115
3.8 REFERENCES	116

4	PYRIDO[2,3-D]PYRIMIDINE-BASED CUG LIGANDS	121
4.1	INTRODUCTION	123
4.1.1	<i>Scaffold definition</i>	123
4.1.2	<i>Novel pyrido[2,3-d]pyrimidin-7(8H)-one structures</i>	125
4.2	OBJECTIVES	127
4.3	RESULTS	128
4.3.1	<i>Synthesis of DM1 CUG RNA binders using 2 and 4-halopyrido[2,3-d]pyrimidines</i>	128
4.3.1.1	<i>Previous work</i>	128
4.3.1.2	<i>Synthesis of DM1 CUG RNA binders using 2-bromo and 2-chloro pyrido[2,3-d]pyrimidines</i>	128
4.3.1.3	<i>New synthesis of 4.14.1</i>	129
4.3.1.4	<i>Synthesis of pyridone 4.7.1</i>	130
4.3.1.5	<i>Synthesis of 4.15 by cyanamide substitution</i>	130
4.3.1.6	<i>Synthesis 2-halopyrido[2,3-d]pyrimidines 4.16</i>	130
4.3.1.6.1	<i>2-chloropyrido[2,3-d]pyrimidine 4.16.1</i>	131
4.3.1.6.2	<i>2-bromopyrido[2,3-d]pyrimidine 4.16.2</i>	131
4.3.1.7	<i>Dual coupling of 4.16.2 with xylylenediamine</i>	132
4.3.1.8	<i>Study of the Suzuki Coupling on 2-bromopyridopyrimidine 4.16.2</i>	134
4.3.2	<i>Exploring the synthesis of 6-bromomethyl pyrido[2,3-d]pyrimidine</i>	138
4.3.2.1	<i>Hypothesis</i>	138
4.3.2.2	<i>The Bayliss-Hillman reaction to afford 4.26</i>	139
4.3.2.3	<i>Synthesis of methyl 2-bromomethylacrylate 4.28</i>	140
4.3.2.4	<i>Williamson ether coupling between 4.28 and tetraethylene glycol</i>	141
4.3.3	<i>Findings in the synthesis of 6-bromo substituted pyrido[2,3-d]pyrimidines</i>	144
4.3.3.1	<i>Previous work</i>	144
4.3.3.1.1	<i>Stablishing the real structure of intermediate 4.37</i>	145
4.3.3.2	<i>Structural possibilities</i>	149
4.3.3.2.1	<i>Scaffold assembling</i>	150
4.3.3.2.2	<i>Bromination</i>	151
4.3.3.2.3	<i>Transposition of 4.41 to 4.42</i>	152
4.3.3.3	<i>Halogen diversity study</i>	153
4.3.3.3.1	<i>N-Bromosuccinimide</i>	154
4.3.3.3.2	<i>Use of other N-halosuccinimides</i>	155
4.3.3.4	<i>Kinetic study of the formation of the 6-bromo substituted pyridopyrimidines</i>	159

4.3.3.5	<i>Reactivity of 6-halo substituted pyridopyrimidines towards anti-DM1 candidates</i>	161
4.3.4	<i>Findings in the synthesis of 6-carboxymethyl pyrido[2,3-d]pyrimidines</i>	165
4.3.4.1	<i>Hypothesis</i>	165
4.3.4.2	<i>Fischer esterification of itaconic acid 4.50</i>	166
4.3.4.3	<i>Michael addition of malononitrile to 4.51</i>	167
4.3.4.4	<i>Cycloaddition of guanidine to 4.52</i>	168
4.3.4.5	<i>Ester saponification of 4.54</i>	169
4.3.4.6	<i>One-pot multicomponent alternative for the synthesis of 4.54</i>	170
4.4	DISCUSSION	171
4.5	REFERENCES	174

5	BASE RECOGNIZERS	177
5.1	Introduction	179
5.1.1	<i>Overview</i>	179
5.1.2	<i>The base interactor</i>	179
5.1.2.1	<i>Base intercalators</i>	179
5.1.2.2	<i>Base recognizer</i>	180
5.1.3	<i>Candidate selection</i>	181
5.1.3.1	<i>Zimmerman</i>	182
5.1.3.2	<i>Nakatani</i>	183
5.1.3.3	<i>This work</i>	184
5.2	OBJECTIVES	186
5.3	RATIONAL DESIGN	187
5.3.1	<i>Strategy design</i>	187
5.3.2	<i>Receptor preparation</i>	188
5.3.2.1	<i>Molecular Dynamics</i>	188
5.3.2.2	<i>Induced-fit docking</i>	188
5.3.3	<i>Ligand evaluation</i>	189
5.3.3.1	<i>Molecular dynamics and MMPBSA</i>	189
5.3.3.2	<i>Structural parameters analysis</i>	191
5.4	SYNTHESIS	194
5.4.1	<i>Second generation of recognizers</i>	197
5.5	BIOLOGIC EVALUATION	202
5.5.1	<i>AID 2675</i>	202
5.5.2	<i>Cell Biology</i>	204
5.6	CHAPTER DISCUSSION	209
5.7	REFERENCES	210
6	CONCLUSIONS	215

7.	EXPERIMENTAL	219
7.1	BIOCHEMISTRY	221
7.1.1	<i>Protein purification.</i>	221
7.1.2	<i>AID 2675.</i>	222
7.2	RATIONAL DESIGN	222
7.2.1	<i>Base recognizers</i>	222
7.3	CELL CULTURE	230
7.3.1	<i>Groove Binders activity determination</i>	230
7.3.2	<i>Base recognizers activity determination</i>	232
7.4	CHEMICAL SYNTHESIS	237
7.4.1	<i>Instrumental</i>	237
7.4.2	<i>Synthesis of 4,4'-methylenedibenzoyl chloride (3.15).</i>	239
7.4.3	<i>Method 1: Schotten-Baumann amide bond formation.</i>	240
7.4.4	<i>Synthesis of 3,3'-((4,4'-methylenebis(benzoyl))bis(azanediyl))dibenzoic acid(3.6{4.1}).</i>	241
7.4.5	<i>Synthesis of 5,5'-((4,4'-oxybis(benzoyl))bis(azanediyl))diisophthalic acid(3.13{9}).</i>	242
7.4.6	<i>Synthesis of 4,4'-oxybis(N-(3-(1H-tetrazol-5-yl)phenyl)benzamide) (3.13{8.1}).</i>	243
7.4.7	<i>Method 2: Conventional amide bond formation.</i>	244
7.4.8	<i>Synthesis of 4,4'-oxybis(N-(3-cyanophenyl)benzamide)(3.13{7.1}).</i>	245
7.4.9	<i>Synthesis of 4,4'-oxybis(N-(3-carbamoylphenyl)benzamide)(3.13{2.1}).</i>	246
7.4.10	<i>Synthesis of 4,4'-oxybis(N-(N',N''-di-tert-butylloxycarbonyl-3-guanidinophenyl) benzamide) (3.23).</i>	247
7.4.11	<i>Synthesis of tert-butyl 4-aminobenzoate (3.27).</i>	248
7.4.12	<i>Synthesis of 4,4'-ureylene-di-(tert-butyl benzoate) (3.28).</i>	249
7.4.13	<i>Synthesis of 3-(1H-tetrazol-5-yl)aniline hydrochloride (3.20).</i>	250
7.4.14	<i>Synthesis of 4,4'-Ureylene-dibenzoic acid(3.14)</i>	251
7.4.15	<i>Synthesis of N,N'-di-tert-butylloxycarbonyl-3-guanidinoaniline (3.18).</i>	252
7.4.16	<i>Synthesis of 2-methoxy-5-methyl-6-oxo-1,4,5,6-tetrahydropyridine-3-carbonitrile (4.7.1).</i>	253
7.4.17	<i>Synthesis of the sodium salt of N-(3-cyano-5-methyl-6-oxo-1,4,5,6-tetrahydropyridin-2-yl)cyanamide (4.15).</i>	254
7.4.18	<i>Synthesis of 4-amino-2-bromo-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (4.16.2).</i>	255
7.4.19	<i>Synthesis of 4-amino-2-chloro-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (4.16.1).</i>	256
7.4.20	<i>Synthesis of 4,4'-oxybis(N-guanidinophenylbenzamide)(3.13{1.1}).</i>	257

7.4.21	Method 3: Protection of nitrobenzoic acid.	258
7.4.22	Synthesis of tert-butyl 4-nitrobenzoate (3.25).	259
7.4.23	Synthesis of tert-butyl 3-nitrobenzoate (3.30).	260
7.4.24	Method 4: Catalyzed reduction from nitro to amine group.	261
7.4.25	Synthesis of tert-butyl 4-aminobenzoate (3.27).	262
7.4.26	Synthesis of tert-butyl 3-aminobenzoate (3.29).	263
7.4.27	Synthesis of 2,2'-((1,4-phenylenebis(methylene)) bis(azanediyl))bis(4-amino-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one) (4.14).	264
7.4.28	Method 5: EDC mediated amide bond formation.	265
7.4.29	Synthesis of di-tert-butyl 3,3'-((4,4'-(carbonylbis(azanediyl))bis(benzoyl))bis(azanediyl))dibenzoate (3.32).	266
7.4.30	Synthesis of 4,4'-(carbonylbis(azanediyl))bis(N-(3-carbamoylphenyl)benzamide) (3.14{2.1}).	267
7.4.31	Synthesis of 4,4'-(carbonylbis(azanediyl))bis(N-(3-cyanophenyl)benzamide) (3.14{7.1}).	268
7.4.32	Synthesis of 3,3'-((4,4'-(carbonylbis(azanediyl))bis(benzoyl))bis(azanediyl)) dibenzoic acid (3.14{4.1}).	269
7.4.33	Synthesis of dimethyl 2-methylenesuccinate (4.51)	270
7.4.34	Synthesis of dimethyl 2-(2,2-dicyanoethyl)succinate (4.52)	271
7.4.35	Synthesis of methyl 2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetate (4.53)	272
7.4.36	Synthesis of 2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetic acid (4.55)	273
7.4.37	Method 6: Amide bond procedure for ligands 5.1	274
7.4.38	Synthesis of N,N'-(pentane-1,5-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetamide)	275
7.4.39	Synthesis of N,N'-(heptane-1,7-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetamide) (5.1.2)	276
7.4.40	Synthesis of N,N'-(octane-1,8-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetamide) (5.1.3)	277
7.4.41	Synthesis of N,N'-(nonane-1,9-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetamide) (5.1.4)	278
7.4.42	Synthesis of N,N'-(decane-1,10-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetamide)	279
7.4.43	Synthesis of N,N'-((butane-1,4-diylbis(azanediyl))bis(propane-3,1-diyl))bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-6-yl)acetamide) (5.2)	280
7.4.44	Synthesis of 2,4-diamino-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (4.40)	281

7.4.45	Synthesis of 2-amino-4a-bromo-4-(bromoimino)-4,5,6,8-tetrahydropyrido[2,3-d]pyrimidin-7(4aH)-one (4.41)	282
7.4.46	Synthesis of 2,4-diamino-6-bromopyrido[2,3-d]pyrimidin-7(8H)-one (4.42)	283
7.4.47	Synthesis of 2,4-diamino-7-oxo-7,8-dihydropyrido[2,3-d]pyrimidine-6-carbonitrile (4.47)	284
7.4.48	Synthesis of 2,4-diamino-6-iodo-4,8-dihydropyrido[2,3-d]pyrimidin-7(4aH)-one (4.43)	285
7.4.49	Synthesis of 2,4-diamino-7-oxo-7,8-dihydropyrido[2,3-d]pyrimidine-6-carboxylic acid (4.46)	286
7.4.50	Synthesis of N'-(6-bromopyridin-2-yl)-N,N-dimethylformimidamide (4.22)	287
7.4.51	Synthesis of N,N-dimethyl-N'-(6-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)pyridin-2-yl)formimidamide (4.23)	288
7.4.52	Synthesis of N'-(6-(4-amino-6-methyl-7-oxo-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-2-yl)pyridin-2-yl)-N,N-dimethylformimidamide (4.24)	289
7.5	REFERENCES	291

1 Introduction

In this chapter, the pharmaceutical foundations necessary for understanding the objectives of this thesis are proposed.

Orphan diseases

1.1 Orphan diseases

According to FDA criteria, rare diseases are those affecting less than 200.000 people. Furthermore, as a general market trend, pharmaceutical corporations are wary of investing in research in rare diseases as revenues are not always guaranteed. For this reason, rare diseases become orphans, as they lack commercial treatment or cure. Most orphan diseases are genetic diseases that often lack treatment and basic comprehension of their causes and implications.

1.1.1 Genetic disorders

Genetic diseases originate when DNA repair mechanisms cannot restore genetic information properly, either due to the severity of DNA damage or repair enzyme malfunction. This erroneous genetic information can be heritable if the affected gene is present in the reproductive cells or non-heritable if the altered gene is not found in the gametes.¹ There are three main groups:

1.1.1.1 Chromosomal disorders

Chromosome abnormalities determine chromosomal disorders either by variations in the genetic material such as a chromosomal excess (trisomy ($2n+1$)), deficiency (monosomy ($2n-1$)), or nullisomy ($2n-2$)), or by mutations in their structure or location.^{2,3} For instance, Turner syndrome is a genetic disorder that only affects the female gender and is characterized by the absence of one or part of an X chromosome.⁴

1.1.1.2 Complex disorders

They are caused by various altered genes that interact with the surrounding environment influencing polygenic expression. These diseases do not follow the Mendelian inheritance pattern and are the most frequent among genetic pathologies. Many factors, both genetic and environmental, may play a significant role in their development.⁵ Diabetes, hypertension, and arteriosclerosis, among others, are some of the most common examples of this kind of genetic condition.

1.1.1.3 Monogenic disorders

Diseases with Mendelian inheritance patterns are those occurring when there is a single affected gene. Studying the pedigree of large families with many affected individuals can be used to determine whether the affected gene is in an autosome (dominant or recessive) or a

Orphan diseases

sex-related gene.⁶ Within this group, disorder-associated mutations can be classified into punctual and increased trinucleotide repeats.

Point mutations can occur at DNA, RNA, and protein levels, presenting either shifting of bases (transitions, transversions), insertions, or deletions of nucleotides. According to how these mutations appear, they can be classified as:

- Silent mutations: when specific codons give the same amino acid.
- Conservative missense mutations: when specific codons give chemically similar amino acids.
- Non-conservative missense mutations: when the resulting amino acid has chemical properties different from the original.
- Nonsense mutations: result in a premature codon stop.
- Read pattern change mutations: result in a loss of function and are null mutations.

A typical case of monogenic point mutations is those leading to the expansion of regular sequences. Occurring during DNA replication, a specific damaged, a short sequence of nucleotides (commonly 3 or 4 bases long) is miscopied, and the sequence tends to grow via corrections on strand slippage. This leads to trinucleotide repeat expansions (if three), and the most common sequences are CAG, CGG, and CTG. There are a few examples of genetic disorders of this kind, all of which are classified as rare diseases.

1.1.1.3.1 TRED's

Trinucleotide repeat expansion diseases (TREDs) are characterized by an expansion into a disease-causing range of a short homogeneous sequence of nucleotides.⁷ These microsatellite expansions may not be toxic *per se*. In some cases, they are located in intronic or intergenic loci; however, they usually undergo transcription giving rise to aberrant transcripts. In some cases, if these mRNAs are successfully exported outside the nuclei, they may trigger repeat-associated non-AUG (RAN) translation.⁸

During the DNA replication process, a mutation may occur due to denaturation and displacement of the DNA strands, resulting in a change in the complementary bases. This mechanism is known as slipped-strand mispairing (SSM).⁹ Whenever a strand is shifted forward or backward from the synthesized strand, the initial bulge generated turns into a hairpin in one of the strands, resulting in a new DNA strand with one more or one less repeat than the original strand.¹⁰

The most characteristic examples of these types of mutations, summarized in Table 1.1, are Huntington's disease, spinobulbar muscular atrophy (SBMA) or Kennedy's disease, fragile X syndrome, myotonic dystrophy type 1 and type 2, Friedreich's ataxia, and spinocerebellar

ataxia (SCA) which can affect different genes.¹¹ Table 1.1 displays a summary of the characteristic nucleotide repeats and gene locus of each disease and how many repetitions are necessary to develop a symptomatic phenotype.

Tab 1.1. Most characteristic trinucleotide repeat expansion diseases (TREDs).¹² For each one, the repeats or trinucleotides causing the disease and their location are shown. The number of copies of the trinucleotides found in non-affected individuals and the number of copies found in affected individuals is also displayed.¹³

Disease	Repeat	Unaffected	Affected	Locus
Huntington disease (HD)	CAG	6-34	36-121	HD
Spinobulbar muscular atrophy (SBMA)	CAG	9-36	38-62	DRPLA
X Fragile syndrome (FRAXA)	CGC	6-60	> 200	FMRP
X Fragile syndrome (FRAXE)	CCG	4-39	200-900	FMR2
Myotonic dystrophy type 1 (DM1)	CTG	5-37	50-10000	DMPK
Myotonic dystrophy type 2 (DM2)	CCTG	5-26	75-11000	ZNF9
Friedreich Ataxia	GAA	34-100	200-1700	FRDA
Spinocerebellar ataxia 1 (SCA1)	CAG	6-39	40-82	SCA1
Spinocerebellar ataxia 2 (SCA2)	CAG	15-24	32-200	SCA2
Spinocerebellar ataxia 8 (SCA8)	CTG	16-34	>74	SCA8
Spinocerebellar ataxia 10 (SCA10)	ATTCT	10-20	500-4500	SCA10

When SSM occurs in the coding region of a gene that contains short tandem repeats, it will result in an amino acid deletion (if deletion of 3 nucleotides occurs) or changes in the reading pattern (if deletion of 1, 2, 4, or 5 nucleotides occurs).¹⁴

In the following sections, myotonic dystrophies are discussed in-depth as the current thesis is focused on those specific TREDs.

1.2 Myotonic dystrophies

The myotonic dystrophies comprise a group of autosomal dominant hereditary multisystemic diseases. Myotonic dystrophies share common features, such as myotonia, muscle weakness, premature development of cataracts, defects in cardiac function, and endocrine disorders.¹⁵⁻¹⁸

Myotonic dystrophies

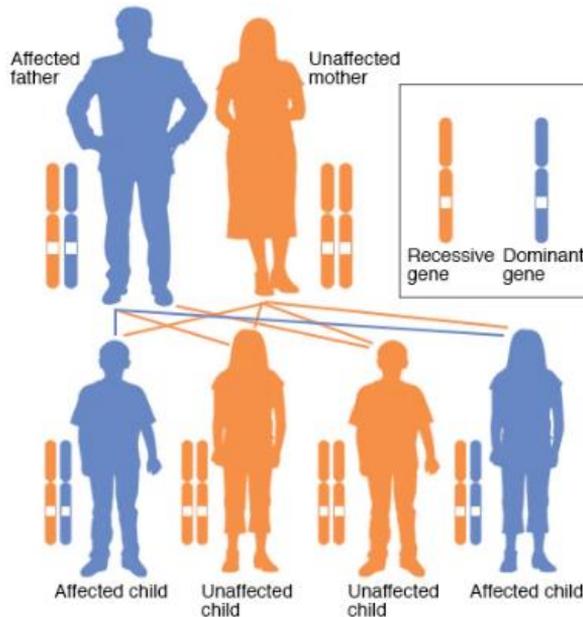


Fig 1.1. A representation of an inheritance diagram of a genetic condition with an autosomal dominant pattern. An affected individual (in this case the father) has a 50 percent chance of passing on the abnormal gene (and having an affected child) and a 50 percent chance of passing on the normal copy of the gene (and having an unaffected child) with each pregnancy.¹⁹

Initially, the two types of myotonic dystrophies were considered a single disease. It was not until 1909 when Hans Steinert's team identified, from the different symptomatology presented by those affected by one characteristic type of myotonia, which became known as Steinert's disease.

Until 1992 it was not known which gene was responsible for Steinert's myotonic dystrophy. It was found that myotonic dystrophy was caused by the expansion of the abnormal CTG triplet repeat in the 3' untranslated region (UTR) of the DMPK gene in chromosome 19.²⁰⁻²² Based on this discovery, it was observed that some patients exhibited the characteristic symptoms but with a greater proximal than distal weakness, in addition to the absence of the gene responsible for Steinert's myotonic dystrophy.

This variant of myotonic dystrophy was called proximal myotonic myopathy (PROMM) in the USA or proximal myotonic dystrophy (PDM) in Europe.^{23,24} It adopted the common term myotonic dystrophy without CTG repeat expansion or myotonic dystrophy type 2 (DM2).²⁵

Tab 1.2. Comparison of the main clinical features of both myotonic dystrophies DM1 and DM2.

Clinical characteristics	DM1	DM2
Epidemiology	worldwide	European
Symptom appearance	any	8-60
Anticipation	yes	rare
Congenital onset	yes	no

Finally, DM2 was described as a unique disease produced by the unstable expansion of the CCTG tetranucleotide repeat in intron 1 of the nucleic acid-binding protein (CNBP) gene, formerly known as ZNF9 (zinc finger 9 gene), on chromosome 3q21.²⁶

1.2.1 Myotonic dystrophy type 1 (DM1)

DM1 is the most common adult-onset muscular dystrophy, an incurable neuromuscular disorder characterized by symptoms such as progressive muscle weakening, myotonia, cardiac arrhythmias, cognitive dysfunction, and cataracts.²⁷ To date, there is no effective treatment against this disease.

It has been reported that the disease worsens from one generation to the next and with the patient's aging. The latter is due to a disorder called mosaicism, which is experienced when two or more populations of cells have different genetic compositions.²⁸ This occurs due to the nature of the mutation that causes this disease, which can undergo dynamic mutation, as it is relatively unstable.²⁹ In myotonic dystrophy type 1, the offspring inherit this disease with a more severe phenotype in the case of reproduction of an affected individual. This phenomenon is known as genetic anticipation.

1.2.1.1 Symptoms, prognosis & prevalence

Like most diseases, the first to be characterized in DM1 was symptomatology.³⁰ Hans Steinert identified a symptomatic pattern in patients with myotonia with a well-defined characteristic phenotype. Among these symptoms, he observed that patients had myotonia, muscular dystrophy, cardiac defects, cataracts, and endocrine disorders (see Figure 1.3).³¹ However, three phenotypic groups of DM1 are categorized by symptoms and their severity. Both are directly related to the moment in which they are detected.³² The phenotypes of DM1 and its symptomatology are summarized in Table 1.3.

Myotonic dystrophies

Tab 1.3. Summary of general DM1 subtypes relating the observed clinical findings to the CTG repeat size in patients DMPK gene.³³

Phenotype	Clinical aspects	CTG repeat size	Age
Congenital	Polyhydramnios in utero Tented/" fish shaped" upper lip Hypotonia, severe weakness Respiratory failure Cardiopulmonary complications Cerebral atrophy Mental retardation/learning disability	> 1000	Birth
Childhood	Facial weakness, dysarthria Myotonia Low intelligence Cardiac conduction abnormalities	50 - 1000	1 - 10
Classic / Adult	Early cataracts Myotonia Distal weakness Ptosis Balding Temporomandibular wasting Excessive daytime sleepiness Central or obstructive sleep apnea Respiratory failure Cardiac conduction abnormalities Insulin resistance Testicular failure GI dysmotility and/or dysphagia Mild intellectual deficits Mood disorders Cluster C personality traits	50 - 1000	10 - 30
Late	Cataracts Mild myotonia	50 - 100	20 - 70

The first group is the most severe phenotype, the so-called congenital. Even before birth, the affected individuals show reduced fetal movements,¹⁵ and once born, many perish during the first months of life.³⁴ Symptoms include hypotonia (excessive muscle weakness), respiratory impairment, and sucking and feeding problems. If the child survives this stage, there is a slight improvement in symptomatology. The rest of the symptoms gradually appear as the patient reaches adulthood, showing cognitive, speech, and hearing impairment.³⁵

The second group is the phenotype in which the symptoms appear in the juvenile stage. It does not develop from birth but manifests in the patient's infancy, and the character of the symptoms is more psychological rather than physical as in the adult phenotype. The most characteristic symptoms are apathy, learning deficits, and cognitive impairments.²³ As the patient ages and reaches adulthood, the signs of the adult phenotype, corresponding to the third group, emerge. Although the adult-onset phenotype could be classified as mild compared to congenital, it can present cardiac events leading to an increase of death ratio or risk

of sudden death, The size of the CTG expansion in the blood of myotonic dystrophy type 1 patients is associated with total and sudden deaths, conduction defects, left ventricular dysfunction, and supraventricular arrhythmias.³⁶

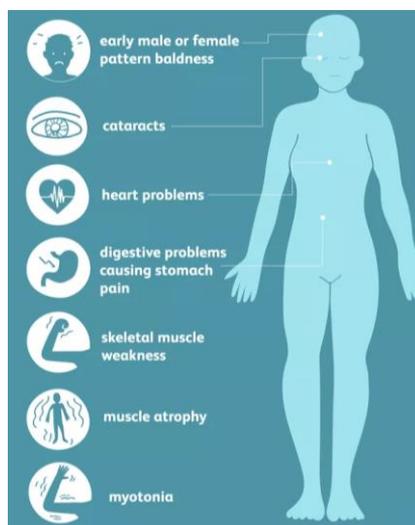


Fig 1.2. A representation of classic DM1 phenotype. The classic onset causes weakness of the skeletal muscles and the internal organs including the heart, the muscles that power breathing, and muscles of the digestive system. Myotonic muscular dystrophy is typically also characterized by delayed muscle relaxation.³⁷

The third group is the adult phenotype, which presents the mildest symptomatology and, at the same time, the most common. As it appears at an advanced age, its diagnosis becomes difficult due to the similarity of the symptoms with other typical age-related conditions (such as cataracts and cardiac deficiencies (see Table 1.3)).³⁸ Among the most characteristic symptoms include myotonia, muscular atrophy, weakness, cataracts, breathing difficulty accentuated at sleeping time, arrhythmias, somnolence, insulin resistance, and weakness in the cervical, particularly head and face muscles. i.e., there is difficulty in closing the eyelids or in chewing and, finally, also in the intestinal tract muscles.³⁹

1.2.1.2 Molecular basis

DM1 is caused by an abnormal expansion of the trinucleotide CTG in the 3' untranslated region (UTR) of the DMPK gene on chromosome 19 (locus 19q13.32)⁴⁰ and anomalous expression of the contiguous genes DMWD and the promoter region of SIX5.^{41,42}

Since the region where the mutation occurs is a non-coding DNA region, it is improbable that this is the sole cause of the disease, resulting in other trinucleotide expansion diseases. Therefore, it is only the trigger of a series of events that collectively explain the complex phenotype of DM1. This expansion of CTG repeats at the 3'UTR end of the DMPK gene is the pathogenic origin of the disease.⁴³ In fact, the longer this expansion becomes, the more

Myotonic dystrophies

severe the symptoms are. For this reason, different subtypes of this disease have been classified according to the size of the expansion, with the congenital subtype having a longer segment of repeats (thousands of repetitions) and the late or asymptomatic subtype having a shorter segment of repeats (50 to 150 repeats).

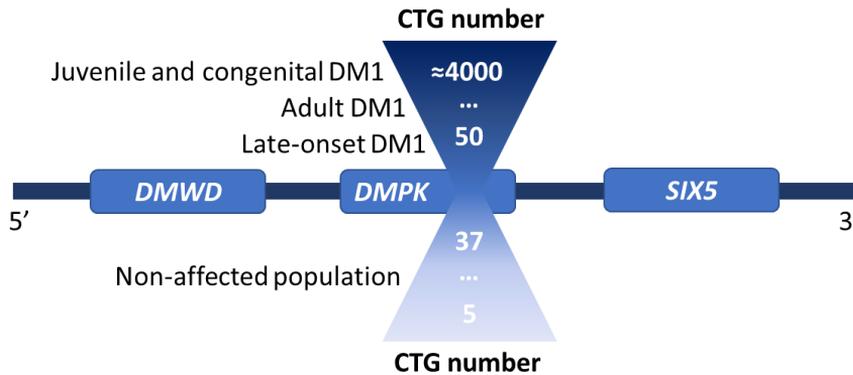


Fig 1.3. Adapted representation of the relationship between CTG repeat size and phenotype onset. Neighbor genes DMWD and SIX5 are also affected in gene expression of the region may be affected by the CTG expansion. Furthermore, can be observed the non-affected population range of CTG repeat size.⁴⁴

As has been described, unaffected individuals have between 5 and 37 CTG repeats (Figure 1.3), while those affected have more than 50. Interestingly, those who have a number of repetitions between these two ranges, these individuals suffer from the so-called premutation. Therefore, it is improbable that they develop any symptoms. However, their offspring may develop them since the number of repetitions increases during the gametogenesis because of the anticipation phenomenon.

1.2.1.3 The direct implication of genetic alteration and cis effect

When transcription of the affected DMPK gene occurs, pre-mRNA is not successfully transported to the cytoplasm due to defects in polyadenylation.⁴⁵ Therefore, this pre-mRNA cannot migrate out of the nucleus. For this reason, their availability of the DMPK protein decreases.^{46,47} On the other hand, it has also been reported that the mutation itself affects the genetic transcription of the region, the so-called cis-effect.^{48,49}

Defects in polyadenylation and anomalous expression in DMPK-contiguous genes such as the promoter region of SIX5,^{16,50} which stems from a family of developmental regulatory transcription factors, are essential factors in the pathogenesis of DM1. However, there are still more factors with major significance underlying the effects caused by DM1.

Apart from the direct consequence of the stalling of CTG transcripts in the nucleus, several phenomena are associated with these non-coding transcripts. Some of them have a significant impact on the characteristic phenotype of DM1. Specifically, the capacity of these

transcripts to group into complex and dynamic structures leads to severe modifications in the cell's metabolism. Moreover, the pre-mRNA CUG itself has potentially toxic characteristics.⁵¹

1.2.1.4 Alternative polyadenylation (APA)

Once transcribed, this toxic RNA should be transported to the cytoplasm through a mechanism initiated by polyadenylation. Alternative polyadenylation (APA) is a cell signaling mechanism that consists of adding a long adenine tail to the sequence to protect this RNA from degradation and facilitate the transport of the pre-mRNA to the cytoplasm. In addition, this phenomenon is linked to the control of gene expression.⁵²

On the other hand, the APA of the CUG RNA in patients affected with DM1 makes impossible the exit of the toxic transcript of the nucleus.⁴⁵ This is due to the intrinsic nature of this transcript. It has been described that GC-rich transcripts (such as DM1-characteristic CUGs) tend to form stable hairpin secondary structures.^{53,54}

The impossibility of tracking out the transcripts of the nucleus causes accumulation. Therefore, other phenomena characteristic of DM1 is triggered.

1.2.1.5 Nuclear foci

As a consequence of the stalling of some pre-mRNAs in the nucleus, it has been shown that this type of structure tends to gelify in *in vivo* systems forming clusters of difficult access for the process and maturation of RNA.⁵⁵ This phenomenon has been widely described given the recent high interest in all non-coding RNAs (ncRNAs) nature and functional diversity, as in most cases are related to genetic disorders.

Recent advances in understanding the complex and transversal role of RNA in all cellular processes have broken with the central dogma of biology. Especially relevant have been the discovery of the ribosome structure (approximately 50% of its structure is RNA, including the active center, recognized with the chemistry Nobel prize in 2009), riboswitches, miRNAs, siRNAs, and other oligonucleotides with the ability to modulate enzymatic activity.⁵⁶ In addition, and especially interesting for the present thesis is the role of long non-coding RNAs (lncRNAs) in forming membrane-free organelles by liquid-liquid phase transitions. Based on the promiscuous interaction capacity and lability of RNA, lncRNA can generate inclusion bodies creating optimal spaces to develop alternative splicing processes and modulation of cell factor activity. Paradigmatic cases of this are nuclear speckles and paraspeckles, among others.⁵⁷

In the specific case of DM1, pre-mRNA CUG tends to have complex secondary structures that provide U•U bulges allowing specific interactions.^{16,47,58–60} As a direct consequence, CUG leads the formation of inclusion bodies along with many essential nucleic proteins. These

Myotonic dystrophies

inclusion bodies are the so-called discrete foci or foci, and apart from CUG RNA, they are formed by many essential alternative splicing proteins such as MBNLs.^{61,62}

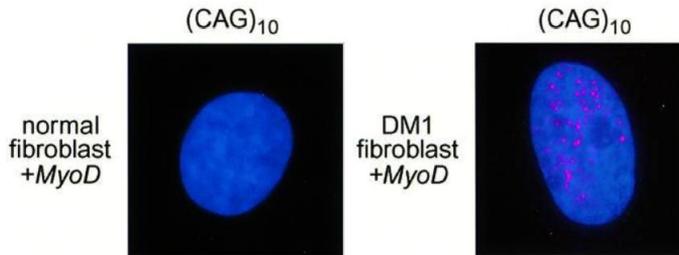


Fig 1.4. FISH analysis of nuclear foci using a (CAG)₁₀ probe (red) in unaffected DM1 cells. It can be observed how the RNA probe detects inclusion bodies in discrete points throughout the nucleus.⁶¹

The formation of foci in a mechanism of toxic RNA gain of function is the most determining factor in the pathogenesis of DM1. As it has been introduced, the nuclear proteins that recruit it are essential factors for the regular pattern of alternative splicing.

1.2.1.6 RAN translation

As described above, defects in the DM1 polyadenylation phenomenon prevent CUG transcripts from being transported outside the nucleus. Therefore, this phenomenon should make translating the sequences affected by the mutation impossible. Nevertheless, antisense transcription occurs when the characteristic mutation of the DM1 locus occurs. This is not an unknown phenomenon, as antisense transcription originates part of the siRNA and RNAi.^{63,64} These antisense transcripts can reach the cytoplasm and be translated without the requirement of an ATG initiator, resulting in polyGln, polySer, and polyAla homopolymers that can be toxic from accumulating inside the cell.^{65,66}

It has been described that the transcription is upregulated for these CAG(n) antisense sequences (the antisense strand from CTG). On the other hand, the concentration of RNAs with that sequence in the cytoplasm and that of the translated RAN homopolymers is low, although detectable.⁶⁷ This is believed to occur because the sequence is found in an intronic region that is effectively spliced.

1.2.1.7 Protein implications

DM1 protein scenario is focused on two main actors: MBNL1 and CELF1. Although not being the only affected proteins by such a multifactorial disease, the effects on these two proteins can explain most of the characteristic phenotype of DM1. Therefore, the following sections will introduce and discuss the effects on those proteins.

1.2.2 Muscleblind-like proteins

Some of the proteins most likely to be sequestered during foci formation are the Muscleblind-like proteins (MBNL).⁶⁸ Four zinc-finger domains and a linker between them form these proteins.⁶⁹ MBNL are RNA binding proteins (RBP). Their activity consists of the splicing of pre-mRNA fragments to attain mature mRNA. These proteins, essential for life, are sequestered by the foci due to the great affinity for zones with a high density of C and G bases, causing their bioavailability to decrease and causing alterations in the alternative splicing pattern.⁷⁰ MBNL proteins belong to a family of RNA metabolism regulators consisting of three isoforms of a characteristic gene, MBNL1, MBNL2, and MBNL3. These three types of proteins have certain structural similarities, such as the presence of the four zinc-finger (ZnF) domains, which are essential in recognizing pre-mRNA and mRNA sequences. The most relevant is MBNL1, as it significantly regulates alternative splicing in most tissues.⁷¹

MBNLs are RNA-associated proteins that are not passively trapped within CUG but are also involved in CUG formation, giving the foci a dynamic nature. Indeed, it has been widely described that CUG tends to form regularly shaped hairpins instead of chaotic and dynamic foci.^{54,58}

The interaction between MBNL and RNA is based on the ZnF domains, these being structured in the form of ZnF1/2 and ZnF3/4 pairs, linked by a linker. Each consists of three cysteines followed by a histidine motif CCCH⁷², presenting a structural symmetry between them with the same spacing between cysteines and histidines between ZnF 1/3 and ZnF 2/4. Either removing one ZnF or the linker that binds them produces a significant decrease in affinity against RNA.⁷³ Therefore, both pairs of ZnF are critical in RNA binding.

One exciting feature is that ZnFs can recognize only single strands of RNA via base recognition (see Figure 1.5). This may explain why regular secondary structures of CUG RNA turn into foci. Several binding mechanisms have been proposed for MBNL to disrupt the GC-rich hairpins.⁷⁴

Myotonic dystrophies

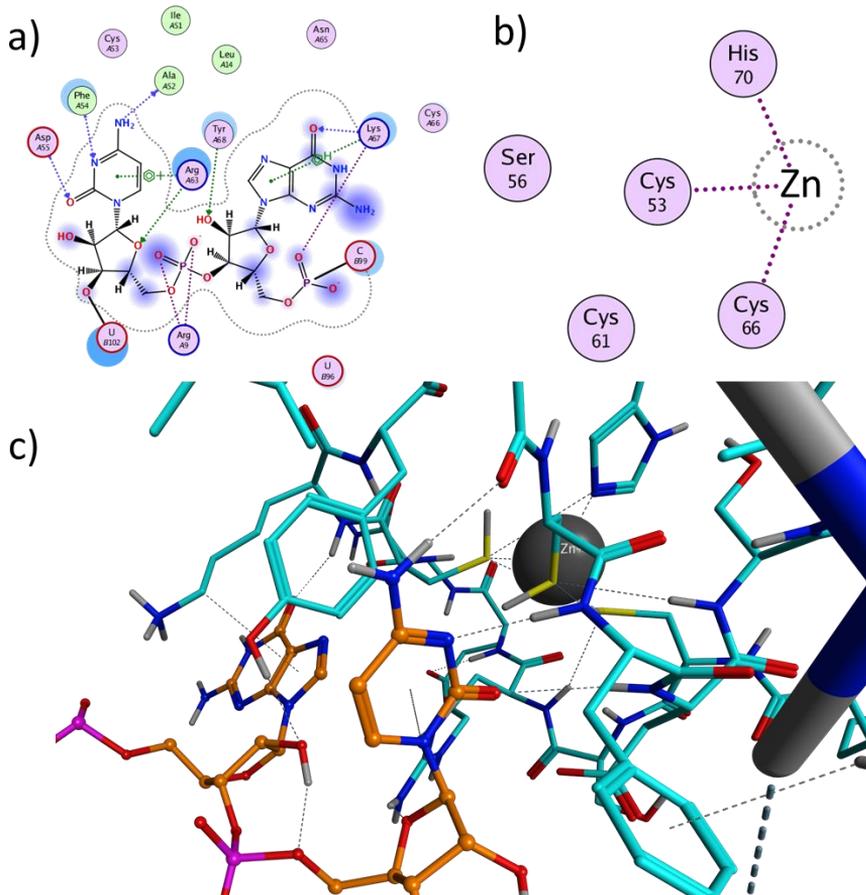


Fig 1.5. Crystal structure of ZnF bound to RNA chain. a) Pharmacophoric representation showing key residue interacting with RNA. b) Pharmacophoric representation of the CCCH motif, showing the stabilization of the Zn atom by these residues. c) 3D representation of the Crystal structure. Interactions between the neighbor residues of CCCH and RNA can be observed. (PDB id:5u9b)

Each ZnF domain recognizes a 5'-GC-3' pair of the YGCY sequence (Y: pyrimidines), presenting different possible binding modes for RNA chains. GC nucleotide pairs present their Watson-Crick interaction zone in ZnF gaps composed of arginine and lysine aromatic side chains (Figure 1.5.a: Arg63 and Lys 67), with two zinc-coordinated cysteines (Figure 1.5.b: Cys66 and Cys 70).⁷⁵ The interaction between MBNL/RNA is based on the multiple hydrogen bonds established between them. Furthermore, pi stacking between arginine and lysine and cytosine and guanine bases stabilizes the union. As a result, this bonding mechanism shows significant interaction of GC-rich RNA and MBNL1, the origin of foci formation.⁷⁶

Literature shows that MBNL1 interacts preferably with structures containing multiple bulges of an undefined secondary structure formed by CHHG or CHG repetitions (H = C, A, U) versus RNA structures that do not present this type of bulges.^{73,77} These U•U mismatches are essential in recognizing MBNL1 of this RNA. Furthermore, if base pairs replace mismatches, the binding of the proteins falls.^{78,79}

All these findings explain the irreversible binding of MBNL1 to CUG RNA. The lack of this free protein and the splicing activity on many pre-mRNAs codifying essential proteins is one of the leading causes of the DM1 phenotype. Furthermore, as a result of the mutation, one of the misspliced genes is the MBNL pre-mRNA, making not only being sequestered but the remaining free MBNL is less active as it is in misspliced variant forms.⁸⁰

1.2.3 CELF1 Overexpression

Another nuclear protein called CUGBP Elav-like family member 1 (CELF1) is an MBNL1 antagonist. It has been reported that CUG repeats bind to CELF1 and stabilize it.⁸¹ CELF1 binds UG-rich RNAs using three RRM domains.⁸² Furthermore, CELF1 has activity outside the nucleus. In addition to being a splicing agent in the nucleus, CELF1 also exhibits translational activity in the cytoplasm. There, protein kinase C phosphorylates CELF1 raising its stability.⁸³

Like other alternative splicing factors, CELF1 dysregulation can affect the expression of many other genes. It has been reported that its overexpression in mice causes DM1 symptoms.^{84,85} So its effect is additive to the loss of available DM1.

A key aspect of CELF1 activity is phosphorylation. The translational activity of this protein is regulated upstream by cyclin D3cyclin-dependent kinase 4 (CDK4) that phosphorylates CELF1 in its serine 302. To attain its regular activity, P-CELF1 need to bind the initiation translation factor 2 α (eIF2 α).⁸⁶ However, in DM1 tissues, CDK4 is downregulated by the protein glycogen synthase kinase 3 β (GSK3 β).⁸⁷ The lack of CDK4 affects the phosphorylation of CELF1, turns it into a repressor of the correct translation, and misregulates its targets by not binding to eIF2 α .⁸⁸ GSK3 β , as CELF1, is upregulated in DM1 tissues by CUG stabilization and marks CDK4 for degradation phosphorylating it in its tyrosine 283.⁸⁹

Both phenomena, the inner nucleus stabilization, and the outside translation decay make CELF1 protein play a critical role in DM1 pathogenic events.

1.2.4 miRNA dysregulation

MicroRNAs (miRNAs) are small, conserved, non-coding RNAs. They have a crucial role in gene regulation. As DM1 has an RNA toxic gain of function, this also affects miRNAs regulation. CUG has been reported to modify the maturation of several pre-mRNAs and pre-miRNAs. An example is the miR1 family, which MBNL1 regulates. It has been proven that the reduction of MBNL1 also causes a depletion of mature miR1.⁹⁰ The dysregulation of this family causes cardiac affections.⁹¹

This is just an example, as many other miRNAs have been discovered to be dysregulated in DM1 and few are used as biomarkers of the disease.⁹²⁻⁹⁴

Myotonic dystrophies

1.2.5 Splicing pattern imbalance

Although the complete characterization of the DM1 basis is yet far to be reached, many events have been described and related. The two most significant features to explain the big picture in DM1 are the MBNL1 sequestration and the CELF1 upregulation. Furthermore, as both have similar roles and can target the same pre-mRNAs, it has been stated that both combined cause what is called missplicing or splicing pattern imbalance.

During a lifetime, the concentration of MBNL1 and nuclear CELF1 is inversely proportional. In the early life stages, a high concentration of active CELF1 and low MBNL1 is observed, and during growth to adulthood, the concentrations are progressively reversed until mostly MBNL1 and inactive CELF1.

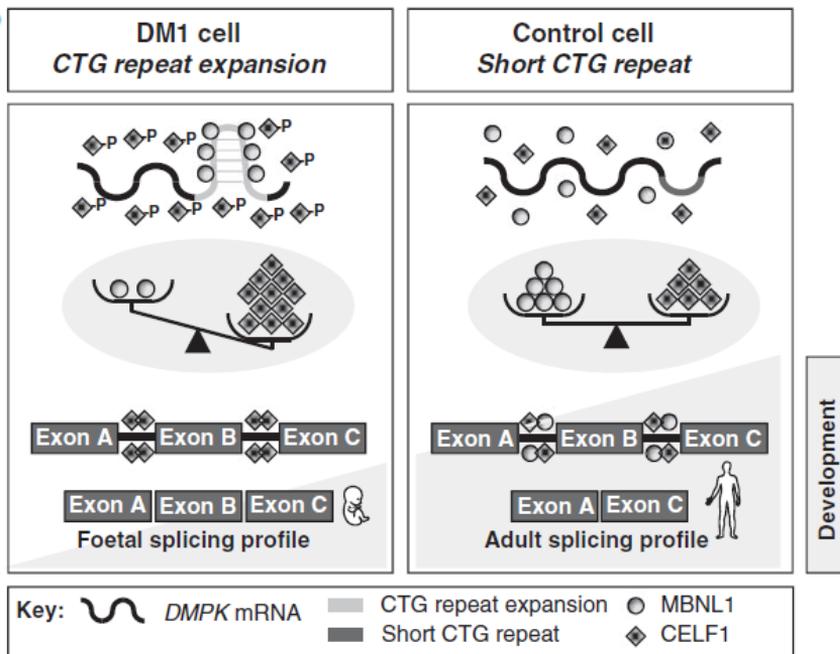


Fig 1.6. Splicing pattern relation between non-affected cells and DM1 cells. On the right side, regular protein levels trigger standard splicing patterns of multiple pre-mRNAs, leading to the expression of the adult splicing profile. On the left side, DM1-related splicing pattern imbalance leads to pre-mRNA maturation and fetal splicing profile expression dysfunction.⁹⁵

In adult DM1 patients, a low concentration of free MBNL1 and an increase in the concentration of CELF1 protein is experienced, provoking an alteration in the alternative splicing pattern of the proteins regulated by these genes (see Figure 1.6).⁴⁴ For this reason, most of the observed symptoms are related to muscle waste, as for obvious reasons, muscle development needs to be downregulated in the fetal stage of life.

Both the MBNL1 depletion and the CELF1 upregulation build this change in the alternative splicing pattern of an adult human, resembling the one that would present in a fetal state.^{96–98} The sum of both is considered the most critical hallmark in the pathogenesis of myotonic dystrophy.^{99,100}

Once discussed the basis of DM1, in the following sections, the different approaches to reach a feasible treatment will be described.

1.3 Therapeutic state of the art

It can be assumed that the phenotypic characteristics of DM1 are reversible via genetic transcription inhibition of the affected region, RNA foci stabilization or degradation, and splicing pattern correction. Several feasible strategies have been explored through the latter years. In this section, a brief description of these approaches to DM1 treatment or cure will be discussed.

1.3.1 DNA targeting

The pathogenesis of DM1 and other diseases of its kind causes them to be incurable by most known techniques. This statement does not include the examples introduced in this section. Although these techniques have great controversy and low social acceptance, they are the best option to establish themselves as a corrective treatment for genetic diseases. Gene therapy is capable of blocking CTG transcription¹⁰¹ or even inducing repeat contractions to pre-mutation levels.

1.3.1.1 CRISPR genomic engineering

One has represented a complete revolution in its field among all gene therapy techniques. Indeed, it has recently been awarded the Nobel Prize (chemistry 2020). It is the identification of cluster regularly interspaced short palindromic repeats (CRISPR) associated protein (Cas) system as a revolutionary tool in genomic engineering.

The origin of this tool, like many others, is the harnessing of the evolutionary attributes of other natural beings. Specifically, the CRISPR systems were first identified in bacteria and *archaea*.¹⁰² They are part of an adaptative immune system to defend these unicellular organisms from phage infections.¹⁰³ CRISPR-associated immunization occurs by recognizing foreign DNA and using it as a single guide RNA biosynthesis (sgRNA) pattern. After assembling with endonuclease Cas proteins, it forms the effector complex. The effector complex can recognize and degrade the foreign DNA upon subsequent infections.^{104,105}

Therapeutic state of the art

The potential of this tool was unveiled when it was discovered that sgRNAs were synthetically feasible. Furthermore, and even more critical, the breaks caused by Cas endonuclease could be rejoined via homology-directed repair (HDR). The directed cut and the effective repair lead to a precise gene excision in a target DNA. This made the tool notorious and controversial in recent years.^{106,107} References involving this technique in genetic disease studies outnumber as a clear sign of this technique's potential. The reason is simple; there is no other classic technique to revert genetic disorders phenotype permanently.

DM1 is a suitable target for the CRISPR system, as the defects in the genetic coding gather specific features that are likely compatible with the system.¹⁰⁸ On the one hand, the CTG expansion is located in a non-coding region. Therefore, if the repair junction of the excised sequence is not perfect, it should not affect the correct expression of the DMPK protein. Furthermore, the regular and specific repetition should present fewer off-targeting issues as this is not a conserved sequence along the human genome. Indeed, several studies have succeeded in cleaving CTG repeats selectively using this technique in-vitro.^{109–112}

1.3.2 Protein targeting

As discussed above, there are two main actors in the DM1 protein scene: MBNL1 and CELF1. Most of the phenotypes originated from the imbalance in those protein activity regulations. Given that MBNL1 is sequestered inside the foci, at least at the protein level, the available target is the inhibition of CELF1 splicing activity. The discovery of Tideglusib, a small molecule able to inhibit the repression activity of the MBNL1 antagonist CELF-1 protein, and the consequent balancing of alternative splicing is a hallmark in the development of a treatment for DM1. This route proposes ATP noncompetitive ligand inhibition of the GSK3-B kinase. This causes a reduction in phosphorylated cyclin D3cyclin-dependent kinase 4 (CDK4). CDK4 regulates CELF1 activity, thus cutting the phosphorylation activity of GSK3 β upstream, and the translational repression of CELF1 falls.^{113,114} In the preclinical studies carried out by Wang *et al.* It can be observed that by tuning GSK3 β activity, mice phenotypic defects are corrected.¹¹⁵ Therefore, the regular splicing pattern is restored. Now, Tideglusib is in clinical phase II. (NCT02858908)

1.3.3 RNA targeting

The classical central dogma of molecular biology establishes few RNA roles, all strictly related to gene expression.¹¹⁶ However, discoveries in the latter decades have probed that the versatility of RNA requires a much broader definition. Ribozymes,¹¹⁷ interference RNAs,^{118–120} RNA sponges¹²¹, and a plethora of new roles emerged, turning ncRNAs into one of the most trending fields in biology.¹²² However, to target such a labile target can be complicated. In

some cases, such as nuclear foci, the specific structure is unclear, as RNA is much harder to crystallize and obtain X-ray diffraction data.

As previously described, CUG RNA plays a critical role in DM1 pathogenesis. Therefore, targeting the RNA itself is one of the most explored strategies in the way of discovering a suitable DM1 treatment. A few different approaches are discussed below, grouped by the intrinsic characteristics of the candidates.

1.3.3.1 Oligonucleotides

A clever strategy to target nucleotides is to use other nucleotides. Among nucleotides, the most promising candidates are antisense oligonucleotides (ASOs). ASOs are a novel class of synthetic single-stranded nucleotide short chains. Antisense agents inhibit the expression of a target gene by coupling it to a specific sequence. In the DM1 case, ASOs can bind to target CUG toxic RNA impeding the formation of complex secondary structures leading to foci formation.^{123,124}

In terms of affinity, there is no other candidate, no protein, and no small molecule able to attain the values of ASOs.¹²⁵ This technology allows significant diminishing in DM1 phenotypic features. Some chemical modifications also can induce gene silencing on CTG or CAG transcripts.^{126,127}

Either way, ASOs can target misspliced genes, inducing silencing of exons that should be spliced out but, in DM1 conditions, are not. An ASO targeting chloride channel 1 (ClC-1) pre-mRNA has been reported, allowing the correct expression of the protein in DM1-affected mice.¹²⁸

Although relatively recent, few of them have already been approved by FDA in recent years for the treatment of genetic diseases such as Duchenne muscular Dystrophy (AMONDYS 45, Viltepso, Vyondys 53 & Exondyss 51).

1.3.3.2 Oligopeptides

Given the idiosyncrasy of DM1 toxic RNA and the type of structure it requires to be targeted, peptides are in themselves a powerful tool to reach potential candidates, either as part of a larger structure or as drug candidates themselves.^{129,130}

On the one hand, these compounds take advantage of being easily synthesized by automated techniques based on solid-phase peptide synthesis (SPPS, chemistry Nobel prize 1984) and are therefore ideal candidates for HTS evaluation techniques. Although peptides do not bear all the ideal features for a drug candidate, they exhibit specific beneficial properties such as high specificity, affinity, biocompatibility, and tissue penetration. On the other hand, they are

Therapeutic state of the art

easily degraded by proteases, a characteristic not always desired in a drug candidate (reduces bioavailability). This phenomenon can be corrected by few strategies, such as chemical modification by adding groups that hinder cleavage by proteases.¹³¹ Another option is the substitution of L amino acids by D amino acids, which are not recognized by proteases, thus considerably increasing the lifetime of these entities in the organism.^{132,133}

Several studies lead to candidates by the screening of peptide libraries. García-López A. *et al.* discovered hexapeptide ABP1 (Ac-ppyawe-NH₂) by HTS. ABP1 exhibits the ability to prevent the formation of the secondary structure of CUG. Therefore, it provides a reduction in both the number of foci formed in cell models and a reduction in histological defects in HSA^{LR} mice models.¹³⁴ Based on these results, it is concluded that this structure could stabilize linear CUG structures, giving reversibility to the binding of MBNLs, which would avoid the sequestration in the aggregates of essential alternative splicing factors and other biomolecules. One of the crucial characteristics of ABP1 is the solution to the problem of reduced bioavailability due to protease degradation. ABP1 consists exclusively of *D*-aminoacids, which give stability and long-lasting effects.

The use of peptides or short amino acid sequences as part of a drug candidate is another example of using these structures in the DM1 field. In the studies, these structures are used as spacers and vehicles that give the candidate selectivity, and, on the other hand, they also facilitate the penetration of tissues, thus increasing their effectiveness *in vivo*. In the studies of Angelbello *et al.*, the presence of peptides is recurrent in most of its candidates, used as a spacer for structures that bind the CUG RNA.¹³⁵ Furthermore, in the studies described by the Zimmerman group, the use of these oligomers fulfills a function similar to the previous one. As a distinguishing feature, the use of longer chains can be seen, providing a greater capacity for tissue transfer and the possibility of carrying several CUG recognizing structures.¹³⁶

1.3.3.3 Small molecular entities

Finally, the most influential structures in the present thesis are small molecules. Currently, this approach is undoubtedly the preferred pharmacological strategy to treat any kind of disease due to its robustness and cost-effectiveness ratio. In the last few years, some molecular entities have been described to exert a significant activity against DM1.

Some naturally occurring molecules with strong RNA interaction have been described as hits or used in derivative synthesis. Most of them are used as antibiotics such as erythromycin¹³⁷ or kanamycin^{138,139} and other natural structures such as harmine¹⁴⁰ or (di)lomofungin.¹⁴¹

Described small molecules can be classified into a few significant groups based on their characteristic mechanisms of action: compounds able to bind DNA and repress transcription,

inhibitors able to recover missplicing events by CELF1 repression, and compounds able to recover MBNL1 from the foci by targeting RNA.^{142–144}

Focusing on the foci, only three strategies have led to compounds able to achieve selectivity and potency: groove binding, base intercalation, and base recognition.^{142,143,145} Although not all reported structures are designed with a known mechanism of action as most are selected through high throughput screening assays, the basis underlying these mechanisms is discussed in the following chapters, and novel designed structures are described and tested against DM1.

1.3.4 The study of Myotonic Dystrophy at IQS School of Engineering

Our group of research, Grup de Química Farmacèutica (GQF) of IQS School of Engineering, entered in the field of myotonic dystrophy in 2010 through the project entitled “Development of novel treatments for myotonic dystrophy: in vivo drug discovery” granted by La Fundació Marató de TV3 (REF. 100231), the catalan telethon. Such project, leaded by Prof. Ruben Artero Allepuz from the University of Valencia, was carried out in collaboration between the University of Valencia, IQS and IUCT (Institut Universitari de Ciència i Tecnologia).

The main objectives of the project were the identification of new chemical scaffolds with potential biological activity against DM1 in *Drosophila* models and the design and synthesis of new drug candidates, able to revert the pathogenic phenotype. During the project, several families of compounds were assessed as CUG-binding small molecules, including pentamidine-like compounds and pyrido[2,3-*d*]pyrimidines.¹⁴⁶ Molecular modeling methods were applied to analyze the druggability of the candidates. Compounds **1-2** and **2-5** were identified as promising CUG repeat binders by polarization assays (Fig 1.5. A, B) and their biological activity was tested in DM1 mutant cells and *Drosophila* DM1 models with an impaired locomotion phenotype. These compounds enhanced the levels of free MBNL1 in patient-derived myoblasts in vitro and greatly improved DM1 fly locomotion in climbing assays (Fig. 1.5.C).

Therapeutic state of the art

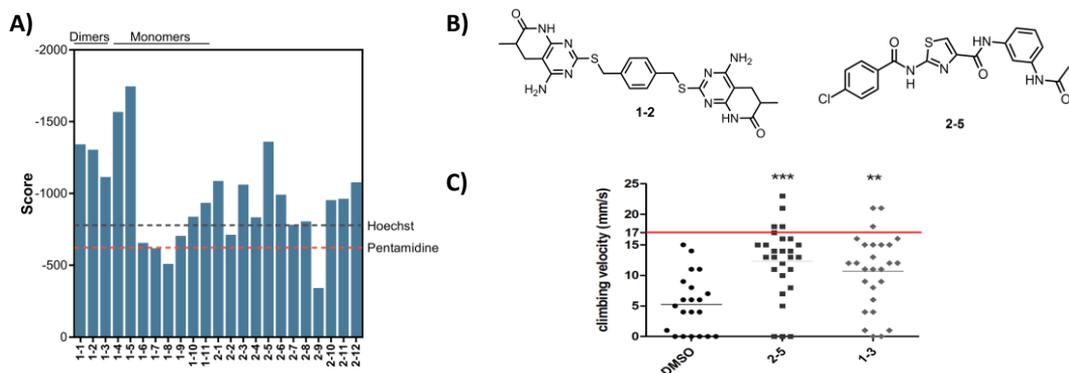


Fig 1.7. A) LigandRNA function score reported for the studied families. Pentamidine and Hoechst 33258 reference energies are included as red and grey lines, respectively. B) chemical structures of compounds 1-2 and 2-5. C) Climbing assay performed on 30 adult male flies fed for five days with the indicated compound. Climbing velocity was significantly higher in flies receiving 40 μ M compound 2-5 or 100 μ M compound 1-3 compared to DM1-model flies receiving only 1%DMSO solvent carrier. The average climbing velocity for wild type flies is marked as a red line. Statistics were calculated using the Student's t-test (** $p < 0.01$, *** $p < 0.001$).¹⁴⁶

Although promising, the results of the biological activity of selected compounds were not good enough to consider them as lead compounds. For this reason, this doctoral thesis aims to review the strategy for designing new anti-DM1 compounds and to improve the biological activity of the selected candidates. Apart from the usual difficulties involved in any drug design process, the first problem we identified was the lack of an in-house biological test able to identify promising hits fast and cheaply.

For this reason, the general objectives of this PhD are:

- 1) Development of a primary test to screen chemical libraries in-house.
- 2) Design and selection of new compounds with potential anti-DM1 activity.
- 3) Synthesis and evaluation of the biological activity of selected candidates.

Each one of these objectives will be developed in the subsequent chapters that will include, when necessary, an introduction and revision of the state of the art to center the subject.

1.4 References

1. Griffiths, A. J. F. & Ayllón Gómez, F. *Genética*. (McGraw-Hill Interamericana, 2008).
2. Alliance, G. Understanding Genetics: A New York, Mid- Atlantic Guide for Patients and Health Professionals. *Genet. Alliance, New York - Mid-Atlantic Consort. Genet. Newborn Screen. Serv.* 105 (2008).
3. NIH. Chromosome Abnormalities Fact Sheet. <https://www.genome.gov/about-genomics/fact-sheets/Chromosome-Abnormalities-Fact-Sheet>.
4. Bondy, C. A. Turner syndrome 2008. *Horm. Res.* **71**, 52–56 (2009).
5. Hunter, D. J. Gene-environment interactions in human diseases. *Nat. Rev. Genet.* **6**, 287–298 (2005).
6. Antonarakis, S. E. & Beckmann, J. S. Focus on Monogenic Disorders. *Nat. Rev. Genet.* **7**, 277–282 (2006).
7. Ashley, C. T. & Warren, S. T. Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.* **29**, 703–728 (1995).
8. Cleary, J. D. & Ranum, L. P. W. Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum. Mol. Genet.* **22**, 45–51 (2013).
9. Levinson, G. & Gutman, G. a. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221 (1987).
10. Pearson, C. E. *et al.* Slipped-strand DNAs formed by long (CAG)·(CTG) repeats: Slipped-out repeats and slip-out junctions. *Nucleic Acids Res.* **30**, 4534–4547 (2002).
11. Robitaille, Y., Lopes-Cendes, I., Becher, M., Rouleau, G. & Clark, A. W. The neuropathology of CAG repeat diseases: Review and update of genetic and molecular features. *Brain Pathol.* **7**, 901–926 (1997).
12. Orr, H. T. & Zoghbi, H. Y. Trinucleotide Repeat Disorders - annurev.neuro.29.051605.113042. *Annu. Rev. Neurosci.* **30**, 575–623 (2007).
13. Boucher, C. A. *et al.* A novel homeodomain-encoding gene is associated with a large CPG Island interrupted by the myotonic dystrophy unstable (CTG)_n repeat. *Hum. Mol. Genet.* **4**, 1919–1925 (1995).
14. Zoghbi, H. Y. Trinucleotide Repeat Disorders. in *Principles of Molecular Medicine* 1114–1122 (Humana Press, 2006). doi:10.1007/978-1-59259-963-9_116.
15. Meola, G. Clinical aspects, molecular pathomechanisms and management of myotonic dystrophies. *Acta Myol.* **32**, 154–165 (2013).
16. Liquori, C. L. *et al.* Myotonic dystrophy type 2 caused by a CCTG expansion in intron I of ZNF9. *Science (80-)*. **293**, 864–867 (2001).
17. Day, J. W. *et al.* Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. *Neurology* **60**, 657–664 (2003).
18. Turner, C. & Hilton-Jones, D. Myotonic dystrophy: Diagnosis, management and new therapies. *Curr. Opin. Neurol.* **27**, 599–606 (2014).
19. How Genetic Disorders are Inherited. <https://www.mayoclinic.org/> (2020).
20. Brook, J. D. *et al.* Molecular basis of myotonic dystrophy: Expansion of a trinucleotide

References

- (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**, 799–808 (1992).
21. Fu, Y. H. *et al.* An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science (80-.)*. **255**, 1256–1258 (1992).
 22. Mahadevan, M. *et al.* Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science (80-.)*. **255**, 1253–1255 (1992).
 23. Maurage, C. a *et al.* Similar brain tau pathology in DM2 / PROMM and DM1 / Steinert disease. *J. Neurol.* **2**, 1636–1639 (2007).
 24. Newman, B., Meola, G., O'Donovan, D. G., Schapira, A. H. V. & Kingston, H. Proximal myotonic myopathy (PROMM) presenting as myotonia during pregnancy. *Neuromuscul. Disord.* **9**, 144–149 (1999).
 25. Meola, G., Sansone, V., Radice, S., Skradski, S. & Ptacek, L. A family with an unusual myotonic and myopathic phenotype and no CTG expansion (proximal myotonic myopathy syndrome): A challenge for future molecular studies. *Neuromuscul. Disord.* **6**, 143–150 (1996).
 26. Bachinski, L. L. *et al.* Confirmation of the Type 2 Myotonic Dystrophy (CCTG)_n Expansion Mutation in Patients with Proximal Myotonic Myopathy/Proximal Myotonic Dystrophy of Different European Origins: A Single Shared Haplotype Indicates an Ancestral Founder Effect. *Am. J. Hum. Genet.* **73**, 835–848 (2003).
 27. Thomas, J. D., Oliveira, R., Sznajder, Ł. J. & Swanson, M. S. Myotonic dystrophy and developmental regulation of RNA processing. *Compr. Physiol.* **8**, 509–553 (2018).
 28. Biesecker, L. G. & Spinner, N. B. A genomic view of mosaicism and human disease. *Nat. Publ. Gr.* **14**, 307–320 (2013).
 29. Van Eyk, C. L. & Richards, R. I. Dynamic mutations: Where are they now? *Adv. Exp. Med. Biol.* **769**, 55–77 (2012).
 30. Wagner, A., Gustav, H. & Steinert, W. Hans Steinert (1875–1911). 1607–1608 (2008) doi:10.1007/s00415-008-0004-x.
 31. Harper, P. S. Congenital myotonic dystrophy in Britain. II. Genetic basis. *Arch. Dis. Child.* **50**, 514–21 (1975).
 32. Amack, J. D. & Mahadevan, M. S. Myogenic defects in myotonic dystrophy. *Dev. Biol.* **265**, 294–301 (2004).
 33. Smith, C. A. & Gutmann, L. Myotonic Dystrophy Type 1 Management and Therapeutics. *Curr. Treat. Options Neurol.* **18**, (2016).
 34. Schoser, B. & Timchenko, L. Myotonic Dystrophies 1 and 2: Complex Diseases with Complex Mechanisms. *Curr. Genomics* **11**, 77–90 (2010).
 35. Schara, U. & Schoser, B. G. H. Myotonic Dystrophies Type 1 and 2: A Summary on Current Aspects. *Semin. Pediatr. Neurol.* **13**, 71–79 (2006).
 36. Chong-Nguyen, C. *et al.* Association between Mutation Size and Cardiac Involvement in Myotonic Dystrophy Type 1: An Analysis of the DM1-Heart Registry. *Circ. Cardiovasc. Genet.* **10**, 1–8 (2017).
 37. Moawad, H. Overview of Myotonic Muscular Dystrophy. 2022 <https://www.verywellhealth.com/>.

38. Matthews, E. *et al.* The non-dystrophic myotonias: Molecular pathogenesis, diagnosis and treatment. *Brain* **133**, 9–22 (2010).
39. Rudnik-Schöneborn, S. & Zerres, K. Outcome in pregnancies complicated by myotonic dystrophy: A study of 31 patients and review of the literature. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **114**, 44–53 (2004).
40. Udd, B. & Krahe, R. The myotonic dystrophies: Molecular, clinical, and therapeutic challenges. *Lancet Neurol.* **11**, 891–905 (2012).
41. Wansink, D. G. & Wieringa, B. Transgenic mouse models for myotonic dystrophy type 1 (DM1). *Cytogenet. Genome Res.* **100**, 230–242 (2003).
42. Ranum, L. P. W. & Day, J. W. Myotonic Dystrophy: RNA Pathogenesis Comes into Focus. *Am. J. Hum. Genet* **74**, 793–804 (2004).
43. Kaliman, P. & Llagostera, E. Myotonic dystrophy protein kinase (DMPK) and its role in the pathogenesis of myotonic dystrophy 1. *Cell. Signal.* **20**, 1935–1941 (2008).
44. Gomes-Pereira, M., Cooper, T. A. & Gourdon, G. Myotonic dystrophy mouse models: Towards rational therapy development. *Trends Mol. Med.* **17**, 506–517 (2011).
45. Batra, R. *et al.* Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell* **56**, 311–322 (2014).
46. O’Cochlain, D. F. *et al.* Transgenic overexpression of human DMPK accumulates into hypertrophic cardiomyopathy, myotonic myopathy and hypotension traits of myotonic dystrophy. *Hum. Mol. Genet.* **13**, 2505–2518 (2004).
47. Mankodi, A. *et al.* Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.* **10**, 2165–2170 (2001).
48. Amack, J. D., Paguio, A. P. & Mahadevan, M. S. Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum. Mol. Genet.* **8**, 1975–1984 (1999).
49. Signor, S. A. & Nuzhdin, S. V. The Evolution of Gene Expression in cis and trans. *Trends Genet.* **34**, 532–544 (2018).
50. Pearson, C. E., Edamura, K. N. & Cleary, J. D. Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet.* **6**, 729–742 (2005).
51. Mulders, S. A. M. *et al.* Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc. Natl. Acad. Sci.* **106**, 13915–13920 (2009).
52. Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and Consequences of Alternative Polyadenylation. *Mol. Cell* **43**, 853–866 (2011).
53. Krzyzosiak, W. J. *et al.* Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* **40**, 11–26 (2012).
54. van Cruchten, R. T. P., Wieringa, B. & Wansink, D. G. Expanded CUG repeats in DMPK transcripts adopt diverse hairpin conformations without influencing the structure of the flanking sequences. *Rna* **25**, 481–495 (2019).
55. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
56. Rohilla, K. J. & Gagnon, K. T. RNA biology of disease-associated microsatellite repeat

References

- expansions. *Acta Neuropathol. Commun.* **5**, 63 (2017).
57. Clemson, C. M. *et al.* An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
 58. Tian, B. *et al.* Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *Rna* **6**, 79–87 (2000).
 59. Sobczak, K. *et al.* Structural diversity of triplet repeat RNAs. *J. Biol. Chem.* **285**, 12755–12764 (2010).
 60. Mooers, B. H. M., Logue, J. S. & Berglund, J. A. The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci.* **102**, 16626–16631 (2005).
 61. Miller, J. W. Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy. *EMBO J.* **19**, 4439–4448 (2000).
 62. Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T. & Thornton, C. A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13**, 3079–3088 (2004).
 63. Kawasaki, H. & Taira, K. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431**, 211–217 (2004).
 64. Morris, K. V. Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells. *Science (80-)*. **305**, 1289–1292 (2004).
 65. Zu, T. *et al.* Non-ATG – initiated translation directed by microsatellite expansions. *Pnas* **108**, 260–265 (2010).
 66. Cho, D. H. *et al.* Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20**, 483–489 (2005).
 67. Gudde, A. E. E. G. *et al.* Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)_n repeat. *RNA Biol.* **14**, 1374–1388 (2017).
 68. Fardaei, M. *et al.* Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* **11**, 805–14 (2002).
 69. Kanadia, R. N. *et al.* Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. *Gene Expr. Patterns* **3**, 459–462 (2003).
 70. Ho, T. H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J.* **23**, 3103–3112 (2004).
 71. Sznajder, L. J. *et al.* Mechanistic determinants of MBNL activity. *Nucleic Acids Res.* **44**, 10326–10342 (2016).
 72. Pascual, M., Vicente, M., Monferrer, L. & Artero, R. The Muscleblind family of proteins: An emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* **74**, 65–80 (2006).
 73. Kino, Y. *et al.* Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.* **13**, 495–507 (2004).

74. Konieczny, P., Stepniak-Konieczna, E. & Sobczak, K. MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.* **42**, 10873–10887 (2014).
75. Teplova, M. & Patel, D. J. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.* **15**, 1343–1351 (2008).
76. Cass, D. *et al.* The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol. Biol.* **12**, 14–17 (2011).
77. Park, S. *et al.* Structural Basis for Interaction of the Tandem Zinc Finger Domains of Human Muscleblind with Cognate RNA from Human Cardiac Troponin T. *Biochemistry* **56**, 4154–4168 (2017).
78. Warf, M. B. & Berglund, J. A. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA Soc.* **13**, 2238–2251 (2007).
79. Kiliszek, A., Kierzek, R., Krzyzosiak, W. J. & Rypniewski, W. Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res.* **39**, 7308–7315 (2011).
80. André, L. M., Van Cruchten, R. T. P., Willemse, M. & Wansink, D. G. (CTG)_n repeat-mediated dysregulation of MBNL1 and MBNL2 expression during myogenesis in DM1 occurs already at the myoblast stage. *PLoS One* **14**, 1–17 (2019).
81. Timchenko, N. A. *et al.* RNA CUG Repeats Sequester CUGBP1 and Alter Protein Levels and Activity of CUGBP1. *J. Biol. Chem.* **276**, 7820–7826 (2001).
82. Teplova, M., Song, J., Gaw, H. Y., Teplov, A. & Patel, D. J. Structural Insights into RNA Recognition by the Alternate-Splicing Regulator CUG-Binding Protein 1. *Structure* **18**, 1364–1377 (2010).
83. Kuyumcu-Martinez, N. M., Wang, G. S. & Cooper, T. A. Increased Steady-State Levels of CUGBP1 in Myotonic Dystrophy 1 Are Due to PKC-Mediated Hyperphosphorylation. *Mol. Cell* **28**, 68–78 (2007).
84. Koshelev, M., Sarma, S., Price, R. E., Wehrens, X. H. T. & Cooper, T. A. Heart-specific overexpression of CUGBP1 reproduces functional and molecular abnormalities of myotonic dystrophy type 1. *Hum. Mol. Genet.* **19**, 1066–1075 (2010).
85. Ward, A. J., Rimer, M., Killian, J. M., Dowling, J. J. & Cooper, T. A. CUGBP1 overexpression in mouse skeletal muscle reproduces features of myotonic dystrophy type 1. *Hum. Mol. Genet.* **19**, 3614–3622 (2010).
86. Timchenko, N. A., Wang, G. L. & Timchenko, L. T. RNA CUG-binding protein 1 increases translation of 20-kDa isoform of CCAAT/enhancer-binding protein β by interacting with the α and β subunits of eukaryotic initiation translation factor 2. *J. Biol. Chem.* **280**, 20549–20557 (2005).
87. Jones, K. *et al.* GSK3 β mediates muscle pathology in myotonic dystrophy. *J. Clin. Invest.* **122**, 4461–4472 (2012).
88. Huichalaf, C. *et al.* Expansion of CUG RNA repeats causes stress and inhibition of translation in myotonic dystrophy 1 (DM1) cells. *FASEB J.* **24**, 3706–3719 (2010).

References

89. Naderi, S. *et al.* cAMP-induced degradation of cyclin D3 through association with GSK-3 β . *J. Cell Sci.* **117**, 3769–3783 (2004).
90. Rau, F. *et al.* Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.* **18**, 840–845 (2011).
91. Sayed, D., Hong, C., Chen, I. Y., Lypowy, J. & Abdellatif, M. MicroRNAs play an essential role in the development of cardiac hypertrophy. *Circ. Res.* **100**, 416–424 (2007).
92. Perbellini, R. *et al.* Dysregulation and cellular mislocalization of specific miRNAs in myotonic dystrophy type 1. *Neuromuscul. Disord.* **21**, 81–88 (2011).
93. Fernandez-Costa, J. M. *et al.* Expanded CTG repeats trigger miRNA alterations in Drosophila that are conserved in myotonic dystrophy type 1 patients. *Hum. Mol. Genet.* **22**, 704–716 (2013).
94. Kalsotra, A. *et al.* The Mef2 transcription network is disrupted in myotonic dystrophy heart tissue, dramatically altering miRNA and mRNA expression. *Cell Rep.* **6**, 336–345 (2014).
95. Gomes-Pereira, M., Cooper, T. A. & Gourdon, G. Myotonic dystrophy mouse models: towards rational therapy development Myotonic dystrophy: a paradigm of RNA toxicity NIH Public Access. *Trends Mol Med* **17**, 1–21 (2011).
96. Lin, X. *et al.* Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15**, 2087–2097 (2006).
97. Kalsotra, A. *et al.* A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20333–20338 (2008).
98. Du, H. *et al.* Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol.* **17**, 187–193 (2010).
99. Ranum, L. P. W. & Day, J. W. Pathogenic RNA repeats: An expanding role in genetic disease. *Trends Genet.* **20**, 506–512 (2004).
100. Cho, D. H. & Tapscott, S. J. Myotonic dystrophy: Emerging mechanisms for DM1 and DM2. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1772**, 195–204 (2007).
101. Pinto, B. S. *et al.* Impeding Transcription of Expanded Microsatellite Repeats by Deactivated Cas9. *Mol. Cell* **68**, 479-490.e5 (2017).
102. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-)*. **315**, 1709–1712 (2007).
103. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
104. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-)*. **337**, 816–821 (2012).
105. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2013**, 1–9 (2013).
106. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280–284 (2017).

107. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).
108. Raaijmakers, R. H. L., Ripken, L., Ausems, C. R. M. & Wansink, D. G. CRISPR/Cas Applications in Myotonic Dystrophy: Expanding Opportunities. *Int. J. Mol. Sci.* **20**, 3689 (2019).
109. Provenzano, C. *et al.* CRISPR/Cas9-Mediated Deletion of CTG Expansions Recovers Normal Phenotype in Myogenic Cells Derived from Myotonic Dystrophy 1 Patients. *Mol. Ther. - Nucleic Acids* **9**, 337–348 (2017).
110. van Agtmaal, E. L. *et al.* CRISPR/Cas9-Induced (CTG-CAG)_n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing. *Mol. Ther.* **25**, 24–43 (2017).
111. Dastidar, S. *et al.* Efficient CRISPR/Cas9-mediated editing of trinucleotide repeat expansion in myotonic dystrophy patient-derived iPS and myogenic cells. *Nucleic Acids Res.* **46**, 8275–8298 (2018).
112. Yanovsky-Dagan, S. *et al.* Deletion of the CTG Expansion in Myotonic Dystrophy Type 1 Reverses DMPK Aberrant Methylation in Human Embryonic Stem Cells but not Affected Myoblasts. *bioRxiv* (2019) doi:10.1101/631457.
113. Höglinger, G. U. *et al.* Tideglusib reduces progression of brain atrophy in progressive supranuclear palsy in a randomized trial. *Mov. Disord.* **29**, 479–487 (2014).
114. Lovestone, S. *et al.* A phase II trial of tideglusib in alzheimer's disease. *J. Alzheimer's Dis.* **45**, 75–88 (2015).
115. Wang, M. *et al.* Correction of Glycogen Synthase Kinase 3 β in Myotonic Dystrophy 1 Reduces the Mutant RNA and Improves Postnatal Survival of DMSXL Mice. *Mol. Cell. Biol.* **39**, 1–20 (2019).
116. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).
117. Emilsson, G. M., Nakamura, S., Roth, A. & Breaker, R. R. Ribozyme speed limits. *RNA* **9**, 907–918 (2003).
118. Sobczak, K., Wheeler, T. M., Wang, W. & Thornton, C. A. RNA interference targeting CUG repeats in a mouse model of myotonic dystrophy. *Mol. Ther.* **21**, 380–387 (2013).
119. Seow, Y., Sibley, C. R. & Wood, M. J. A. Artificial mirtron-mediated gene knockdown: Functional DMPK silencing in mammalian cells. *Rna* **18**, 1328–1337 (2012).
120. Cerro-Herreros, E. *et al.* MiR-23b and miR-218 silencing increase Muscleblind-like expression and alleviate myotonic dystrophy phenotypes in mammalian models. *Nat. Commun.* **9**, (2018).
121. Cerro-Herreros, E., Fernandez-Costa, J. M., Sabater-Arcis, M., Llamusi, B. & Artero, R. Derepressing muscleblind expression by miRNA sponges ameliorates myotonic dystrophy-like phenotypes in *Drosophila*. *Sci. Rep.* **6**, 1–13 (2016).
122. Matsui, M. & Corey, D. R. Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* **16**, 167–179 (2017).
123. Gao, Z. & Cooper, T. A. Antisense Oligonucleotides: Rising Stars in Eliminating RNA Toxicity in Myotonic Dystrophy. *Hum. Gene Ther.* **24**, 499–507 (2013).

References

124. Sardone, V., Zhou, H., Muntoni, F., Ferlini, A. & Falzarano, M. S. Antisense oligonucleotide-based therapy for neuromuscular disease. *Molecules* **22**, (2017).
125. Mulders, S. A. M. *et al.* Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13915–13920 (2009).
126. Nakamori, M., Gourdon, G. & Thornton, C. A. Stabilization of expanded (CTG).(CAG) repeats by antisense oligonucleotides. *Mol. Ther.* **19**, 2222–2227 (2011).
127. Kurreck, J. Antisense technologies: Improvement through novel chemical modifications. *Eur. J. Biochem.* **270**, 1628–1644 (2003).
128. Wheeler, T. M., Lueck, J. D., Swanson, M. S., Dirksen, R. T. & Thornton, C. A. Correction of CIC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy. *J. Clin. Invest.* **117**, 3952–3957 (2007).
129. Patel, D. J. Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. Struct. Biol.* **9**, 74–87 (1999).
130. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. *Science (80-.).* **287**, 820–825 (2000).
131. Renukuntla, J., Vadlapudi, A. D., Patel, A., Boddu, S. H. S. & Mitra, A. K. Approaches for enhancing oral bioavailability of peptides and proteins. *Int. J. Pharm.* **447**, 75–93 (2013).
132. Bruno, B. J., Miller, G. D. & Lim, C. S. Basics and recent advances in peptide and protein drug delivery. *Ther. Deliv.* **4**, 1443–1467 (2013).
133. Sachdeva, S. Peptides as ‘Drugs’: The Journey so Far. *Int. J. Pept. Res. Ther.* **23**, 49–60 (2017).
134. García-López, A., Llamusi, B., Orzáez, M., Pérez-Payá, E. & Artero, R. D. In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11866–11871 (2011).
135. Angelbello, A. J. *et al.* Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7799–7804 (2019).
136. Bai, Y. *et al.* Integrating Display and Delivery Functionality with a Cell Penetrating Peptide Mimic as a Scaffold for Intracellular Multivalent Multitargeting. *J. Am. Chem. Soc.* **138**, 9498–9507 (2016).
137. Nakamori, M., Taylor, K., Mochizuki, H., Sobczak, K. & Takahashi, M. P. Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy. *Ann. Clin. Transl. Neurol.* **3**, 42–54 (2016).
138. Disney, M. D. *et al.* Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. *J. Am. Chem. Soc.* **130**, 11185–11194 (2008).
139. Disney, M. D. & Childs-Disney, J. L. Using selection to identify and chemical microarray to study the RNA internal loops recognized by 6'-N-acylated kanamycin A. *ChemBioChem* **8**, 649–656 (2007).

140. Herrendorff, R. *et al.* Identification of plant-derived alkaloids with therapeutic potential for myotonic dystrophy type I. *J. Biol. Chem.* **291**, 17165–17177 (2016).
141. Hoskins, J. W. *et al.* Lomofungin and dilomofungin: Inhibitors of MBNL1-CUG RNA binding with distinct cellular effects. *Nucleic Acids Res.* **42**, 6591–6602 (2014).
142. Lee, J. Y. *et al.* Intrinsically cell-penetrating multivalent and multitargeting ligands for myotonic dystrophy type 1. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8709–8714 (2019).
143. Coonrod, L. A. *et al.* Reducing Levels of Toxic RNA with Small Molecules. *ACS Chem. Biol.* **8**, 2528–2537 (2013).
144. Palomo, V. *et al.* Subtly Modulating Glycogen Synthase Kinase 3 β : Allosteric Inhibitor Development and Their Potential for the Treatment of Chronic Diseases. *J. Med. Chem.* **60**, 4983–5001 (2017).
145. Jahromi, A. H. *et al.* A novel CUGexp-MBNL1 inhibitor with therapeutic potential for myotonic dystrophy type 1. *ACS Chem. Biol.* **8**, 1037–1043 (2013).
146. González, À. L. *et al.* In silico discovery of substituted pyrido[2,3-d]pyrimidines and pentamidine-like compounds with biological activity in myotonic dystrophy models. *PLoS One* **12**, e0178931 (2017).

References

2 Adapting AID 2675

Disposing of a quick test capable of analyzing any drug candidate's potency is crucial in developing new APIs. In this project, the test plays a pivotal role in screening multiple candidates and giving discriminatory results before studying more complex systems. This chapter describes the implementation and tuning of the AID 2675 biochemical test.

Introduction

2.1 Introduction

2.1.1 Need for a test

The peculiarity of this project generates a need that most drug design studies would never meet. This necessity is to design or adapt a primary screening biological test.

Orphan diseases, including DM1, lack an accessible infrastructure of commercial dedicated resources. This is not surprising, given that it would not be economically viable for corporations to invest in such small markets. This fact affects every stage of the development process, from the early discovery stages to the most advanced clinical studies.

We intend to test the biological activity and validate the computationally performed selection for the current project. To achieve this goal, different possibilities of studying drug candidates' potency are evaluated to select and incorporate the most suitable one into the project.

2.1.2 Biological evaluation assays

Since drug candidates cannot be tested in patients, disease models are required to evaluate new drug candidates effectively. The model to be used must be selected according to the information sought and the resources and time willing to invest in obtaining the data. Therefore, different options are presented with decreasing levels of complexity:

2.1.2.1 Animal models

Despite being initially designed for disease characterization, mouse models have been widely used to evaluate drug candidates' potency or any other anti-DM1 therapeutic approach.

On the one hand, the most widely used is the HSA^{LR} model.¹ This transgenic mouse expresses an aberrant CTG repeat inserted in a different locus from the characteristic of DM1 (3' UTR DMPK). Despite being at another locus, this mutation causes an RNA gain of function and foci appearance with relatively few CTG repeats (~250).

However, this is not the only model designed for mice. Apart from all KO models of MBNLs or CELF1 due to their limited use for drug screening, other models such as DMSXL (>1000 CTG),^{2,3} EpA960 (960 CTG),⁴ DM5 (5 CTG),⁵ or DM200 (200)⁵ are relevant for the study of drug candidate viability.

On the other hand, models in other simpler animals such as *D. Melanogaster*,⁶ Zebrafish,⁷ or *C. Elegans*⁸ have proven to help screen drug candidates. The widespread use of *Drosophila* models is noteworthy among them due to the rapid offspring generation.

Introduction

2.1.2.2 Cell models

At a lower level of representativity, there are the DM1 cell models. Their high reliability and speed in obtaining biological activity data, on the contrary, make them an ideal and necessary component of any candidate selection process.

However, difficulties with these models arise in finding the cell type able to express the disease's whole phenotype. Since DM1 is a multisystemic disease, the representation of a single cell type is limited.

Nevertheless, there are very relevant cell models, particularly for muscle cells (fibroblasts, myoblasts, and myotubes).^{9,10} Other specific cell types derived from stem cells have also been studied, cardiac cells,¹¹ and even neuronal cells.¹²

Another critical factor in the use of cells for drug screening is their origin. The most successful models are currently based on immortalized cells from biopsies and treated for culture in the laboratory. However, cells expressing exogenous CTG repeats have been widely used in the past decade.¹³

2.1.2.3 In vitro screening

There are not many different approaches to calculating a candidate's efficacy in a system more straightforward than a cell. Most of the candidates end up being evaluated for their ability to interact with nucleotide sequences. The most commonly used techniques are isothermal titration calorimetry (ITC)¹⁴ and circular dichroism (CD).¹⁵

However, these studies lack the ability to observe whether the most characteristic phenomenon of DM1 is affected: the sequestering of MBNL1. A study published in 2012 finally managed to adapt this process to be able to perform high-throughput screening. This test was recorded in the Bioassay database with ID 2675.

2.1.3 AID 2675 Bioassay

Chen et al. described two methods to perform high-throughput assays¹⁶ and detect the protein-RNA interaction since, to date, the most common methods used to detect these interactions were low-throughput such as gel mobility shift, filter-binding, and yeast-three-hybrid (Y3H).¹⁷ The described assays are based on proximity to measure the binding of the MBNL1 protein (in this case, a recombinant MBNL1- Δ 105-His₆) to an oligo (CUG)₁₂ RNA. To do so, a time-resolved fluorescence energy transfer (HTRF) assay was performed to quantify the binding of MBNL1 protein to CUG RNA and an AlphaScreen-based test.

qHTS Assay for Inhibitors of MBNL1-poly(CUG) RNA-binding (AID2675)¹⁸ test is based on the fact that MBNL1-His₆ is captured by the anti-His6 antibody conjugated to a tertiary fluorophore (Anti-His-Tb cryptate), and the RNA-bound biotin (Biot-(CUG)₁₂) is captured by an avidin, i. e. streptavidin conjugated to XL-665 (SA-XL665). The proximity of Anti-His-Tb-cryptate to XL-665 allows the phenomenon of Förster Resonance Energy Transfer (FRET). Non-radiative energy transfer occurs because the antibody contains a terbium fluorophore, which acts as a FRET donor. Once excited, if its pair is sufficiently distant, energy is transferred to XL-665, which acts as an acceptor, emitting light at a wavelength of 665 nm in response.

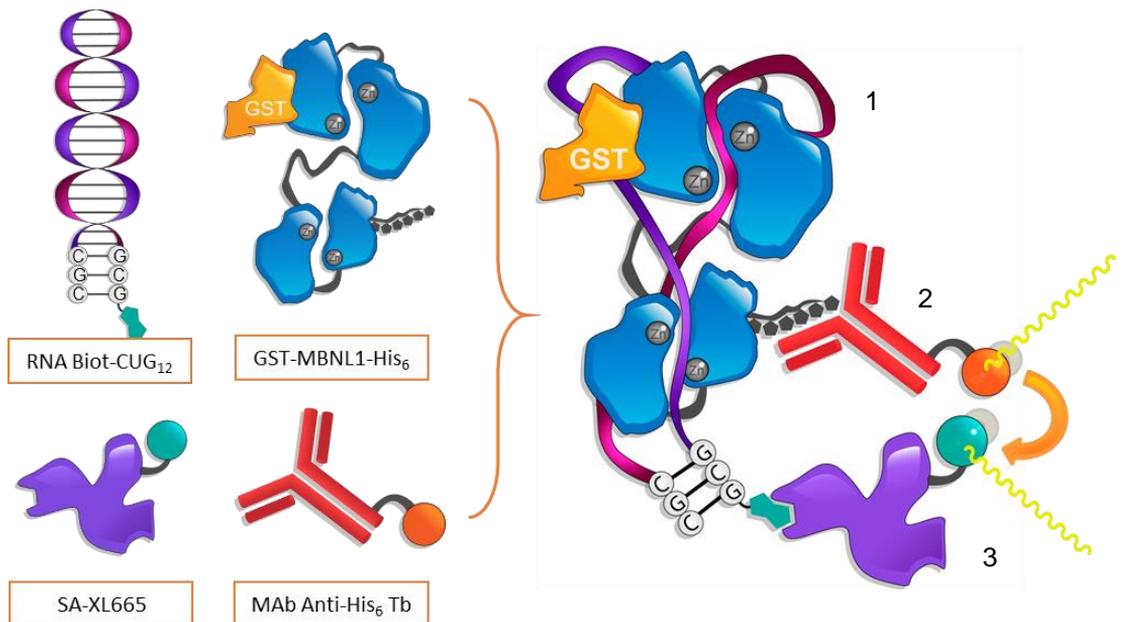


Fig 2.1. Graphical representation of AID 2675 FRET test molecular basis, where: (1) represents the complex formed by 5'-biotinylated RNA oligo Biot-(CUG)₁₂ and the recombinant GST-MBNL1-His₆. (2) FRET donor anti-His₆-Terbium cryptate. (3) FRET acceptor streptavidin-conjugated XL665. While no active compound is present in the reaction media, 340 nm beam light hits Tb conjugate (2), and this is close enough to transfer the energy to XL665 conjugate (3), emitting at 665 nm. If a compound can free the protein, no FRET occurs, and Tb fluorophore emits at 545 nm.

The components included in the AID 2675 test are the following: a short strand of CUG RNA. It consists of 12 repeats, a clip to stabilize the hairpin conformation, and a biotin tag. Although 12 repetitions would not be considered pathogenic in DNA, it has been described that the number of repeats in RNA required to interact with MBNL1 can be reduced to four.¹⁹ Furthermore, the recombinant protein MBNL1- Δ 105-His₆. In addition to a 105 amino acid size that allows enhanced stability under exogenous conditions, the sequence used contains six histidine that perform a dual purpose: on the one hand, they will enable the purification of the protein, and on the other hand, it is used at the time of the test due to the binding of the Anti-His₆ antibody.

The protein is obtained from the PGEX-6P-MBNL1- Δ 105-His plasmid. The pair of FRET fluorophores selected are Tb-cryptate and XL665. Each is conjugated to a factor that allows binding to the corresponding component. The donor (Tb-cryptate) is conjugated to an anti-his

Introduction

antibody, allowing binding to the protein. On the other hand, the FRET acceptor is conjugated to avidin to take advantage of these proteins and biotin interactions. In this case, streptavidin is used.

To sum up, when a candidate able to release the protein from the RNA is introduced into the system, the distance between the two FRET fluorophores is increased, stopping the transfer, and observing the fluorescence of the donor (Anti-His-Tb cryptate). Conversely, when there is no active candidate, the transfer phenomenon causes acceptor fluorescence (SA-XL665).

2.1.3.1 Time-resolved fluorescence energy transfer (HTRF)

The AID 2675 test implemented in the project is based on Time-resolved fluorescence energy transfer (HTRF). The HTRF results from the combination between the quantum phenomenon Förster Resonance Energy Transfer (FRET) and the measurement of time-resolved fluorescence, a technique that suppresses background fluorescence.^{20,21}

FRET is based on the transfer of excitation energy between two fluorophores, one being a donor and an acceptor when they are close enough. When two molecules are sufficiently close, donor excitation, provided by an energy source, triggers non-radiative energy transfer to the acceptor. With a non-irradiated excitation wavelength, this acceptor emits its specific fluorescence at its specific wavelength (less energetic than that of the donor) that allows detecting this donor-acceptor interaction.

As a transfer phenomenon between molecules, the phenomenon has a slight delay compared to the usual fluorescence emission. This delay can be harnessed to avoid the appearance of signals not due to the excitation between donor and acceptor. It is an advantage that gives selectivity to the measurement. For this reason, FRET is combined with time-resolved fluorescence. Therefore, signals from the emissions that are not specific are eliminated, as all these interferences have a much shorter emission time.²⁰

2.2 Objectives

Design and obtain a plasmid that allows the production of MBNL1.

Reproduce and, if necessary, adapt the protocol for the production of MBNL1.

Tune the AID 2675 test to determine the biological activity of the designed candidates.

Obtain a straightforward and reproducible protocol so that others can benefit from this new development in the future.

Results

2.3 Results

The main objective of this chapter is to describe the implementation process of the AID 2675 biochemical test for providing biological activity data. For this purpose, it has been necessary to obtain several components for its implementation.

All the components are commercially available and reasonably affordable except for the CUG oligo-RNA, which can be purchased via custom synthesis, and the recombinant MBNL1 protein. The possibility of acquiring recombinant MBNL1 was studied. However, due to economic limitations, it was finally decided to produce it *ad hoc*. For this purpose, designing and purchasing the plasmid encoding the desired protein was necessary.

2.3.1 Plasmid design

To design the plasmid pGEX-6P-MBNL1- Δ 105-H required for this study, it is necessary to consider the features that the resulting protein must exhibit. First, the MBNL1 protein must be obtained with a glutathione S-transferase (GST) tag at the N-terminal. Besides, it must have a 105 aa deletion at the opposite (C-terminal) end. Finally, at the cut end, it must incorporate a tag of six histidine residues.

Initially, it was attempted to obtain the plasmid from its source, the authors. Despite the refusal, it was decided to proceed with the adaptation of the AID 2675 test. Since it was necessary to start from scratch, the available information was compiled. It was essential to have the genetic information (DNA) encoding the recombinant MBNL1 protein to synthesize the plasmid.

The source sequence of the MBNL1 gene extracted from Genbank (HGNC:6923) contains 5997 base pairs (bp). This sequence includes exons and introns. So, it is necessary to discriminate which segment is coding and verify whether there are intronic regions in the middle of the coding sequence.

Since the isolated coding sequence is not described, it was decided to study the amino acid sequences of the human isoforms of MBNL1. Several amino acid sequences with different lengths were retrieved (392, 388, 382, 377 aa, respectively. Source: Genebank).

Unfortunately, the total length of the MBNL1 sequence to be used in AID 2675 is not reported in the literature,¹⁶ which only specifies the size of the C-terminal cut-off. When that restriction is applied, discrepancies are observed in the last amino acids of the potential sequences.

Since the search via the amino acid sequence seems to be misleading, it was decided to study the possibilities from the DNA sequence itself thoroughly. A partial DNA sequence encoding 340 aa was found in the UniProt database (Q9NR56-7). The UniProt sequence fragment fits into the complete Genebank sequence.

Results

green. Finally, in magenta, the final part of this segment is highlighted, the controversial section where the discrepancies between the different sequences found were located.

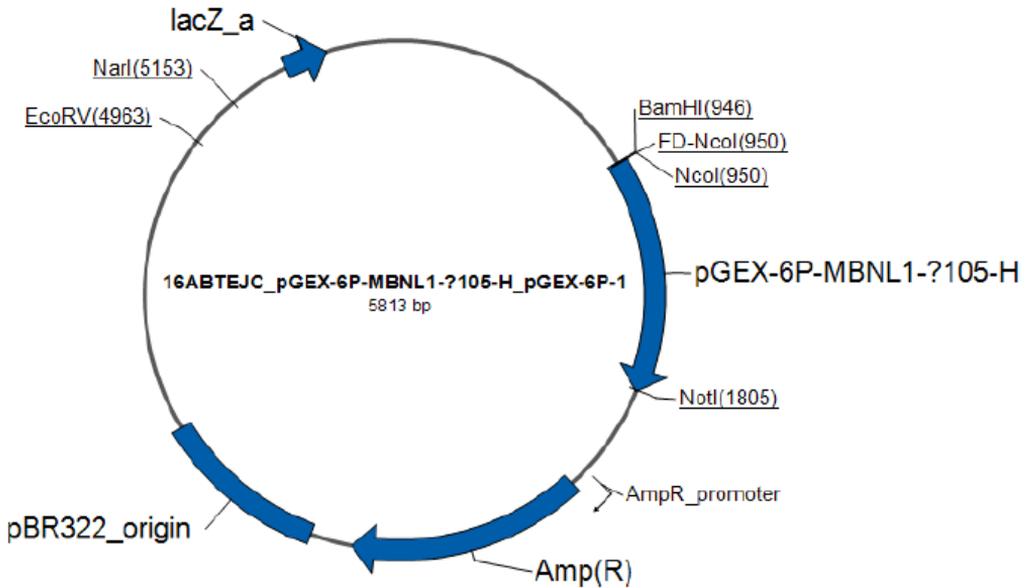


Fig 2.4. pGEX-6P-MBNL1-Δ105-His plasmid map.

The synthetic gene pGEX-6P-MBNL1-Δ105-H (from now on pGEX) was assembled from synthetic oligonucleotides and PCR products by GE-healthcare. The fragment was inserted into pGEX-6P-1. The plasmid DNA was purified from transformed bacteria, and the concentration was determined by UV spectroscopy. The final construct was sequenced to verify the correct insertion. The sequence congruence within the insertion sites was 100%.

2.3.2 Bacteriological culture studies

Different steps and studies are carried out with the final objective of producing recombinant MBNL1 protein and its use in the AID2675 biological test.

The pGEX plasmid is inserted into the competent bacterial strains BL21 and DH5α. Once transformed, the MBNL1 protein is expressed and purified to obtain the recombinant protein, which is quantified and analyzed.

2.3.3 BL21(DE3)pLysS transformation

The bacteria used during all experiments were competent, both BL21 and DH5α. This fact allows the transformation by thermal impact to be faster and more reliable.

Once the strains have been transformed, plasmid extraction is performed to verify the correct transformation. A qualitative result of each strain's transformation with the plasmid is obtained by means of an electrophoresis gel.

2.3.3.1 Bacterial strains transformation

Strains are transformed in duplicate. In BL21, three controls are made: one negative control and two positive controls (one of them with a control plasmid). In the case of DH5 α , two positive controls are made, one of them with a control plasmid. The two strains are exposed to heat shock, and the plasmid is inserted inside the bacteria. (Fig 3.5 and Fig 3.6).

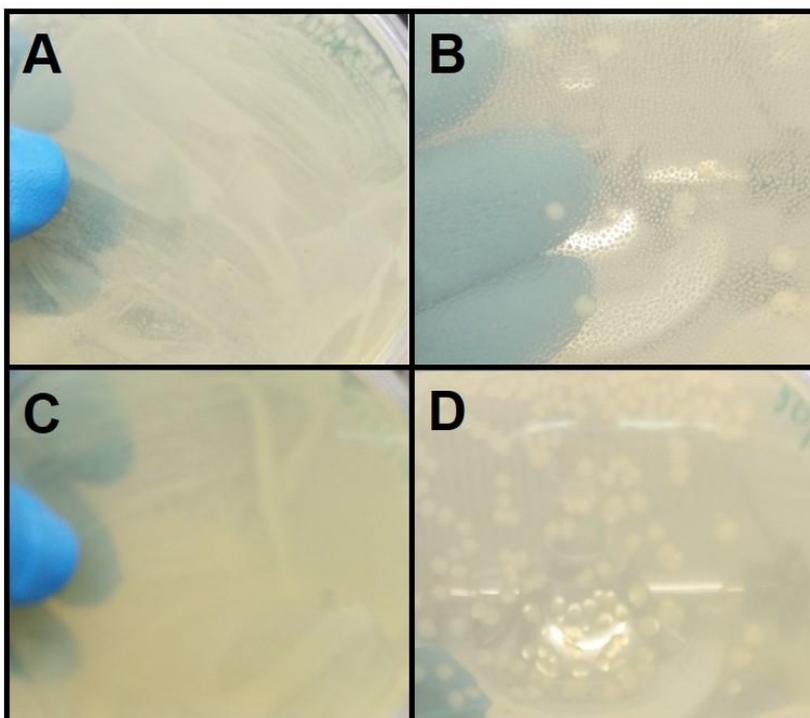


Fig 2.5. Growth of the BL21(DE3)pLysS strain under different conditions on agar media. In all conditions, the bacteria were transformed with the pGEX plasmid except in condition B. A: negative control, growth with agar medium without antibiotic; B: positive control with pUC19 as a plasmid and agar medium with Ampicillin as an antibiotic; C: positive control with agar medium and chloramphenicol as an antibiotic; D: BL21(DE3)pLysS objective with agar medium and Ampicillin and chloramphenicol as antibiotics. It can be seen the differential growth in the different plates.

The BL21 negative control (Fig 3.5.A) contains no antibiotic in the culture media. A dense growth of bacteria is observed, resulting in the plate being 90-95% confluence. Since there is no antibiotic selection, all types of bacteria grow on this plaque. Therefore, it is possible to confirm that no bacteriophage contamination occurred, which is essential in this chapter.

The first positive control of BL21 (Fig 3.5.B) is performed with the bacteria transformed with the pUC19 plasmid and Ampicillin added to the culture medium. Since pUC19 contains the

Results

ampicillin resistance gene (Amp^R), only those bacteria correctly transformed with the plasmid will grow in this control. Thus, it is used to check the efficiency of the transformation.

The second positive control by BL21 (Fig 3.5.C) contains chloramphenicol as an additive to the culture medium. This leads to the selection of pLysS. The pLysS plasmid is already found in BL21 bacteria and produces the T7 lysozyme.

Finally, target bacteria transformed with the pGEX plasmid (Fig 3.5.D) are seeded on a medium containing chloramphenicol and Ampicillin. Thus, bacteria able to grow are those incorporating the plasmid since it includes the Amp^R gene.

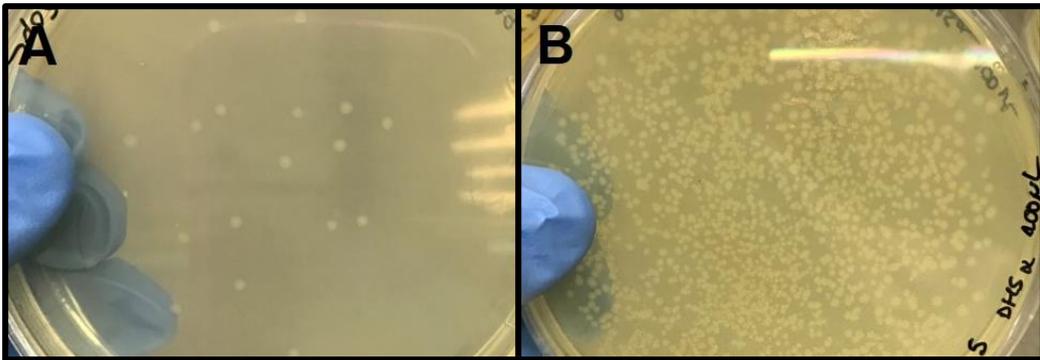


Fig 2.6. Growth of DH5 α strain under different agar media conditions. A: positive control with pUC19 as a plasmid and agar medium with Ampicillin as an antibiotic; B: BL21(DE3)pLysS target with pGEX plasmid and agar medium with Ampicillin as an antibiotic.

Analogously, the positive control of DH5 α strain (Fig 3.6.A) is transformed with plasmid pUC19 containing Ampicillin as an additive to the culture.

The plaque containing the DH5 α target bacteria (Fig 3.6.B) is carried out with the plasmid pGEX. Therefore, since it includes the Amp^R gene, the colonies able to grow have incorporated the exogenous plasmid into their genetic material.

2.3.3.2 Verifying transformation

Plasmid extraction is performed with the Gen Elute™ Plasmid Miniprep Kit (Sigma-Aldrich). This Miniprep kit contains all the plasmid extraction solutions (Resuspension Solution, Lysis Solution, Neutralization Solution, Optional Wash, Wash column, Wash Solution, and Elution Solution).

Since plasmids are circular nucleotide sequences, a phosphodiester bond cleavage is required to isolate this plasmid from the rest of the genomic DNA. For this reason, a Resuspension Solution containing RNase A, an endoribonuclease that hydrolyzes RNA 3' of pyrimidine residues explicitly, is added.

Cell lysis is necessary to isolate plasmid DNA. For this reason, the Lysis Solution contains sodium dodecyl sulfate (SDS) and sodium hydroxide. As a detergent, SDS dissolves the phospholipid membranes of the bacterial cell wall. Sodium hydroxide increases the pH and denatures bacterial cellular proteins, single-stranded DNA, and high-molecular-weight RNA. As a result of this denaturation, circular plasmid DNA is obtained since it remains in its super-coiled form in solution without interacting with the denatured molecules.

SDS, proteins, and lipids must be precipitated to separate the denatured molecules from the mixture. When denatured, chromosomal DNA binds to different sites on cell membrane proteins and precipitates upon centrifugation of the sample. To achieve this precipitation, add a Neutralization Solution (containing potassium acetate) to neutralize the pH.

Keeping in mind that RNase A has been added to the first solution, an Optional Wash must be performed to avoid this nuclease's presence in the final product. Thus, with a Wash Column and Wash Solution, it is possible to wash the solution of possible cellular residues that might have remained.

To isolate pure DNA after washing the column, Elution Solution is used, containing 10 mL Tris and 1 mL EDTA at pH 8, facilitating the complexation of Mg^{2+} cations.

The plasmid's molecular weight obtained in the transformation of the two strains of *E. coli* is determined to verify its proper transformation. An agarose gel electrophoresis is performed, a technique used to identify DNA or RNA samples.

With the aid of a DNA marker, the extracted plasmid's weight is determined using the Miniprep samples.

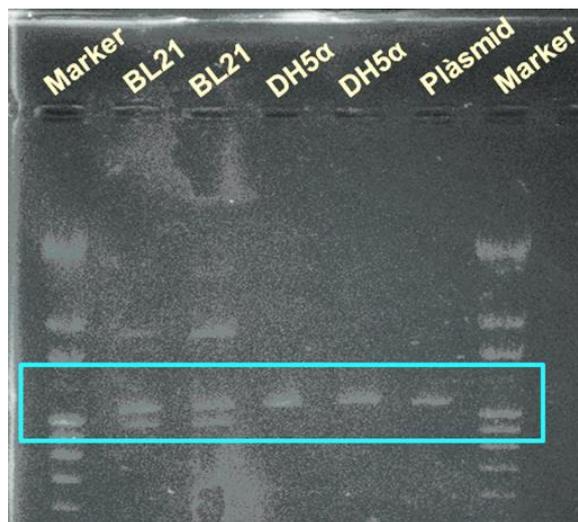


Fig 2.7. Agarose gel of the pGEX plasmid run using the GeneElute™ Plasmid Miniprep Kit of BL21 and DH5α, both duplicated. Each sample was run on a 0.8% agarose gel and allowed to run for 1 h 15 min at 100 mV.

Results

As a result of the electrophoresis gel (Fig 3.7), it is possible to observe that the two bacterial strains have been correctly transformed. Indeed, comparing the DNA marker IV's molecular weight with the molecular weight of the samples, the band corresponding to 1134 b is obtained in all the samples (the duplicates of BL21 and DH5 α).

DH5 α strain shows a single band at 1134 b corresponding to the linear form of the plasmid DNA, while BL21 shows the three conformations that the plasmid can present: the relaxed circular form (2829 b), the linear form (1134 b), and the supercoiled form (970 b) Furthermore, it can also be observed the presence of a band at 4886 b corresponds to the pLysS plasmid already found in the BL21. The fact that DH5 α shows a single band confirms that it is a strain with a storage profile. Indeed, it is selected for its low mutation rate. On the other hand, the different bands that originated in BL21 indicate the genomic DNA's capacity to interact with exogenous DNA. Therefore, this strain has been selected for plasmid expression and production of MBNL1.

2.3.3.3 Induction studies

The concentrations of IPTG that are suitable to produce MBNL1 protein are studied.

This study is carried out because an excess of the inducer IPTG, being a strong promoter, can cause an excess of protein production and lead to the generation of inclusion bodies that can be toxic for *E. coli*.

E. coli batches are seeded with different conditions, and growth is monitored (optical density, OD₆₀₀, is measured).

Once all the data have been obtained, they are plotted (Fig 2.8). It is observed that IPTG causes a slowdown in the growth of BL21. Therefore, the reduction in growth is an effect of the recombinant protein production competing with bacterial proliferation.

IPTG is an inducer that releases the LacI repressor, which is a transcriptional regulation factor of the MBNL1 gene. Thus, IPTG activates the synthesis of the gene of interest.

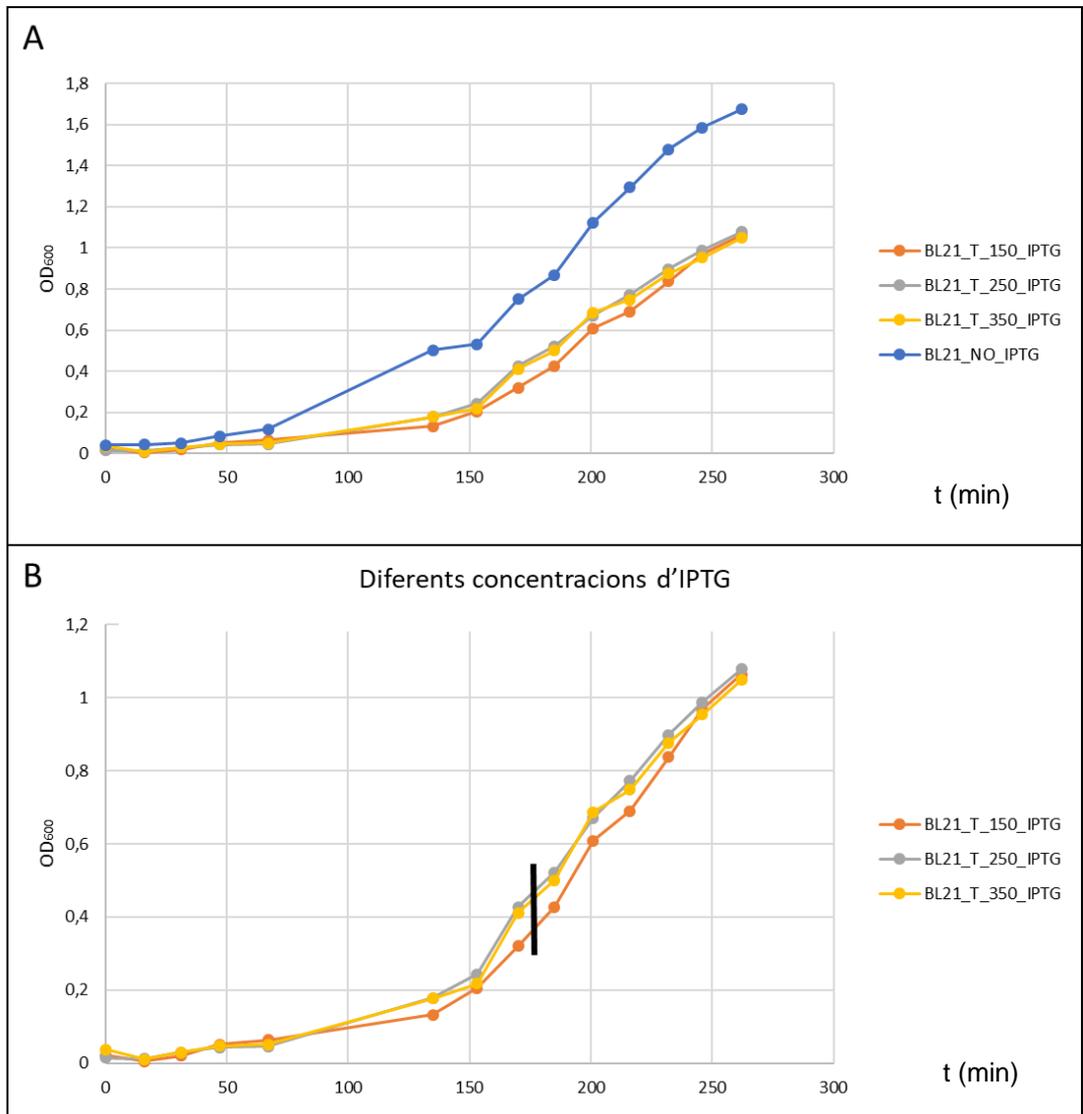


Fig 2.8. Data representation for different growth conditions of *E. coli* bacteria. A) Comparison between the growth curve without IPTG induction and the growth curves with different IPTG concentrations (0, 150, 250 and 350 μM). B) Representation of the different IPTG concentrations (150, 250 and 350 μM). The point at which IPTG was added is marked in black.

It is observed that the concentration that allows a higher production of the strain is 250 μM IPTG, although the differences are minor.

The growth rate without IPTG and 250 μM IPTG is also determined, with more significant differences being observed. The OD_{600} was measured during the exponential phase, and with this, it is possible to calculate the specific growth rate (μ , h^{-1}), i.e. the growth rate per biomass unit. Thereafter, the generation time (g) is calculated, defined as the time it takes for the population to be doubled. To obtain afterward the growth rate (K), which shows the number of generations per hour. Finally, it is determined that the *E. coli* strain used without IPTG

Results

induction has a growth rate of 1.23 h^{-1} and, if we induce BL21 with $250 \mu\text{M}$ IPTG, the growth rate is reduced to 1 h^{-1} .

2.3.4 Recombinant protein production

Once it is confirmed that the transformation is correctly performed, the recombinant protein expression is carried out, reproducing the conditions described by Chen et al.¹⁶ It was necessary to introduce some initial methodology changes after identifying some issues described hereafter.



Fig 2.9. Initial appearance of the culture media after inoculation (left) and appearance after 6 h (right).

The salt concentration initially established was 150 mM NaCl in the buffers and was increased to 300 mM to facilitate cell lysis. By increasing the osmolarity, a variation in osmotic pressure was achieved, facilitating cell lysis.

The use of both phenylmethylsulfonyl fluoride (PMSF) and 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF) as protease inhibitors was studied, and no differences were observed between them, and both are suitable for MBNL1 production.

In the described methodology, many of the protein isolation steps are performed at low temperatures, indicating that the recombinant target protein presents stability issues.

It has to be bear in mind that working with unstable protein requires performing all the steps relatively fast. An increase in temperature may cause protein aggregation. For this reason, work is performed at $4 \text{ }^\circ\text{C}$.

For efficient lysis, the incubation time of the lysozyme is a crucial factor. It is therefore left to act for at least 30 minutes. It is performed centrifugation after lysis to remove the cell debris.

After the sonication process, there is still membrane debris, and this is removed by another centrifugation (10000 xg for 30 minutes).

The control gels show no signs of proteolysis; therefore, the protease inhibitors are no longer added after purification.

First, a volume of 2.5 ml (not 5 ml as stated in literature) of Ni-agarose resin is used since, being a soluble protein, it binds a significant amount of protein per milliliter of resin. Both the incubation and elution of the affinity column are performed at 4 °C.

A purification study is performed with a second affinity column. It is performed because impurities are detected in the acrylamide gel after the first purification. A Glutathione Sepharose, 4B column, is used since the protein contains the GST Tag at the N-terminal.

As a result, the recombinant MBNL1 protein is obtained in high purity. Finally, the sample is eluted through a PD-10 exclusion column to separate the protein from the elution buffer components used in the purification steps and obtain it in conditions to be used in the AID 2675 test.

Once the isolation process is completed, its concentration is determined by absorbance at 254 nm, and the characteristic extinction coefficient of MBNL1-105His calculates the resulting amount of protein.²³ Once the samples are purified, adding glycerol up to 50% is decided to preserve the protein at -80 °C.

2.3.4.1 SDS-PAGE check

The method used to verify protein purification is an acrylamide gel (SDS-PAGE). The cross-linking state of the acrylamide determines the mobility in the gel. Thus, due to the construct having a low molecular weight of approximately 58 kDa (sequence), an acrylamide percentage of 10% is used. With the Precision Plus Protein Standards Dual Color molecular weight marker (Bio-Rad), the different samples' weight is determined, and it is observed whether the different aliquots contain the protein of interest.

In the different lanes of the gel, it can be observed the different steps performed in one of the protein purification tests. Finally, it is observed how comparing to the marker (M), MBNL1 would be about 58 kDa of molecular weight as we expected to obtain based on what was calculated by its sequence.

Results

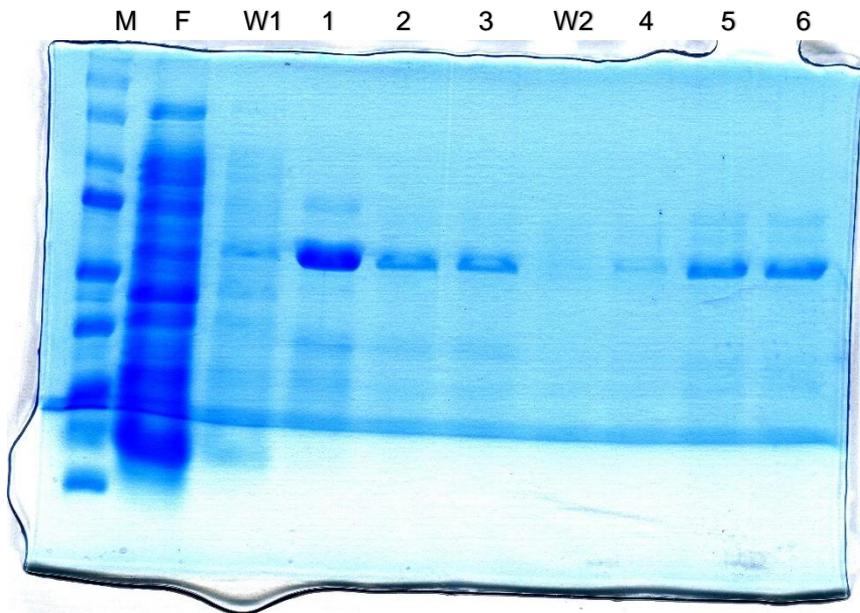


Fig 2.10. Acrylamide gel (SDS-PAGE) with Coomassie stain of aliquots taken during protein expression. M) Precision Plus Protein Standards Dual Color F: Flow-through. W1) Wash Ni column. 1, 2 and 3) Ni Elutes. W2) Wash PD-10 column. 4, 5 and 6) PD-10 Elutes.

The weight showed higher than that found in the literature references for MBNL1 wildtype (UniProt: Q9NR56-1). This is due to the protein having a different number of amino acids, with a 105 amino acid cleavage, the GST tag, and six histidine, and, for this reason, its final weight has been increased.

As expected, proteins with different molecular weights can be found in the flow-through since the first step from bacterial lysis is traced and no biomolecule has been previously removed. The rest of the lanes show the protein of interest in different concentration ranges. In this case, the purification test depicted in the figure only represents a two-step purification using a Ni-Agarose affinity column and subsequent purification by PD-10.

With these results, it is concluded that the protein needs an extra further purification or a different affinity column to obtain it in the required purity. As observed in the lanes after the molecular exclusion column, bands not corresponding to MNBL1 are present.

2.3.4.2 MBNL1 quantification

A UV-Visible light spectrophotometer (Figure 2.11) is used to quantify the sample concentration obtained from the purification. Knowing our protein's extinction coefficient and the spectrophotometer's absorbance, it is possible to determine the protein amount.

It is decided to use this technique for the reported protein instability. Furthermore, this technique allows fast and non-destructive measurements.

To know the extinction coefficient of MBNL1, we use the ExPASy ProtParam Tool,²³ which allows us to calculate the extinction coefficient from the protein sequence.

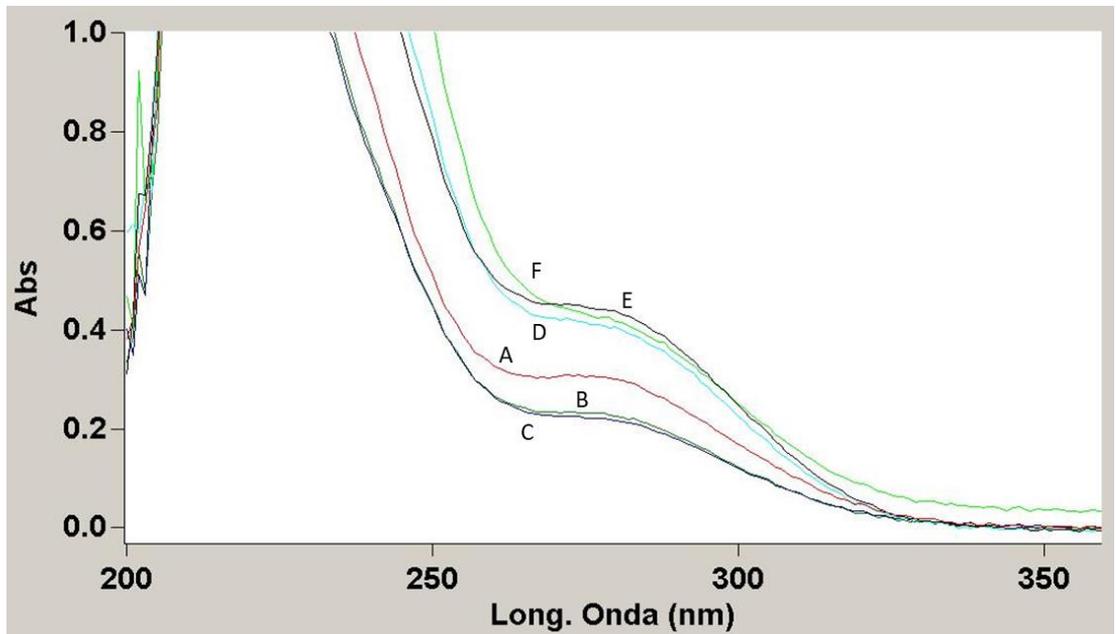


Fig 2.11. Overlay of absorption spectra of the different eluates of the PD-10 desalting column.

Using the Lamber-Beer law, the concentration of MBNL1 in the different samples is obtained by measuring its absorbance at 280 nm. As shown in Fig2.11, the curve shows a maximum at the specific wavelength of aromatic amino acid absorbance, although GST-MBNL1- Δ 105-H does not have a high number of these.

2.3.5 Troubleshooting

During the production process of MBNL1, some problematic points were detected for developing the test and obtaining the protein at the desired purity degree.

2.3.5.1 Ultracentrifugation.

First, a progressive decrease in protein quantifications overtime was initially observed. This was originally assigned to an aggregation problem caused by some deficiency in handling the protein or exposing it to freeze-thawing processes. Thereafter, it was observed that the protein aggregated as soon as it was purified, although carrying out the process at 4 °C.

Results



Fig 2.12. Aggregated MBNL1 at the bottom of the Eppendorf tube.

After several experiments, it was decided to incorporate an extra step before the use of the protein in order to reduce the amount of aggregated protein produced in the test. Since it is unknown whether the aggregated protein would be renatured under test conditions, it was decided to avoid using samples in this state to prevent deviations in the amount of protein, given the importance of stoichiometry in AID 2675.

The stage introduced is ultracentrifugation. This step is incorporated together with a quantification of the protein prior to any use of MBNL1. An approximate 10% reduction of protein per day and use is observed.

2.3.5.2 Cobalt affinity resin

One of the observed and modified aspects is the type of affinity columns used in the purification. Initially, the procedure described is performed, consisting of a nickel affinity column followed by a size exclusion column.

As can be observed in previous sections (Figure 2.10), additional bands besides MBNL1 are observed in the gel. Initially, it was decided to add an extra purification step, using an affinity column exploiting the GST tag incorporated in the protein used. On the other hand, it was decided to test using a TALON® Metal Affinity Resin column (Takara Bio, Kusatsu, JP). This column harnesses the same histidine tag affinity phenomenon as the nickel analog but with two differences: it has a higher selectivity and a lower binding capacity (g protein/mL resin).

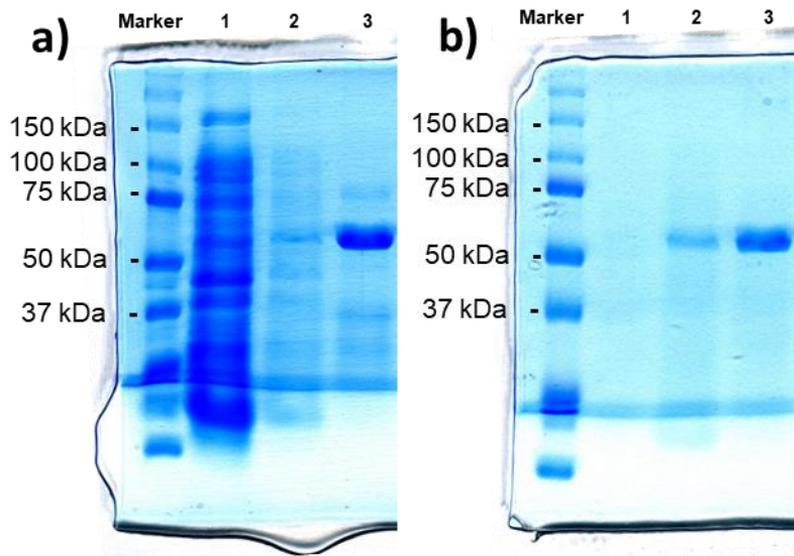


Fig 2.13. Comparison of acrylamide gels of His-affinity purification using Ni (a) and Co (b) resins.

As can be observed in the previous figure, the cobalt column's use eliminates the number of bands to one, MBNL1. Finally, given the results obtained with this column, it was decided to modify the protein purification protocol.

2.3.5.3 Avoiding bacteriophages

A critical factor in this project was the outbreak of a bacteriophage infection in the laboratory, where the production studies were initiated. Even though there is no risk in the described process, the outbreak forced us to move the protein production to another laboratory outside the university. Finally, the study was completed at the Biophysics Unit of the Universitat Autònoma de Barcelona under the supervision of Dr. Àlex Peràlvarez-Marín.

2.3.5.4 RNase

Another critical factor, which may undoubtedly impact the test's performance, is the degradation of the oligo-RNA used in this project. Since it is a short synthetic RNA sequence, it is likely to be extremely sensitive to the presence of RNase in the medium.

To avoid possible degradation, extreme care is taken. For this purpose, RNase-free reagents are used. Water is treated with diethylpyrocarbonate (DEPC), added at 0.1%, and autoclaved before use in buffers and test material rinsing. The fume hood and labware are treated with an RNase removal agent. Finally, the well plates used in the test are submerged in 3% hydrogen peroxide solution for a minimum of 16 h before their use.

Results

2.3.6 Tuning of test conditions

The test conditions were slightly adapted since neither the instrument used nor the intention was the same as the original bioassay since we did not intend to apply it for high-throughput screening.

Therefore, the first modification on the described procedures is focused on increasing the reaction size, thus improving the measurements' precision by increasing its signal. It has been changed from 384 or 1536 well-plates to 96 well-plates.

Another way to increase the detected FRET signal is by modifying the delay times and the integration window. The count delay is the factor that provides selectivity to the measurement since it is the period between the irradiation and the fluorescence measurement. As already described, FRET consists of an energetic transmission between two molecules. Therefore, it has a significantly longer emission time than a single molecule fluorescence relaxation.

Tab 2.1. Reading conditions of the TR-fluorescence comparison.

	count delay / μs	count window / μs
Described	70	400
This work	30	1500

As can be observed in Tab 2.1, the delay time used is significantly lower than described. It could be established by successive readings of the same plate that this value was appropriate in the working conditions to obtain a good signal-to-noise ratio. Likewise, the count window increase also seeks to maximize the accumulation of signals in the detector. The value selected for the count window is the maximum value allowed by the instrument used.

2.4 Chapter Discussion

Implementing AID 2675 would allow a primary screening of the synthesized candidates with a low time cost before assessing their *in vivo* activity.

The availability of this tool allows the design of new candidate families, both within the current thesis and afterward. An important consideration is that the AID 2675 test does not determine whether the active compounds affect the RNA or the protein. However, the ideal situation would be to target the pathogenic RNA, and the protein would maintain its affinity and splicing pattern. Thus, the assay would allow identifying false positives previous to any *in vivo* assay. However, active AID 2675 candidates will need further biologic studies to confirm their activity.

Regarding the purification and handling of GST-MBNL1- Δ 105-H, it has been possible to obtain it and provide knowledge and infrastructure to continue producing the protein whenever necessary. Some modifications, such as the purification column replacement, have shown to be a possible improvement over the purities described (~60%). Since the cobalt column's purification is sufficient, it was finally decided not to incorporate the GST column. Even though producing a similar purity product, the GST column prolonged the purification process and reduced the purification yield. Besides, in that experiment, an increase in the aggregation phenomenon was detected.

One possibility to improve the loss of MBNL1 by aggregation could be to perform a study of storage buffers that stabilize this protein. Thus, the protein should be eluted by the PD-10 column before each use to exchange the buffer and quantify it.

Other interesting studies to be carried out could be incorporating other proteins of the same family, such as MBNL2 and MBNL3, whose effect on DM1 has been described.²⁴

2.4.1 Final standard procedure

- A single colony of the transfected BL21 is picked into 10 mL of LB with 100 µg/mL ampicillin and grown for 16 hours at 37 °C and 200 rpm.
- 1 mL of the resulting culture mixture is inoculated to 1 L of LB with 100 µg/mL ampicillin and incubated at 37 °C and 150 rpm in a shaking incubator until OD_{600nm} reached 0.5-0.6.
- Induction with 500 µM IPTG for 3 hours at 30 °C.
- Centrifuge at 4000 xg and 4 °C for 30 minutes.
- The pellet is resuspended in 50 mL of cell lysis buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.00), 5 mM imidazole, 0.1 % IGEPAL CA-630, 5 % glycerol and freshly added: 0.1 M PMSF, 5 mM benzamidine, 1 EDTA-free complete protease inhibitors tablet, 1 mM DNase and 1 mg/mL lysozyme).
- Lysis is achieved by alternating 15 cycles of sonication (1 min sonication – 1 min icing).
- Centrifuge at 10000 xg and 4 °C for 50 min.
- The supernatant is added to 2 mL of a previously equilibrated TALON® Metal Affinity Resin (Takara Bio, Kusatsu, JP) and incubated at 4 °C for 1 hour.
- wash 3x with 10 column volumes (30 mL) of washing buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.00), 25 mM imidazole, 0.1% IGEPAL CA-630, 5% glycerol).
- The protein is eluted with 5mL of elution buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.00), 250 mM imidazole, 0.1% IGEPAL CA-630, 5% glycerol).
- Eluates are added to PD-10 Desalting Columns (GE Healthcare, Chicago, IL) that is previously preequilibrated using fluorescence test buffer (25 mM HEPES pH 7.4, 110 mM KCL, 10 mM NaCl, 1 mM MgCl₂, 15 µM ZnCl₂, 0.02%Tween-20).
- Quantify GST-MBNL1-His6 via direct absorbance spectroscopy at 280 nm using the ExPASy Protparam (protein parameter) tool to assess extinction coefficient (57300 M⁻¹·cm⁻¹).

- Prior to each use, samples need to be ultracentrifuged at 150000 xg, 4 °C for 1 h and quantified to avoid aggregation issues.
- GST-MBNL1-His6 samples must be stored at -80 °C with 50% glycerol.

References

2.5 References

1. Mankodi, A. *et al.* Myotonic Dystrophy in Transgenic Mice Expressing an Expanded CUG Repeat. *Science* (80-.). **289**, 1769–1772 (2000).
2. Gomes-Pereira, M. *et al.* CTG trinucleotide repeat 'big jumps': Large expansions, small mice. *PLoS Genet.* **3**, 0488–0491 (2007).
3. Seznec, H. *et al.* Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum. Mol. Genet.* **9**, 1185–1194 (2000).
4. Wang, G. S., Kearney, D. L., De Biasi, M., Taffet, G. & Cooper, T. A. Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J. Clin. Invest.* **117**, 2802–2811 (2007).
5. Mahadevan, M. S. *et al.* Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nat. Genet.* **38**, 1066–1070 (2006).
6. Garcia-Lopez, A. *et al.* Genetic and chemical modifiers of a CUG toxicity model in *Drosophila*. *PLoS One* **3**, (2008).
7. deLorimier, E. *et al.* Modifications to toxic CUG RNAs induce structural stability, rescue mis-splicing in a myotonic dystrophy cell model and reduce toxicity in a myotonic dystrophy zebrafish model. *Nucleic Acids Res.* **42**, 12768–12778 (2014).
8. Wang, L. C. *et al.* Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in *Caenorhabditis elegans*. *Cell. Mol. Life Sci.* **68**, 1255–1267 (2011).
9. Provenzano, C. *et al.* CRISPR/Cas9-Mediated Deletion of CTG Expansions Recovers Normal Phenotype in Myogenic Cells Derived from Myotonic Dystrophy 1 Patients. *Mol. Ther. - Nucleic Acids* **9**, 337–348 (2017).
10. Arandel, L. *et al.* Immortalized human myotonic dystrophy muscle cell lines to assess therapeutic compounds. *DMM Dis. Model. Mech.* **10**, 487–497 (2017).
11. Kalra, S., Montanaro, F. & Denning, C. Can Human Pluripotent Stem Cell-Derived Cardiomyocytes Advance Understanding of Muscular Dystrophies? *J. Neuromuscul. Dis.* **3**, 309–332 (2016).
12. Han, S. S. W., Williams, L. A. & Eggan, K. C. Constructing and Deconstructing Stem Cell Models of Neurological Disease. *Neuron* **70**, 626–644 (2011).
13. Konieczny, P. *et al.* Myotonic dystrophy: candidate small molecule therapeutics. *Drug Discov. Today* **22**, 1740–1748 (2017).
14. Wong, C. H. *et al.* Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. *J. Am. Chem. Soc.* **136**, 6355–6361 (2014).
15. Li, J. *et al.* A Dimeric 2,9-Diamino-1,10-phenanthroline Derivative Improves Alternative Splicing in Myotonic Dystrophy Type 1 Cell and Mouse Models. *Chem. - A Eur. J.* **24**, 18115–18122 (2018).
16. Chen, C. Z. *et al.* Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* **402**, 1889–1898 (2012).

17. Warf, M. B. & Berglund, J. A. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *Rna* **13**, 2238–2251 (2007).
18. AID 2675 - qHTS Assay for Inhibitors of MBNL1-poly(CUG) RNA binding - PubChem.
19. Warf, M. B. & Berglund, J. A. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA Soc.* **13**, 2238–2251 (2007).
20. Jares-erijman, E. A. & Jovin, T. M. FRET imaging. *Nat. Biotechnol.* **21**, 1387–1395 (2003).
21. Piston, D. W. & Kremers, G. Fluorescent protein FRET : the good , the bad and the ugly. **32**, (2007).
22. Yuan, Y. *et al.* Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **35**, 5474–5486 (2007).
23. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, 597–603 (2012).
24. Fardaei, M. *et al.* Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* **11**, 805–14 (2002).

References

3 Groove Binders

One of the most studied therapeutic pathways among DM1 studies is using small molecules to stabilize the characteristic toxic transcripts, thus enabling the release of MBNL1. In this chapter, the design of a new chemical library of candidates with the potential ability to fit the dsRNA's minor groove is described. Furthermore, the synthesis of the selected compounds is addressed, and finally, their biological potency is determined both *in vitro* and *in vivo*.

3.1 Introduction

3.1.1 Overview

The discovery of the CUG secondary structure¹ made such a revolution in the basic science field of DM1 as it gave the first sight of a possible therapeutic target, which was a very early disease characterization step at that moment. Nowadays, one of the most recognizable and reproduced strategies to attain DM1 candidates are groove binding on those hairpin-like CUG secondary structure.^{2,3}

RNA stabilization and, therefore, normalization of protein levels is one of the main hallmarks of DM1; the idea of externally "clamping" RNA dual-stranded conformations was adopted by many of the main groups developing novel structures and rational design approaches, including ours. There are few scaffolds still being investigated by notorious scientific groups that are based on this strategy.⁴⁻⁶

This strategy parts of two assumptions extracted from previous studies: CUG tends to adopt double-stranded secondary structures ended in hairpin loops and the fact that by stabilizing such conformations, the sequestered alternative splicing proteins are relieved.⁷

As an orphan disease, DM1 lacks a vast scientific community publishing novel crucial data frequently, and therefore, that first study led to a decade of struggle for a few groups that tried to target RNA and stabilize those hairpins from scratch. Some groups started to test already commercial chemical libraries using high throughput screening (HTS) biological testing,⁸ searching for the first anti-DM1 hits. These studies led to a few hits, yet their action mechanism remained unclear.^{4,9} In some cases, the rationale underlying plenty of active molecules in the DM1 field remains ambiguous or has been described years after being discovered.

Many biological trials have been conducted on candidates (~0.3 M of HTS-tested candidates). (<https://pubchem.ncbi.nlm.nih.gov/bioassay/2675>) Furthermore, significant progress has also been made in the development of the rational characterization of DM1. The disease's pathogenic mechanism and possible therapeutic pathways are partially described. However, several groups have developed candidate design strategies or described much better the rationale underlying their candidates using this information.¹⁰⁻¹²

In this chapter, a specific mechanism of interaction with RNA secondary structure is studied: Groove binding molecular basis and a few remarkable examples will be discussed. Specifically, two of those studies led to a significant number of derivatives being the critical element of probably two of the most successful research groups, at least in terms of publications, in small molecule drug discovery in the field.

3.1.2 The groove bendir approach

Groove binding of small molecules on nucleic acids is a widely described strategy, and research groups have been using it for decades.^{5,13,14} Nevertheless, most of the structures exhibiting this characteristic do target DNA helices. It must be born in mind that DNA mostly adopts a regularly paired double-stranded helical conformation, unlike RNA. For instance, some relevant studies on drug design oriented to RNA used to decline the possibility of using its double-stranded secondary structure as a possible drug target.^{15–18} To understand what impelled those groups to make that statement, readers must understand the helical RNA targeting rationale.¹⁹

3.1.2.1 The grooves

The pivotal concept in this chapter is the groove, the actual receptor for drug candidates. Taking DNA A-helix as an example, two different grooves can be clearly distinguished (Fig 3.1). The chiral nature and the nonlinear geometry of the sugars forming DNA provoke a lack of symmetry between the two sides of the chain. Their properties are tightly related to the base pairs inside of them. They can differ by the composition of the sequence, mispairing issues, proximity to other secondary structure features (bulges, hairpins, pseudoknots, G-quadruplexes, etc.)²⁰ or interaction with exogenous structures (metal ions, metabolites, proteins, or any other structure able to interact with the nucleic acid)²¹. All these features are measured in two big data groups, the sugar puckers,²² and the structural parameters.²³

On one hand, the major groove, deep and wide, exposes the stacked base-pairs of nucleotides. It is usually discarded to groove binding strategy in DNA as it is too wide and most of ligands would rather interact with the nucleobases destabilizing the helix by originating mispairing issues.

On the other hand, the minor groove, narrower and extremely polar due to the presence of the side phosphates, are composed by the deoxyribose units and the side functional groups of the nucleotides. It has been assigned as the actual target for small molecules as most of the intrinsic helical properties can be modulated via interactions in this groove (even the wideness of the major groove).

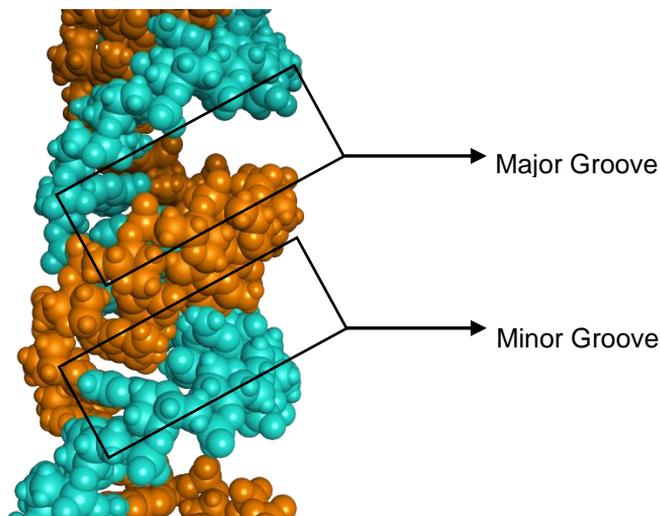


Fig 3.1. Modelled DNA A-helix secondary structure. Major and minor grooves are highlighted. Opposite strands are colored to easily differentiate.

Unfortunately, the picture turns blurrier when RNA enters the equation. Even though RNA exhibit only two structural changes to DNA (the 2'-hydroxy group at the ribose and the methyl converting thymine into uracil) the whole helical structure is deeply modified. Well established comparisons between DNA and RNA helical conformations shed light on the reason why most scientist refused to use dsRNA as a practical receptor. Well-paired helical RNA exhibit much smaller differences between the major and minor grooves.²⁴ Simulated full-paired RNA A-helix show a much deep and narrow major groove while minor groove is depicted as shallow and barely exposing nucleotide functional groups, precluding selective targeting. However, in the specific case of DM1, the target RNA is a regular repetition of a trinucleotide. Therefore, a full-paired dsRNA conformation is impossible to achieve. This characteristic impairment issue is known as a mismatch and makes groove binding a feasible strategy, as this regular mismatch phenomenon turns DM1 dsRNA a groove less structured. Thus, DM1 dsRNA is readily available for external structures to interact with their grooves.^{25,26}

3.1.3 Candidate selection

In this section we analyze the main chemical features that a drug candidate should present to act as a groove binder. First, it is necessary to define by which of the grooves the structure is supposed to interact. In this case, we study those that interact with the minor groove of the RNA. Two structures are taken as a reference as both have been leads of series of candidates reporting high anti-DM1 activity in their multiple and diverse forms: Hoechst derivatives and pentamidine.

Introduction

3.1.3.1 Hoechst

Hoechst 33258 (Fig 3.2) and its derivatives, bioconjugates, and macromolecular inclusions are probably the most described bioactive scaffold in DM1 field. This dye and their family are widely known as they bind tightly and selectively to nucleotides. For this reason, they are commonly used as nuclear stain (Hoechst 33342). What is less known about this structure is its mechanism of interaction with RNA.²⁷ However, they are crystal data of some Hoechst derivatives binding nucleotides secondary structure, specifically, binding grooves in dual-stranded conformations of DNA.

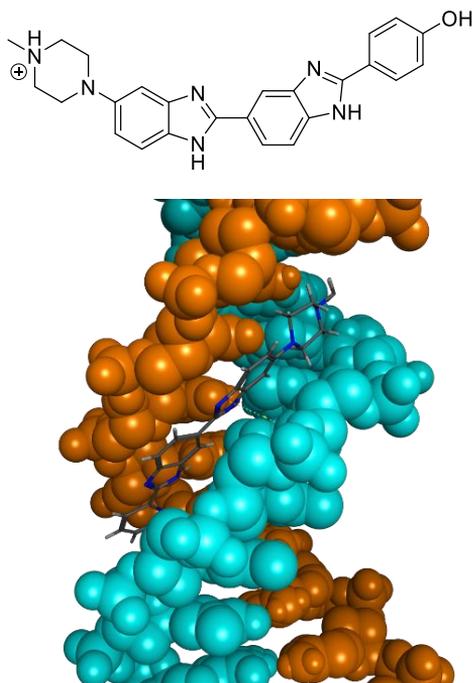


Fig 3.2. Top: Hoechst 33258 structure. Bottom: crystal structure of Hoechst bound to dsDNA. (PDBID: 4Z4B)

Due to the similarity between A-form DNA double helix and the hairpin like CUG RNA stable conformation, DM1 toxic transcripts seem a logic choice to develop new groove binders to inhibit sequestering of splicing proteins in foci formation.

Before starting to perform their studies with DM1 and DM2, Disney *et al.* had already tested the efficacy of Hoechst 33258 in RNA therapy as the treatment of the fungus *Candida albicans*.²⁸ However, the emergence of the CUG new therapeutic target attracted their attention and, using the experience achieved in previous investigations, they contributed to identify new families of ligands and new strategies of inhibition that enriched the knowledge around the DM1. One of its main contributions is the rational design and synthesis of modularly assembled ligands.²⁹ For instance, candidates consisting of a series of Hoechst binding groups with a terminal azide used to anchor it to a core spacer with peptide-like structure. Some determining factors in creating these structures include determining the optimal distance between

binding moieties or the addition of small structural differences to Hoechst, providing specificity against RNA transcripts.²⁸ One exciting feature displayed by Disney *et al.* is the nature of spacers. Spacers are seldomly used in classic drug design approaches. However, attaining selectivity in nucleotide drug design becomes a crucial feature. These attempts to attain candidates and bring chemical diversity to the spacer paved the road to many studies in the latter years, including this thesis. Controversial studies on the nature of the spacers attempt to describe their optimal profile. However, flexible, rigid spacers bound to a suitable binder have been described to have biological activity against DM1.³⁰

3.1.3.2 Pentamidine

On the other hand, one of the earliest hits describing remarkable *in vivo* recovery values of MBNL1 was pentamidine. This compound has been widely studied in the field and related to A-DNA.⁴ For this reason, Berglund *et al.* selected it as a possible candidate due to the significant similarity between the DNA type helix and the desired conformation of CUG. These studies generated a starting point for the rest of the researchers because, for the first time, a compound was capable of both inhibiting transcription and MBNL1 recovery.³¹

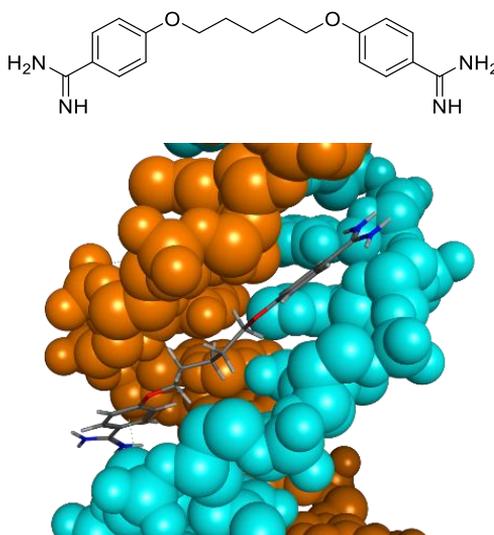


Fig 3.3. Top: Pentamidine structure. Bottom: crystal structure of pentamidine bound to dsDNA (PDBID: 3EY0).³²

One of the main drawbacks of this compound (with antifungal activity) was the narrow margin between activity and toxicity. Although pentamidine mechanism of action is not well-known, there are two consistent hypotheses about its inhibition mechanism. The first, that pentamidine interacts with the DNA bases inhibiting the transcription, and therefore, the toxic RNA is not formed.³³ The second is that it interacts with CUG repetitions, binding the minor groove and stabilizing the structure; thus, the splicing factors such as MBNL1 are released.¹³ However, the actual mechanism is likely to be the addition of both. Despite not having the certainty of how it acts, its activity values drove many groups (among them us) to search for analogs

Introduction

that could present a similar activity without its high toxicity. In subsequent studies, Berglund et al. extended the family of analogous compounds varying the terminal binding groups' spacer length. In these studies (see Fig 3.4), they observed that the longer the spacer, the more active the candidates were. However, heavier candidates also proved to be more insoluble and toxic. Finally, in a recent study, more rigid structures have been studied, giving similar pentamidine activity levels with no toxicity. The new hit, furamidine, has also proven interesting synergies with other drugs to treat DM1.³⁴

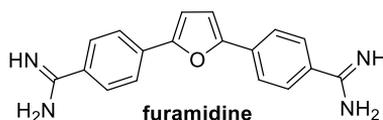


Fig 3.4. Structure of furamidine.³⁴

3.1.3.3 This work

Joining both concepts and harnessing their knowledge, our group started designing possible candidates against DM1 in collaboration with University of Valencia and the Institute Univ. of Science and Technology (IUCT), being endorsed by a project granted by La Marató de TV3 (ref: 100231).

In such framework, Àlex López-González performed his PhD thesis, based on the computational characterization of the disease along with the target identification and first candidate designs.^{35,36} The results obtained by Dr. López impelled us to give a step forward, using all the existing data, and the modelled *in silico* RNA and to attempt full rational design strategies for novel groove binders as long with the synthesis and biological evaluation of them.

3.2 Objectives

- Build a methodology allowing the rational design of novel groove binders and evaluate their possible biological activity.
- Design synthetic pathways for all selected compounds.
- Synthesis of a novel groove binder family with potential activity against DM1.
- Evaluate the activity of obtained compounds in both in vitro and in DM1 disease cellular models.

3.3 Rational design of potential groove binders

This chapter's main objective is to obtain a novel structure with a potential groove binding affinity for CUG hairpins. Therefore, a rational model needs to be built to predict candidates' interaction obtained via combinatorial design and grouped in large virtual chemical libraries. This is not an unexplored path since our group already has experience and knowledge regarding the rational design approach against DM1.³⁶ A systematic procedure composed of a series of stages was developed. The whole process could be described cyclically, and that successive iteration would lead to a list of possible ligands to be synthesized.

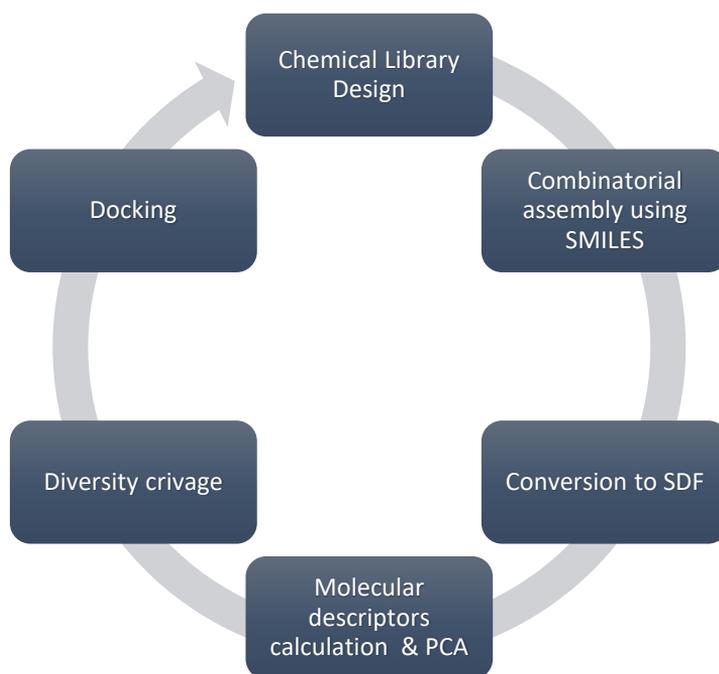


Fig 3.5. Schematic representation of the computational procedure to design the novel groove binder structures.

Getting feedback on each cycle is crucial to improve the candidates in terms of affinity and improve the whole method. Moreover, this itinerary is thought to be used in the design of *de novo* structures never synthesized in our lab, and therefore, synthetic criteria were also included in this process to avoid designing non-feasible structures or too complicated in synthetic terms. The whole study aims to design and synthesize a set of novel structures and see whether they can be set as lead compounds for further generations of compounds.

3.3.1 Assembly of the chemical library

As stated previously, our group has already worked on the design of active small molecules against DM1. In Dr. López PhD thesis³⁷ and the work performed by Miquel Gallart³⁸ during his master thesis, predictive models and structures were generated in analogy to pentamidine

using ligand-based drug design (LBDD). The willingness to explore new possibilities and gain further knowledge on the ligand-RNA interaction impelled us to hop to structure-based drug design for the first time.

The starting point of this chapter is to build a chemical library of novel chemical entities able to bind the minor groove of the RNA. The chemical library must be wide enough to be representative of any possible interaction with the purposed receptor. However, it has to be short enough to enable the use of structure-based methods, which are, in fact, much more demanding in terms of computational resources. As a starting point, the most successful literature candidates were considered.^{4,39,40} Furthermore, the chemical library was enriched with the candidates synthesized by M. Gallart. Those candidates were designed by analogy to pentamidine. Therefore, similar derivatives were included in the first design of the new chemical library.

3.3.1.1 Scaffold definition

Attempting to fit the ligands in the RNA groove secondary structure and the intrinsic complexity of the receptor, all scaffolds must fulfill a list of features, including a slightly curved structure, or a stable conformation with this shape and polar groups at both ends, aiming to interact with the RNA backbone in a selective manner.

The curved structure or conformation allows candidates to fit in the minor groove of DM1 toxic RNA. This feature can be achieved via several approaches. It has been described as a relationship between ligand rigidity and the ability to reach minor groove.³⁰ Furthermore, polar groups in the central part of the molecule can interact with the RNA backbone, stabilizing the curved conformation in flexible structures,⁴¹ even aliphatic chains can adopt such conformation if the candidate has a selective and potent binding group at both ends of it.⁴² A graphic representation of this concept is exhibited in the latter figure.



Fig.3.6. Designed groove binder scaffold graphical representation.

In most cases, the structure consists of a central spacer that can present greater or lesser flexibility and can or not incorporate polar groups capable of interacting with the base pairs of the hairpin. At the ends, we can find symmetrical terminal groups, usually aromatic rings replaced by functional groups capable of acting as electron acceptors or donors of the hydrogen bond.

Rational design of potential groove binders

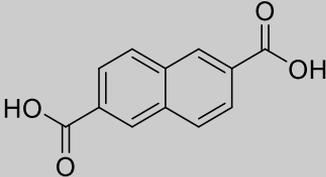
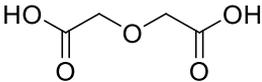
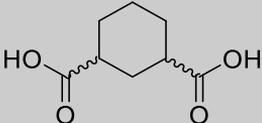
Once this structural peculiarity has been recognized, reagents available in the laboratory or commercially available were collected, which could be used as precursors of both spacers and terminal groups. To obtain this kind of compound, we decided that all spacers had to be dicarboxylic acids (or even better acid chlorides), and the terminal groups had to have an alcohol or terminal amine group that allowed the acyclic substitution coupling (forming ester or amide bonds). Consequently, we prioritized the compounds with terminal amine due to the possible hydrolysis of the ester link.

3.3.1.2 Linker selection

A selection of compounds consisting of a series of laboratory-available or commercially affordable carboxylic acids led to the first chemical library.

Tab 3.1. Dicarboxylic acids were selected to be the structural spacers in the chemical library. *) SBDD considers diastereomers, so they are represented separately.

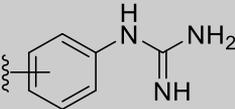
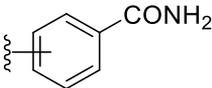
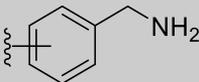
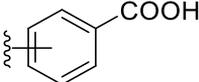
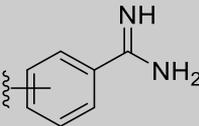
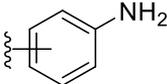
Spacer structure	Code
 <chem>OC(=O)CC(=O)O</chem>	3.1
 <chem>OC(=O)CCCC(=O)O</chem>	3.2
 <chem>OC(=O)CCCCCC(=O)O</chem>	3.3
 <chem>OC(=O)CCCCCCC(=O)O</chem>	3.4
 <chem>OC(=O)c1ccc(cc1)C(=O)O</chem>	3.5
 <chem>OC(=O)c1ccc(cc1)Cc2ccc(cc2)C(=O)O</chem>	3.6
 <chem>OC(=O)COCCOCCOCCOCCOCC(=O)O</chem>	3.7
 <chem>OC(=O)c1ccccc1C(=O)O</chem>	3.8

	3.9
	3.10
	3.11*
	3.12*

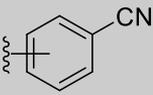
3.3.1.3 Binding groups selection

If we take a closer look to the pentamidine structural profile, two benzamidine groups are placed at both ends of the structure. We chose to keep that feature, assuming they would improve the affinity for RNA due to their positive charge density. Therefore, aryl systems were selected. All these systems are commercially available or can be synthesized from others with a reasonable number of synthetic steps (i.e. guanidinophenyl can be synthesized in a single step from the corresponding aniline). The chosen groups are displayed in the table below.

Tab 3.2. Arylic groups were selected to act as binding groups of novel bivalent candidates.

Ligand structure	Regio	Code
	<i>m</i> -	{1.1}
	<i>p</i> -	{1.2}
	<i>m</i> -	{2.1}
	<i>p</i> -	{2.2}
	<i>m</i> -	{3.1}
	<i>p</i> -	{3.2}
	<i>m</i> -	{4.1}
	<i>p</i> -	{4.2}
	<i>m</i> -	{5.1}
	<i>p</i> -	{5.2}
	<i>m</i> -	{6.1}
	<i>p</i> -	{6.2}

Rational design of potential groove binders

	<i>m</i> -	{7.1}
	<i>p</i> -	{7.2}

Once defined the chemical space of this study, it was time to start the computational process that would lead to the definitive list of candidates to synthesize.

3.3.2 Combinatorial library enumeration

The first step to select the candidates was to build the chemical library. As can be seen above (Tab.3.1 & 3.2), data is segmented in fragmented libraries. As shown in Fig 3.6, candidates are formed by two terminal groups and a binding functional group separated by a central spacer. Therefore, it was necessary to assemble the fragments before the realization of any prediction of their properties. Library enumeration was conducted using Smlib v 2.0, a Java-based open source software developed by A. Schüller.⁴³ However, before using this software, it was necessary to perform a preliminary job since Smlib only recognizes the SMILES format as an input for molecule fragments. Thus, we had to transcribe this format to all the fragments.

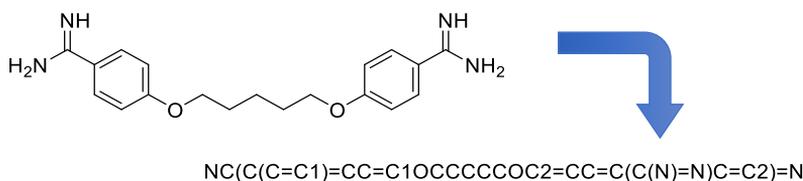


Fig 3.7. SMILES string representation of pentamidine.

After getting the SMILES notation for all fragments, the combinatorial library is enumerated. Binding points in spacers and terminal groups have to be indicated as [A] or [R], respectively

After this operation, the output obtained was a SMILES list of all the compounds that made up our chemical library. Besides, reference compounds such as pentamidine, Zimmerman Z1 or Hoechst and the candidates previously synthesized in our group were added. Finally, the resultant chemical library consisted of 339 (12 x 14 x 2 +3) compounds.

3.3.3 Molecular conformation calculation

The next step in the computational study was to calculate the molecular descriptors of the candidates. The conversion from the SMILES list obtained in Smlib to a format that could be interpreted by descriptors calculating software was needed. Connectivity information is crucial for calculating both the stability of conformations and the molecular descriptors. The chosen format was *SDF (SD-file)*. This format, based on the molfile format, is composed of atomic coordinates, a table of connections between the atoms composing the structure and supports the insertion of extra data in the form of physicochemical properties.

Rational design of potential groove binders

Once these descriptors are calculated, those containing no relevant information (same data for all candidates) were removed, and principal component analysis (PCA) was performed to group significant descriptors. The resulting library comprised 339 structures between candidates and reference compounds. Given the high number of compounds, we decided to perform a selection based on chemical diversity. This selection would be representative of the whole chemical space that defines the chemical library. Selected compounds should be as distant as possible in this chemical space, ensuring that the selection is representative.

3.3.5 Diversity selection

PRALINS was used to perform the diversity selection.⁴⁷ This software was developed in ANSI-C language to develop component selection operations, both in diversity and similarity. In this work, we used genetic algorithms to optimize a search in diversity within a specific chemical space.

Given a chemical space, we performed a chemical diversity selection. PRALINS calculates the distance between the points and maximizes the minimum distance between a subset of them. Therefore, a reduction in the total number of entries can be done without losing a whole part of the chemical space (maintaining as much as possible the representativity of the sub-selection). To avoid performing all combinatorial calculations, it is possible to use a genetic algorithm to optimize the calculus. This method consists of mimicking the replicating of a chromosome in the sexual reproduction of living beings. Given a virtual chromosome (a combination of ligands, i.e., a possible solution) of length equal to the number of compounds desired for the user (twenty in our case), a replication process is performed. Two parental chromosomes are crosslinked to form two new chromosomes, which will combine part of the parental chromosomes and random point mutations (all these factors are controlled by statistical parameters in the software's input file). The new generation is compared to the parental generation, and from among the chromosomes of both generations, the two best solutions are selected to beget the subsequent generation. Finally, a truncation parameter is included, consisting of completing the search after a defined number of iterations without improvement in the distance value.

After performing the calculation several times, most of the answers gave an identical selection of compounds (up to 19 of the 20 candidates were present in all selections).

3.3.6 Prediction of the binding mechanism of selected compounds

The binding mechanism of selected compounds was predicted by means of molecular docking

A docking process was performed with the 19 compounds selected. This calculation had to provide each ligand's affinity using a London DG scoring function (which estimates the ligand Gibbs energy) in two hundred conformations. A CUG modeled hairpin RNA previously designed in our research group was used as the receptor.³⁵ To carry out this calculation, we used Molecular Operating Environment (MOE), 2013.08 of the Chemical Computing Group Inc.⁴⁸

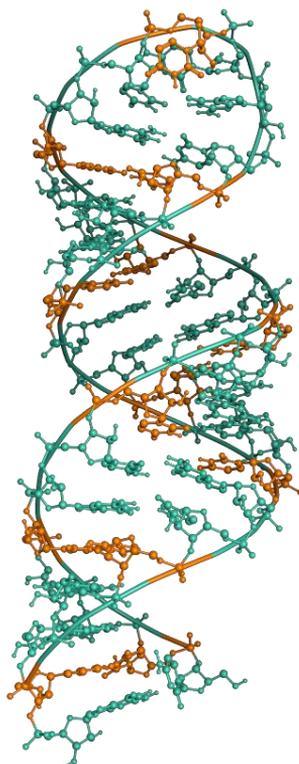


Fig 3.9. shRNA CUG receptor. U-U mismatches are highlighted in orange.³⁵

As this receptor is quite singular, it was not defined any active site. The only restriction applied to the calculus was the addition of pharmacophoric keys through the grooves and exclusion surfaces in the inner part of the hairpin to force candidates to interact with the receptor's minor groove.

On the other hand, and to reduce the calculation's computational cost, it was decided to use a rigid receptor conformation to dock ligands. The result obtained is displayed in the priority list presented below.

Rational design of potential groove binders

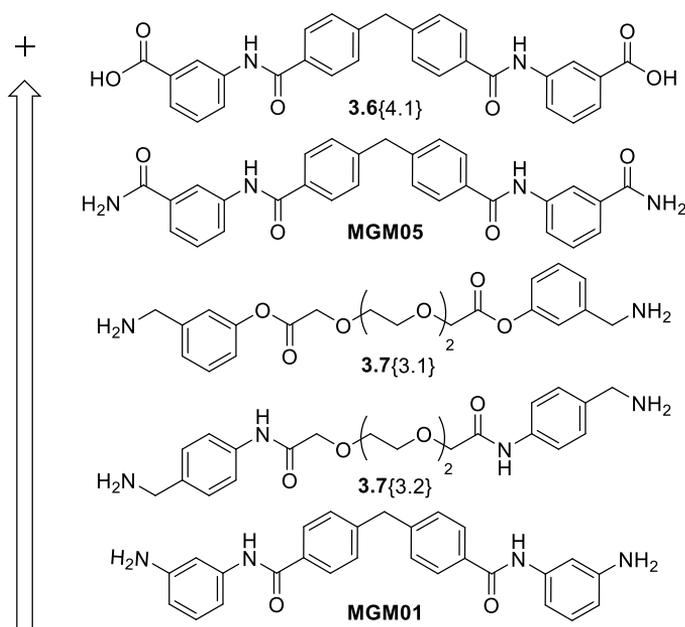


Fig 3.10. First results obtained in the docking process. Docking produced a result of affinity and here are the 5 ligands with the best value of scoring sorted in descending order. MGM compounds were already synthesized previously.

This list was the top of the resulting list. The first active conformation of a reference structure was Hoechst appearing at position 26 and pentamidine at 231.

Looking at the docking results, the best results are only represented by two spacers (**3.6** and **3.7**). This fact can be interpreted as these two spacers provide a greater affinity with the minor groove than the other spacers proposed. For this reason, in a second *docking* process, we decided to explore these two spacers' chemical space, eliminating the other spacers from the chemical library. It was also decided to remove the ester link from the current selection due to the possibility of presenting the hydrolysis phenomenon. The results of this second calculation are shown below.

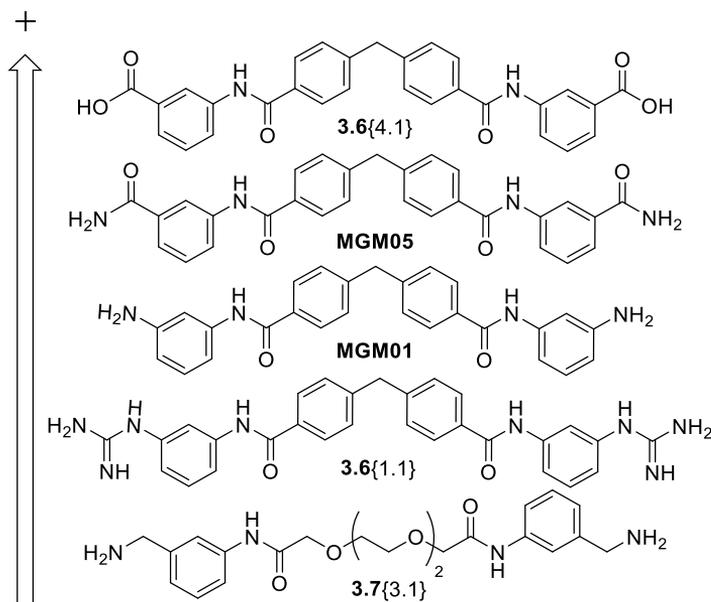


Fig 3.11. Second results obtained in the docking process. Docking produced a result of affinity and here are the 5 ligands with the best value of scoring sorted in descending order (in absolute scale).

The results obtained in this second docking are consistent with those obtained in the first docking attempt and gave new relevant data. It is possible to see that the terminal groups with orientation in *-meta* (1,3-disubstituted) lead to higher *scoring* values than their isomers in *-para* (1,4-disubstituted). Furthermore, after observing **3.6** derivatives in the top of the priority list, it was decided to explore other spacers like that one. A bibliographic search contributed with new spacers to our chemical library by introducing new spacers similar to **3.6**.^{49–51}

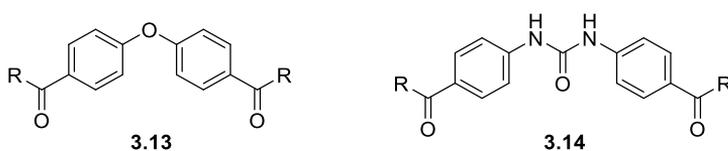


Fig 3.12. Second generation of spacers.

The previous chemical library added the combination of the new two spacers with all the existing terminal groups. Using the same docking protocol, the following priority list is obtained:

Rational design of potential groove binders

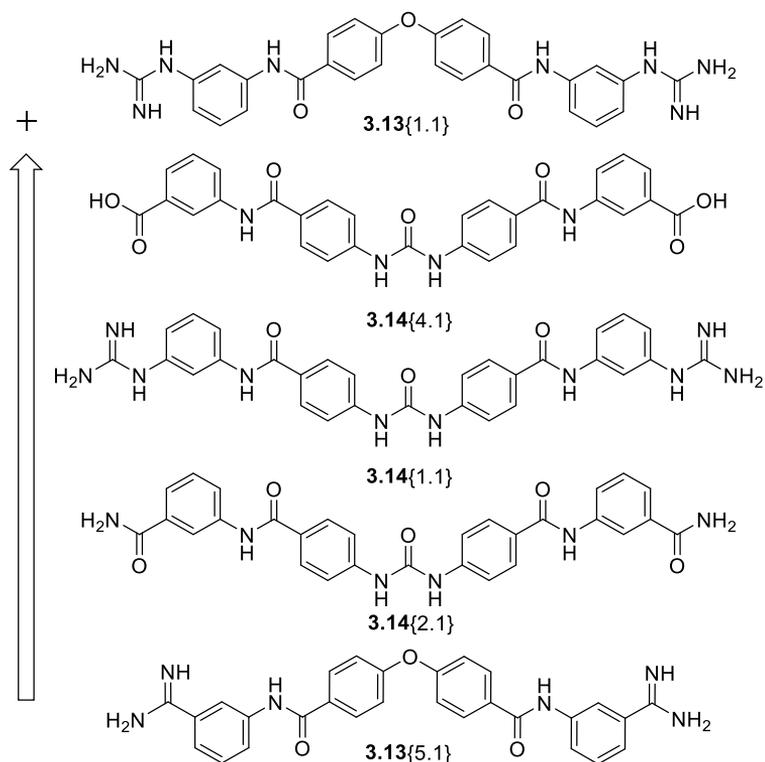


Fig 3.13. Third docking produced a result of affinity the 5 ligands with the best scoring value are sorted in descending order.

The new structures added to the existing chemical library were positioned on top of the last docking priority list. A possible reason is that incorporating high polarity groups in the middle of the structure allows polar interactions with the receptor. This radical change at the top of the list made us wonder if candidates bearing **3.13** and **3.14** spacers should prioritize the synthesis process.

The three-dimensional model showed the candidates fit in a relaxed conformation (without tensioned angles) with a practically perfect fit with the space generated in the minor groove (Fig 3.14).

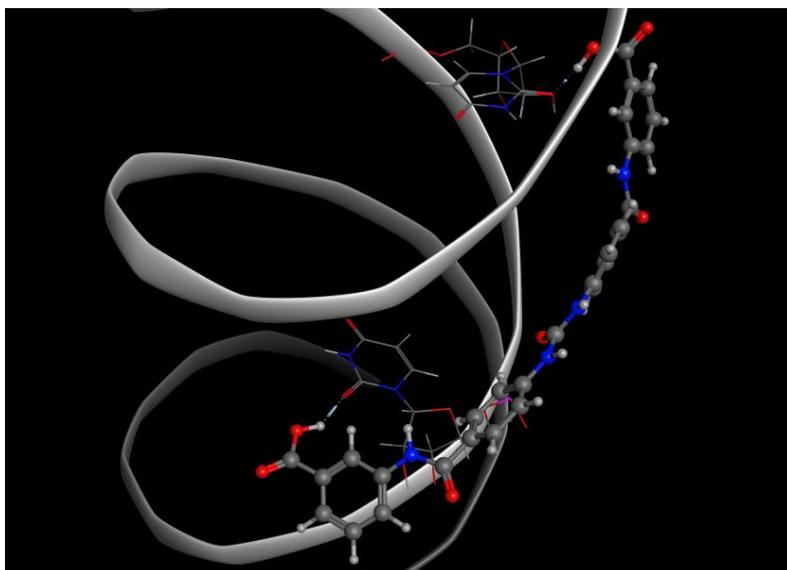


Fig 3.14. Interactions between two uracil bases and the 3.14{4.1} showing hydrogen-bonding.

A specific interaction between two non-consecutive uracil bases in the minor groove was observed for **3.14**{4.1} (Fig 3.14). This interaction is particularly interesting because, given the similarity of CUG with DNA, any specific interaction is highly valuable. Furthermore, as previously described, uracils in CUG are mismatched and more accessible than in any other regular structure.

Finally, drug-like properties were calculated using Qikprop 3.5 (Schrödinger) for all compounds. Unfortunately, the results were not conclusive.

We realized that compound **3.13**{4.1} was protected by a patent. Therefore, the last docking process was performed, using the same conditions as the previous studies, without the conflictive structure and including some isosteric groups for carboxylic acid.⁵²

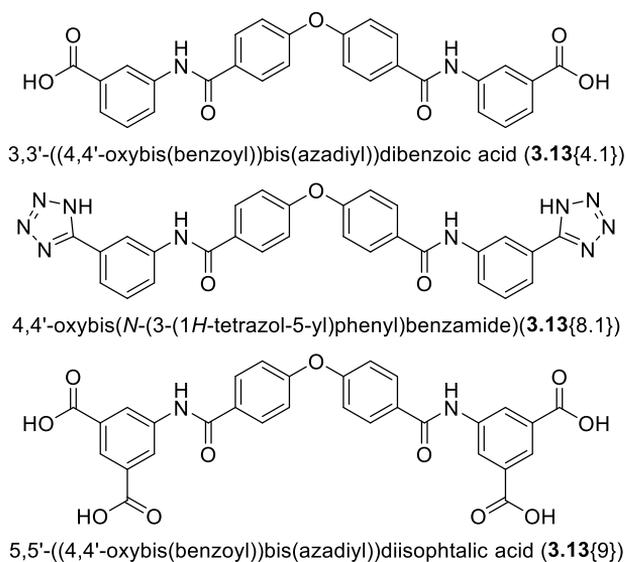


Fig 3.15. Proposed structures to replace the protected **3.13**{4.1}.

3.4 Synthesis

After obtaining a prioritized list of candidates, the next step was to address the synthetic pathway. The resulting candidates of the docking processes were divided into three categories. These categories bring together compounds bearing the same spacer, and therefore they share the synthetic precursor. The following figure presents the result of this classification.

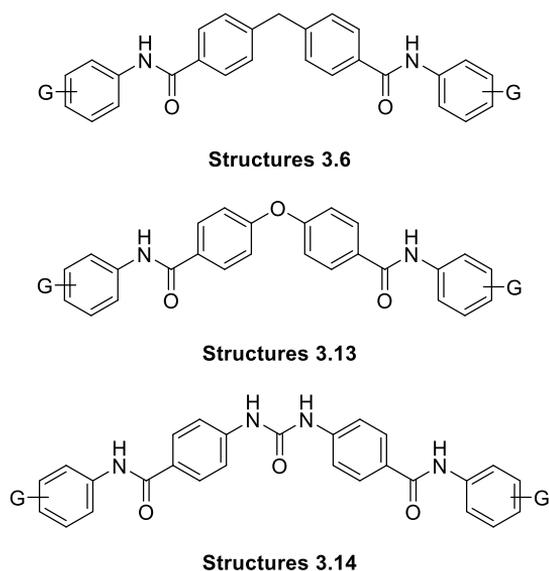
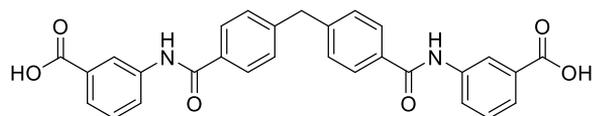


Fig 3.16. Markush structures of the ligands that were synthesized.

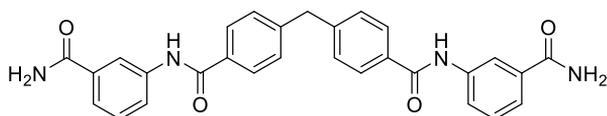
Candidates synthesis are discussed in the following pages, introducing the general strategy to achieve them and, after that, specifying every candidate's particular synthetic itinerary.

3.4.1 Structures 3.6: methylenebisphenyl spacer

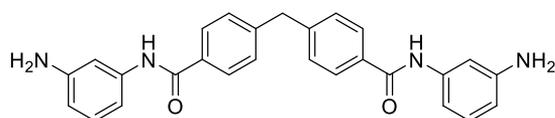
The first of the categories studied consists of a single candidate synthesized during this project (**3.6{4.1}**). On the other hand, the other two were previously synthesized in our research group by M.Gallart. Both were tested biologically against DM1 resulting in no DM1 phenotypic reversion. Since there was still enough sample of this spacer's precursor to obtain any further derivative, we decided to incorporate the **3.6** spacer in the first computational selections, giving preliminary high-affinity results.



3,3'-((4,4'-methylenebis(benzoyl))bis(azadiyl))dibenzoic acid (**3.6{4.1}**)



4,4'-methylenebis(*N*-(3-carbamoylphenyl)benzamide)(**MGM05**)



4,4'-methylenebis(*N*-(3-aminophenyl)benzamide)(**MGM01**)

Fig 3.17. Studied candidates bearing spacer **3.6**.

Therefore, the most favored compound was synthesized (**3.6{4.1}**). After that, we decided to prioritize the other group's candidate's synthesis because, in the latter docking process, the candidates of groups 2 and 3 scored higher affinity values than structures bearing **3.6** spacer. The strategy and results obtained from the synthesis of **3.6{4.1}** acid are shown below.

Synthesis

3.4.1.1 Retrosynthetic analysis and synthesis

compound's synthesis, a strategy was designed to describe with reasonable disconnections its obtention. In the following figure the retrosynthetic analysis that led to the synthetic pathway for the synthesis of **3.6**{4.1} is shown.

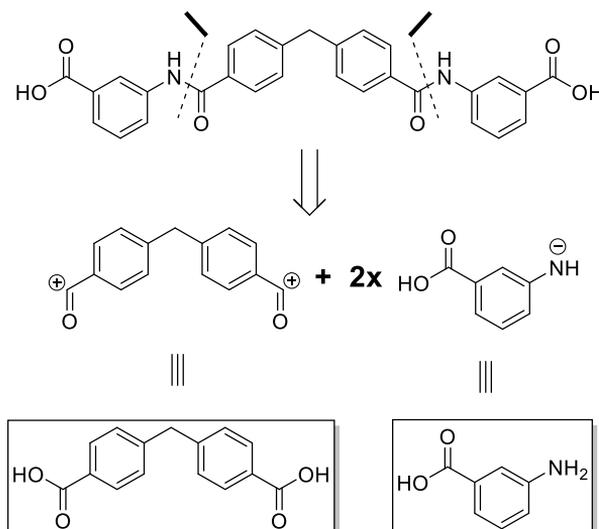


Fig 3.18. Retrosynthetic analysis of **3.6**{4.1}.

The proposed disconnection presented in the figure above takes advantage of the molecule symmetry, disconnects the two aromatic moieties simultaneously, giving rise to some known synthons. These synthon precursors could be a carboxylic acid or an activated species and a primary aromatic amine.

However, it presented a problem. In the second stage of this itinerary, the coupling between **3.15** and **3.16**, an acid group in **3.16**, could result in unwanted reactions leading to a series of by-products that were difficult to remove. Given this situation, two possibilities arise: first, to add protection/protection stages for this carboxylic group, increasing the total number of steps required to obtain **3.6**{4.1}, or searching a methodology allowing the coupling under without further reactions. Conditions described by C. Schotten and E. Baumann in the late 19th century^{53,54} in what has subsequently become known as the Schotten-Baumann reaction. The method consists in performing the reaction in aqueous media, dissolving the amine into a slight excess of an inorganic base (NaOH, NaHCO₃, or Na₂CO₃), and adding an acid chloride dropwise. The higher nucleophilicity of amines than water gives a kinetic advantage over acid chloride hydrolysis. The presence of a base and the usual organic heavy molecules lack solubility in an aqueous medium, finally displace the equilibrium to product formation. In the next figure the reader could find the synthetic proposal for the synthesis of **3.6**{4.1}.

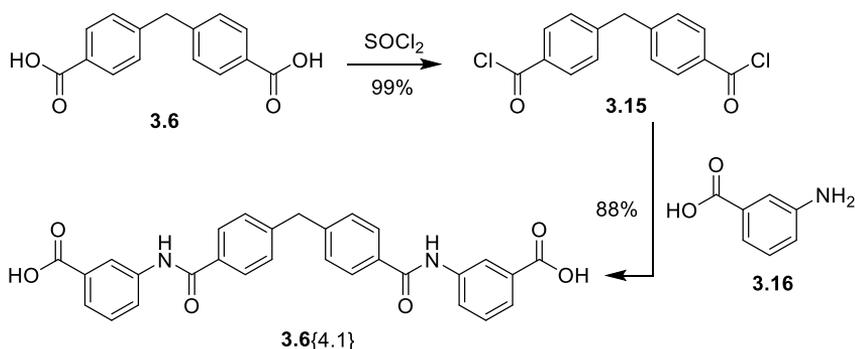


Fig 3.19. Synthetic pathway to attain **3.6{4.1}**, divided in two synthetic steps: carboxylic acid activation to acid chloride and double amidation coupling under Schotten-Baumann conditions.

In this case, a modified version of the conditions typical of the Schotten-Baumann method allows affording **3.6{4.1}** using the methodology developed by Chattopadhyay *et al.*⁵⁵

3.4.1.1.1 Synthesis of **3.15**: acid chloride formation

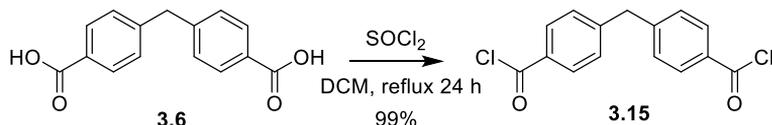


Fig 3.20. Synthesis of intermediate **3.15**.

The first of the two stages that make up the synthesis of **3.6{4.1}** consists of activating the two-terminal carboxylic acid groups. Therefore, the acid chloride was synthesized. It was necessary to add an excess of the halogenating reagent, in this case, thionyl chloride. Once synthesized, and even though ¹H-NMR was consistent with what was described in the literature,⁵² it was decided to analyze the product obtained by IR spectroscopy. The presence of carbonyl bands between 1780 and 1696 cm⁻¹ corresponding to the symmetrical and asymmetrical stretches of a pair of halo-substituted carbonyls confirmed the desired compound's formation. **3.15** is obtained in the form of a brown solid with a yield of 99%.

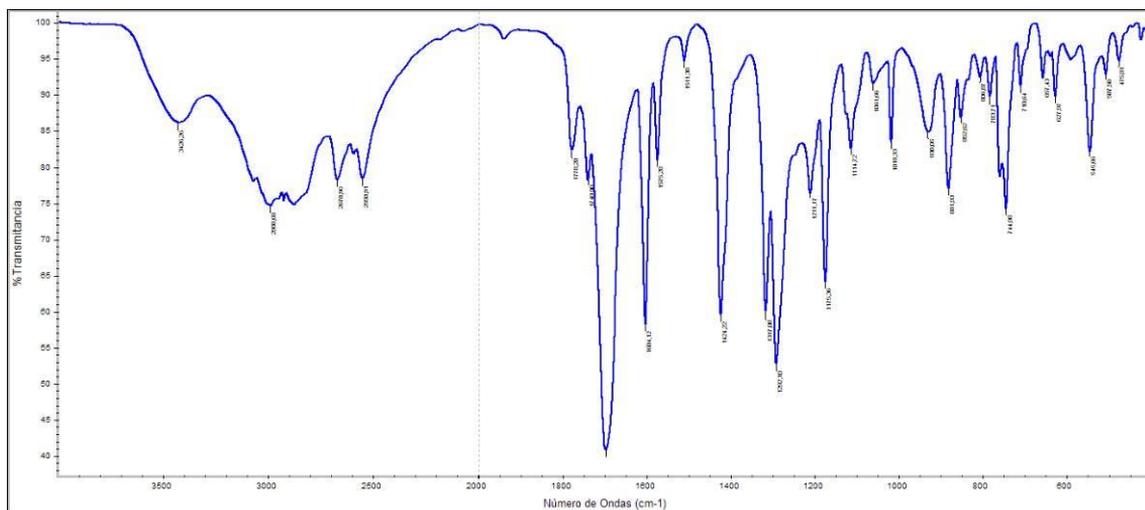
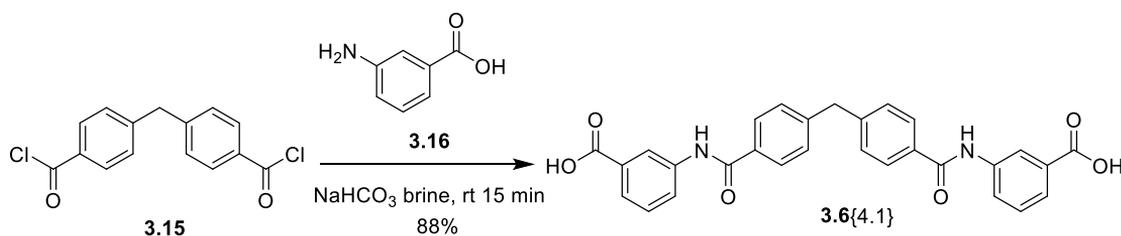


Fig 3.21. IR spectrum of 3.15.

3.4.1.1.2 Synthesis of **3.6{4.1}**: Schotten-Baumann double amidation

Fig 3.22. Schotten-Baumann double amidation affording **3.6{4.1}**.

In the second synthetic step affording **3.6{4.1}**, the Schotten-Baumann method is required.

$^1\text{H-NMR}$ showed the slight presence of the hydrolysis product in the final product in the first attempt. To avoid this issue, changes in the isolation process were introduced for ad hoc that would precipitate the reaction's product, maintaining dissolved the hydrolyzed carboxylate of acid dichloride. After successive solubility tests with different cations (Li^+ , K^+ , Ca^{2+} , NH_4^+ , NBut_4^+), it was observed that by precipitating the reaction mixture with a saturated LiCl solution instead of using hydrochloric acid, the ionic form of the desired product was obtained. However, it was also found to precipitate a remarkable amount of NMR-undetectable LiHCO_3 . Therefore, instead of changing the original methodology, we decided to perform an extra digestion step in a solvent in which the product was barely soluble. Thus, the reaction mixture was dissolved in methanol and was heated to $120\text{ }^\circ\text{C}$ using μW for an hour. A solid crystallized upon cooling to room temperature, which was isolated by filtration. **3.6{4.1}** is obtained as a white solid in 88% yield. This reaction is the last step in the synthetic pathway. Therefore, candidate cumulative yield is calculated.

$$\eta(\mathbf{3.6\{4.1\}}) = (0,99 \cdot 0,88) \cdot 100 = 87\%$$

3.4.2 Structures 3.13: oxybisphenyl spacer

The second of the subcategories studied is derived from **3.13 spacer**. This spacer was not incorporated into the first *docking processes* its precursor was not available in the laboratory. However, with the finding of **3.6** derivatives among the most favored structures during the first computational selection processes, it was decided to enrich the variety of this chemical space by testing other similar spacers. One of the first proposed groups was the provision of the ether group.

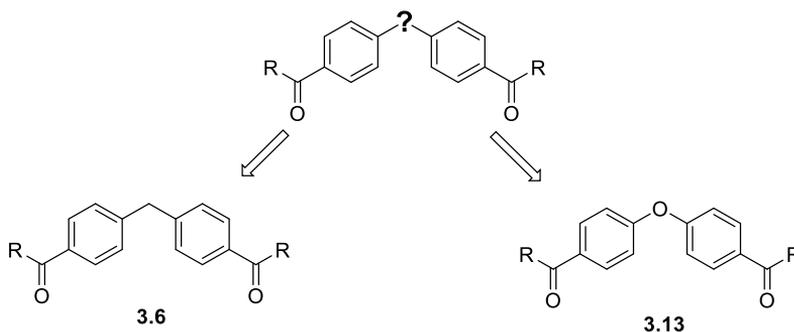


Fig 3.23. General scheme of rational design of spacers.

Besides, the oxygen in the middle structure allowed potential polar interactions with the receptor. This hypothesis was not new, as there are plenty of examples in the bibliography of candidates that enrich their binding affinity with polar groups in the spacer zone.^{5,56} Therefore, it is assumed that the spacer is an essential part of the interaction with the receptor and not just a separator between the terminal polar groups. It was decided to synthesize a wide variety of terminal groups maintaining this spacer to confirm or reject this hypothesis. In the following figure, there are structures of the candidates bearing this spacer:

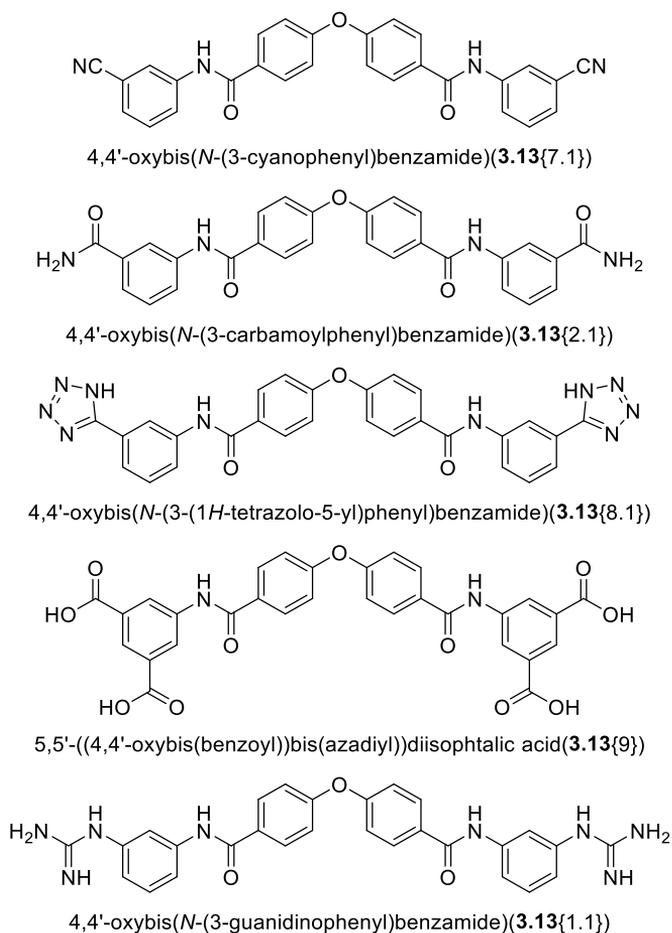


Fig 3.24. Selected compounds with **3.13** spacer.

3.4.2.1 Retrosynthetic analysis

In the following figure the general strategy for obtaining candidates **3.13** is displayed.

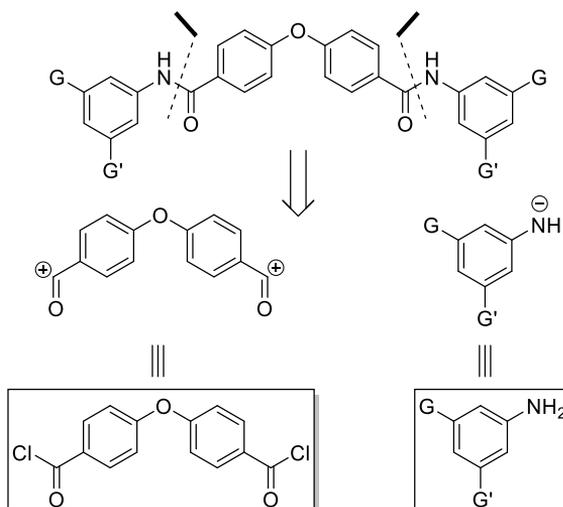


Fig 3.25. Retrosynthetic analysis to obtain candidates 3.13.

It should be borne in mind that not all candidates can be synthesized following the strategy resulting from the previous figure retrosynthetic analysis. Some groups (**G**, **G'**) that constitute binding groups are not compatible with the coupling chemistry. In the table displayed below, compounds that meet the strategy presented in the previous figure without adding any extra synthetic stage:

Tab.3.3. Functional groups in G and G' positions of the generic structure of Fig 3.23.

Candidate	G	G'
3.13{7.1}	-CN	-H
3.13{2.1}	-CONH ₂	-H
3.13{9}	-COOH	-COOH

In addition to the previous candidates, those incorporating guanidine (**3.13{1.1}**) and tetrazole (**3.13{8.1}**) functional groups require pre-transformation (and subsequent) steps and are therefore not synthesized in one-pot synthesis, adapting the whole synthetic pathway avoiding byproduct formation.

3.4.2.2 Synthesis of binding moieties

The following pages describe the binding groups synthesis before its assembly:

Synthesis

3.4.2.2.1 Synthesis of **3.18**

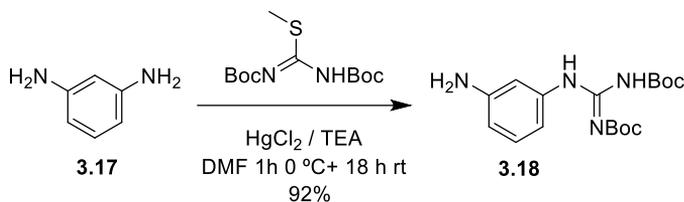


Fig 3.26. 1-(3-aminophenyl)-N,N-bis-(*tert*-butoxycarbonyl)guanidine.

Guanidine classic synthesis procedure is dismissed in **3.18**⁵⁷ as the product is obtained as a salt. Therefore, as the reactant has two reactive groups (amines), hard to separate byproducts are expected. Alternatively, using thiourea or thioseoudourea can introduce the amidino group to the primary amine to form the corresponding guanidine.⁵⁰ It is also necessary to emphasize mercury chloride's role in this process. It is responsible for capturing the sulfides, allowing the reaction to advance by displacing the equilibrium towards the product formation.

As expected, spectroscopy determines that the compound is impurified by double substitution byproduct and by remains of **3.17**. The resulting mixture is purified by flash chromatography. The chromatographic profile of the separation of the three products provided by the ISCO Teledyne: CombiFlash Rf team is displayed in Fig 3.27. **3.18** is obtained as a white solid in 92% yield.

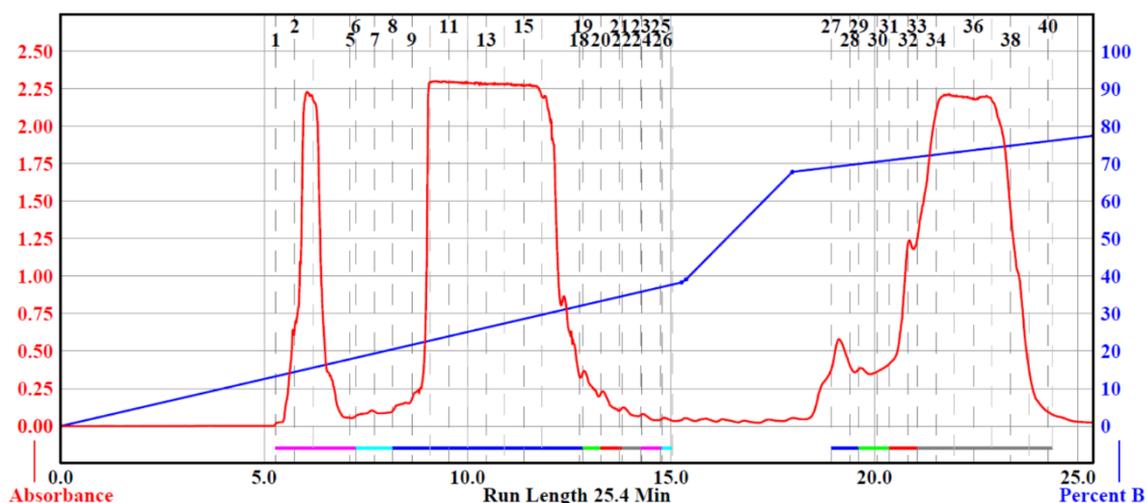


Fig 3.27. Chromatographic profile of the **3.18** purification (central peak).

3.4.2.2.2 Synthesis of 3.20

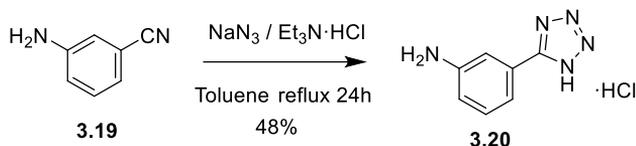


Fig 3.28. Synthesis of 3-(1*H*-tetrazol-5-yl)aniline (**3.20**).

We also decided to introduce the tetrazole group, as a carboxylic acid isosteric functionality, we expected to attain similar biologic results avoiding patent conflicts. The usual strategy to form a tetrazole group is the 3+2 cycloaddition between a nitrile and sodium azide. This type of pericyclic reaction is included among the group of so-called click reactions. This concept refers to those reactions meeting the characteristics described by Sharpless,⁵⁸ i.e., reactions with a high atom economy, and the formation of this heterocycle is an excellent example of it.

A proper methodology for synthesizing **3.20** is attained by adapting the bibliographic methodology.⁵⁹ The reagents dissolved in toluene and were led to reflux. After 24 hours, it was observed that the crude oil consisted of two phases, a colorless upper fraction and a lower color of intense red color and a murky appearance. Water was added to the lower fraction, which, when acidified and cooled, began a rapid crystallization process, affording pure **3.20** in 48% yield.

NMR Spectroscopy showed a change in the benzene proton chemical shifts, indicating a chemical transformation in the functional groups connected to it. However, more consistent evidence of the reaction success is fading the characteristic nitrile band in the IR spectrum.

Synthesis

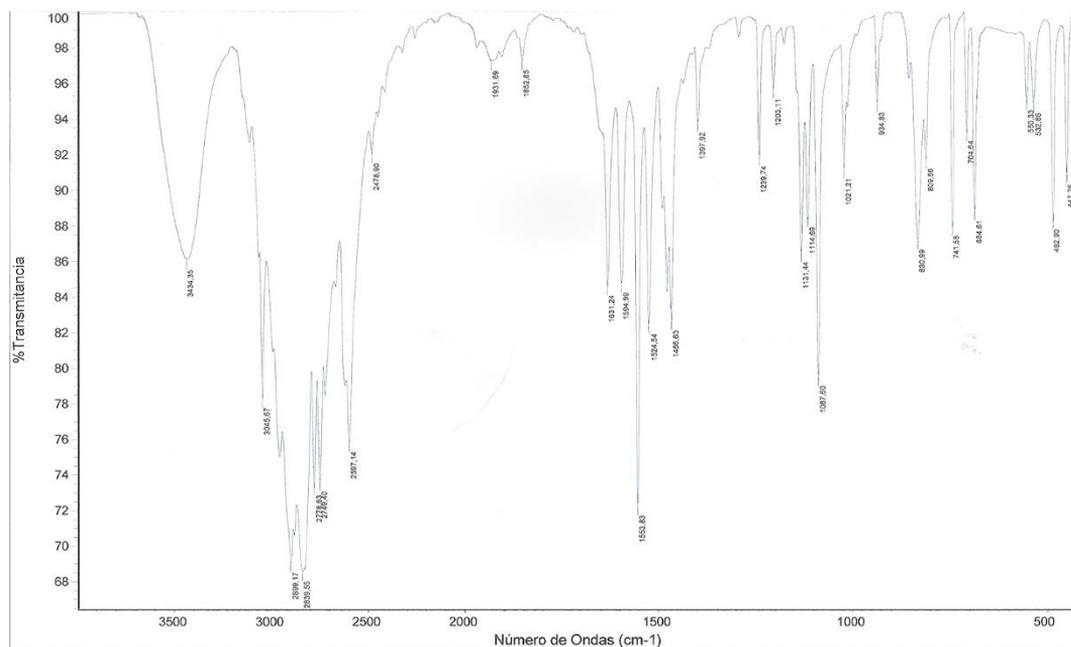


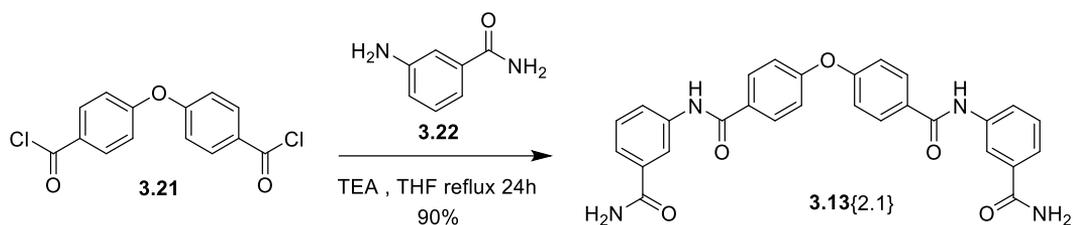
Fig 3.29. Infrared Spectrum of 3.20. No -CN stretch is observed.

3.4.2.3 Amidation couplings

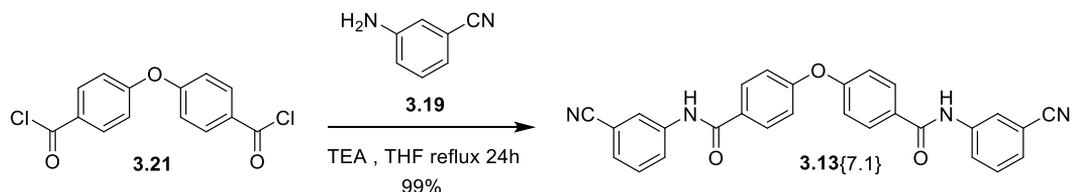
In the following pages, two different methodologies are described to perform couplings between recognizing moieties and spacers.

3.4.2.3.1 Acyclic substitution

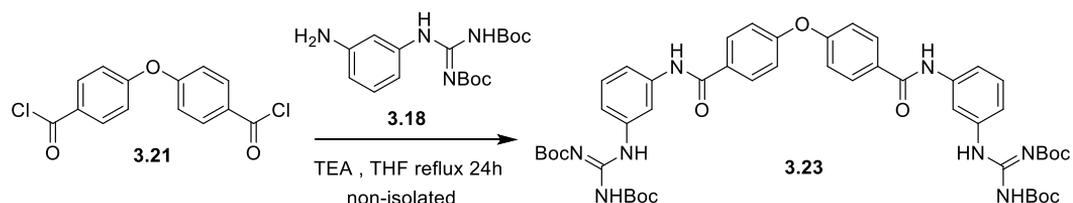
The following methodology was selected for those candidates bearing functional groups that could not interfere with the coupling reaction. The reagents were dissolved in a polar aprotic solvent (i.e., THF) in the presence of a base. Reaction evolution was monitored by TLC. Upon reaction completion, the solvent was removed under reduced pressure. The desired product was afforded by adding an aqueous hydrochloric acid solution (1M) and filtering.

3.4.2.3.1.1 4,4'-oxybis(*N*-(3-carbamoylphenyl)benzamide) **3.13**{2.1}Fig 3.30. Synthesis of 4,4'-oxybis(*N*-(3-carbamoylphenyl)benzamide) **3.13**{2.1}.

An equimolar ratio of acid chloride and TEA was sufficient to displace the equilibrium and afford **3.13**{2.1} without any detectable byproduct via NMR. **3.13**{2.1} is obtained as a white solid in 90% yield.

3.4.2.3.1.2 4,4'-oxybis(*N*-(3-cyanophenyl)benzamide) **3.13**{7.1}.Fig 3.31. Synthesis of 4,4'-oxybis(*N*-(3-cyanophenyl)benzamide) (**3.13**{7.1}).

As hydrogen chloride reportedly reacts with nitriles, a slight excess of the base was added to avoid any possible by-product formed this way in this specific reaction. **3.13**{7.1} was successfully achieved with an excess of 1:3 of triethylamine to the acid dichloride. **3.13**{7.1} is obtained as a white solid in 99% yield.

3.4.2.3.1.3 4,4'-oxybis(*N*-(*N'*,*N''*-di-*tert*-butoxycarbonyl-3-guanidinophenyl)benzamide) **3.23**Fig 3.32. Synthesis of 4,4'-oxybis(*N*-(*N'*,*N''*-di-*tert*-butoxycarbonyl-3-guanidinophenyl)benzamide)(**3.23**).

Synthesis

Boc groups (tert-butoxycarbonyl) are susceptible to acidic conditions. Therefore, the reaction was conducted using the same base approach as the dinitrile derivative **3.13**{7.1}. It is observed by NMR spectroscopy that the reaction crude contains the desired product. However, to avoid partial deprotection and maximize the yield, it is decided directly fully deprotect instead of isolating this intermediate.

3.4.2.3.2 Schotten-Baumann Couplings

For those candidates bearing acid protons in their binding moieties, a different approach is needed. Since these groups may hamper the reaction, a method analogous to methylene candidate **3.6**{4.1} (Schotten-Baumann) is proposed.

3.4.2.3.2.1 5,5'-((4,4'-oxybis(benzoyl))bis(azadiyl))diisophthalic acid **3.13**{9}

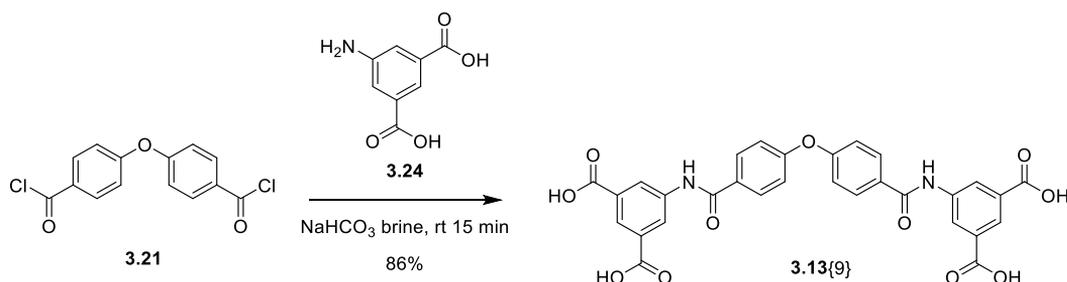


Fig 3.33. Synthesis of 5,5'-((4,4'-oxybis(benzoyl))bis(azadiyl))diisophthalic acid (**3.13**{9}).

A small variation of the aforementioned methodology was used to afford **3.13**{9}. The product intended to synthesize bears one more carboxylic acid moiety per binding group. Therefore, an increase in the volume of saturated solution sodium bicarbonate is needed. For this reason, inorganic salts remain at the end of the standard isolation process, and a digestion process with acetone is required to remove them. **3.13**{9} is obtained as a white solid in 86% yield.

3.4.2.3.2.2 4,4'-oxybis(*N*-(3-(2*H*-tetrazole-5-yl)phenyl)benzamide) **3.13**{8.1}.

As the tetrazole group represents an isostere of the carboxylic group (see 3.3.7 Docking iterations), and its hydrogen is acidic enough to interfere in the coupling, it is decided to synthesize **3.13**{8.1} via Schotten-Baumann coupling.

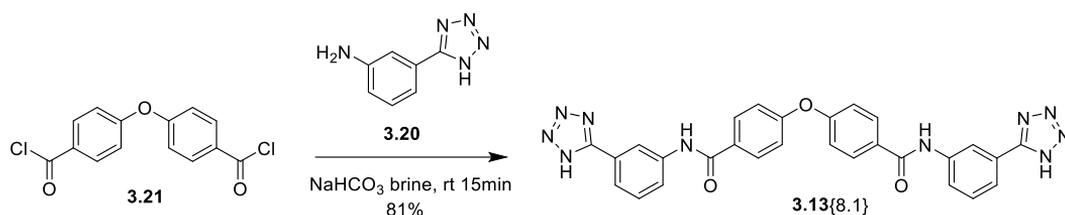


Fig 3.34. Synthesis of 4,4'-oxybis(*N*-(3-(2*H*-tetrazole-5-yl)phenyl)benzamide)(**3.13{8.1}**).

It was attempted to obtain **3.13{8.1}** exchanging reaction order. First, by coupling the acid dichloride with *m*-aminobenzonitrile and then trying to make the 3+2 cycloaddition with sodium azide. However, only the monotetrazole compound is partially formed. There is literature describing how to increase the yield of this reaction by metallic catalysis.⁶⁰ However, it was prioritized the synthesis of new candidates rather than to explore alternative pathways to an already obtained candidate.

$$\eta(\mathbf{3.13}\{8.1\}) = (0,48 \cdot 0,81) \cdot 100 = 39\%$$

3.4.2.4 Post coupling transformations

3.4.2.4.1 4,4'-oxybis(*N*-(3-guanidinophenyl)benzamide) (**3.13{1.1}**).

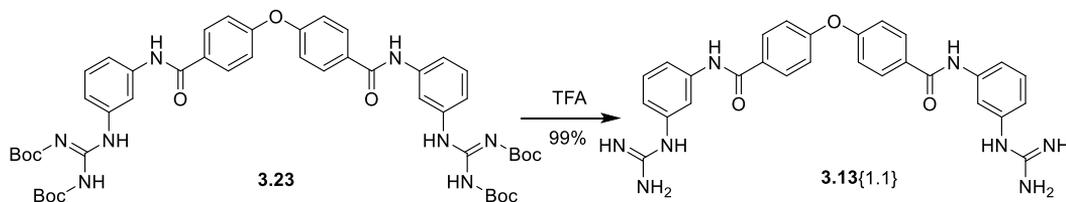


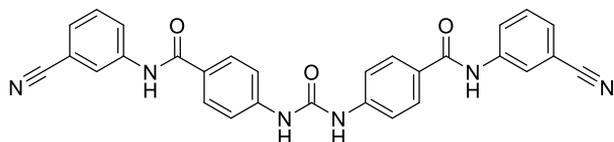
Fig 3.35. Synthesis of 4,4'-oxybis(*N*-(3-guanidinophenyl)benzamide) (**3.13{1.1}**).

This synthetic step was carried out by dissolving the prior reaction mixture in a 1:1 mixture of trifluoroacetic acid and methylene chloride. The reaction was stirred at room temperature for three hours. The resulting mixture was concentrated under reduced pressure, affording **3.13{1.1}** in 81% overall yield.

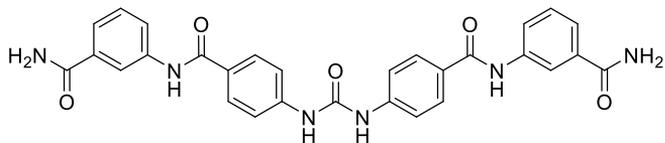
$$\eta(\mathbf{3.13}\{1.1\}) = (0,92 \cdot 0,89 \cdot 0,99) \cdot 100 = 81\%$$

3.4.3 Structures 3.14: urea bisphenyl spacer

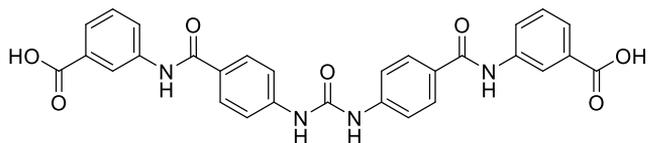
The last family of synthesized candidates is **3.14** derivatives. Although it is a similar structure to those proposed previously in this chapter, this spacer is not commercially available and must be synthesized. The structures of the candidates containing this spacer synthesized are displayed in the figure below.



4,4'-(carbonylbis(azadiyl))bis(*N*-(3-cyanophenyl)benzamide)(**3.14**{7.1})



4,4'-(carbonylbis(azadiyl))bis(*N*-(3-carbamoylphenyl)benzamide)(**3.14**{2.1})



3,3'-((4,4'-(carbonylbis(azadiyl))bis(benzoyl))bis(azadiyl))dibenzoic acid(**3.14**{4.1})

Fig 3.36. Candidates containing the spacer **3.14**.

3.4.3.1 Retrosynthetic análisis

This section discusses the strategy followed for the synthesis of candidates bearing spacer **3.14**. The following figure proposes a generic retrosynthetic analysis for all of them.

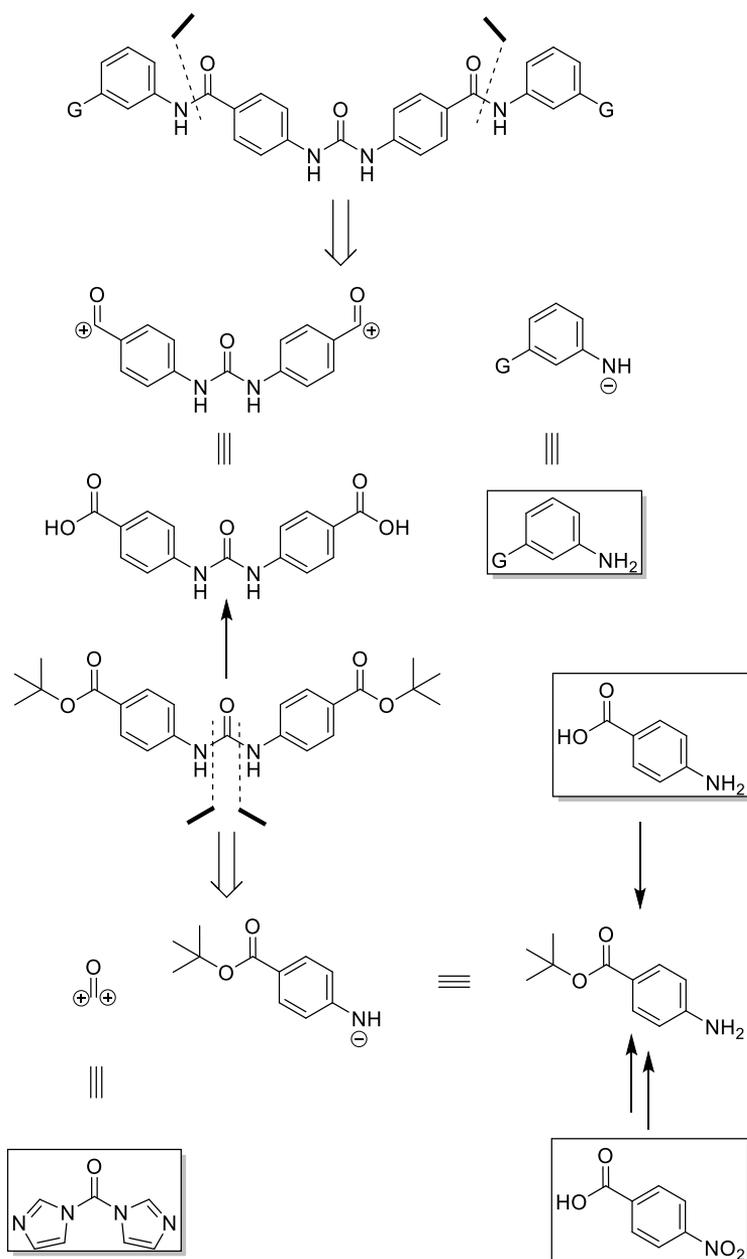


Fig 3.37. Generic retrosynthetic analysis for candidates with spacer 3.14.

As shown in the previous figure, the retrosynthetic analysis allows different ways to obtain the desired candidate via key synthetic intermediates. It can also be modified depending on the nature of the functional groups located at position **G**. The table below displays the different possibilities studied of functional groups replacing position **G**.

Synthesis

Tab. 3.4. Functional groups in G positions of the generic structure of the Fig 3.35.

Candidate	G
3.14{7.1}	-CN
3.14{2.1}	-CONH ₂
3.14{4.1}	-COOH

3.4.3.2 Spacer synthesis

To achieve the synthesis of **3.15**, two milestones are crucial: first, obtaining *tert*-butyl *p*-aminobenzoate, and second, its coupling with CDI (carbonyldiimidazole) to form *N*-substituted urea. The synthetic itinerary for the spacers is displayed below, based on the previous section's strategy.

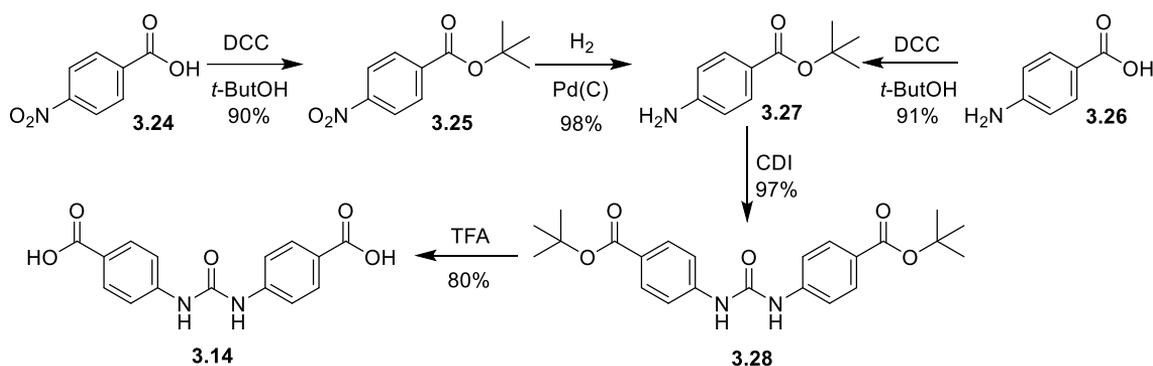


Fig 3.38. Synthetic itinerary to obtain spacer 4,4'-(carbonylbis(azadiyl))dibenzoic acid(**3.14**).

Two possible paths are proposed in Fig 3.38 for obtaining the critical intermediate **3.27**. Initially, the shortest pathway was adopted, starting from *m*-aminobenzoic acid. However, this option meant a significant waste of resources since, to avoid nucleophilic competition (*tert*-butyl alcohol and the amino group of **3.26**), the alcohol was used as a solvent, and the reaction was carried out in high dilution. For these reasons, we decided to improve synthesis efficiency by reducing the necessary resources. In conclusion, adding one more synthetic step and using nitrobenzoic acid as a starting product eliminates the possible nucleophilic competence.

3.4.3.2.1 Synthesis of **3.27**

This first step describes the two reactions used to protect the carboxylic acid with a *tert*-butoxy group.

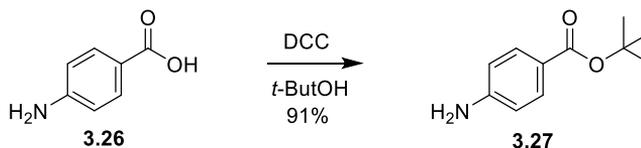


Fig 3.39. Protection of *p*-aminobenzoic acid.

The first synthetic alternative consists of an esterification reaction using an unprotected amino group in the same structure where the acid group is located. For this reason, this synthesis presents extra difficulty due to the possibility of connecting this amine to the carboxylic group instead of the *tert*-butyl alcohol. To avoid this phenomenon, *tert*-butyl alcohol is used as a solvent and in significant excess. Thus, in these conditions, it is improbable the formation of the byproduct. The crude obtained was purified by flash chromatography to obtain **3.27** in 91% of yield. The chromatographic profile of the purification is displayed below.

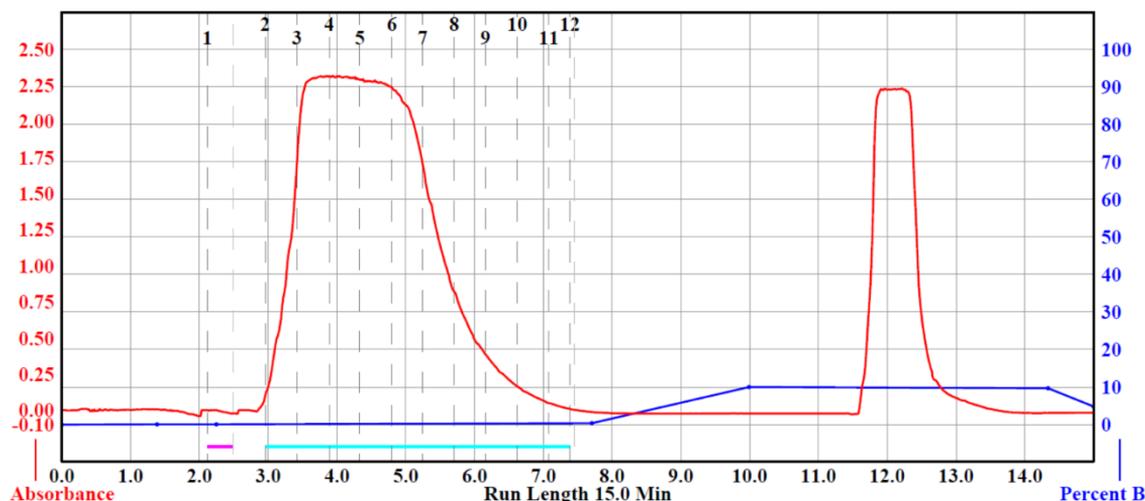


Fig 3.40. Chromatographic profile of the purification of *tert*-butyl *p*-aminobenzoate (first peak).

As mentioned above, although this reaction is carried out successfully, it was decided to change the synthetic strategy to obtain the critical intermediate **3.27**, reducing the consumption of available resources. An itinerary that started from *p*-nitrobenzoic acid with two steps is found in the literature.⁴⁹The first of the two steps is described below.

Synthesis

3.4.3.2.2 Synthesis of **3.25**

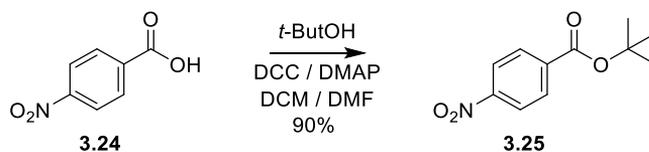


Fig 3.41. Protection of p-nitrobenzoic acid affording **3.25**.

This reaction is equivalent to the previous reaction, but in this case, it is unnecessary to use an excess of tert-butyl alcohol since the nitro group does not interfere with the reaction. Reaction conditions are both milder, as using a solvent mixture (DCM 12:1 DMF) and DMAP as a catalyst, the reaction evolves at room temperature in 4 hours. In most of these Steglich-type reactions, the coupling reagent's residue (DCU) forces a chromatographic purification stage to be separated from the desired product.

3.4.3.2.3 Synthesis of **3.27** (II)

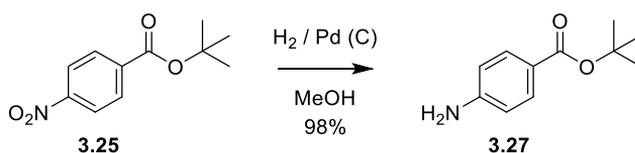
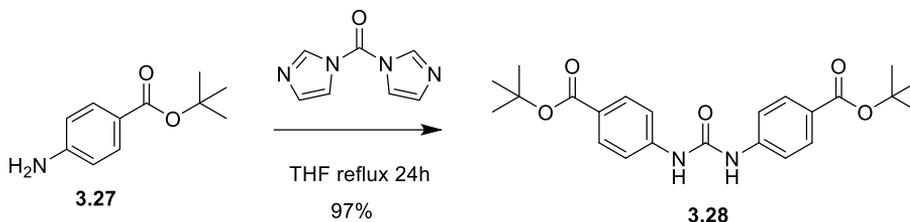
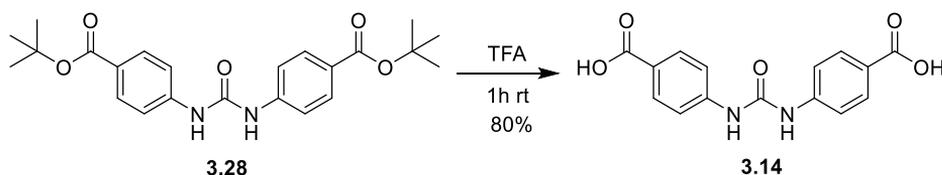


Fig 3.42. Reduction of the nitro group affording **3.27**.

To attain **3.27** in this step, it was necessary to use hydrogen (gas) and a metal catalyst (palladium) while the reagent is dissolved in methanol. Therefore, this is a paradigmatic example of heterogeneous phase reactions. Thin-layer chromatography (TLC) was used to monitor the reaction progress. Once the starting reactant had been entirely consumed, the reaction crude was filtered through the celite coarse to remove the catalyst. Finally, the elimination of methanol in vacuo affords to the product tert-butyl 4-aminobenzoate.

3.4.3.2.4 Synthesis of **3.28**Fig 3.43. Coupling with CDI affording **3.28**.

Once **3.27** is obtained, the next step is assembling it to obtain the core structure of **3.14**. A reagent capable of forming symmetrical ureas by substitution is needed to achieve this goal. To perform this stage, **3.27** and CDI were dissolved in THF in proportion 2:1, and the resulting mixture was heated to reflux for 24 h. To isolate the target product, the solvent was removed in vacuo once the heating period elapsed. The crude oil was redissolved in ethyl acetate, washed with a solution of hydrochloric acid (1 M), dried, filtered, and finally concentrated to afford the target compound.

3.4.3.2.5 Synthesis of **3.14**Fig 3.44. Saponification of *tert*-butyl esters of **3.28**.

Finally, once the core structure of the **3.14** spacer has been built, it is necessary to deprotect the previously formed esters to prepare the structure with the corresponding terminal groups for posterior couplings. To eliminate the *tert*-butoxy groups, the product was dissolved in TFA and stirred an hour at room temperature. It is observed that instantaneously, once the starting product is dissolved, the reaction product starts precipitating by turning the reaction medium into a whitish suspension. A volumetric equivalent of diethyl ether was added, and the resulting suspension was filtered and washed, affording **3.14** without further purification.

3.4.3.3 Obtaining the binding groups

Once the **3.14** spacer is synthesized, it is time to consider the terminal group synthesis. The three compounds used to perform the coupling are displayed below.

Synthesis

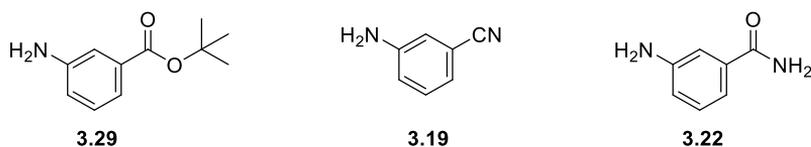


Fig 3.45. Aniline-based binding groups selected for 3.14 series.

Of the three compounds, only **3.29** requires to be synthesized as described in the strategy section. For this reason, the next pages only reflect the results of obtaining the *tert*-butyl *m*-aminobenzoate.

3.4.3.3.1 Synthesis of **3.28**

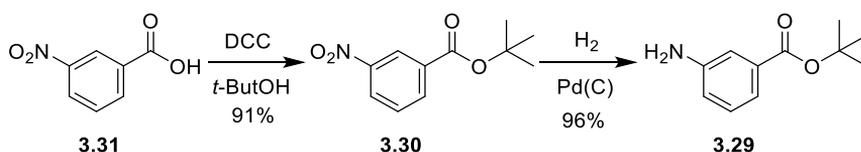


Fig 3.46. Synthetic pathway leading to *tert*-butyl 3-aminobenzoate.

To obtain **3.29**, the development of any new synthetic itinerary was not required. During the synthesis of **3.14**, a geometric isomer of this compound (**3.27**) is synthesized. Thus, this compound was obtained by reproducing the previous section's methodologies (see 3.4.3.2. Spacer synthesis). The final product is obtained as white solid, and spectroscopy corroborates the structure.

3.4.3.4 Couplings

The differentiating step in the synthesis of any of those candidates is the coupling. This section describes what possibly proved most complex of all those described in the present chapter. The literature describes methodologies for coupling compounds similar to the chosen candidates.⁴⁹ The activation of the corresponding diacids with thionyl chloride was unsuccessful. Therefore, A Steglich type methodology was implemented.

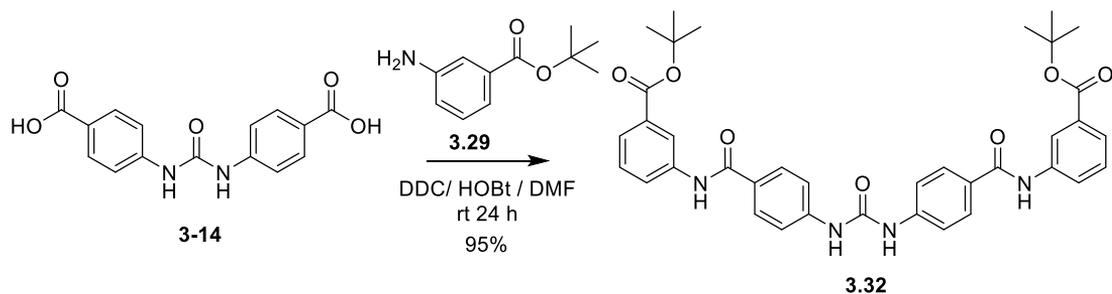


Fig 3.47. DCC mediated binding group coupling.

Nuclear magnetic resonance (NMR) spectroscopy showed the formation of the amide bond. The only drawback of this first test was that the DCU formed due to the product's formation was very difficult to eliminate. Given the chemical properties and the similarity with synthesized products (they are substituted ureas as well). However, through a purification step by flash chromatography, the target compound is isolated with low yields due to the lack of solubility of the **3.14** structures in organic solvents. Despite being already obtained, it was tested to change the coupling reagent for another to facilitate the isolation process. The chosen one was the EDC, part of the carbodiimide reagent family but presenting a peculiarity that makes it protrude. EDC coupling byproduct is soluble in water. This fact, together with the fact that HOBT is also partially soluble in water, significantly simplifies candidates' isolation.

3.4.3.4.1 Synthesis of **3.14**{7.1}

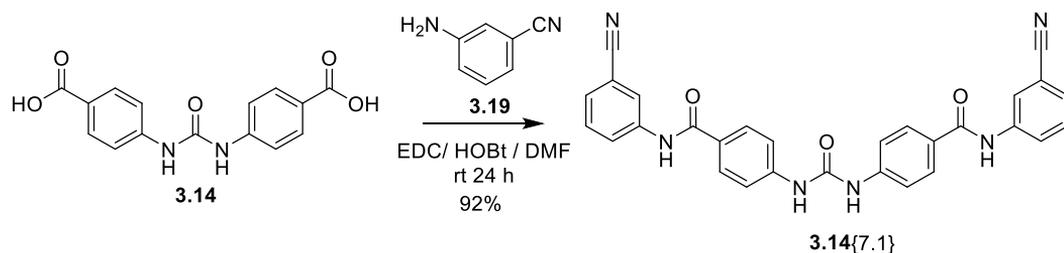


Fig 3.48. 4,4'-(carbonylbis(azadiyl))bis(*N*-(3-cyanophenyl)benzamide)(**3.14**{7.1}).

Once the reagents and isolation procedure were selected, the next step was to perform the corresponding coupling with each of the three binding groups. The first of the three reactions was the coupling of **3.14** with **3.19**. After a couple of hours of stirring, it was observed how the reaction medium, which was initially completely transparent, began to turn cloudy, which was interpreted as the product's formation. The product was isolated by filtration after eliminating DMF and washing the crude oil with water.

$$\eta(\mathbf{3.14}\{7.1\}) = (0,91 \cdot 0,97 \cdot 0,80 \cdot 0,92) \cdot 100 = 65\%$$

Synthesis

3.4.3.4.2 Synthesis of **3.14**{2.1}

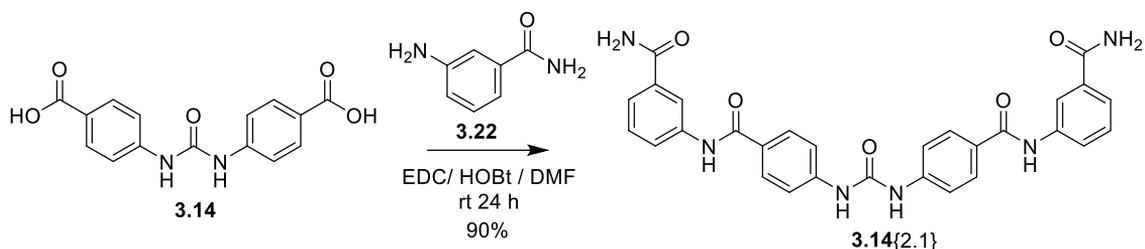


Fig 3.49. 4,4'-(carbonylbis(azadiyl))bis(*N*-(3-carbamoylphenyl)benzamide)(**3.14**{7.1}).

3.14{7.1} was synthesized using the same conditions. Spectroscopy corroborated the formation of the desired bond between the spacer and the two terminal groups.

$$\eta(\mathbf{3.14}\{7.1\}) = (0,91 \cdot 0,97 \cdot 0,80 \cdot 0,90) \cdot 100 = 63\%$$

3.4.3.4.3 Synthesis of **3.32**

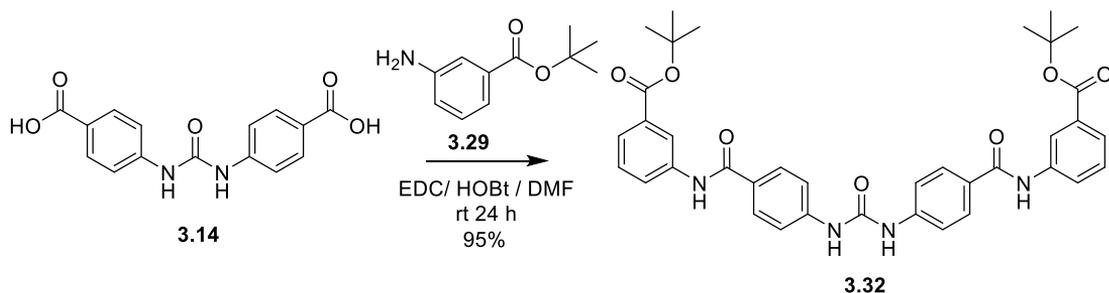


Fig 3.50. di-*tert*-butyl 3,3'-(4,4'-(carbonylbis(azadiyl))bis(benzoyl))bis(azadiyl)dibenzoate(**3.32**).

The last of the compounds synthesized using this methodology is **3.32**. This compound had already been synthesized using DCC, but it was still decided to perform the coupling to verify that the methodology was adequate with all selected terminal groups. Moreover, the **3.32** is formed and isolated without incident.

3.4.3.5 Post coupling transformations

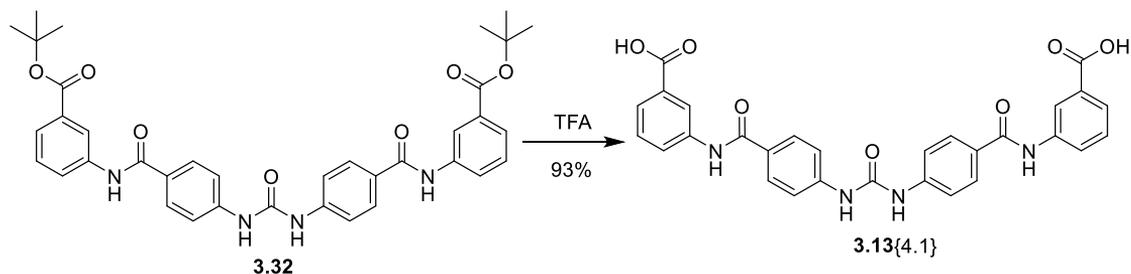


Fig 3.51. 3,3'-((4,4'-(carbonylbis(azanediyloxy))bis(benzoyloxy))bis(azanediyloxy))dibenzoic acid(3.13{4.1}).

Finally, to obtain the last candidate of this group, **3.13{4.1}**, a last synthetic step is necessary to deprotect the tert-butyl groups located at both ends of the molecule. The reagent is suspended in TFA to perform this step, like the other deprotections described in the present work. **3.13{4.1}** instantly precipitates when dissolved in TFA and is isolated by filtration.

$$\eta(\mathbf{3.13}\{4.1\}) = (0,91 \cdot 0,97 \cdot 0,80 \cdot 0,95 \cdot 0,93) \cdot 100 = 62\%$$

3.5 Biological evaluation

Once the first family of candidates for DM1 described in this thesis has been obtained, it is decided to assess its biological activity. In vitro measurements were performed for this purpose. Besides, samples of these candidates were analyzed by Dr. Artero at the University of Valencia.

3.5.1 AID 2675

First, the selected and synthesized candidates are studied by the biochemical test developed by Chen et al. and adapted by us as described in the previous chapter (see Chapter 2).

In this specific case, the study's conditions are slightly different from those described in the final methodology (see chapter 2.4.1). First, the concentration of the candidates is notably higher than the one finally described. This means that some candidates with optical properties compatible with the test's FRET donors can somehow contribute to the result of time-resolved fluorescence. For **3.13**{8.1} and **3.14**{7.1} candidates, this phenomenon is evidenced as their fluorescence intensity values soar above the negative value (in theory, the maximum).

Moreover, these results occurred before observing the recombinant protein aggregation phenomenon GST-MBNL1-105-His used in the test (see chapter 2). However, it cannot be confirmed that in that batch of protein and at the time of testing that affected the results obtained, it probably was.

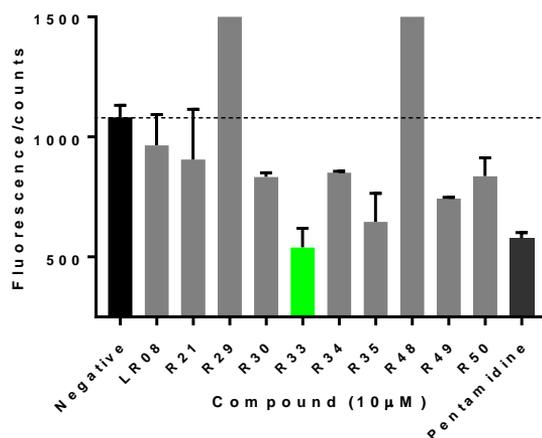


Fig 3.52. TR-FRET results for chapter 3 candidates.

In vitro tests showed positive results for some of the compounds studied, such as **3.13**{7.1} (R33 in Fig 3.52) and **3.13**{1.1} (R35), whose results are comparable to those of the positive control chosen to evaluate Groove binders, pentamidine. The most promising candidates were chosen to conduct a more thorough study using both polarization and cellular models.

3.5.2 Polarization fluorescence assays

With the aid of our colleagues from the University of Valencia, 4 candidates were studied in vitro and in cellulo.

First, the in vitro affinity of the selected compounds to CUG repeats was determined by performing fluorescence polarization spectroscopy experiments. A fluorescent (6-FAM-labeled) RNA probe containing 23 CUG repeats was incubated with increasing concentrations of the compounds to be tested. Whereas 6-FAM-CUG RNA molecules do not fluoresce in any polarization axis, binding to a molecule slows down the rotational movement of the molecule and increases its polarization values. In these experiments, we normalized the relative polarization to the values we obtained in the same experiment but using 100 μ M pentamidine as reference.

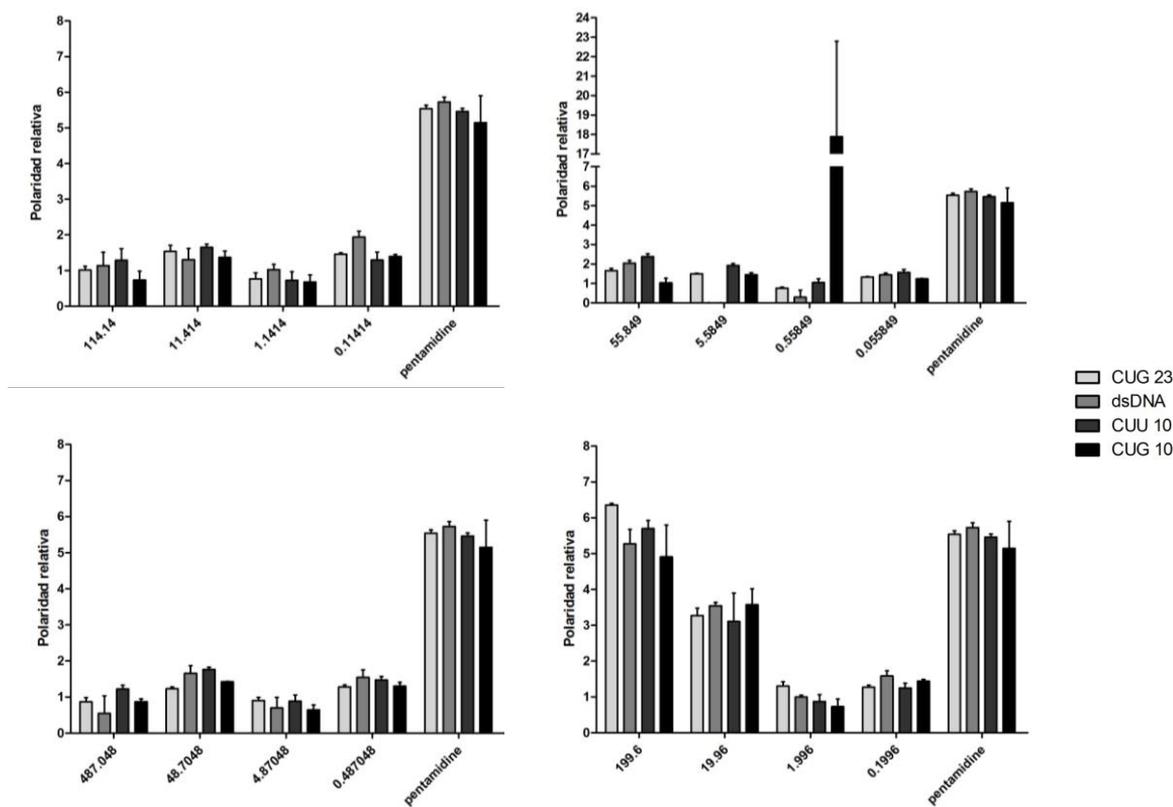


Fig 3.53. Polarization fluorescence results for: 3.14{4.1}, 3.13{7.1}, 3.13{9}, 3.13{1.1} respectively.

Simultaneously, the potency and selectivity of candidates were evaluated because they were tested against different nucleotide fragments, some related to DM1 (CUG), others not (CUU),

Biological evaluation

and their RNA-DNA selectivity (dsDNA). The results obtained show how most candidates do not significantly affect any of the nucleotide fragments presented, only in the case of **3.13**{1.1}, which has a particular ability to alter nucleotide sequences in a manner directly proportional to concentration. Besides, in the case of **3.13**{1.1}, values like those obtained for control are observed. However, these values are only obtained using high concentration value.

An aberrant value is observed in **3.13**{7.1}, without an associated trend in the concentration range that data is discarded for analysis.

Finally, one effect observed for all tested candidates (including control) is low selectivity. In fact, in all of them except for the aberrant value of **3.13**{7.1}, it can be observed that the values obtained for each concentration and candidate against the nucleotides studied (including dsDNA) are similar.

3.5.3 Cell biology evaluation

Finally, the compounds are studied using cell biology techniques to evaluate the ability to present significant activity against DM1 in a living system and the compound's possible toxicity. To do this, studies are carried out on the DM1 cell model developed by Dr. Furling's group.⁶¹

To investigate whether candidates had any impact on the number of foci present in a cell model of the disease, Artero group performed fluorescence in situ hybridization (FISH) to detect CUG RNA accumulations in the nuclei of normal and DM1 fibroblasts. Using an RNA fluorescent labelled probe, the CUG are highlighted in fluorescent microscopy. Then the nuclei and foci are manually counted and analyzed to obtain the results shown in Fig.3.54.

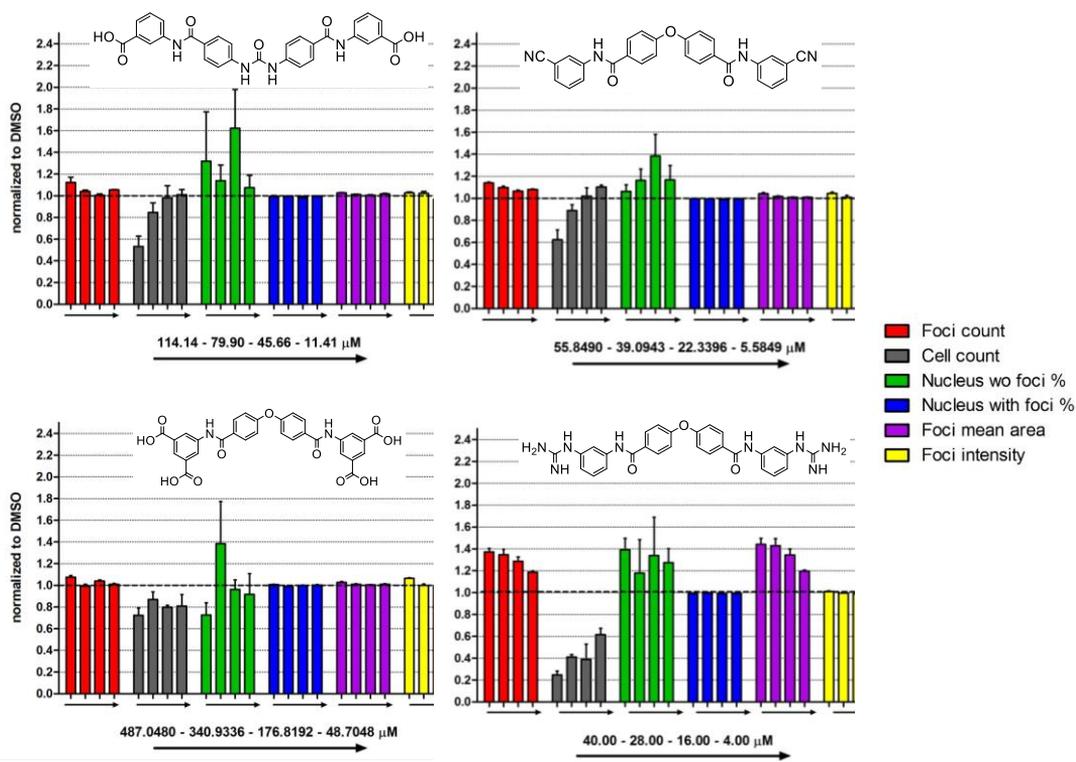


Fig 3.54. FISH results for: 3.14{4.1}, 3.13{7.1}, 3.13{9}, 3.13{1.1} respectively.

Preliminary results show some reduction in the number of cells that do not have foci for some compounds. Although none seem to be promising enough to be considered a lead different mechanisms of action are observed.

To sum up, these results suggest that, the interaction of these compounds with the CUG repeats is not sufficient to dissolve the foci. However, a slightly increase in the number of nuclei without any foci is observed in some cases. However, some candidates show an important decrease in cell count on microscopy fields. Therefore, a toxicity study is performed.

Finally, after observing low cell count results in the previous study, a toxicity study (MTT) is conducted to observe the actual effect on the proliferation of candidates for the cell model studied.

Biological evaluation

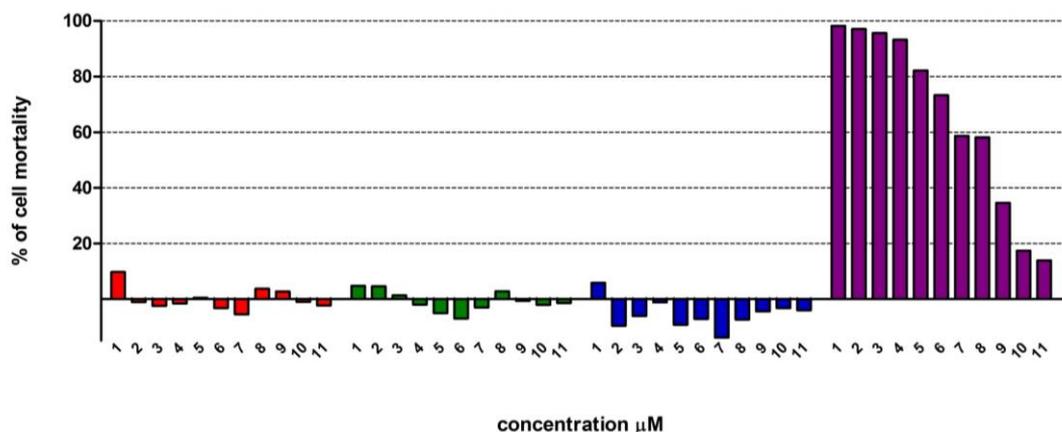


Fig 3.55. MTT values for:3.14{4.1}, 3.13{7.1}, 3.13{9}, 3.13{1.1}respectively

Tab 3.5. Concentrations used in the MTT assay represented in Fig 3.51. 1 refers to the maximum concentration in culture media without observing precipitation of the compound.

	3.14{4.1} / µM	3.13{7.1} / µM	3.13{9} / µM	3.13{1.1} / µM
1	114.14	55.85	487.05	199.60
2	103.87	50.82	443.21	181.64
3	93.59	45.80	399.37	163.67
4	83.32	40.77	355.54	145.71
5	73.05	35.74	311.71	127.74
6	62.78	30.72	267.88	109.78
7	52.50	25.69	224.04	91.82
8	42.23	20.66	180.21	73.85
9	31.96	15.64	136.37	55.89
10	21.68	10.61	92.54	37.92
11	11.41	5.59	48.70	19.96

The results show non-significant mortality for three of the candidates studied at relatively low concentrations. On the other hand, high toxicity is observed for the **3.13{1.1}** candidate at typical concentrations of a study range (1-10 µM). In addition, this result occurs in a concentration-dependent manner.

Since none of the candidates gets a promising activity result and probably the one with the best overall results is the same one with a high toxicity value, it is decided to end the preliminary study of these compounds and reevaluate such structures.

3.6 Further Studies

After analyzing the result obtained in the biological trials carried out on the structures presented in this chapter and its low anti-DM1 activity, it is decided to complete the study using an alternative design technique published in the research. Unlike the proposed in this chapter, the basis of rational design is using the results published in the largest HTS trial conducted in the DM1 (<https://link.springer.com/article/10.1007/s00216-011-5604-0>). These results were used to conduct a principal component analysis (PCA) study based on radial fingerprints. This involves switching from a structure-based design (SBDD) to a ligand-based design (LBDD) strategy.

Data from compounds tested by HTS in AID2675 were processed to try to obtain predictive structural information in that study.

To obtain a visual representation of the molecular space, a principal component analysis (PCA) was carried out using Canvas from the Schrödinger 2014 suite, considering the radial fingerprints properties.⁶²

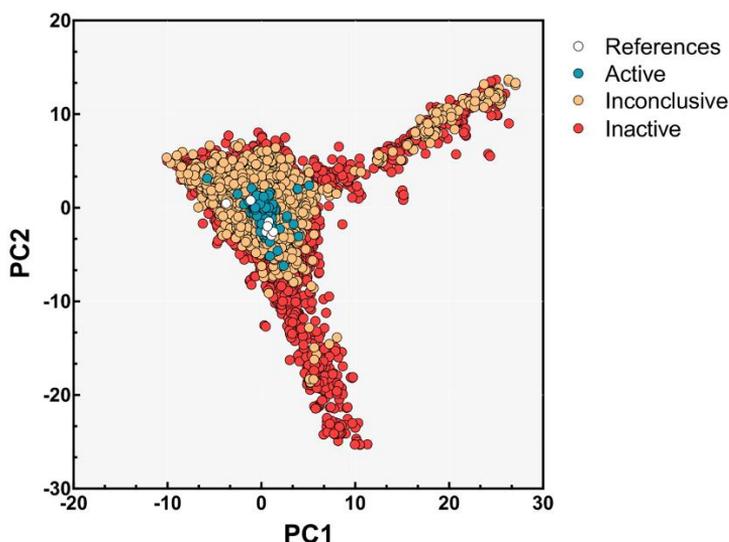


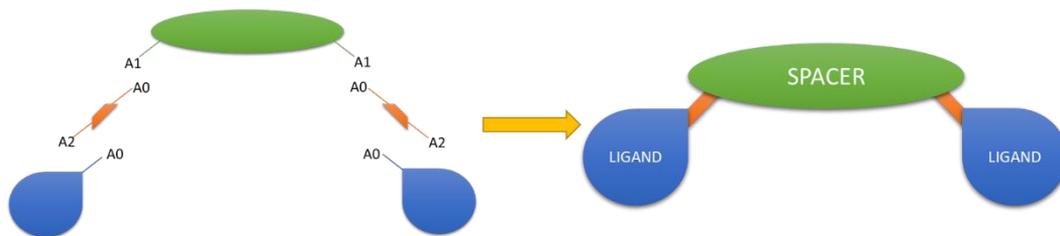
Fig 3.56. 2D radial fingerprint representation using AID 2675 HTS Database.³⁶

This technique allows rapid clustering of compounds based on binary molecular descriptors called fingerprints. The basis of the calculation is not directed by descriptors but based on

Further Studies

fingerprints. Thus, it is not possible to obtain the information that contributes in a more significant way to the positioning of the structures in the chemical space generated based on analysis of principal components.

A new chemical library is designed from scratch, following the same steps as at the beginning of this chapter. The new chemical library includes a new factor, the connector, given the possibility of generating different types of links and that they interact in some way with the nucleotides, decides to add this variable to the startup scheme. On the other hand, the symmetrical structure is maintained to obtain compounds that are somehow affordable in the laboratory.



$$169 \text{ spacers} \times 15 \text{ connectors} \times 37 \text{ ligands} = 93795 \text{ compounds}$$

Fig 3.57. New scaffold definition scheme.

Once the new chemical library of potential candidates is built using 169 ligands, 15 connectors, and 37 spacers (see Annexes), 93795 compounds are obtained. Simultaneously, the compounds classified as positive against DM1 are extracted from the study of A.López *et al.*³⁶ with those extracted from the combinatorial calculation, a two-dimensional representation based on radial fingerprints is generated in the same way as in the published study.

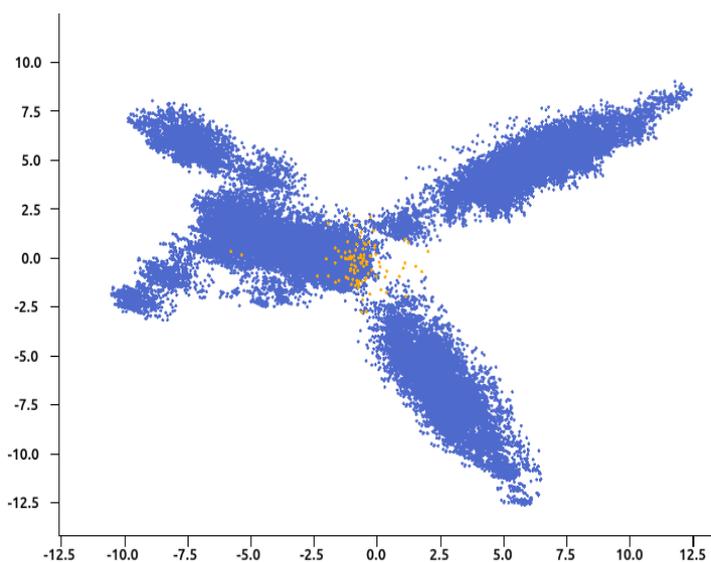


Fig 3.58. 2D radial fingerprint representation the new chemical library + reference compounds extracted from AID HTS assay. Blue spots are the actual chemical library. Orange spots are reference active compounds.

The obtained results show a scatter pattern like that described in the article. Interestingly, a new area of unexplored chemical space appears in the study to the left of the reference structures cluster. It was intended to analyze both those areas and those of the structures around the reference structures. Given the nature of the generated library, subgroups made up of thousands of potential candidates were obtained in both cases. For this reason, it was proposed to subdivide the chemical library into subfamilies, grouped by chemical properties.

1. Only ligands
2. 1 Ligand connected to a polar spacer
3. Non-ionizable dimeric compounds
4. ionizable dimeric compounds
5. rigid spacer (arylic) dimeric compounds

Fig 3.59. Segmentation of the chemical library into families. The overall chemical library has 85795 compounds.

The new chemical library was built so that compounds with potential different CUG interactions were classified separately. First, all binding groups were grouped, ignoring spacers and connectors to establish whether they could generate an interaction similar to the structures described than other modular structures with different chemical characteristics. Secondly, it was intended to include structures with a single end binding group and thus study the effect dimeric structures versus structures with one single binding group. Finally, the last three families of compounds were designed as in the previous stage, separating by properties of the spacer. The differential characteristics chosen were the polarizability of the spacer and aromatic groups that confided structural rigidity to the candidate.

After obtaining the new chemical library of 85785 compounds, the calculation and representation of radial fingerprints were carried out.

Further Studies

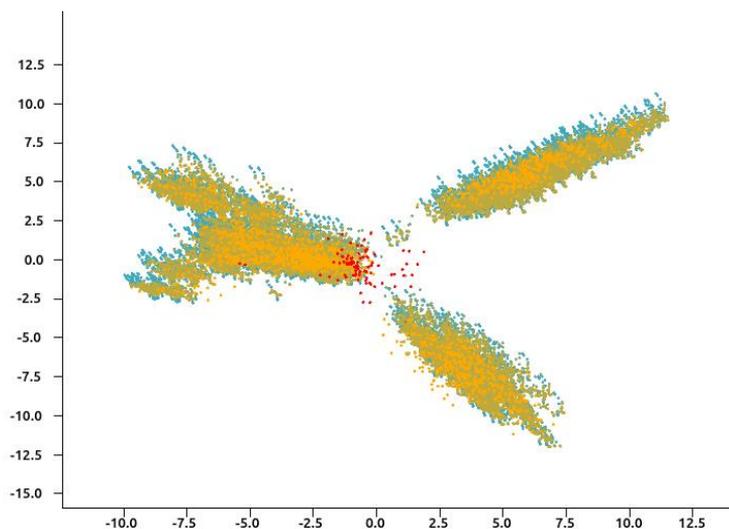


Fig 3.60. 2D radial fingerprint representation the new chemical library separated into five different families + reference compounds extracted from AID HTS assay.

As a result of this calculation, a complete dispersion of the five families was obtained throughout the chemical space, including the reference structures' area. This data was statistically analyzed using cluster analysis, verifying this same observation. This result terminates the search for structures by ligand-based design and discourages following the study of groove binder compounds.

3.7 Chapter discussion and perspectives

In this first approach to the design of candidates against DM1, a first family of compounds designed by analogy to pentamidine with potential groove binder activity was obtained. Throughout the study, some molecular design limitations were discovered when dealing with RNA. It is well known the challenge for these techniques to address the flexibility and structural variability of RNA.³⁰ On the other hand, many groups, including ours, have made advances in these studies, generating new routines, software, or specific scoring functions to treat this peculiar therapeutic target.^{5,63–65}

In this chapter, even using the most demanding docking parameters, many limitations in this technique's ability to reproduce a hairpin RNA's characteristics arose. This difficulty was solved with the addition of pharmacophoric keys guiding the structures to the correct anchor point.

At the synthetic level, most of the proposed structures could be successfully obtained, except for those with the amidine functional group, which, given the impossibility of having dry ammonia in gas form or generating it safely, could not be obtained. This is especially critical given that the reference compound used, pentamidine, bears two amidine groups in its structure.

On the other hand, there is an inherent problem with groove binders: their poor selectivity. This phenomenon is explainable given the way of interacting with such structures. Groove binders bind to the receptor through polar interactions in the grooves. These areas are composed mostly of sugars and phosphates, making it challenging to generate selective interactions.

For those reasons, it is decided to explore new strategies to obtain new candidates for DM1.

3.8 References

1. Tian, B. *et al.* Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *Rna* **6**, 79–87 (2000).
2. Lee, J. J. E. & Cooper, T. T. A. Pathogenic mechanisms of myotonic dystrophy. *Biochem. Soc. Trans.* **37**, 1281–1286 (2009).
3. Ranum, L. P. W. & Day, J. W. Pathogenic RNA repeats: An expanding role in genetic disease. *Trends Genet.* **20**, 506–512 (2004).
4. Warf, M. B., Nakamori, M., Matthys, C. M., Thornton, C. A. & Berglund, J. A. Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci.* **106**, 18551–18556 (2009).
5. Wong, C. H. *et al.* Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. *J. Am. Chem. Soc.* **136**, 6355–6361 (2014).
6. Kumar, A. *et al.* Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. *ACS Chem. Biol.* **7**, 496–505 (2012).
7. deLorimier, E. *et al.* Modifications to toxic CUG RNAs induce structural stability, rescue mis-splicing in a myotonic dystrophy cell model and reduce toxicity in a myotonic dystrophy zebrafish model. *Nucleic Acids Res.* **42**, 12768–12778 (2014).
8. Chen, C. Z. *et al.* Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* **402**, 1889–1898 (2012).
9. Herrendorff, R. *et al.* Identification of plant-derived alkaloids with therapeutic potential for myotonic dystrophy type I. *J. Biol. Chem.* **291**, 17165–17177 (2016).
10. Disney, M. D. Rational design of chemical genetic probes of RNA function and lead therapeutics targeting repeating transcripts. *Drug Discov. Today* **18**, 1228–1236 (2013).
11. Rzuczek, S. G. *et al.* Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **13**, 188–193 (2017).
12. Baisden, J. T., Childs-disney, J. L., Ryan, L. S. & Disney, M. D. Affecting RNA biology genome-wide by binding small molecules and chemically induced proximity. *Curr. Opin. Chem. Biol.* **62**, 119–129 (2021).
13. Cooper, T. A. Chemical reversal of the RNA gain of function in myotonic dystrophy. *Proc. Natl. Acad. Sci.* **106**, 18433–18434 (2009).
14. Disney, M. D., Childs, J. L. & Turner, D. H. Hoechst 33258 selectively inhibits group I intron self-splicing by affecting RNA folding. *ChemBioChem* **5**, 1647–1652 (2004).
15. Thomas, J. R. & Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **108**, 1171–1224 (2008).
16. Tor, Y. Targeting RNA with small molecules. *ChemBioChem* **4**, 998–1007 (2003).
17. Gallego, J. & Varani, G. Targeting RNA with Small-molecule drugs: Therapeutic promise and chemical challenges. *Acc. Chem. Res.* **34**, 836–843 (2001).

18. Meyer, S. T. & Hergenrother, P. J. Small Molecule Ligands for Bulged RNA Secondary Structures. *Org. Lett.* **11**, 4052–4055 (2009).
19. Guan, L. & Disney, M. D. Recent advances in developing small molecules targeting RNA. *ACS Chem. Biol.* **7**, 73–86 (2012).
20. LEONTIS, N. B. & WESTHOF, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, S1355838201002515 (2001).
21. Matsui, M. & Corey, D. R. Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* **16**, 167–179 (2017).
22. Lu, X. J., Shakked, Z. & Olson, W. K. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840 (2000).
23. Definitions and nomenclature of nucleic acid structure parameters. *EMBO J.* **8**, 1–4 (1989).
24. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).
25. Mulders, S. A. M., van Engelen, B. G. M., Wieringa, B. & Wansink, D. G. Molecular therapy in myotonic dystrophy: Focus on RNA gain-of-function. *Hum. Mol. Genet.* **19**, 90–97 (2010).
26. Miller, J. W. Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy. *EMBO J.* **19**, 4439–4448 (2000).
27. Strekowski, L. & Wilson, B. Noncovalent interactions with DNA: An overview. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **623**, 3–13 (2007).
28. Childs-Disney, J. L., Hoskins, J., Rzuczek, S. G., Thornton, C. A. & Disney, M. D. Rationally Designed Small Molecules Targeting the RNA That Causes Myotonic Dystrophy Type 1 Are Potently Bioactive. *ACS Chem. Biol.* **7**, 856–862 (2012).
29. Lee, M. M., Pushechnikov, A. & Disney, M. D. Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. *ACS Chem. Biol.* **4**, 345–355 (2009).
30. Disney, M. D., Lee, M. M., Pushechnikov, A. & Childs-Disney, J. L. The Role of Flexibility in the Rational Design of Modularly Assembled Ligands Targeting the RNAs that Cause the Myotonic Dystrophies. *ChemBioChem* **11**, 375–382 (2010).
31. Chakraborty, M. *et al.* Pentamidine rescues contractility and rhythmicity in a Drosophila model of myotonic dystrophy heart dysfunction. *Dis. Model. Mech.* **8**, 1569–1578 (2015).
32. Moreno, T., Pous, J., Subirana, J. A. & Campos, J. L. Coiled-coil conformation of a pentamidine–DNA complex. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 251–257 (2010).
33. Coonrod, L. A. *et al.* Reducing Levels of Toxic RNA with Small Molecules. *ACS Chem. Biol.* **8**, 2528–2537 (2013).
34. Jenquin, J. R. *et al.* Furamidine Rescues Myotonic Dystrophy Type I Associated Mis-splicing through Multiple Mechanisms. *ACS Chem. Biol.* **13**, 2708–2718 (2018).
35. González, À. L., Teixidó, J., Borrell, J. I. & Estrada-Tejedor, R. On the applicability of elastic network models for the study of RNA CUG trinucleotide repeat overexpansion. *PLoS One* **11**, 1–20 (2016).

References

36. González, À. L. *et al.* In silico discovery of substituted pyrido[2,3-d]pyrimidines and pentamidine-like compounds with biological activity in myotonic dystrophy models. *PLoS One* **12**, e0178931 (2017).
37. López González, A. In silico strategies for the design of RNA binders: focus on nucleotide repeat expansion disorders and HIV-1. (Universitat Ramon Llull, 2016).
38. Gallart, M. Estudi computacional de la interacció ARN-ligand orientat al disseny de fàrmacs per a la Distrofia Miotònica de tipus 1. (IQS, 2014).
39. Wong, C. H. *et al.* Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. *J. Am. Chem. Soc.* **136**, 6355–6361 (2014).
40. Pushechnikov, A. *et al.* Rational Design of Ligands Targeting Triplet Repeating Transcripts That Cause RNA Dominant Disease: Application to Myotonic Muscular Dystrophy Type 1 and Spinocerebellar Ataxia Type 3. *J. Am. Chem. Soc.* **131**, 9767–9779 (2009).
41. Jahromi, A. H. *et al.* Developing Bivalent Ligands to Target CUG Triplet Repeats, the Causative Agent of Myotonic Dystrophy Type 1. *J. Med. Chem.* **56**, 9471–9481 (2013).
42. Coonrod, L. A. *et al.* Reducing levels of toxic RNA with small molecules. *ACS Chem. Biol.* **8**, 2528–2537 (2013).
43. Schüller, A., Hähne, V. & Schneider, G. SmiLib v2.0: A Java-based tool for rapid combinatorial library enumeration. *QSAR Comb. Sci.* **26**, 407–410 (2007).
44. O'Boyle, N. M. *et al.* Open Babel: An Open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
45. Todeschini, R. & Consonni, V. *Molecular Descriptors for Chemoinformatics. Molecular Descriptors for Chemoinformatics* vol. 2 (2010).
46. Steffen, C. *et al.* PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **31**, 2967–2970 (2010).
47. Pascual, R., Borrell, J. I. & Teixido, J. Analysis of selection methodologies for combinatorial library design. *Sect. Title Hist. Educ. Doc.* **6**, 121–133 (2003).
48. Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2015.
49. Drewe, W. C. *et al.* Rational design of substituted diarylureas: A scaffold for binding to G-quadruplex motifs. *J. Med. Chem.* **51**, 7751–7767 (2008).
50. McKeever, C., Kaiser, M. & Rozas, I. Aminoalkyl derivatives of guanidine diaromatic minor groove binders with antiprotozoal activity. *J. Med. Chem.* **56**, 700–711 (2013).
51. Diez-Cecilia, E. *et al.* Guanidinium-based derivatives: Searching for new kinase inhibitors. *Eur. J. Med. Chem.* **81**, 427–441 (2014).
52. Hu, D. & Kluger, R. Functional cross-linked hemoglobin bis-tetramers: Geometry and cooperativity. *Biochemistry* **47**, 12551–12561 (2008).
53. Schotten, C. Ueber die Oxydation des Piperidins. *Berichte der Dtsch. Chem. Gesellschaft* **17**, 2544–2547 (1884).
54. Baumann, E. Ueber eine einfache Methode der Darstellung von Benzoësäureäthern. *Berichte der Dtsch. Chem. Gesellschaft* **19**, 3218–3222 (1886).
55. Chattopadhyay, G., Chakraborty, S. & Saha, C. Brine-Mediated Efficient Benzoylation

- of Primary Amines and Amino Acids. *Synth. Commun.* **38**, 4068–4075 (2008).
56. Reddy, K., Jenquin, J. R., Cleary, J. D. & Berglund, J. A. Mitigating RNA toxicity in myotonic dystrophy using small molecules. *Int. J. Mol. Sci.* **20**, 1–20 (2019).
 57. Ha, H. H., Kim, J. S. & Kim, B. M. Novel heterocycle-substituted pyrimidines as inhibitors of NF- κ B transcription regulation related to TNF- α cytokine release. *Bioorganic Med. Chem. Lett.* **18**, 653–656 (2008).
 58. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **40**, 2004–2021 (2001).
 59. Makovec, F. *et al.* Antiallergic and Cytoprotective Activity of New N-Phenylbenzamido Acid Derivatives. *J. Med. Chem.* **35**, 3633–3640 (1992).
 60. Himo, F., Demko, Z. P., Noodleman, L. & Sharpless, K. B. Why is tetrazole formation by addition of azide to organic nitriles catalyzed by zinc(II) salts? *J. Am. Chem. Soc.* **125**, 9983–9987 (2003).
 61. Arandel, L. *et al.* Immortalized human myotonic dystrophy muscle cell lines to assess therapeutic compounds. *DMM Dis. Model. Mech.* **10**, 487–497 (2017).
 62. Sastry, M., Lowrie, J. F., Dixon, S. L. & Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J. Chem. Inf. Model.* **50**, 771–784 (2010).
 63. Disney, M. D. *et al.* Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem. Biol.* **11**, 1720–1728 (2016).
 64. Ursu, A., Vézina-Dawod, S. & Disney, M. D. Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). *Drug Discov. Today* **24**, 2002–2016 (2019).
 65. Nguyen, L. *et al.* Rationally Designed Small Molecules That Target Both the DNA and RNA Causing Myotonic Dystrophy Type 1. *J. Am. Chem. Soc.* **137**, 14180–14189 (2015).

References

4 Pyrido[2,3-*d*]pyrimidine-based CUG ligands

This chapter describes the development of novel synthetic strategies and mechanisms related to the discovery of novel CUG ligands

Introduction

4.1 Introduction

4.1.1 Scaffold definition

After completing our initial studies on novel myotonic dystrophy ligands, we selected pyrido[2,3-*d*]pyrimidin-7(8*H*)-ones **4.1** (Fig 4.1) as an interesting and promising scaffold for the next generation of anti-DM1 that could promote the recovery of natural levels of MBNL1. The structural complexity of the molecules needed for the interaction with the CUG repetitions, often based on modular repetitive scaffolds, pointed a direction of novel synthetic opportunities.

In this context, our initial strategy for using pyrido[2,3-*d*]pyrimidin-7(8*H*)-ones for the design of structures capable to interact with the CUG repetitions was to study the possibility that the binding groups could interact specifically with ribose moieties and or nucleobases in a selective manner increasing the potency and/or selectivity of those compounds.

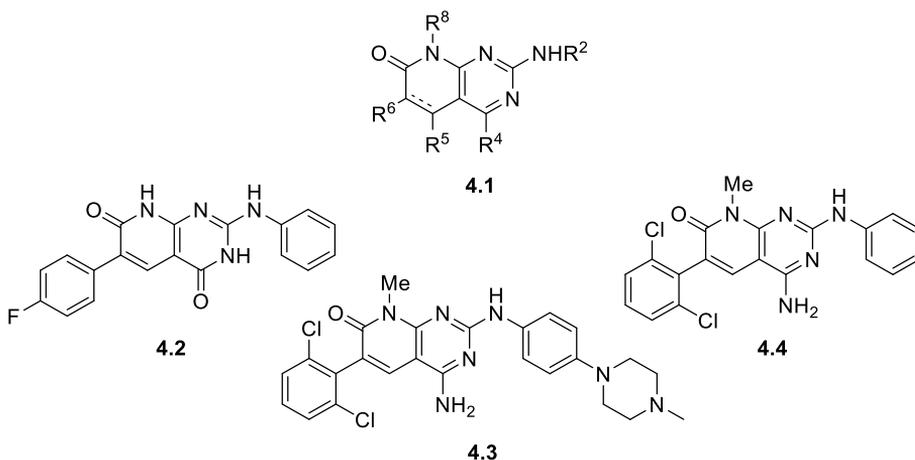


Fig 4.1. General pyrido[2,3-*d*]pyrimidinic scaffold **4.1** and biologically active pyrido[2,3-*d*]pyrimidin-7(8*H*)-ones **4.2**, **4.3** and **4.4**.

Pyrido[2,3-*d*]pyrimidin-7(8*H*)-ones are a bicyclic heterocyclic scaffold for which very interesting inhibitory activities have been described in the field of protein kinase inhibitors.¹⁻⁸ However, we have to take into account that in this work context, the actual target is not a protein. For this reason, some new chemical modifications to the basic pyrido[2,3-*d*]pyrimidin-7(8*H*)-one scaffold are required. Our group has described in the past years several straightforward strategies for the synthesis of 4-amino and 4-oxo substituted pyrido[2,3-*d*]pyrimidin-7(8*H*)-ones **4.1** (R⁴ = NH₂, OH) (Fig 4.1). From different starting materials up to 5 diversity centers and two possible degrees of unsaturation in the pyridone ring can be attained.⁹⁻¹²

Introduction

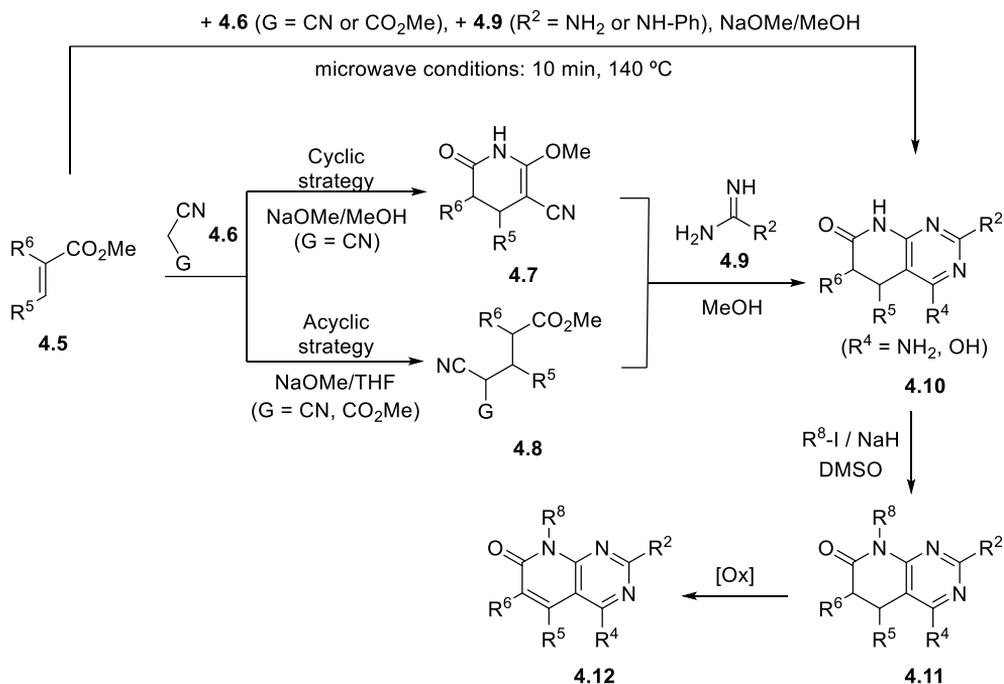


Fig 4.2. Scheme showing the already existing strategies pyrido[2,3-*d*]pyrimidine synthesis.

The synthesis of 5,6-dihydropyrido[2,3-*d*]pyrimidin-7(8*H*)-ones **4.11** is afforded from α,β -unsaturated esters **4.5** through three synthetic pathways as depicted in Fig 4.2. Therefore, in the so-called cyclic strategy 2-methoxy-6-oxo-1,4,5,6-tetrahydropyridin-3-carbonitriles **4.7** are obtained by reaction of an α,β -unsaturated ester **4.5** and malononitrile **4.6** (G = CN) in NaOMe/MeOH. Treatment of pyridones **4.7** with guanidine or aryl guanidines affords 4-amino-pyrido[2,3-*d*]pyrimidines **4.10** (R² = NH₂, NHAr) whether treating with amidine systems **4.9** (R² = H, alkyl, aryl, heteroaryl) afford non-4-amino derivatives. Alternatively, our group described an acyclic variation of the above protocol for the synthesis of pyridopyrimidines **4.10** (R⁴ = NH₂) based on the isolation of the corresponding Michael adduct **4.8** (G = CN, CO₂Me)¹³. Therefore, this protocol achieves 4-oxopyrido[2,3-*d*]pyrimidines **4.10** (R⁴ = OH) by treatment of intermediates **4.8** (G = CO₂Me), synthesized by Michael addition of α,β -unsaturated esters **4.5** and methyl cyanoacetate **4.6** (G = CO₂Me), with an amidine building block **4.9**¹⁴. One last approach implemented by our group is an efficient multicomponent reaction providing 4-amino or 4-oxopyrido [2,3-*d*]pyrimidines **4.10** (R⁴ = NH₂, OH) in a one-pot microwave-assisted cyclocondensation of α,β -unsaturated esters **4.5**, amidine or guanidine systems **4.9** and malononitrile or methyl cyanoacetate **4.6** (G = CN, CO₂Me) in NaOMe/MeOH.^{15,16}

Some post scaffold building modifications can be obtained to reach even more diversity. First, alkylation of lactam position N8 can be attained for non-heavy alkylic derivatives by treatment of **4.10** with the corresponding alkyl halide and NaH. Furthermore, an extra unsaturation

between the α - β carbonyl positions in the pyridone ring to yield **4.11**, the oxidation can be afforded by diverse methodologies including MnO_2 , Se/SeO , NaH/DMSO or $\text{Pd(C)}/\text{decaline}$.¹⁷

Although the synthetic methodologies for such structures are well established, there is room to expand their modifications. One of the requirements for designing structures that bind to RNA is that they must be modular to be attached, in one side, to a backbone structure or spacer and, on the other side, they have to contain a functional group capable of interacting with the backbone of the RNA. Such requirements are in agreement with the characteristics of pyrido[2,3-*d*]pyrimidin-7(8*H*)-ones which allow the synthesis of structures **4.12** with several orthogonal reactive positions that can be used to obtain any desired core. Nevertheless, it must be pointed out that such scaffold is highly substituted with highly polar groups, not being in consequence trivial to introduce further reactive groups to the pyridopyrimidine skeleton without affecting the pre-existing functionalities.

4.1.2 Novel pyrido[2,3-*d*]pyrimidin-7(8*H*)-one structures

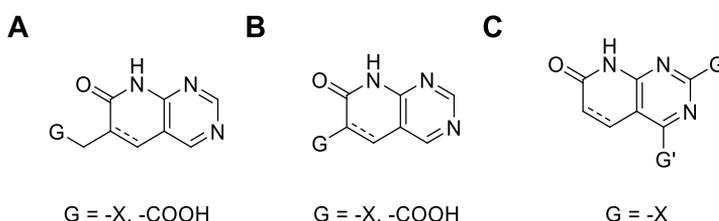


Fig. 4.3. Schematic representation of the three novel strategies developed in this chapter.

In the following pages, some strategies leading to novel pyridopyrimidines bearing reactive moieties are discussed. One interesting option is to add a reactive point at R^6 on structures **4.11** or **4.12** (IUPAC C6 position). There are several possibilities to be added in this position:

The first option is to add a methylene and a reactive group such as a halogen or even a carboxylic acid (Fig. 4.3.A). The resulting compound should allow a great variety of novel chemical derivatizations via nucleophilic substitution on a sp^3 carbon or acyclic reactions in case of the carboxylic derivative. The $-\text{CH}_2\text{-X}$ group could allow theoretically easy reactions with nucleophilic species creating a plethora of novel compounds via direct substitution or open the door to click chemistry by the introduction of the corresponding reactive groups.

The second option would be the direct attachment of a reactive group at the C6 position (Fig. 4.3.B). Halogens or carboxylic acids would allow reactivity without enlarging the system. However, directly bound to the heterocycle, the introduced groups may be less reactive.

Finally, synthesis of pyrido[2,3-*d*]pyrimidines with reactive groups at R^2 or R^4 of **4.12** could be another option (Fig 4.3.C). One example of that could be to obtain products bearing a halogen

Introduction

at R² (Br, Cl). An already proven viable synthesis developed in our group led to the selective synthesis of such kind of substituted compounds or the C4-Br derivative with regioselectivity. This later type of compounds are crucial intermediates to a novel orthogonal synthesis approach leading to the diversification of the available compounds via organometallic cross-coupling reactions. To reach this step, first a deep study on the halogenation of these structures is mandatory, as the previous described data only is able to obtain such scaffold halogenated in R² and R⁴ positions.¹⁴ In the following chapter, all these options will be analyzed.

4.2 Objectives

To obtain a new generation of DM1 CUG RNA binders using pyrido[2,3-d]pyrimidines as main scaffold.

To design novel strategies to introduce reactive groups to the well-established pyrido[2,3-d]pyrimidine core.

To study the intrinsic reactivity of these structures, raising knowledge on their behavior in unexplored reaction conditions.

Specifically study the way to obtain halogenated or carboxylic derivatives to allow the connection to a spacer.

Results

4.3 Results

4.3.1 Synthesis of DM1 CUG RNA binders using 2 and 4-halopyrido[2,3-*d*]pyrimidines

4.3.1.1 Previous work

Within the L. Ros PhD thesis framework,¹⁸ a group of symmetric pyrido[2,3-*d*]pyrimidine compounds **4.14** were synthesized. A *p*-xylylene spacer separated two heterocyclic scaffolds. These compounds were designed to target HIV as CXCR4 receptor inhibitors.¹⁹ However, further studies within the current project showed that one of these compounds was relatively active in DM1 therapy.²⁰ The most promising candidate was a derivative from (**4.14.1**, R¹ = H, R² = Me).

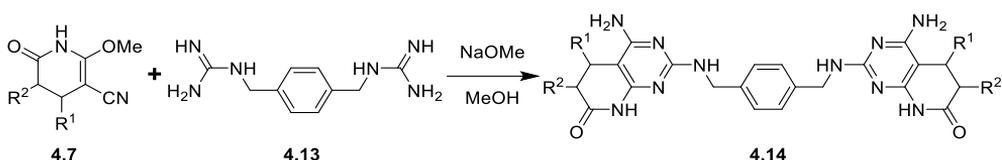


Fig 4.4. General procedure for the obtention of **4.14** structures.

However, this itinerary led to a complex isolation process of the final compound since the double cycloaddition of the pyridine ring to the *p*-xylylenbisguanidine **4.13** produced a series of byproducts of difficult separation.

The bisguanidine bidentate spacer was obtained by reaction of *p*-xylylenediamine **4.17** with *S*-methylthiourea sulfate, yielding the salt of the compound **4.13** displayed in Fig 4.4. On the other hand, the 2-methoxy-6-oxo-1,4,5,6-tetrahydropyridin-3-carbonitrile **4.7** was obtained by the classical methodology developed by the group.²¹

Given the need to obtain higher amounts of this compound for further biological activity testing, we decided to redesign the synthetic pathway to improve the reproducibility of the procedure.

4.3.1.2 Synthesis of DM1 CUG RNA binders using 2-bromo and 2-chloro pyrido[2,3-*d*]pyrimidines

The new strategy for the synthesis of **4.14.1** wants to skip the double cycloaddition step by avoiding the synthesis of the corresponding bisguanidine **4.13**. The proposed alternative was to afford **4.14.1** through a double nucleophilic substitution. Specifically, xylylenediamine **4.17** is used as a nucleophile, and the halogenated derivative at position C2 of the corresponding

pyrido[2,3-d]pyrimidine acts as an electrophile in the final reaction. The following scheme displays the retrosynthetic analysis corresponding to the alternative proposed for the synthesis of **4.14.1**.

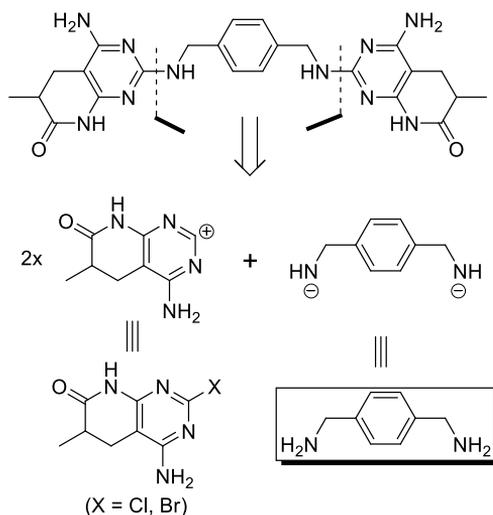


Fig 4.5. Retrosynthetic analysis to afford **4.14.1** using a nucleophilic di-substitution step.

Our group has previously described the synthesis of the corresponding C2 halogenated pyrido[2,3-d]pyrimidine, key intermediate of this protocol.¹⁴

4.3.1.3 New synthesis of 4.14.1

The new synthetic pathway for the synthesis of **4.14.1** is displayed in Fig 4.6.

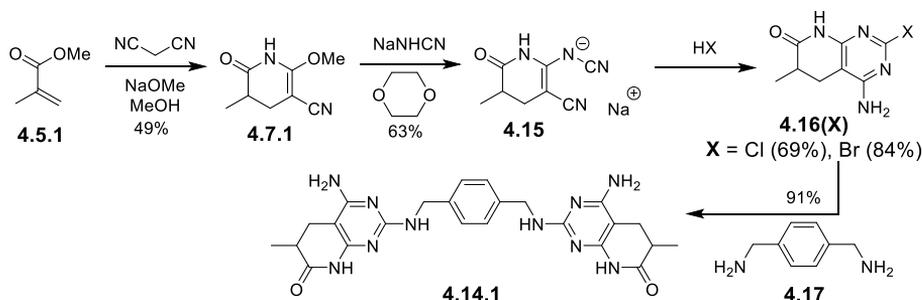


Fig 4.6. Reformulated synthesis of **4.14.1**.

The designed route consists of four synthetic steps, from which only the last one has not been previously explored. The results obtained in this new approach for the synthesis of **4.14.1** are described below.

Results

4.3.1.4 Synthesis of pyridone **4.7.1**

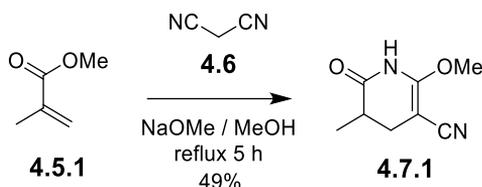


Fig 4.7. Synthesis of **4.7.1** following the cyclic strategy.

In this first step, methyl methacrylate **4.5.1** is allowed to react with malononitrile in the presence of sodium methoxide using anhydrous methanol as a solvent. Under these conditions a multistep reaction occurs in a single synthetic step. The Michael-type addition between malononitrile and methacrylate is followed by the cyclization of one of the nitrile groups onto the carboxylic ester group with the simultaneous addition of methoxide to the carbon atom of this nitrile to afford the final pyridone ring. Using this reaction conditions **4.7.1** is obtained in 49% yield. As stated above, our group has experience in both single step transformations leading to final pyrido[2,3-*d*]pyrimidines and the multistep shortcuts (See cyclic strategy in 4.11).^{16,22}

4.3.1.5 Synthesis of **4.15** by cyanamide substitution

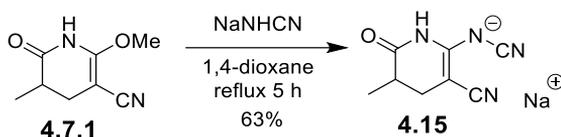


Fig 4.8. Reaction affording the sodium salt **4.15**.

The second reaction to obtain **4.14.1** is the replacement of the methyl ether in **4.7.1** position C2 by sodium cyanamide through and ethylene nucleophile substitution. The reaction occurs without the presence of an additional base. However, a small amount of methanol must be added to initiate the reaction. Upon transformation, the salt product formed is not soluble. Filtering the reaction mixture and washing with dioxane and diethyl ether is sufficient to afford the compound **4.15** in 63% yield.

4.3.1.6 Synthesis 2-halopyrido[2,3-*d*]pyrimidines **4.16**

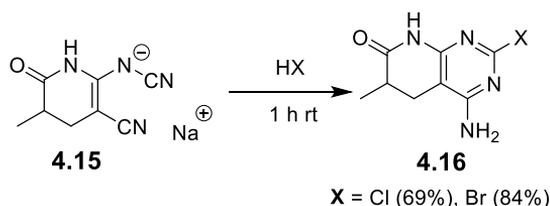


Fig 4.9 2-halopyrido[2,3-*d*]pyrimidine **4.16** formation.

The last synthetic step for the construction of the pyrido[2,3-*d*]pyrimidine **4.16** substituted in position C2 by a halogen atom (chlorine or bromine) consists on the cyclization with the corresponding hydrogen halide of the cyanamide substituted compound **4.15**. Since this is a synthetic step with two variants, these are described separately.

4.3.1.6.1 2-chloropyrido[2,3-*d*]pyrimidine **4.16.1**

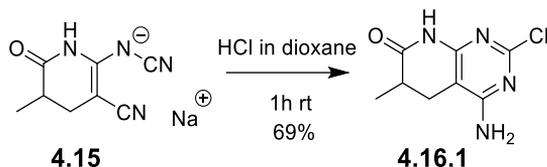


Fig 4.10. 4-amino-2-chloro-6-methyl-5,8-dihydropyrido[2,3-*d*]pyrimidin-7(6*H*)-one **4.16.1**.

To synthesize compound **4.16.1**, **4.15** is dissolved into a solution of hydrogen chloride (4M) in 1,4-dioxane. The reaction flask is immersed in a water-ice bath to absorb the excess heat generated by the reaction. Once stabilized the temperature, the mixture is stirred for an hour at room temperature. The reaction mixture is neutralized with aqueous ammonia, filtered, washed with deionized water, and oven-dried to afford **4.16.1** in 69% yield. The spectroscopy of the obtained product is in accordance with the previously described.²¹

4.3.1.6.2 2-bromopyrido[2,3-*d*]pyrimidine **4.16.2**

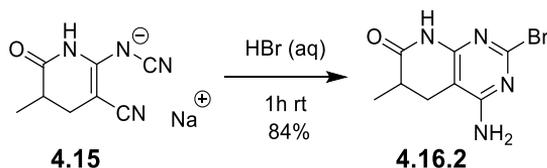


Fig 4.11. 4-amino-2-bromo-6-methyl-5,8-dihydropyrido[2,3-*d*]pyrimidin-7(6*H*)-one **4.16.2**.

The synthesis of the 2-bromo substituted pyridopyrimidine **4.16.2** is carried out analogously to the **4.16.1** but using 48% hydrobromic acid instead of the hydrogen chloride solution.

This reaction exhibits a thermodynamic-kinetic control phenomenon. Temperature is, therefore, a critical factor in the formation of the desired compound. The kinetic isomer (**4.16.2**, the one needed for our purpose) is the product with the bromine atom at position C2. On the other hand, the group has also described reaction conditions to obtain the thermodynamic isomer, the 4-bromo-derivative **4.18**.¹⁴ The following scheme shows the two compounds that may result from the reaction between **4.15** and hydrobromic acid.

Results

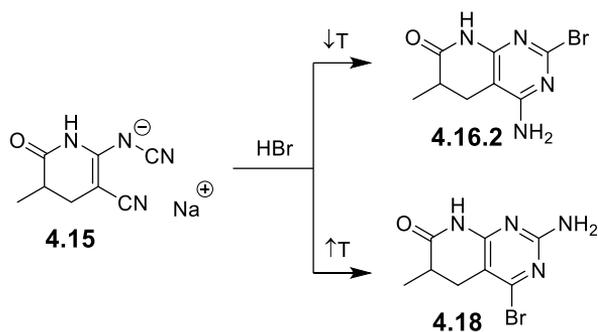


Fig 4.12. Pyrido[2,3-*d*]pyrimidine isomers obtained in this kinetic-thermodynamic effect controlled transformation.

The reaction occurs very fast, consequently it is necessary to carry it in a low-volume flask to control the temperature during the reaction to maximize heat transfer outwards and avoid overheating issues. Using this reaction conditions **4.16.2** was obtained in 84% yield.

4.3.1.7 Dual coupling of **4.16.2** with xylylenediamine

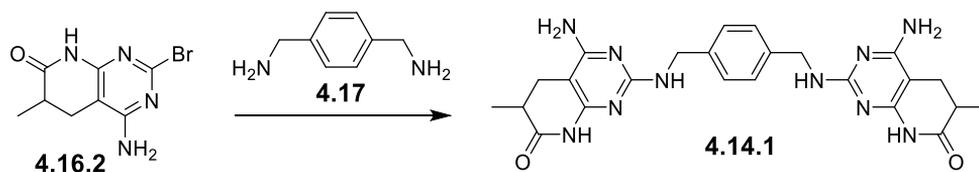


Fig 4.13. Reaction scheme for **4.14.1** obtention.

The last of the steps corresponding to the redesigned synthesis of **4.14.1** is the coupling between the symmetric nucleophilic amine *p*-xylylenediamine **4.17** and the 2-halosubstituted pyridopyrimidine. Finally, we decided to use the 2-bromosubstituted compound **4.16.2** due to its greater reactivity than the 2-chloro substituted derivative. To design the reaction conditions, the physicochemical properties of the reactants play a critical role. Thus, the low solubility of both starting materials, of the possible products and by-products could difficult the purification of the final product as it happened in the previous synthesis of **4.14.1**. Such low solubility is caused by the high number of hydrogen bond donors and acceptors included in the pyrido[2,3-*d*]pyrimidine scaffold. The high polarity of such structures renders also difficult their chromatographic separation. Therefore, the desired reaction conditions should be selective, forming only the disubstitution product. Thus, if the advance of the reaction stalls in the monosubstitution or the reaction produces a mixture of two or three components (starting materials, monosubstituted and disubstituted compound), those reaction conditions should be discarded. Bearing this in mind, the only fixed variable of this reaction condition is the equivalents of the two reactants (2 equivalents of **4.16.2** and 1 of xylylenediamine). The variables that can be modified are the reaction time, the temperature, the type of base and equivalents, the reaction solvent, and the heating method. After some attempts, DMSO was selected as the

solvent and sodium carbonate as the base. The use of sodium carbonate as base simplifies the control of the advance of the reaction by NMR.

Although the complete transformation was not obtained in the first trial at 160 °C during 6 h, the signals in the NMR spectrum are reduced to two groups, those of **4.16.2** and those corresponding to a second compound, which appear in all dioxane tests in small amounts. After that, it was decided to conduct a trial in high energetic conditions with DMSO under MW irradiation at 190 °C for 10 hours.

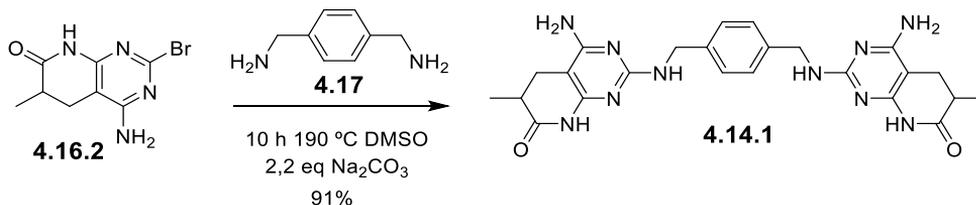


Fig 4.14. Final reaction conditions for **4.14.1** synthesis.

The NMR spectrum showed the complete transformation of the reactants into a single product, which after a process of digestion in methanol could be verified to be **4.14.1** (91% yield, 23% overall).

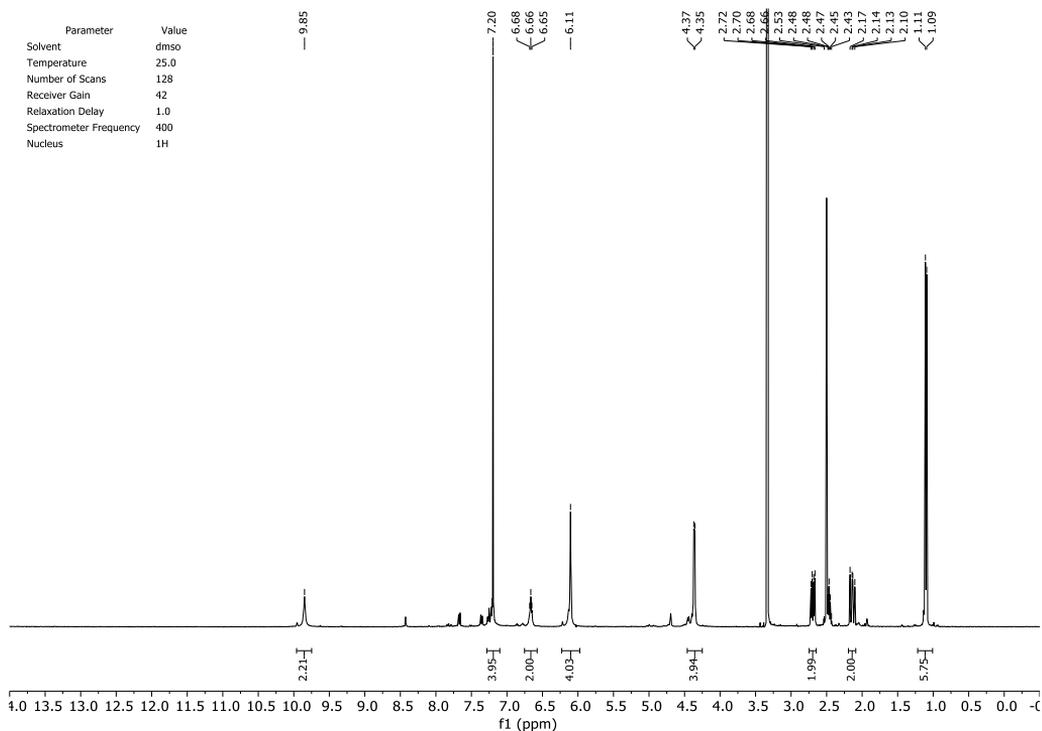


Fig 4.15. NMR spectrum of **4.14.1**.

Results

4.3.1.8 Study of the Suzuki Coupling on 2-bromopyridopyrimidine **4.16.2**

Having compound **4.16.2** in hand, we decided to explore its possible use in cross-coupling reactions with the aim of preparing other possible building blocks for the synthesis of DM1 CUG RNA binders. Among all described reactions, the Suzuki-Miyaura coupling was selected as the obtention of bis-pyridine or bis-pyrimidine moieties proven is of interest as potential anti-DM1 hits.

First, a proof of concept using a simple pyridine boronic acid and the 2-bromo pyrido[2,3-*d*]pyrimidine was set.

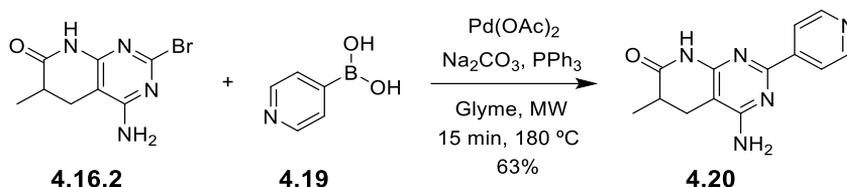


Fig 4.16. Concept Suzuki-Miyaura cross-coupling reaction using pyrido[2,3-*d*]pyrimidine scaffold **4.16.2** as halogenated starting material.

The transformation was successful in the following conditions: palladium acetate as the catalyst source, triphenylphosphine as ligand and reducer for the palladium, sodium carbonate as the base, and glyme as the solvent. The reaction mixture was heated in a microwave reactor at 180 °C for 15 minutes. The product was isolated by liquid-liquid extractions using dichloromethane, and although this product is not exceptionally soluble in such solvent, a 63% yield was obtained after several extractions.

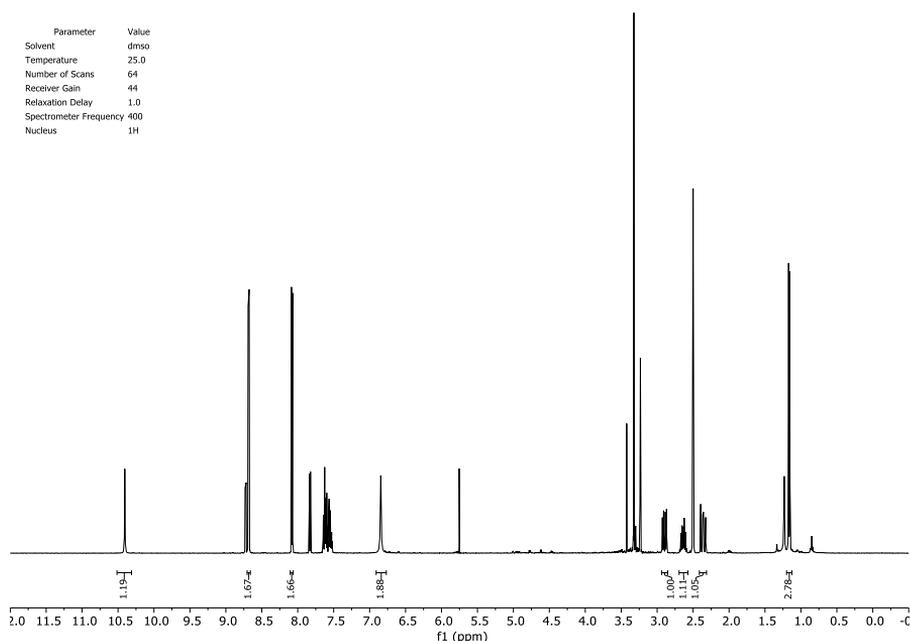


Fig 4.17. NMR Spectrum of **4.20**.

Although the product was not fully isolated of the triphenylphosphine oxide observed in Fig 4.17, bond formation was observed in 2-D NMR correlations. After that, we decided to synthesize compound **4.24** to be used in our spacer strategy in bidentate DM1 ligands.

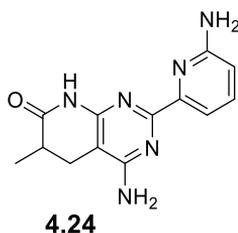


Fig 4.18. Structure of the objective base recognizer **4.24**.

To achieve compound **4.24**, the corresponding boronate must be prepared bearing the extra amino group protected. Therefore, 2-amino-6-bromopyridine **4.21** was selected as starting material (Fig 4.19).

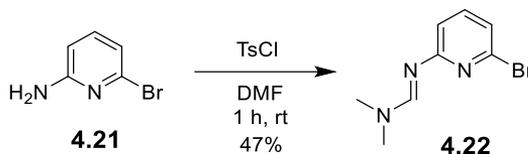


Fig 4.19. DMF-triggered protection of primary pyridinamine **4.21**.

The protection of such amino group is performed using DMF. to form the corresponding amidine **4.22** (Fig 4.19). Thus, TsCl is dissolved in DMF and stirred for 5 minutes. 2-amino-6-bromopyridine **4.21** is then added and allowed to react for 2 hours. After this time, the reaction mixture is basified with K_2CO_3 and the product is extracted with ethyl acetate. The solvent is then evaporated and **4.22** is afforded as a white solid in 47% yield.

The next step is the borylation the 6-bromo position of **4.22**. A Miyaura borylation is performed using B_2Pin_2 to attain the pyridineboronate **4.23**.

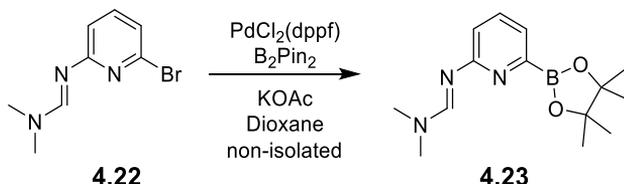


Fig 4.20. Boronate **4.23** synthesis through $\text{PdCl}_2(\text{dppf})$ catalyzed Miyaura borylation.

Then, a mixture of **4.22**, $\text{PdCl}_2(\text{dppf})$, B_2Pin_2 and KOAc are dissolved in dioxane and allowed to react at 80°C for 16 hours. After this time, the reaction mixture is diluted with AcOEt, and the suspension is filtered through a celite pad to remove palladium. The solvent was removed by evaporation, and a black oil was obtained. This oil was analyzed by NMR, and although other signals are observed due to various impurities, it was possible to identify the signals corresponding to the product. To purify the product, an attempt was made via flash

Results

chromatography. However, this mixture was not suitable for this kind of purification. The hydrolysis of the boronate to afford the corresponding boronic acid was also attempted but with no results.

Consequently, we decided to purify the desired final coupling product **4.24** (Fig 4.21) expecting that all impurities that may carry over from this reaction would be removed. We decided to react the pinacol boronate **4.23** with the pyrido[2,3-*d*]pyrimidine **4.16.2** since, although less reactive, boronates are suitable scaffolds for Suzuki reaction. Therefore, the Suzuki-type coupling reaction between **4.16.2** and boronate **4.23** was set. For this purpose, the procedure aforementioned with the synthesis of **4.20** was used.²³

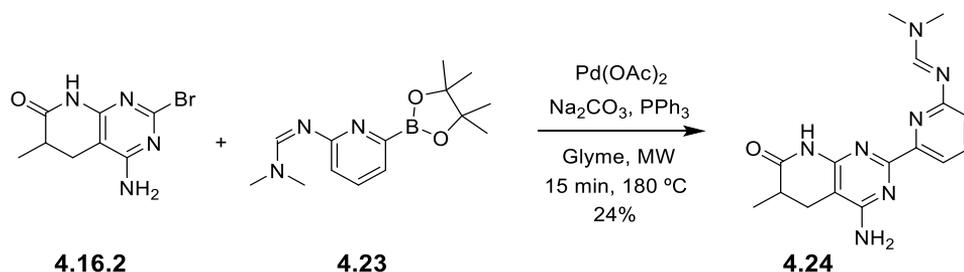


Fig 4.21. Suzuki cross-coupling C-C formation affording binder candidate **4.24**.

To attain **4.24**, sodium carbonate is dissolved in the minimum amount of water possible and added to a microwave vial containing the rest of the reactants (**4.16.2**, **4.23**, $\text{Pd}(\text{OAc})_2$, PPh_3) dissolved in glyme. The mixture is heated at 180 °C for 15 minutes. After that, the solvents are evaporated, the crude is dissolved in AcOEt and washed with water. The organic extract is evaporated, and a white solid is obtained. Finally, the crude product was purified by silica flash chromatography, yielding a white solid that was confirmed to be **4.24** (24% yield). This compound will allow the synthesis of a whole new family of DM1 drug candidates through binding two or more of these moieties to a central spacer.

Pyrido[2,3-d]pyrimidine-based CUG ligands

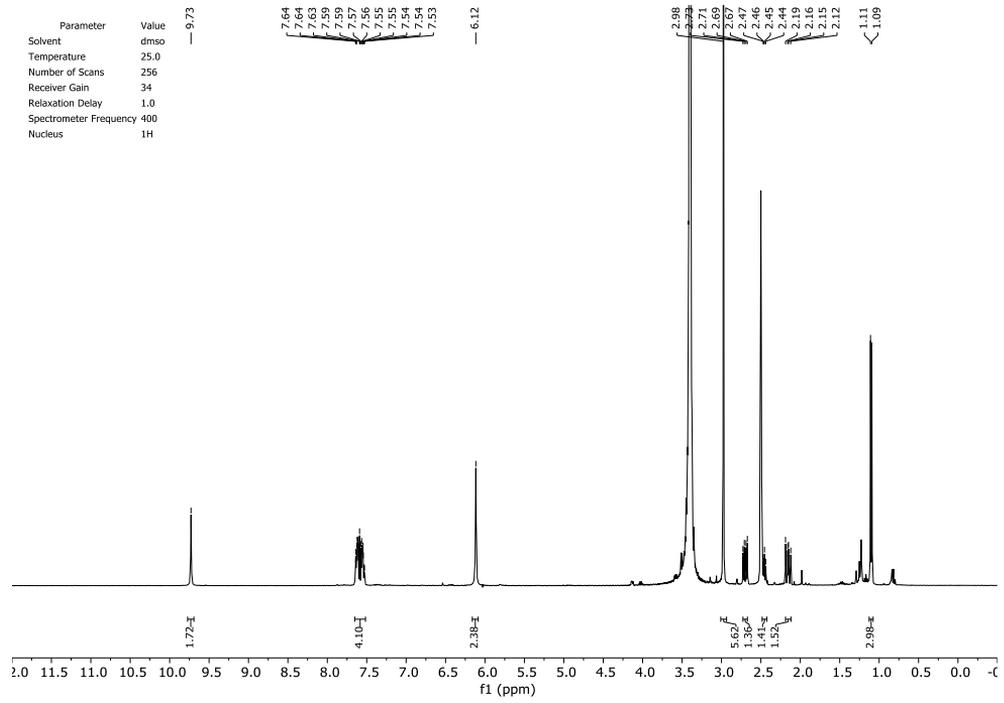


Fig. 4.22. ¹H-NMR spectrum of structure 4.24.

Results

4.3.2 Exploring the synthesis of 6-bromomethyl pyrido[2,3-*d*]pyrimidine

4.3.2.1 Hypothesis

A second proposal was designed to establish a reactive point in the skeleton of the pyrido[2,3-*d*]pyrimidine core **4.1**. This second approach studied the introduction of an electrophilic reactive point at position C6 with a methylene moiety at that position. Nevertheless, if that functionality is incompatible with performing the reactions required to form the pyrido[2,3-*d*]pyrimidine structure, the alternative is to approach the construction from inside to outside. First, attach the spacer to the pyrido[2,3-*d*]pyrimidine precursor and, afterward, assemble the heterocycle.

The two options proposed in this study require a multistep synthesis starting from methyl acrylate **4.25**. **4.25** will be condensed with formaldehyde in a Bayliss-Hillman umpolung coupling to afford the hydroxymethyl substituted acrylate **4.26**. Subsequently, the pyridone **4.29** could be directly obtained using the protocol described by our group, or alternatively the alcohol formed could be replaced by an active group for nucleophilic substitution, such as the bromo derivative depicted in **4.28**.

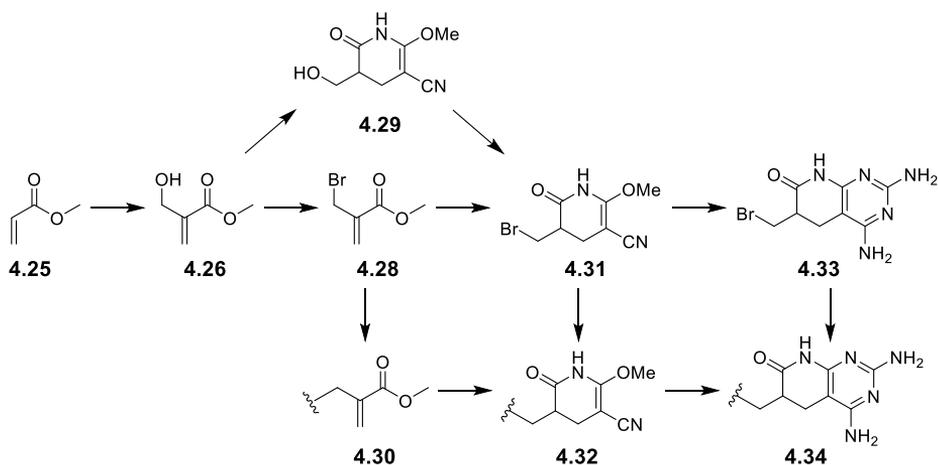


Fig.4.23. Scheme of the possible synthesis of C6-substituted pyrido[2,3-*d*]pyrimidines **4.34**.

At that point, the pathways could converge in obtaining the pyridone with the bromomethyl group at C6 **4.31**. As an alternative, if the construction of this structure is not possible, it is proposed to make the coupling with the spacer to form **4.30** (spacers with terminal nucleophiles such as alcohols or amines are used). Finally, the pyrido[2,3-*d*]pyrimidine **4.33** could be assembled and bonded to the central spacer to afford **4.34**, compound that could also be formed from the pyridone **4.32** bearing the previously attached spacer.

4.3.2.2 The Bayliss-Hillman reaction to afford **4.26**

The first step of our strategy is the formation of compound **4.26** in a carbonyl condensation between two a priori electrophilic compounds, **4.25** and formaldehyde. In order to carry out this umpolung coupling, the intervention of 1,4-diazabicyclo[2.2.2]octane (DABCO) is necessary. Its catalysis allows the polarity of the acrylate to be reversed, converting it into a nucleophilic enolate. After condensation of enolate and carbonyl, the bond with DABCO quaternary ammonium, is broken, returning it to its initial state and thus being available to perform its function again.²⁴

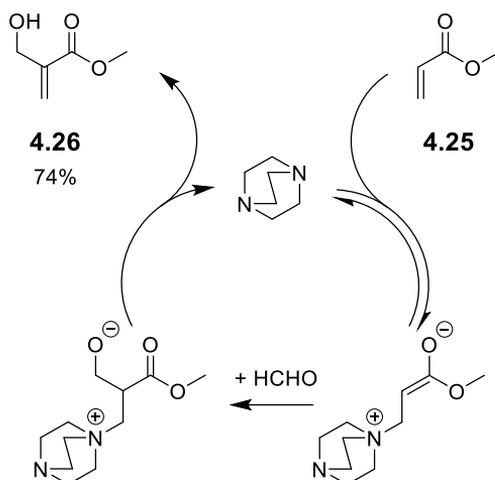


Fig 4.24. Catalytic Bayliss-Hillman umpolung enolic condensation.

This reaction is carried out in a mixture of 1,4-dioxane:H₂O (1:1). Furthermore, as the source of formaldehyde is selected a 0,1 M solution of formaldehyde (37% wt. in water). The reaction mixture is stirred under inert atmosphere for 24 hours at room temperature. The product is extracted from the reaction mixture with DCM and concentrated under reduced pressure to afford methyl 2-(hydroxymethyl)-acrylate **4.26** as a yellow oil (74 % yield). The spectroscopic data matches the reported in bibliography.²⁵

Results

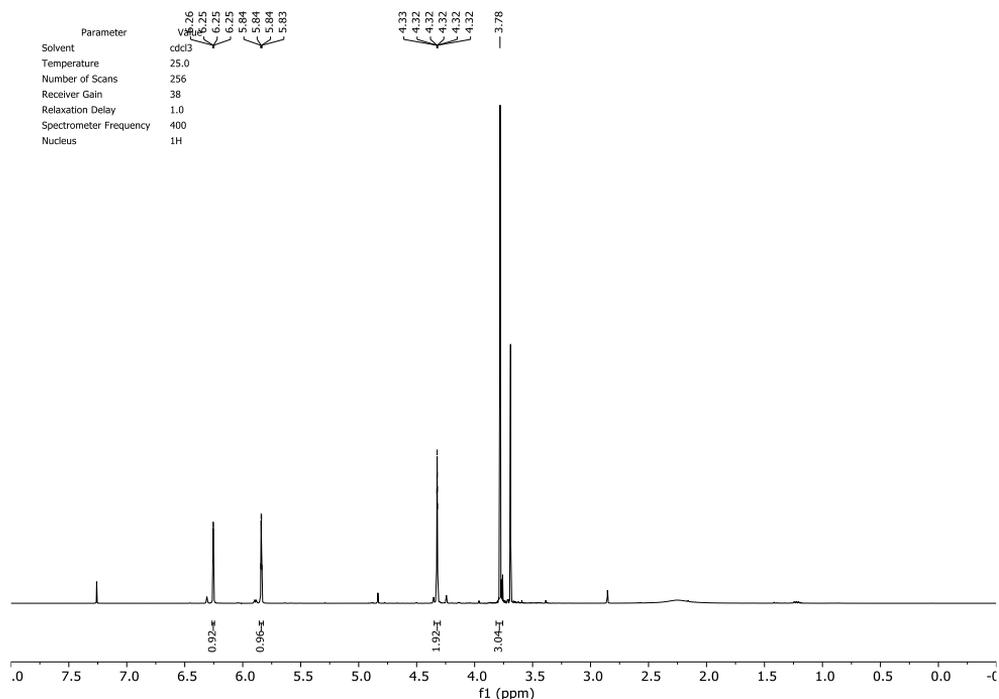


Fig 4.25. $^1\text{H-NMR}$ spectra of **4.26**.

The signal at 4.32 ppm in the $^1\text{H-NMR}$ confirms that the reaction has occurred as intended. It is a doublet that corresponds to the methylene formed in the formaldehyde condensation. Therefore, a new carbon-carbon bond has been formed.

As any synthetic acrylate, it must be kept away from light, as without the presence of radical scavengers they tend to polymerize with any radical initiator. Despite that and under the proper conditions, the product can be stored up for months.

After the synthesis of **4.26**, we attempted to obtain the 6-hydroxymethyl substituted pyridone **4.29**. However, the synthesis of this product is not feasible using the established methodology. For this reason, we decided to proceed with the synthesis of the bromomethyl derivative **4.28**.

4.3.2.3 Synthesis of methyl 2-bromomethylacrylate **4.28**

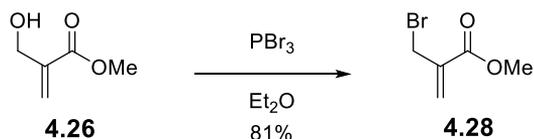


Fig 4.26. Phosphorous tribromide-driven alcohol substitution.

The conversion of the hydroxymethyl group of **4.26** in a bromomethyl group was carried out by stirring it with phosphorus tribromide in Et_2O for 3 hours at room temperature. The resulting

product is extracted out of the reaction mixture with DCM and concentrated under reduced pressure to afford methyl 2-(bromomethyl)-acrylate **4.28** as a yellow oil (81% yield). The spectroscopic data matches the reported in bibliography.²⁵

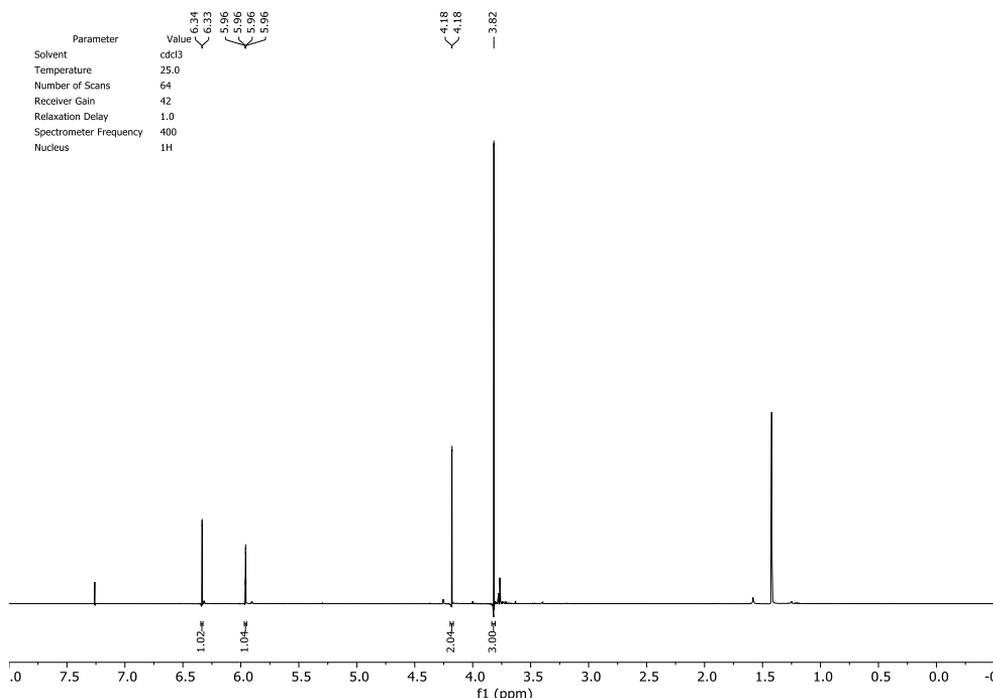


Fig 4.22. ¹H-NMR spectra of **4.28**.

The signal at 4.18 ppm in the ¹H-NMR spectrum confirms that the reaction has occurred as intended. It is a doublet that corresponds to the same methylene in carbonyl β position present in the spectrum of **4.26** but the signal suffers an upfield shift as bromine is less electronegative than oxygen. Therefore, the carbon hydrogen bond in this position is more shielded.

4.3.2.4 Williamson ether coupling between **4.28** and tetraethylene glycol

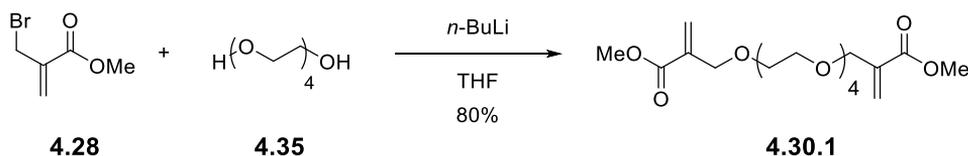


Fig 4.27. Dual S_N2 type substitution affording **4.30.1**.

Analogously to what happens with the hydroxymethyl substituted acrylate **4.26**, obtaining the bromomethyl substituted pyridone **4.31** from compound **4.28** is not feasible with the established methodologies. For this reason, of all the possible routes initially proposed, the only one that remains feasible is the one that links acrylate **4.28** to the spacer and then builds the pyrido[2,3-d]pyrimidine core.

Results

The spacer selected for this pilot synthesis should be a product with two reactive functionalities that do not preclude the subsequent reactions needed for the construction of the pyridopyrimidine heterocycle. Tetraethylene glycol **4.35** is selected assuming to be unreactive once the ether is formed. However, alcohols are not potent nucleophiles, so to link an alcohol to **4.28** in a Williamson reaction a strong base without reducing character is needed, in this case, *n*-BuLi. The S_N2 between the formed alkoxide and the bromomethyl group of **4.28** must occur dually at both ends of **4.35**.

Consequently, to a solution of tetraethylene glycol in anhydrous THF a 1.6 M solution of *n*-BuLi is added dropwise and the mixture allowed to react at 0 °C for 1 hour. After this period, 2 equivalents of **4.28** are added and the resulting mixture stirred at room temperature overnight. The resulting product is extracted off the mixture with DCM and concentrated under reduced pressure to afford dimethyl 2,18-dimethylene-4,7,10,13,16-pentaoxonadecanedioate **4.30.1** as a yellow oil (80% yield).

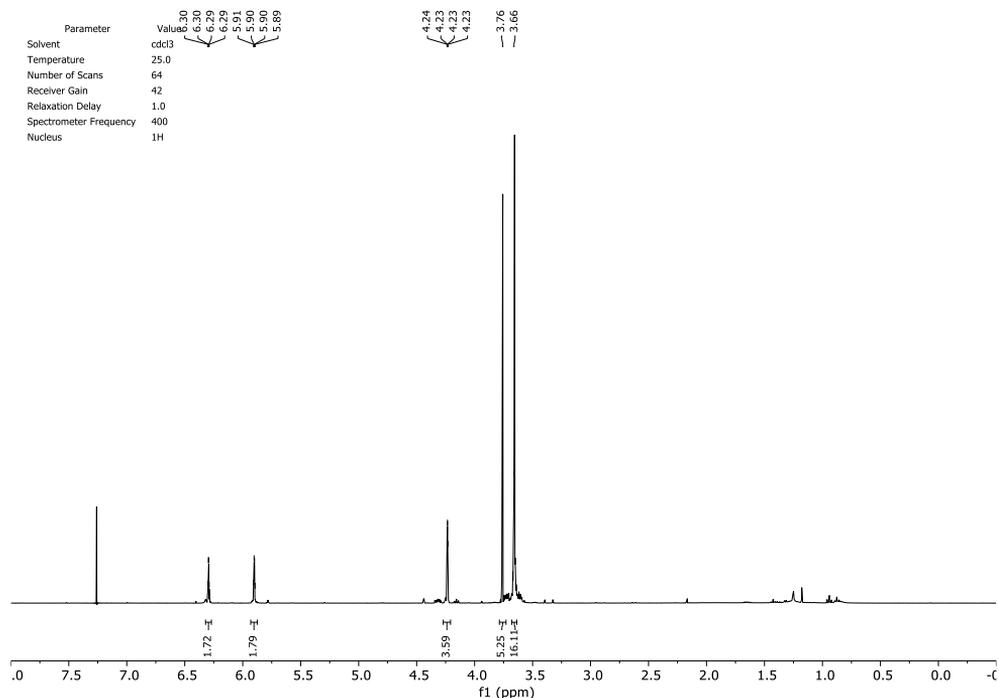


Fig 4.28. ¹H-NMR spectrum of **4.30.1**.

The signal at 3,66 ppm in the ¹H-NMR spectrum corresponding to 16H is the sum of the 8 methylene moieties present in **4.35**. Furthermore, the methylene group in carbonyl β position is again shifted to downfield as it is again bonded to an oxygen atom. **4.30.1** is characterized by NMR, IR and OEA.

After preparing **4.30.1**, several attempts were made to obtain the product containing pyridone rings at both ends of the spacer using the cyclic and acyclic strategies described by our group but without success. Furthermore, the synthesis was also attempted under one-pot conditions

that gave good results in similar cases. However, all attempts fail to reach the objective. For this reason, it was decided to conclude this part of the study and focus on obtaining halogenated compounds directly at the C6 position of core **4.1**.

Results

4.3.3 Findings in the synthesis of 6-bromo substituted pyrido[2,3-*d*]pyrimidines

4.3.3.1 Previous work

Given the need to establish new systems with potential nucleotide recognition activity, we decided to study the C6-halo substituted pyrido[2,3-*d*]pyrimidine systems. 6-bromopyrido[2,3-*d*]pyrimidine systems were previously described in the PhD of I. Galve,²⁶ focusing on systems with C2-phenylamino substituent and using bromine as a halogenating agent (see Fig 4.25, **4.36**). At that time, the product obtained from the reaction of pyrido[2,3-*d*]pyrimidine **4.36** with bromine was assigned to the formation of the ionic intermediate **4.37** whose reactivity was so high that it evolved spontaneously, even during the NMR spectroscopy recording, to the 6-bromosubstitutedpyridopyrimidine **4.38**. However, the structure of **4.37** was not clearly and unequivocally established. Therefore, and given the interest of the versatility of **4.38**, we decided to revisit these studies in order to establish the structure of the intermediate obtained, its synthetic possibilities and, if possible, use such approach to obtain anti-DM1 candidates.

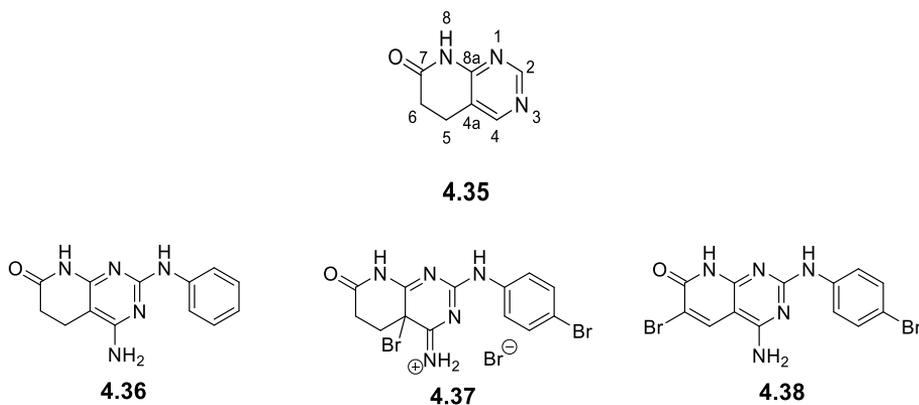


Fig 4.29. General structure of the pyrido[2,3-*d*]pyrimidine, focusing on the IUPAC numeration of each carbon **4.35**. Structure of the starting material for this study **4.36**, **4.37** structure initially assigned for the bromination of **4.36**, and transposition final product obtained **4.38**.

First, the unsubstituted pyrido[2,3-*d*]pyrimidine **4.36** is synthesized using the microwave-assisted multicomponent reaction protocol previously described by our group.^{15,27} Thus, a 2:1:1 molar mixture of methyl acrylate, malononitrile, and phenylguanidine carbonate in MeOH is heated at 140 °C under microwave irradiation in a sealed vial for 10 min. The reaction affords **4.36** in 51% yield. Such multicomponent microwave protocol allows the synthesis of a pyrido[2,3-*d*]pyrimidin-7(8*H*)-one not substituted at the pyridone ring. In our hands, this is the only protocol capable of affording such kind of structure because all other synthetic protocols developed by our group do not afford **4.36** due to the undesired reactivity of methyl acrylate. Once **4.36** was obtained, the bromination step was studied starting by using one equivalent

of bromine in AcOH for 3 h at room temperature. The reaction affords the *p*-bromo substituted derivative **4.39** in 97% yield. The structure is established based on the spectral data, particularly the presence in the $^1\text{H-NMR}$ spectrum of the characteristic pattern of *p*-substituted phenyl derivatives, of relative integral 2 H each, at 7.83 and 7.31 ppm, clearly showing that bromination has taken place in such position of the ring. No *o*-substitution is observed probably by steric hindrance of the pyridopyrimidine scaffold.

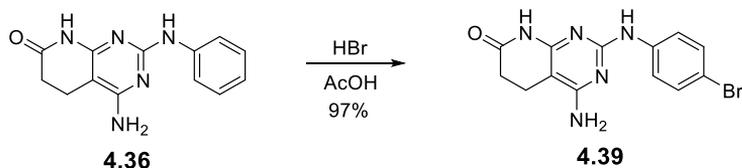


Fig. 4.30. Scheme of the monobromination reaction of **4.36**.

4.3.3.1.1 Establishing the real structure of intermediate **4.37**

Next, **4.36** was treated with 2 equivalents of bromine yielding very unstable products that do not evolve to **4.38**. Then, increasing the amount of Br_2/AcOH to 3 equivalents, an orange-red precipitate was abundantly formed. Such product evolves in a fast manner to **4.38** when dissolved, as stated by I. Galve, but is stable at room temperature if stored in solid state.

It has to be remembered that for such orange compound I. Galve proposed the structure of the salt **4.37** (see Fig 4.25), but in this work we were able to carry out a more accurate structural assignment on the basis of the full spectroscopy of such compound (see Figs 4.32 and 4.33 below). The real structure of **4.37** was finally assessed on the basis of a signal at 44.2 ppm in the $^{13}\text{C-NMR}$ spectrum assignable to the C4a carbon of the pyrido[2,3-d]pyrimidine skeleton which suffers a great upfield shift (from 85.9 ppm in **4.36**) due to the presence of the bromine atom. Furthermore, the HRMS (ESI-TOF) was correct for $\text{C}_{13}\text{H}_{11}^{79}\text{Br}_3\text{N}_5\text{O}$ ($[\text{M}+\text{H}]^+$: calculated: 489.8514; found: 489.8497) and the elemental analysis was also correct for $\text{C}_{13}\text{H}_{10}\text{Br}_3\text{N}_5\text{O}$. With such information in hand we can exclude the salt proposed in the PhD of I. Galve and to propose a totally covalently bonded compound **4.37** (Fig 4.31). Such structure would correspond to the Wheland intermediate of the bromination of the 2,4-diaminopyrimidine ring present in **4.36**, in a position that is the most favored for the SEAr in such kind of diamino substituted pyrimidines²⁸.

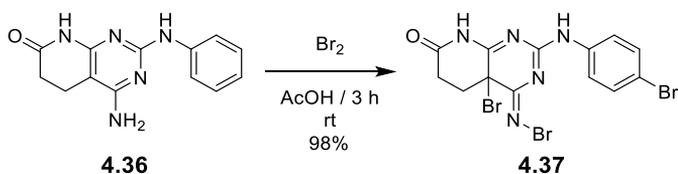


Fig 4.31. Scheme of **4.37** formation conditions.

Results

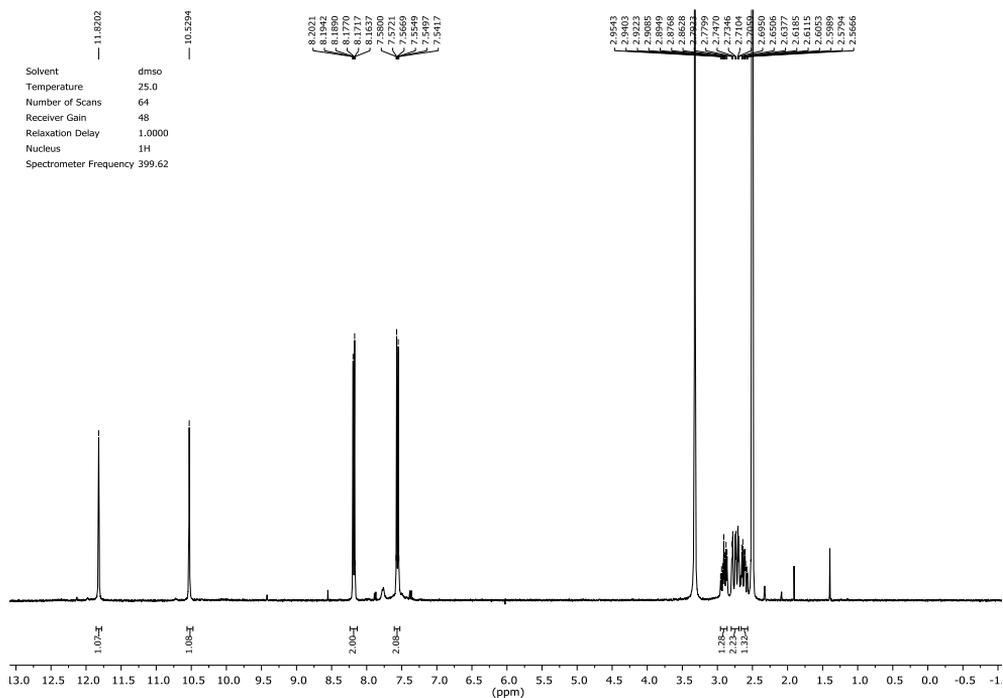


Fig. 4.32. ^1H -NMR spectrum of **4.37**.

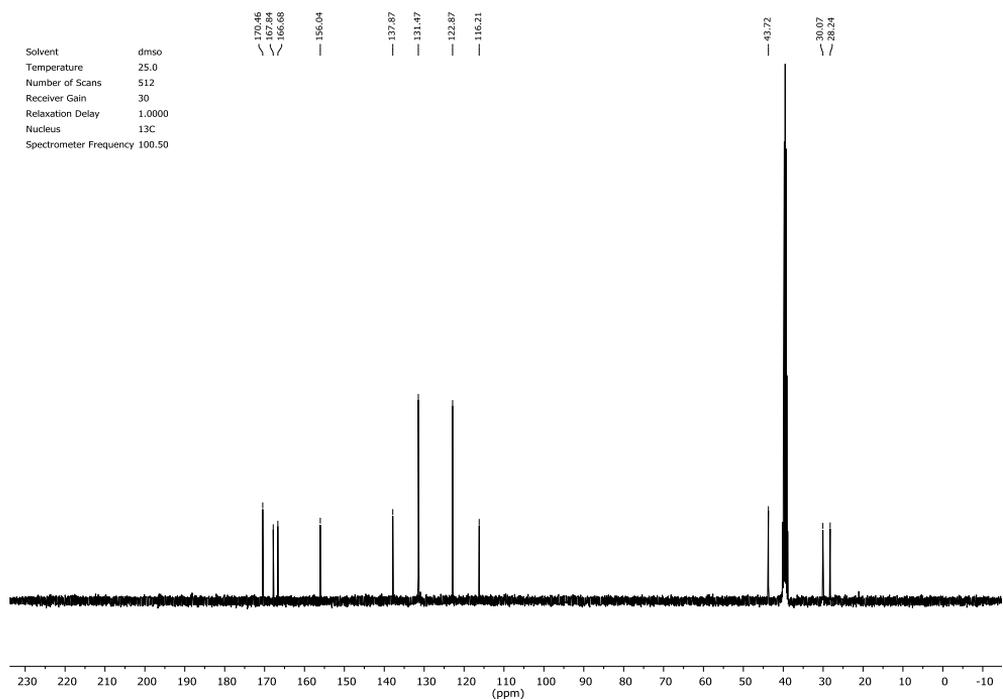


Fig. 4.33. ^{13}C -NMR spectrum of **4.37**.

The formation of **4.37** could be rationalized (Fig. 4.34) by the bromination of the monobromo derivative **4.39** at position C4a of the 2,4-diaminopyrimidine ring, which installs a bromine

atom at the bridgehead carbon of the bicyclic system (the first one ever described to the best of our knowledge), to afford the corresponding Wheland intermediate (**W1**) that corresponds with the structure proposed by I. Galve for such intermediate, perfectly referable to the one depicted for the bromination of aniline (**W2**).

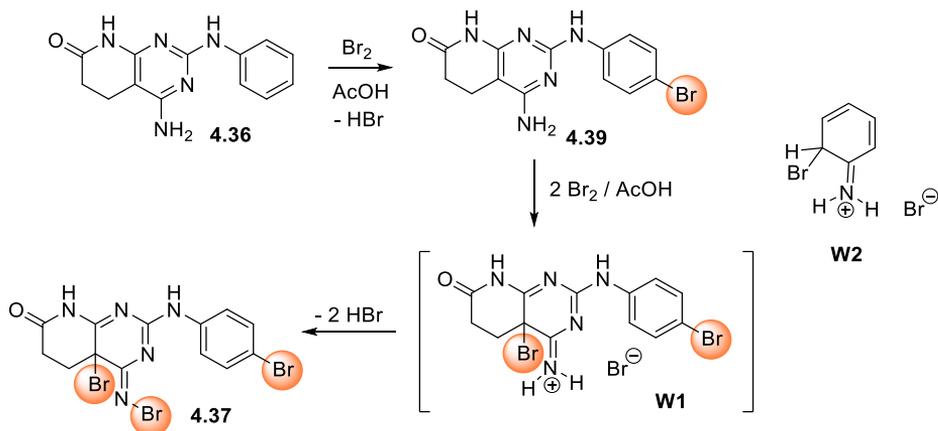


Fig. 4.34. Monobrominated compound **4.39** and capture of the Wheland intermediate **W1** as the corresponding *N*-bromoimino derivative **4.37** and comparison of **W1** with the Wheland intermediate of the aniline bromination **W2**.

Contrary to **W2**, in the case of **W1** the aromaticity cannot be recovered by losing a proton, so such a Wheland intermediate is captured by the formation of the rare *N*-bromoimino derivative **4.37**.

To complete the characterization of **4.37**, we tried to obtain a single crystal due to the theoretical interest of such structure. However, **4.37** evolves in the presence of any solvent to a different structure being impossible to obtain a suitable crystal. Thus, the X-ray powder diffraction pattern was determined at the MSPD beamline of the ALBA synchrotron²⁹ and, surprisingly, the orange solid is not amorphous as expected but presents a microcrystalline structure.

Despite the broad peaks and limited d-spacing available, a triclinic unit cell was indexed with DICVOL06³⁰ and the obtained cell parameters were refined with DAjust software. A promising structure candidate was obtained with the direct-space strategy TALP which underwent a final restrained Rietveld refinement with RIBOLS. The final unit cell parameters are: $a = 8.757(1)$, $b = 9.668(2)$, $c = 10.106(1)$ Å, $\alpha = 63.6(4)$, $\beta = 77.3(7)$ and $\gamma = 82.0(7)^\circ$, $V = 747(3)$ Å³, space group P-1.

The solved structure confirmed the *N*-bromoimino structure **4.37** (Fig 4.35), that is to say, the Wheland intermediate has been stabilized by the subsequent bromination of the imino tautomer of the amino group at C4 of the pyridopyrimidine skeleton.

Results

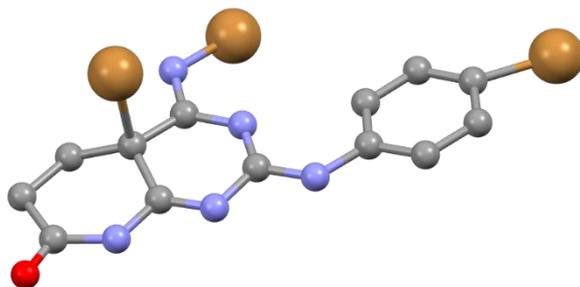


Fig 4.35. Graphical representation of the crystal structure attained for **4.37** through x-ray powder diffraction.

Two unexpected features arise from the crystalline structure. The above mentioned C4a-bromo substitution, breaking the aromaticity of the pyrimidine ring as that carbon turns to sp^3 hybridization and, on the other hand, the *N*-bromoimino group at C4. The bromoimino functional group is scarce, few are described, and none crystallized. It is remarkable in the crystal structure of **4.37** that the formation of a stereocenter due to the introduction of the bromine atom at the C4a bridgehead carbon is reflected in the internal structure by the presence of the two enantiomers of **4.37** associated in a self-complementary ADAD-DADA quadruple hydrogen-bonding centrosymmetric motif, which is in turn associated with a second pair of enantiomers through a π - π interaction of the *p*-bromophenyl rings thus forming a ribbon of pairs of enantiomers (Fig. 4.36).

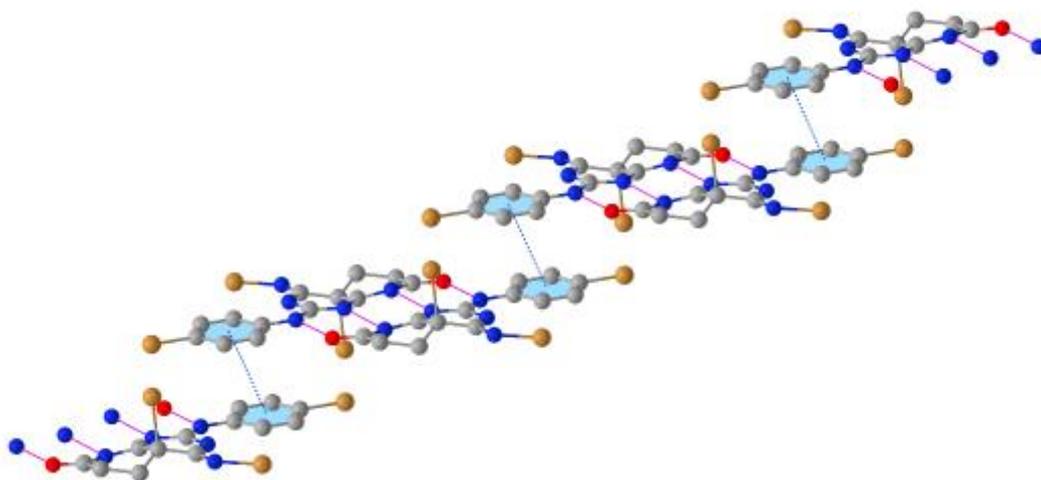


Fig. 4.36. Graphical representation of the crystal structure of **4.37** through x-ray diffraction. The spatial distribution of the structure is due to the hydrogen bond naturally originated between two molecules of **4.37** and also the π -stacking observed in the phenyl ring in C2.

Although being of great theoretical interest, **4.37** still has an extra turn to add scope and become a key intermediate in pyrido[2,3-*d*]pyrimidine synthesis. During the characterization of this compound and the crystallization studies, an evolution phenomenon occurs. Seems to be crucial for this step the dissolution of the compound, and **4.37** is very insoluble in most of the

solvents, being DMSO the only organic solvent with a good ability to dissolve the compound. To accomplish a complete transformation, **4.37** is heated at 80 °C in DMSO under vacuum (50 mbar) to afford the 6-bromo substituted pyrido[2,3-*d*]pyrimidine **4.38** in almost quantitative yield. The structure was confirmed by single crystal X-ray diffraction of a 1:1 solvate with acetone (Fig 4.37).

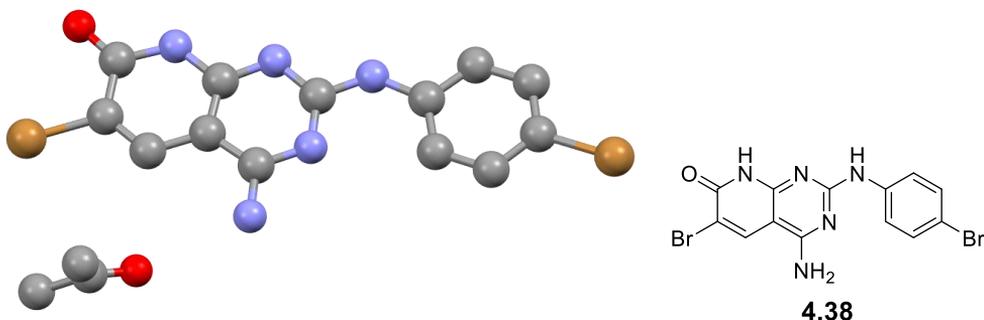


Fig 4.37. Graphical representation of the crystal structure attained for the transposition product **4.38** co-crystallized with acetone through x-ray diffraction.

Going back to our initial objective of obtaining a dibromo substituted structure which should allow the orthogonal decoration of the pyridopyrimidine nucleus, the synthesis of **4.38** is a hallmark in the way to obtain that objective. The bromine atom is placed at position C6 of the pyridopyrimidine core, additionally, a double bond has been introduced between C5 and C6, a step that normally is carried out by dehydrogenation of the C6 aryl substituted compound to achieve the required biological activity.¹⁷

Although the synthesis of **4.38** opens interesting possibilities for the synthesis of pyrido[2,3-*d*]pyrimidines from a common core, the presence of two bromine atoms attached to sp^2 do not allow enough selectivity in the forthcoming decoration of the bicycle. Consequently, we decided to carry out a complete study on the possibility of adding other halogen atoms that could allow the orthogonal synthesis of diverse pyrido[2,3-*d*]pyrimidines.

4.3.3.2 Structural possibilities

On the one hand, we intended to perform scaffold hopping to a more polar and DM1 suitable candidate. For nucleotide recognition, a 2,4-diamino pyrido[2,3-*d*]pyrimidine is a far more suitable than the one bearing a phenyl at position C2, so we decided to add structure **4.40** to the whole study. Such structure also adds the value to be the less substituted scaffold of the study and, therefore, the one requiring the minimum amount of halogen source to obtain both the Wheland and transposition products if they are formed.

Results

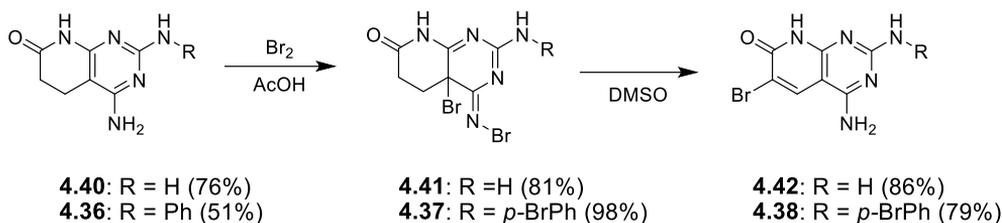


Fig 4.38. Schematic comparison between the synthetic itinerary to afford **4.39** and **4.42**.

4.3.3.2.1 Scaffold assembling

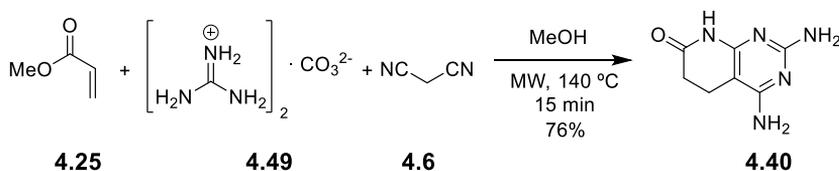


Fig 4.39. One-pot multistep reaction affording **4.40**.

Using the same multi-step methodology as for **4.36**, methyl acrylate **4.25**, malononitrile **4.6** and guanidine carbonate **4.49** are mixed and heated at 140 °C for 15 min in a microwave inside a sealed vial. After completing the reaction, water is added to precipitate the product that is filtered, yielding **4.40** as a brownish solid in 76% yield. The ¹H-NMR spectrum (Fig 4.40) shows no acrylic system but two new amine groups at ~6 ppm and the ring closing lactam group formed at 10.3 ppm.

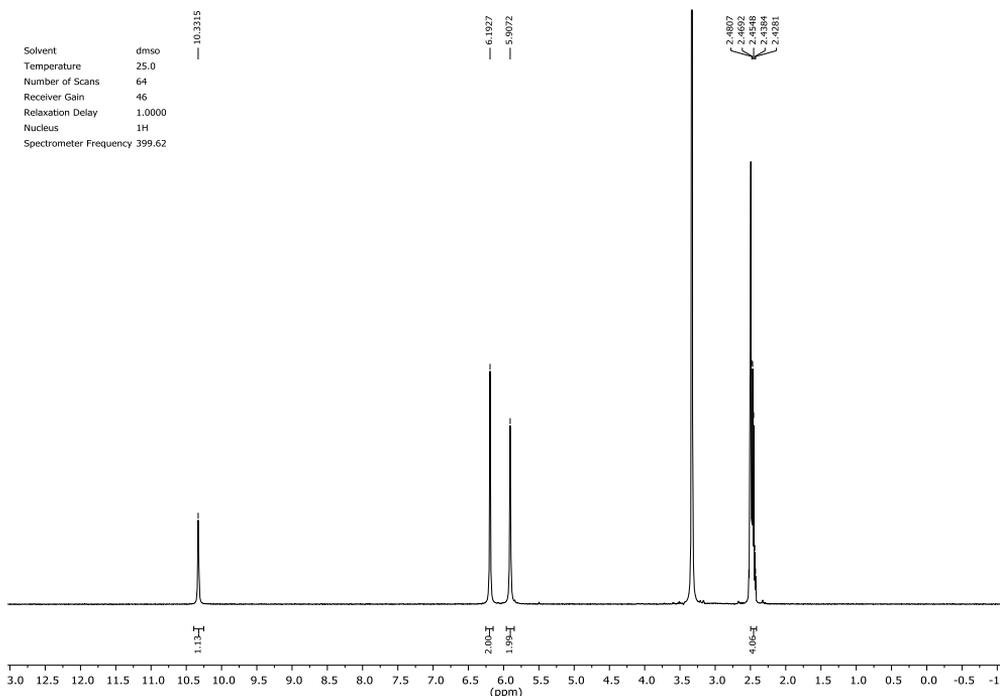


Fig 4.40. ¹H-NMR spectra of pyrido[2,3-*d*]pyrimidine **4.40**.

4.3.3.2.2 Bromination

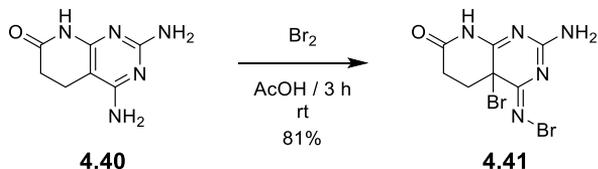


Fig 4.41. Reaction conditions for the formation of the intermediate **4.41**.

Analogously to the conditions used for the synthesis of **4.37**, molecular bromine is added to a suspension of **4.40** in glacial acetic acid and is stirred at room temperature for three hours, protecting the crude from light. Diversely to the case of **4.37** and due to the structural difference, only two equivalents of bromine are added to perform the transformation. Once the reaction is finished, water is added to the crude, and an orange solid precipitate. The suspension is filtered and gently washed with water to remove excess acetic acid and bromine, affording **4.41** (81%) as a yellow solid.

Like its analog **4.37**, **4.41** is not stable in solution. The evolution phenomenon was verified when performing the ^1H spectrum in DMSO-d_6 , observing by NMR that **4.41** reacts in dissolution. This evolution is, in fact, faster than in **4.37**, making it practically impossible to record a spectrum at $t = 0$ s of the transformation. In the spectra obtained (Fig 4.42), it is possible to observe time-dependent increasing signals that show the evolution of the solid towards the transposition product.

Results

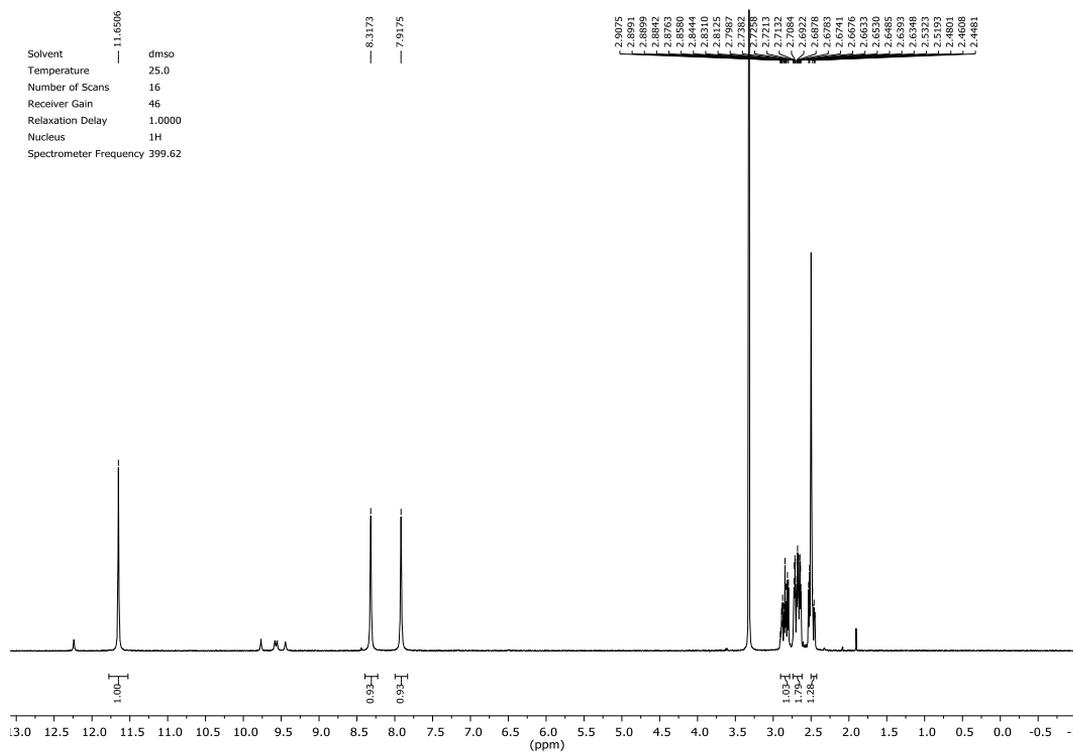


Fig 4.42. ¹H-NMR spectra of pyrido[2,3-d]pyrimidine **4.41**, Evolution can be observed although prepared the solution in and instantly recording the spectra (signals appearing between 10-9 ppm).

4.3.3.2.3 Transposition of **4.41** to **4.42**

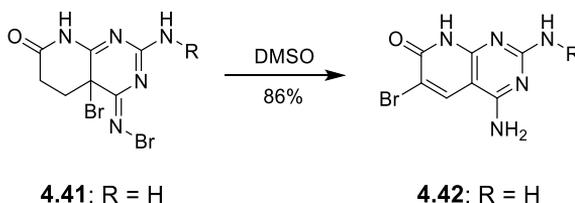


Fig 4.43. Scheme of the evolution of **4.41** in DMSO affording **4.42**.

Once intermediate **4.41** was obtained, the transposition to **4.42** was carried out. The intermediate **4.41** is dissolved in DMSO and placed at the rotatory evaporator at a reduced pressure of 50 mmHg for 3 hours at 60 °C. After that, water is added, and an off-white solid precipitate. The suspension is filtered and washed with further water affording **4.42** (86%). As seen in the NMR spectrum, the water addition is a crucial step to obtain a pure product as the ionic by-products of the reaction tend to accumulate in the reaction mixture as no base is added. For this reason, signals from the evolution spectra and the final products may differ in shift and shape.

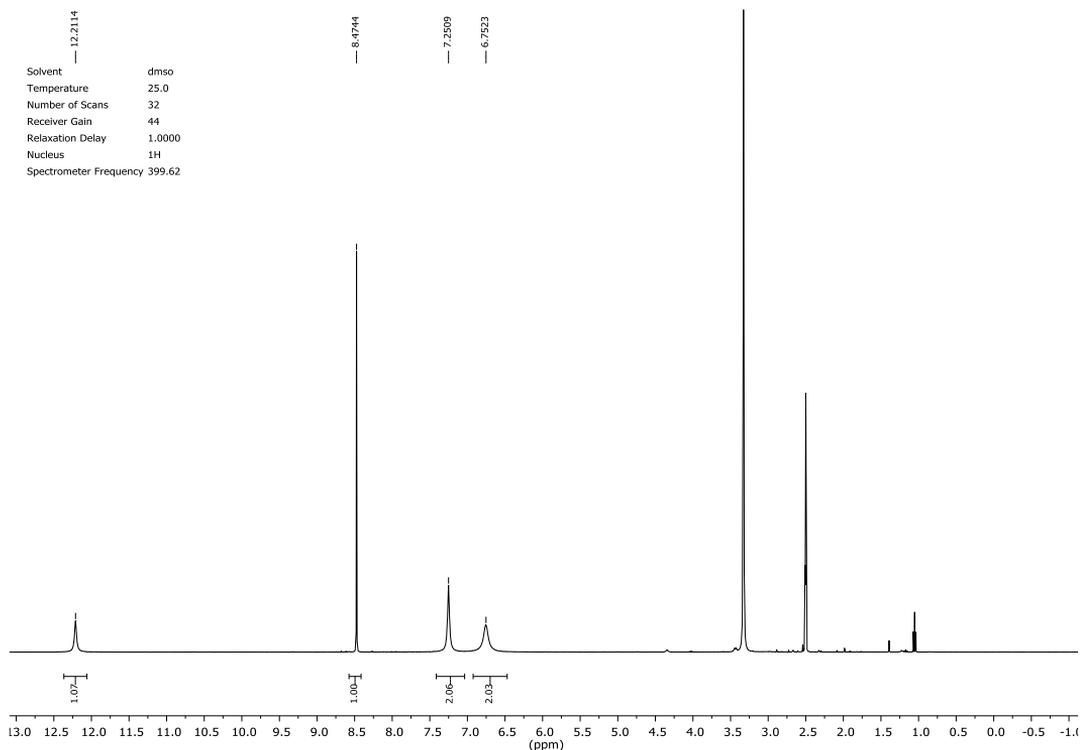


Fig 4.43. $^1\text{H-NMR}$ spectra of transposition pyrido[2,3-*d*]pyrimidine **4.42**.

Remarkably and although the MW of **4.42** is 256 g/mol, only four singlet signals are observed in the $^1\text{H NMR}$. The signal confirming the structure obtained is the C5-H proton (8,47 ppm), the only proton bonded to a carbon atom in this molecule. This data indicates that a double bond is formed between C5-C6. On the other hand, the rearrangement of the bromine atom towards the C6 position is observed as in the case of **4.38**. For the complete characterization of **4.42**, two-dimensional spectroscopy and HRMS techniques were required.

4.3.3.3 Halogen diversity study

Once the possibility of transferring the phenomenon first observed in **4.36** to this analog was verified, we decided to carry out a transversality study between halogen group elements.

After an unsuccessful reaction attempt using molten molecular iodine under conditions analogous to those described above, we decided to try a different approach. Taking into account that the *N*-bromoimino group present in the Wheland intermediates seems to be the primary halogen source that causes the evolution of the Wheland intermediate to the final 6-bromo-substituted compound. Consequently, we decided to test other similar halogenating reagents such as *N*-bromosuccinimide (NBS).

Results

4.3.3.3.1 *N*-Bromosuccinimide

First, we decided to reproduce the reaction conditions of the evolution stage but using **4.40** and two equivalents of NBS to establish whether this reagent can reach a reactivity similar to that of the intermediate **4.41** under the same conditions.

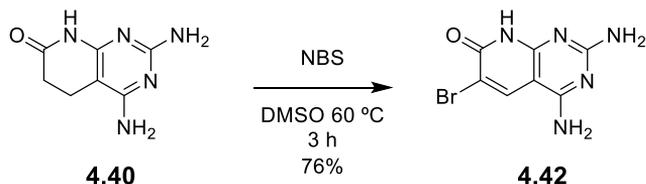


Fig 4.44. Scheme of the reaction of **4.40** and NBS.

The result confirms the hypothesis because the product obtained is also **4.42**, which is reached in two steps from **4.40** via the intermediate **4.41**. In order to isolate the **4.42** from the reaction mixture, we used the same methodology previously described starting from **4.41**. The product is precipitated from the medium by adding water, filtered, and washed with water to yield **4.42** (76%). This finding gives valuable data about the how this reaction proceeds and furthermore, on a practical level, it is easier to accurately measure solids rather than liquids of complex handling such as molecular bromine, especially on a small scale.

The next step of the study is to find out if these reaction conditions are also adequate to brominate the *p*-phenyl position of **4.36** to directly afford **4.38** by increasing the NBS equivalents up to 3.

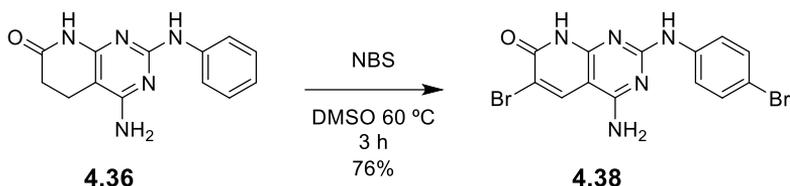


Fig 4.45. Synthesis of **4.38** from **4.36** and NBS.

Using such reaction conditions, **4.38** was obtained in 76% yield (increasing the cumulative yield from the 69% of the protocol via de Wheland intermediate) showing that it is possible to carry both halogenations in one step. To verify completely the study, the reaction is also carried out with **4.36** using 1 equivalent of NBS and, as in the case of molecular bromine, the 2-(*p*-bromophenylamino) substituted compound **4.39** is obtained in 68% yield.

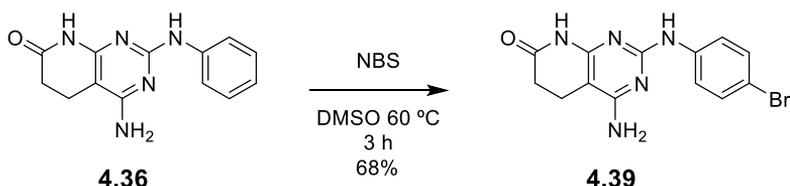


Fig 4.46. Scheme of the reaction of **4.36** and one equivalent of NBS.

As with the rest of the experiments with NBS, the result corresponds to the designed hypothesis. The fact that the reactivity of *N*-bromosuccinimide is equivalent to Wheland intermediates lead us to consider the use of other *N*-halosuccinimides.

4.3.3.3.2 Use of other *N*-halosuccinimides

The main reason for carrying out the following new series of experiments lies in the interest in obtaining products with different halogen reactivity introducing chemical orthogonality in possible subsequent reactions.

First, several experiments were carried out with *N*-chlorosuccinimide (NCS), and in no case could the 6-chloro derivative be obtained, not even by increasing the temperature and time of the reactions.

After attempting the reaction with NCS we decided to try *N*-iodosuccinimide (NIS). We had to slightly alter the reaction conditions since a lower reactivity of this heavier halogenating agent is observed.

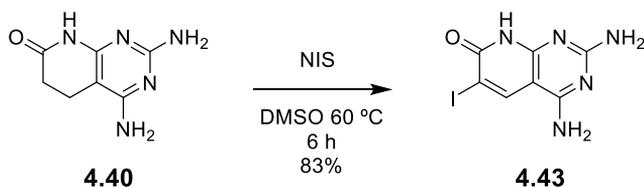


Fig 4.47. Scheme of the reaction of **4.40** and two equivalents of NIS.

4.40 and NIS were dissolved in DMSO and allowed to react for 6 hours at 60 °C at a pressure of 50 mmHg. Then, abundant water was added, the product is filtered and washed, yielding the 6-iodo substituted compound **4.43** (83%). The ¹H-NMR spectrum was registered and the transformation towards **4.44** was verified by the presence of a singlet proton signal at 8.61 ppm corresponding to the H-C5. This is a downfield shift compared to the 6-bromo substituted derivative **4.42** where H-C5 appears at 8.47 ppm. Such spectrum corroborates the formation of a C5-C6 double bond and the introduction of iodine in the desired position, although HRMS was required to verify the structure.

Results

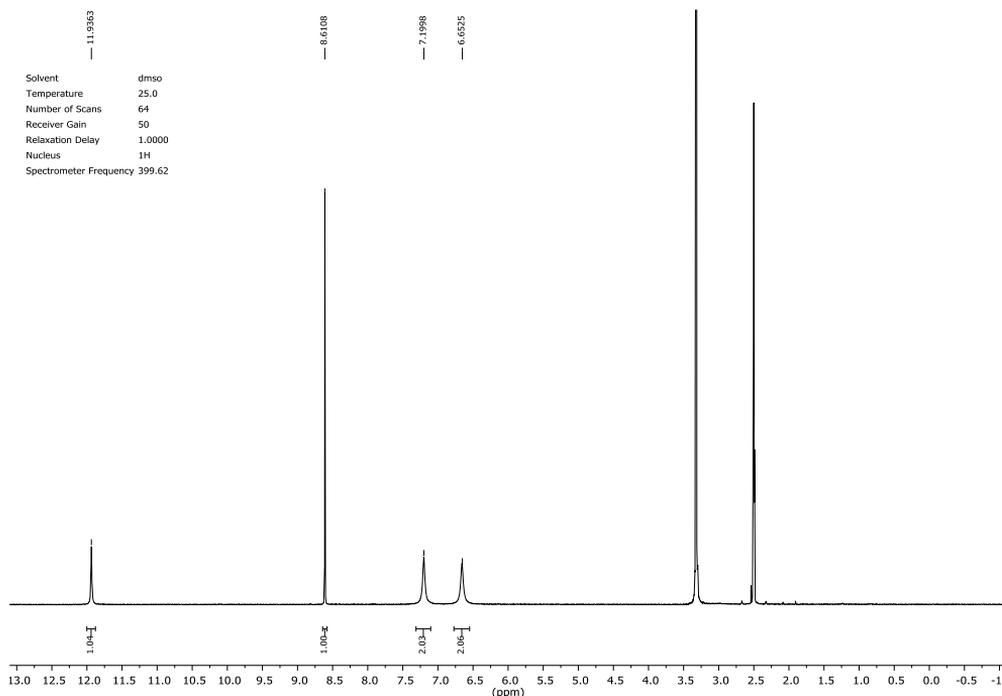


Fig 4.48. $^1\text{H-NMR}$ spectra of 6-iodopyrido[2,3-*d*]pyrimidine **4.43**.

To screen all the possible scenarios, as done in the study carried out with NBS, the possibility of transferring the phenomenon to the 6-phenylamino substituted compound **4.36** was studied. Therefore, an increase from 2 to 3 equivalents of NIS, was used.

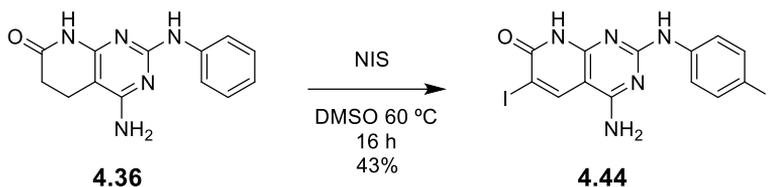


Fig 4.49. Scheme of the reaction of **4.36** and 3 equivalents of NIS.

In such reaction conditions the diiodo substituted compound **4.44** was obtained in 43% yield. The reaction required 16 h to be completed while with NBS was complete in only 3 h, moreover it was challenging to eliminate the residues of succinimide from the reaction mixture since the product tends to occlude it. This product was characterized by two-dimensional NMR, because that both C-I signals appear at practically the same chemical shift in the ^{13}C spectrum (84.27 and 84.25 ppm).

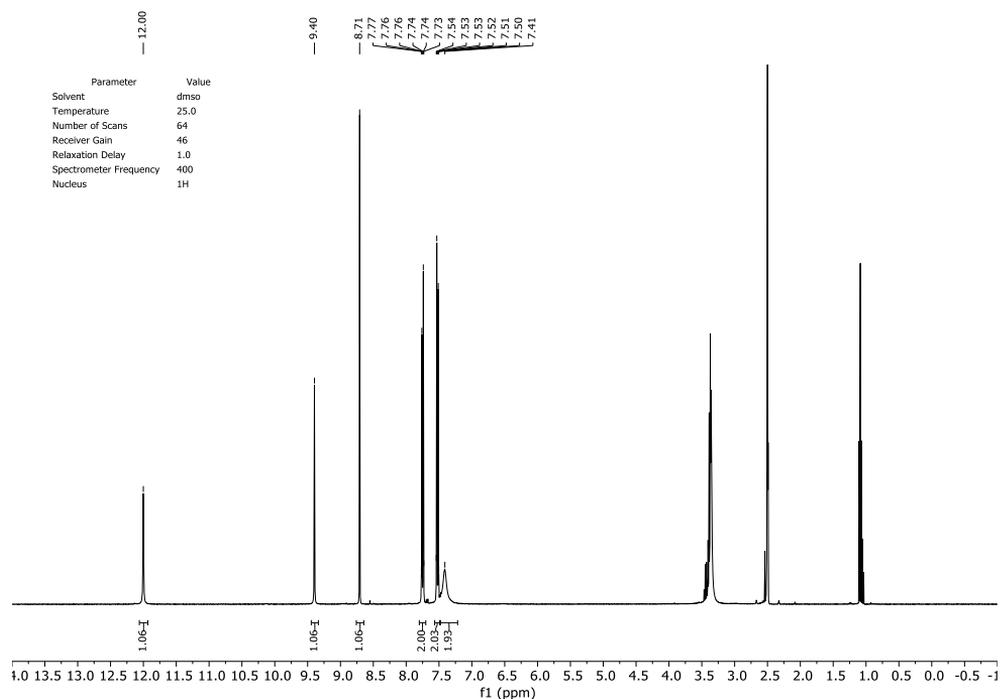


Fig 4.50. ¹H-NMR spectra of 6-iodopyrido[2,3-*d*]pyrimidine **4.44**.

In the ¹H-NMR spectrum, the downfield shifts of C5-H and one of the multiplets of the *p*-phenylamino system are observed. The fact that NIS is also able to halogenate the para position of the aromatic ring triggers the next step in the study.

The first step was to try the mono iodination at the para position of 2-phenylamino substituent of **4.36** in a similar way that is possible to mono brominate such position. However, although different thermal and reaction time conditions were tested by adding a single equivalent of NIS, in no case was possible to form such desired mono iodination compound.

Therefore, the possibility of introducing different halogens at C6 and the *para* position of the pyridopyrimidine skeleton was studied in order to increase the difference of reactivity and achieve orthogonality. Different thermal and time conditions were studied by adding a single equivalent of NIS to the 2-(*p*-bromophenylamino) substituted compound **4.39**, but the results are inconclusive. Therefore, the mono iodinated product at the para position cannot be obtained in this way, and if it is indispensable, it should be synthesized prior to assembling the pyrido[2,3-*d*]pyrimidinic core.

Consequently, the study was focused on the synthesis of a product bearing two different halogens, which may allow certain orthogonality in its subsequent substitutions. **4.39**, the single brominated product using 1 eq NBS or Br₂/AcOH was used as starting material.

Results

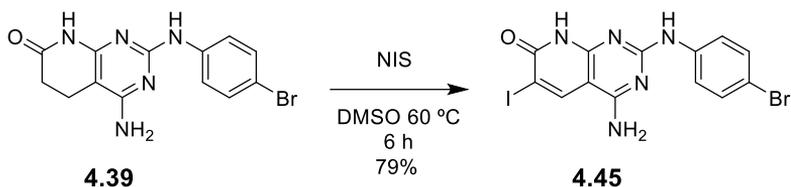


Fig 4.51. Scheme of the reaction of **4.39** and two equivalents of NIS.

The transformation occurred upon dissolving **4.39** and 2 equivalents of NIS in DMSO. After that, the product is isolated by filtration and washing with water yielding **4.45** as a pale-yellow solid.

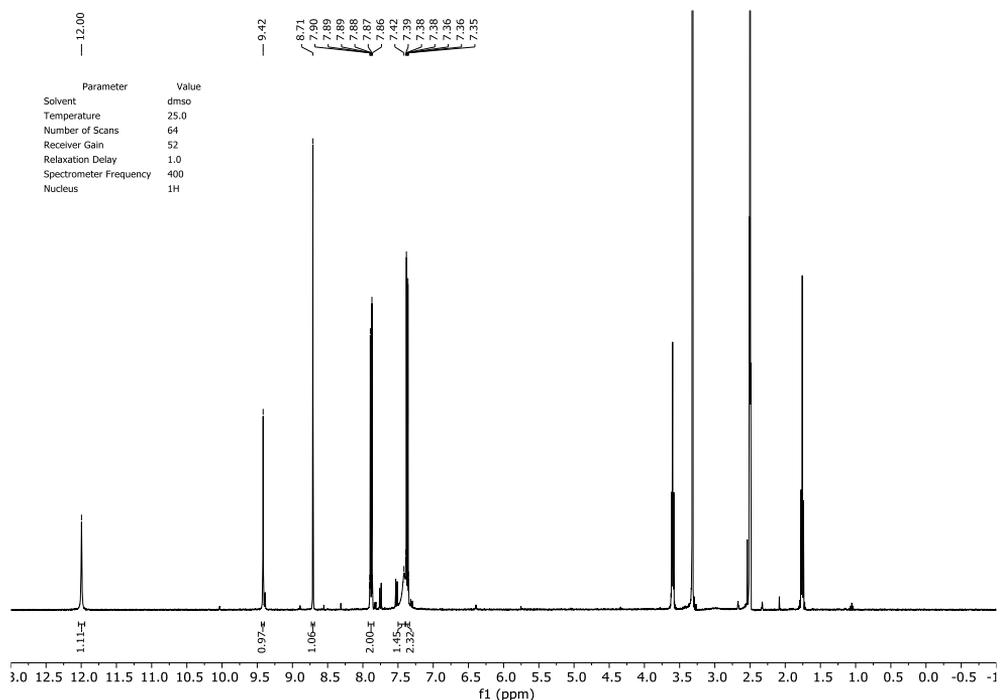


Fig 4.52. ¹H-NMR spectra of 6-iodopyrido[2,3-*d*]pyrimidine **4.45**.

4.45 was characterized using two-dimensional NMR. Analogously to product **4.44**, this compound occludes part of the succinimide, as its multiplets at 3.6 and 1.7 ppm can be seen in the ¹H-NMR spectrum. In this spectrum, the chemical shift of C5-H is similar to other compounds obtained in this chapter bearing C6-I. Whereas the chemical shift of the *p*-substituted carbon in ¹³C-NMR corresponds to a C-Br (112.8, by 84.3 ppm in case of C-I).

In order to evaluate the possible difference of reactivity of the two halogenated positions, two cross-coupling Suzuki reactions were performed with the crude product resulting from this reaction, but the result is inconclusive in both cases. Product **4.45** exhibits high polarity and very low solubility in any solvent other than DMSO, hampering standard purification procedures such as liquid-liquid extractions or flash chromatography. For this reason, in a parallel PhD thesis, the study of the scope of this family of compounds has been performed using N8-

protected pyrido[2,3-d]pyrimidines which improve the solubility of the resulting halogenated compounds allowing the ulterior Suzuki couplings and the purification of the final products .³¹

4.3.3.4 Kinetic study of the formation of the 6-bromo substituted pyridopyrimidines

To continue with the characterization of the phenomenon by which 6-halo substituted pyridopyrimidines are formed, a kinetic and mechanistic study of the evolution reaction of the intermediate **4.37** was carried out. **4.37** was used instead of **4.41** since its reactivity is lower and favors the study of the reaction progress.

Since the product reacts when dissolved in DMSO and its solubility in this solvent is high, a series of experiments of periodic recording of short proton spectra were designed (8 accumulations, recording every 1 min). In order to register these spectra, it is necessary to perform the reaction inside the NMR spectrometer. For this purpose, some changes are forcefully introduced to the original methodology. The first one is that d_6 -DMSO is used as the solvent. We assumed that even if there is a slight change of properties to the non-deuterated DMSO, this should not affect the order of the reaction kinetics. In addition, room temperature and atmospheric pressure are used. These factors undoubtedly affect the reaction rate but should not affect the reaction order.

Results

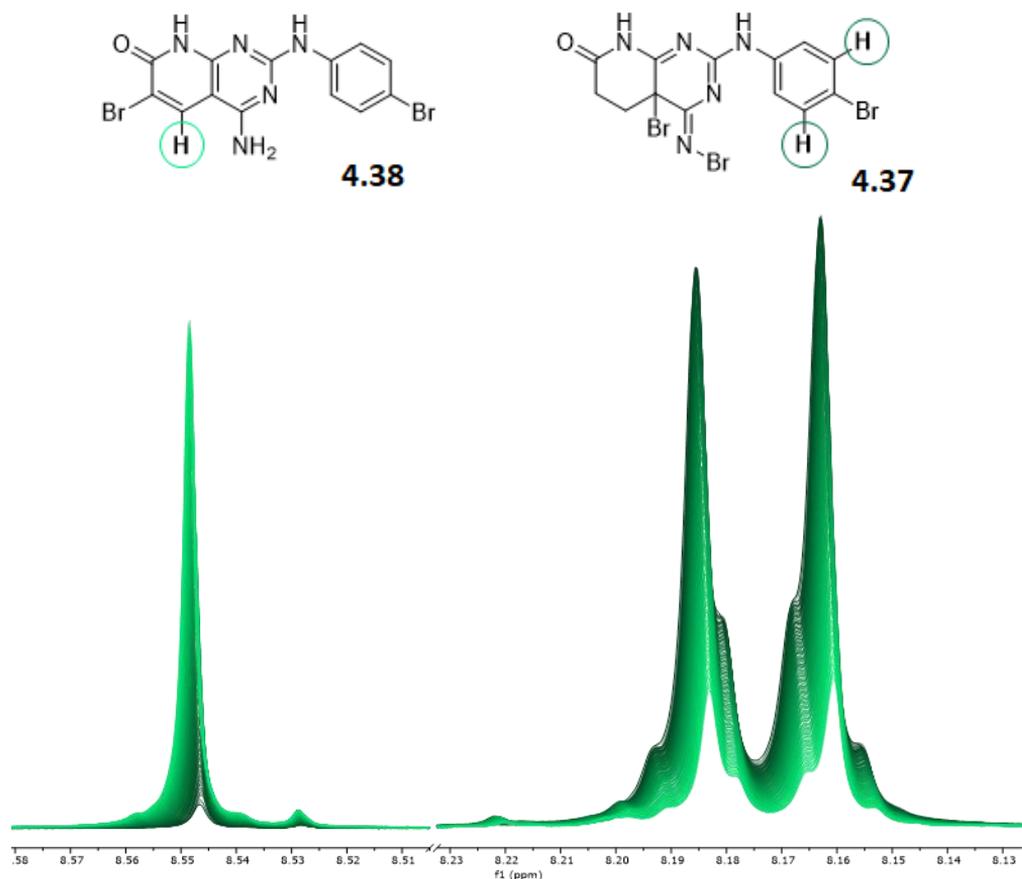


Fig 4.53. Superposition of 250 ¹H-NMR spectra of the evolution of pyrido[2,3-*d*]pyrimidine **4.37**. The figure corresponds to the amplification of two regions showing the increase of the C5-H signal of **4.38** and the decrease of the C3'-H signal of phenyl ring of the reactant **4.37** (dark green: initial spectra, light green: end spectra).

The first 250 spectra were analyzed because the crude turns complex in the last few spectra. It must be borne in mind that the reaction is not carried out in the presence a base, and contrary to the aforementioned experiments the temperature cannot be increased. Consequently, it is impossible to eliminate the hydrogen bromide generated in the evolution. Therefore, as the kinetic study progresses, the reliability of the data decreases. The reason is that they can be affected by the saturation of the medium.

All spectra are stacked in Fig 4.44. Characteristic signals were selected in a clean spectrum areas for both the reactant and the product. The C5-H signal of **4.38** and the proton signal in meta position of the phenylamino ring at C2 of **4.37** were selected (Fig 4.53). Then, the 250 spectra are integrated in two fixed ranges, 8.570-8.535 ppm for **4.38** and 8.215-8.140 ppm for the signal of **4.37**. Once all these data have been recorded, they are plotted using the molar ratio and the reaction time extracted from the spectra metadata. It is necessary to consider the 2:1 stoichiometry between these signals to perform such calculation. The results are plotted in Fig 4.54.

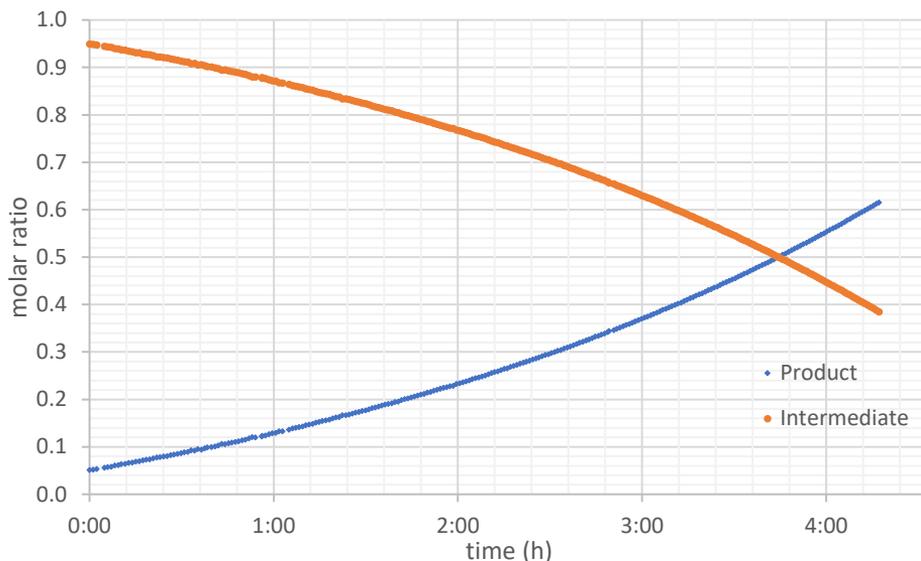


Fig 4.54. Kinetic study of the transformation of **4.37** based on the integration of the 250 spectra recorded.

After observing the results obtained and to determine the order of the kinetics of the reaction, we decided to use only the first hour of results (as close to $t = 0$ s as possible). In that range, a relationship for first-order kinetics could be established. Nevertheless, if the overall reaction decay of product concentration is taken into account, it appears to be a linear function (zero-order) rather than a negative exponential function (first-order) or inverse function (second-order). Interestingly, the reaction rate seems to accelerate with time, which can be affected by the increasing hydrogen bromide concentration in the reaction medium, this might be by the a stabilization of the polar intermediate in the reaction presented in Fig. 4.34.

4.3.3.5 Reactivity of 6-halo substituted pyridopyrimidines towards anti-DM1 candidates

Once developed the synthesis of 6-halo pyridopyrimidines and partially understood the mechanism of formation, we decided to study whether it is plausible to use any of the products obtained as starting material to synthesize anti-DM1 candidates. Getting back to the basis of the interaction with CUG (see section 5.1.2.2), it is necessary to obtain a 2,4-diaminopyridopyrimidine with both amino groups unsubstituted. After the failed Suzuki cross-coupling attempts described in the previous section, we decided to transform the halogen atom of **4.42** or **4.43** into a carboxylic acid to afford **4.46** and thus attempt to perform a Steglich-type coupling with the selected spacers.

Results

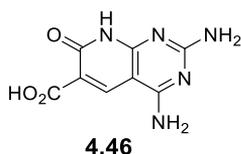


Fig 4.55. Structure of the target compound **4.46**.

is the protocol for accessing **4.46** included two steps starting from the halogenated product **4.42**: first a cyanide coupling to introduce the extra carbon and, second, the hydrolysis of the cyano group to obtain the corresponding carboxylic acid. For the first stage, an Ullmann-type stoichiometric cross-coupling reaction is required. CuCN is used as the source of both the metal and the cyanide.

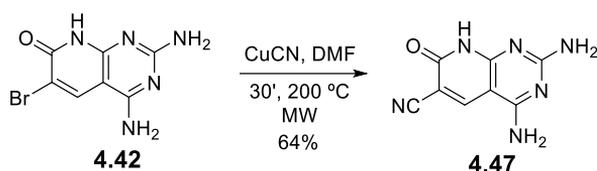
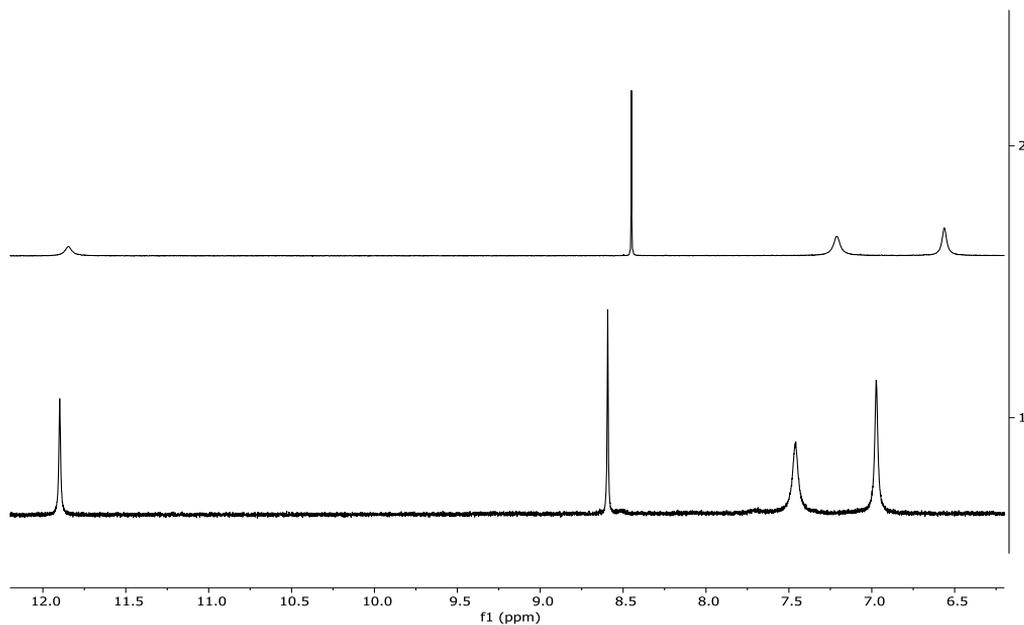


Fig 4.56. Scheme of the copper-catalyzed Ullman coupling affording **4.47**.

4.42 and copper (I) cyanide were suspended in anhydrous DMF in a microwave vial under an inert atmosphere. The mixture was stirred at 200 °C for 30 min under microwave irradiation. Then water was added to the reaction mixture, and a solid precipitate that was collected by filtration. An aqueous solution of EDTA was used as a chelating agent to achieve the elimination of copper. In that way **4.47** was obtained in 64% yield.

Although the product is poorly soluble in DMSO, NMR spectroscopy was recorded. Comparing the NMR spectrum of **4.47** with the spectrum of **4.42**, a slight downfield shift of the C5-H proton is observed.



Fig

4.57. Stacked $^1\text{H-NMR}$ spectra of **4.42** (above) and **4.47** (below).

Since NMR spectroscopy is inconclusive, the IR spectrum of **4.47** is recorded. The spectrum shows the presence of the typical nitrile stretch band at 2213 cm^{-1} .

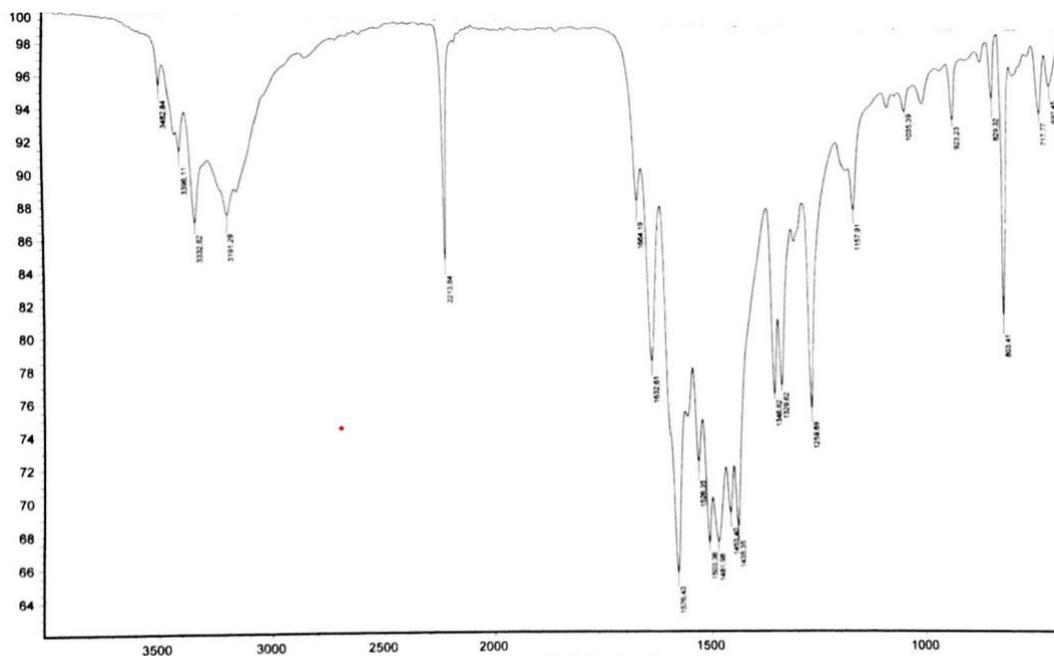


Fig 4.58. IR spectrum of the compound **4.47**. A characteristic nitrile stretch vibration band is observed.

With the structural confirmation of **4.47**, we decided to carry out the hydrolysis of the nitrile group. **4.47** was suspended in 50% aqueous H_2SO_4 and the mixture was stirred at reflux for

Results

16 hours. After this time, the reaction mixture was neutralized with a 10% Na_2CO_3 solution inducing the precipitation of a black solid, which is subsequently collected by filtration.

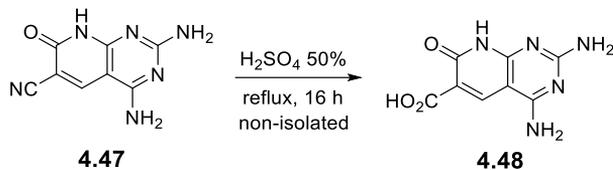


Fig 4.59. Scheme of the attempted hydrolysis of **4.47**.

Given that the solid obtained is insoluble in TFA-*d* and DMSO-*d*₆ so we decided to run the IR spectrum to check if the reaction has worked. Although the product seemed to have evolved, the bands were so broad that it was not possible to conclude if **4.48** was formed or not.

Nevertheless, we tested the Steglich coupling conditions with such solid but with no positive results. Unfortunately, it was not possible to conduct the hydrolysis of **4.47** with any of the methodologies considered in this work. For this reason, this approach was interrupted at this point and left it for future studies.

4.3.4 Findings in the synthesis of 6-carboxymethyl pyrido[2,3-*d*]pyrimidines

4.3.4.1 Hypothesis

The last one of the possibilities considered in this chapter for the construction of a 2,4-diaminopyrido[2,3-*d*]pyrimidine core bearing a substituent at C6 capable to attached to a spacer will consist on the introduction of a $-CH_2-COOR$ group at such position to form compound **4.53** (Fig 4.60). We suppose that the introduction of the methylene group between the heterocyclic core and the carboxylate will favor the synthesis of such structure and, on the other hand, will allow the binding to the selected spacers. With this aim, a retrosynthetic analysis of **4.53** is included in Fig. 4.60.

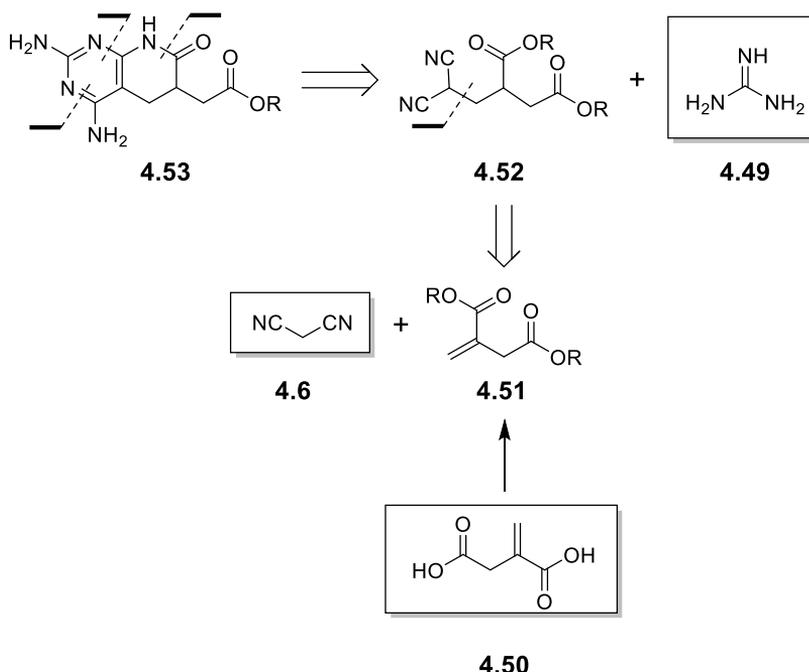


Fig 4.60. Scheme of the summarized retrosynthetic analysis proposed for the synthesis of **4.53**.

Taking into account the strategies developed by our group, the use of the cyclic strategy via a pyridine to access **4.53** will depend in this case on the reactivity of the acrylate **4.51** because during the Michael addition with malononitrile **4.6** and subsequent cyclization of the Michael adduct in NaOMe/MeOH could yield 2 different lactams. The presence in the Michael adduct **4.52** of two ester groups could theoretically afford the corresponding 6-membered and 7-membered lactam rings. Although, in principle the 6-membered ring could be favored, this reaction situation has never been tested. Consequently, we decided to carry out the reaction using the acyclic strategy, first carrying the Michael addition with malononitrile to afford **4.52** and then the cyclization of the open adduct with guanidine **4.49** to directly access the pyridopyrimidine **4.53**.

Results

4.3.4.2 Fischer esterification of itaconic acid **4.50**

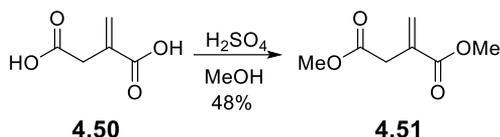


Fig 4.61. Scheme of the Fischer esterification reaction of itaconic acid **4.50** with methanol.

The first step of the synthetic route to **4.53** is a Fischer esterification of both carboxylic acid groups of the commercially available itaconic acid **4.50**. A classical methodology is followed, so H_2SO_4 is added to the solution of itaconic acid **4.50** in MeOH. The reaction is stirred at reflux for 21 h. After that, the product is extracted of the reaction mixture with AcOEt and evaporated under reduced pressure. The $^1\text{H-NMR}$ spectrum evaluation of the crude material (Fig 4.62) showed that the transformation was not complete. Therefore, to attain pure **4.51**, flash chromatography column was performed (cy:AcOEt, 80:20) to afford pure **4.51** (48% yield) as a colorless oil

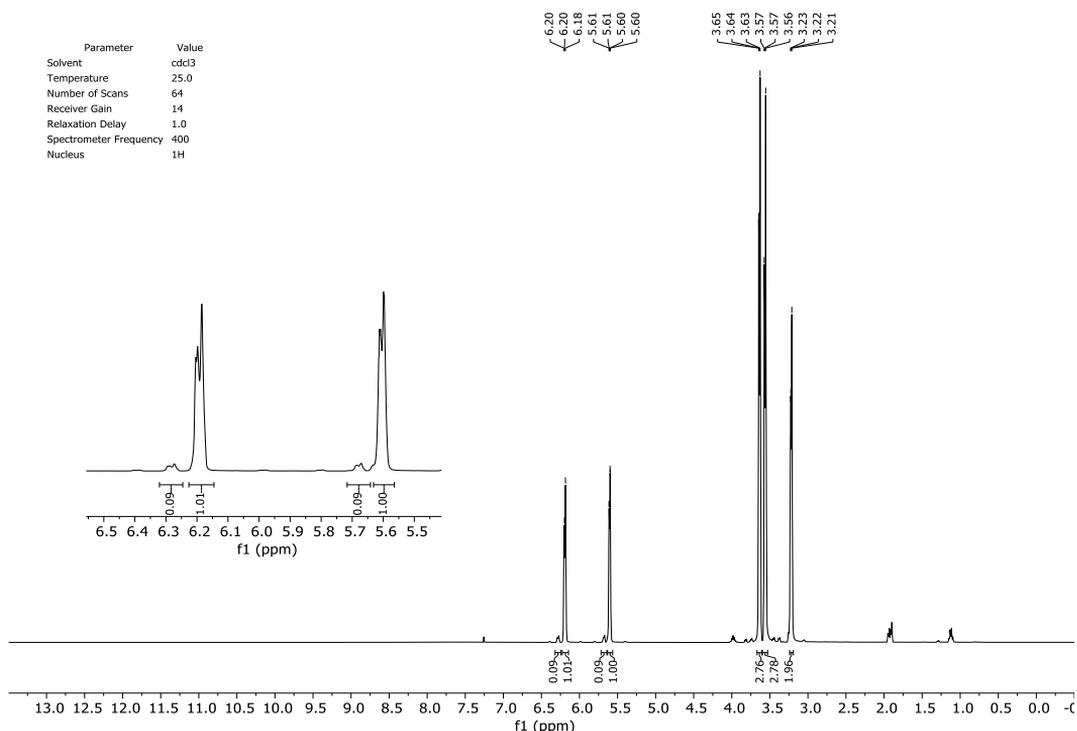
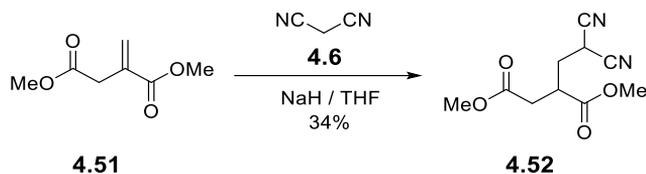
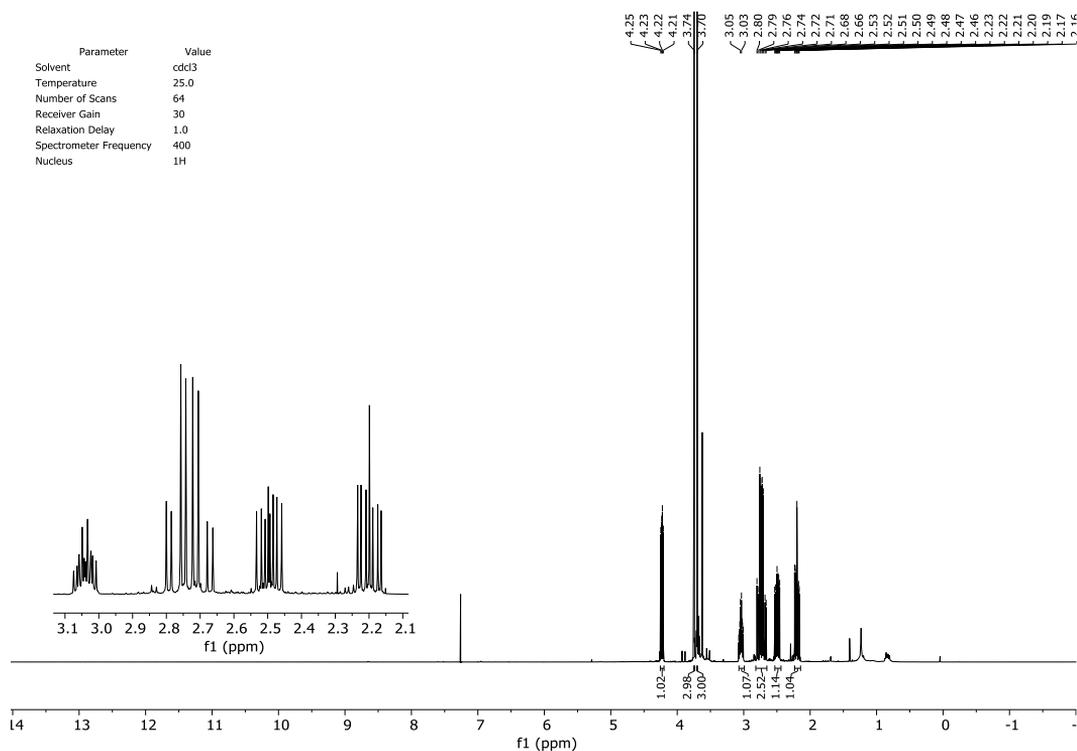


Fig 4.62. $^1\text{H-NMR}$ spectra of **4.51** prior to flash chromatography purification.

4.3.4.3 Michael addition of malononitrile to **4.51**Fig 4.63. Scheme of the Michael addition of malononitrile **4.6** on dimethyl itaconate **4.51**.

As aforementioned, the next step is to perform the Michael addition of malononitrile **4.6** to **4.51** in the acyclic strategy reaction conditions. Thus, malononitrile is ionized with NaH in anhydrous THF, and **4.51** is added to the mixture and stirred for 2 hours at 40 °C. After this period, the reaction is quenched with HOAc, the product extracted with DCM and evaporated under reduced pressure to afford **4.52** (34% yield) as a yellowish oil.

Fig 4.64. ¹H-NMR spectra of **4.52**, Chemical shifts between 2 and 3 ppm are displayed in amplification window to gain resolution.

The methine group present in **4.52** is a stereogenic center and therefore both methylene groups attached to it present interesting diastereotopicity. Using constant coupling analysis (*J*) and 2-dimensional NMR spectra, the structure of **4.52** was characterized.

Results

4.3.4.4 Cycloaddition of guanidine to **4.52**

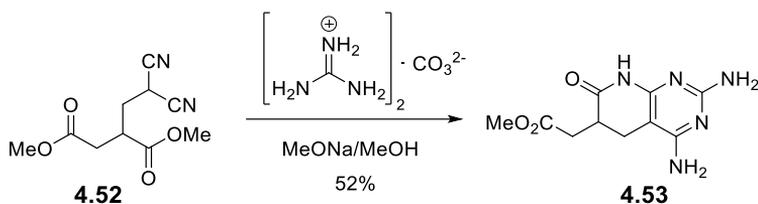


Fig 4.65. Scheme of the cycloaddition of guanidine carbonate **4.43** to **4.52**.

The crucial step in this synthesis is the assembling of the pyrido[2,3-*d*]pyrimidine ring. Following the already described methodology,³² guanidine carbonate is basified with sodium methoxide in MeOH. After that, **4.52** is added and the mixture is refluxed under inert atmosphere overnight. The solid formed is filtered and washed with H₂O, EtOH and Et₂O affording **4.53** as a white powder (52%).

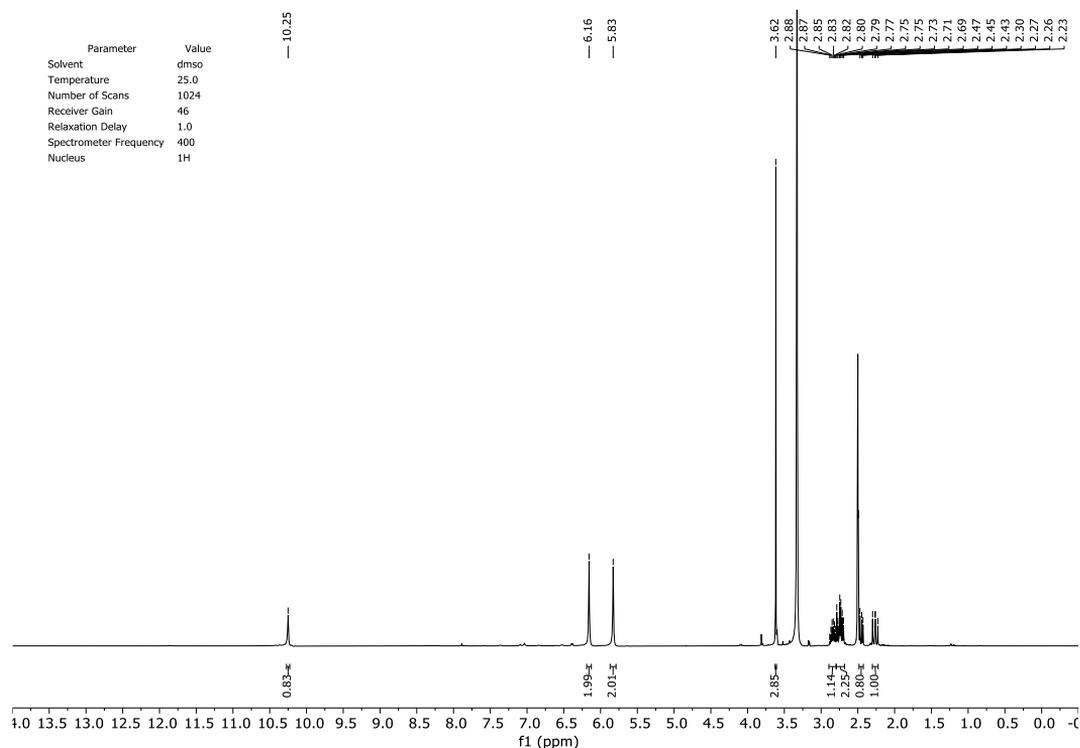


Fig 4.66. ¹H-NMR spectra of **4.53**. Spectrum showing all the characteristic signals of pyridopyrimidine core.

The ¹H-NMR spectrum shows the characteristic signals of a pyrido[2,3-*d*]pyrimidine, such as the lactam NH, the two NH₂ groups, and the C5 and C6 diastereotopic protons. The structure was completely characterized using 2-dimensional NMR spectroscopy and MS. Although being possible to obtain the corresponding 7-membered ring, the spectroscopy confirms that the reaction evolved selectively to the desired pyridopyrimidine structure. Product **4.53** could be

useful itself to attain new DM1 candidates. However, we attempted to transform the ester group to a carboxylic acid to increase its reactivity.

4.3.4.5 Ester saponification of **4.54**

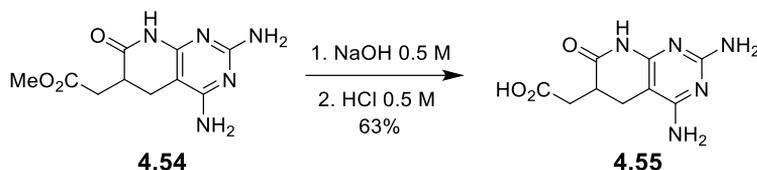


Fig 4.67. Scheme of the saponification reaction affording the RNA recognizing unit **4.55**.

The last transformation in the synthesis of this new pyrido[2,3-d]pyrimidines is the saponification of the methyl ester present in **4.54**. The starting material **4.54** is dissolved in 0.5 M NaOH and the solution is stirred for 21 hours at rt. The reaction mixture is acidified with conc. HCl and the product filtered and washed with H₂O to afford **4.55** (63%) as a white solid.

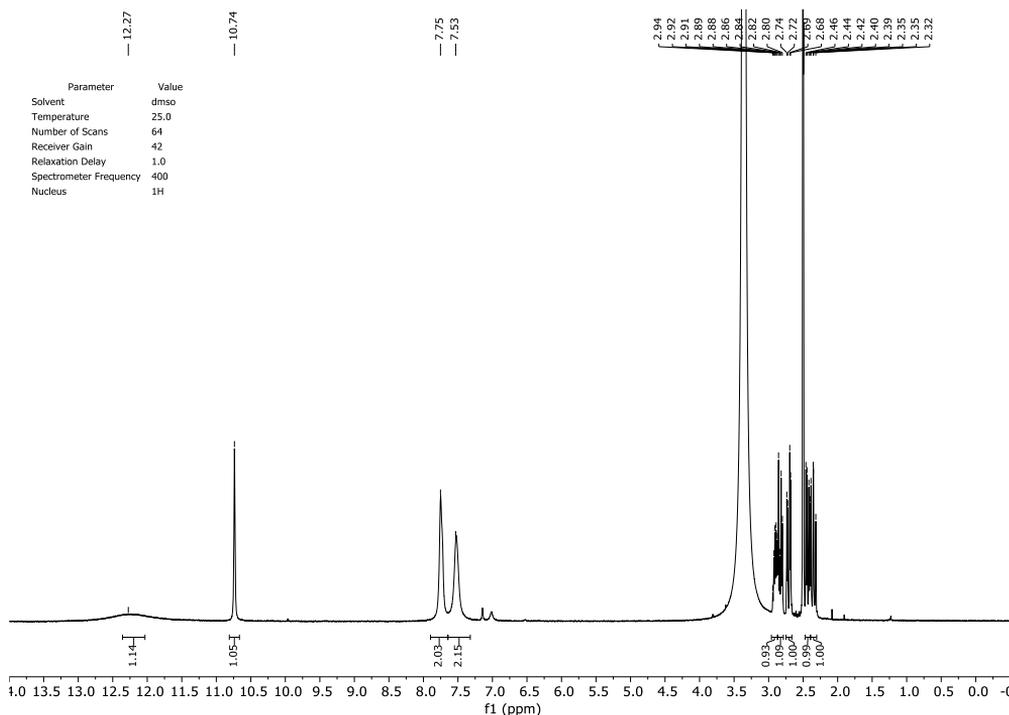


Fig 4.68. ¹H-NMR spectra of **4.55**. Final product of this synthetic itinerary.

The structure of **4.55** is confirmed by 2-dimensional NMR and HRMS. Upon that, preliminary Steglich amidation reactions showed promising reactivity towards aliphatic amines. Therefore, **4.55** was used in the subsequent generation of anti-DM1 candidates as main CUG RNA recognizer.

Results

4.3.4.6 One-pot multicomponent alternative for the synthesis of **4.54**

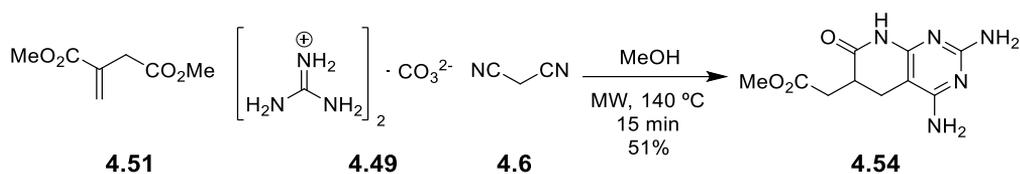


Fig 4.69. Scheme of the one-pot multistep synthesis reaction affording **4.54**.

Once the pyrido[2,3-*d*]pyrimidine **4.54** in hand, we decided to test if such compound was accessible by using our multicomponent reaction that has allowed the synthesis of other similar pyridopyrimidines reducing the total steps to one and usually, increase the overall yield, without compromising the product purity. The synthesis is carried out by mixing **4.51**, guanidine carbonate **4.49** and malononitrile **4.6** in a microwave vial using MeOH as solvent. The mixture is heated at 140 °C under microwave radiation for 15 min. The resultant solid was collected by filtration, washed with H₂O, EtOH and Et₂O and oven dried to afford pure **4.54** (51%) as a yellowish solid.

Given that this alternative reduces the time needed to obtain the product and increases notably the overall yield (from 5.3 to 15.4%) this strategy is the preferred to synthesize the needed amount of **4.55** to perform the Steglich amidation of a series of spacers forming a novel family of DM1 hit candidates that are discussed in the following chapter.

4.4 Discussion

A structural study of 2,4-diaminopyrido[2,3-*d*]pyrimidine systems developed by our group has been carried out to take advantage of the theoretical complementarity between these structures and nucleotides, following a Janus-Wedge interaction model. The designed compounds must fulfill several structural features to comply with this model, i.e. planarity and the correct arrangement of hydrogen bond acceptors and donors. The 2,4-diaminopyrido[2,3-*d*]pyrimidine structures are ideal for interacting with thymine or uracil as it can be seen in Fig 4.62. Since a hallmark in DM1 is the release of MBNL1 from foci via CUG stabilization, selective uracil recognition structures are relevant.

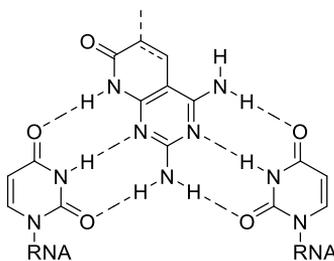


Fig 4.70. Schematic representation of Janus-Wedge interaction mode of pyrido[2,3-*d*]pyrimidines and two uracil moieties via hydrogen bond.

First, a study based on substitution at the C2 position was performed. Several approaches were considered to introduce a spacer linked to C2 and, in this context, the synthesis of compound **4.14.1** was reformulated (see Fig. 4.71).

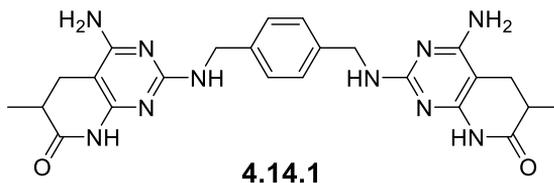


Fig. 4.71. Structure of **4.14.1**.

Compounds like **4.14.1** can present a strong interaction with uracil and, thanks to the presence of two pyridopyrimidines at the end of the linker, selectivity against two sequential U•U mismatches in rCUG sequence could be achieved. This kind of structures preferably interacts with single uracil due to the position where the spacer is bound (C2). Such strategy has been already successfully described in the literature,^{33,34} and therefore, we decided to further investigate the 2-substituted pyrido[2,3-*d*]pyrimidines to achieve uracil monorecognizers. Similar to the studies of Li, Jinxing *et al.*,³³ We decided to increase the distance between the hydrogen bond triggering functionalities in order to reduce the cross-repulsions and, thus, increase the intensity of the classical Watson-Crick interactions.

As a part of this approach, in this chapter, we have been able to construct compound **4.24** by using Suzuki methodologies on 2-bromo substituted pyridopyrimidines thus connecting two

Discussion

deficient aromatic systems. This achievement opens the possibility of synthesizing new families of compounds with potential uracil recognition capacity. However, these compounds had not been developed yet.

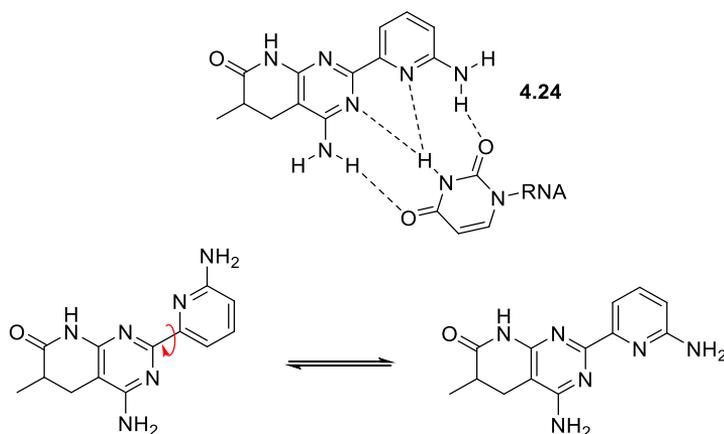


Fig 4.72. Schematic representation the uracil interaction of structure 4.24. Also show the ability of the C2 bond created between the two heterocyclic moieties to rotate allowing dual side recognition.

On the other hand, the synthesis of 6-halo substituted pyrido[2,3-*d*]pyrimidines as intermediates for the synthesis of DM1 candidates substituted at C6. The reactivity of pyrido[2,3-*d*]pyrimidine structures 4.36 and 4.40 in the presence of bromine or *N*-halosuccinimides has been studied. proposing a possible mechanism for the simultaneous halogenation at C6 and formation of a C5-C6 unsaturation, and establishing the scope of the reaction. In the case of 4.36 diverse dihalogenated compounds were obtained (Fig 4.64) opening the door to the orthogonal decoration of such structures.

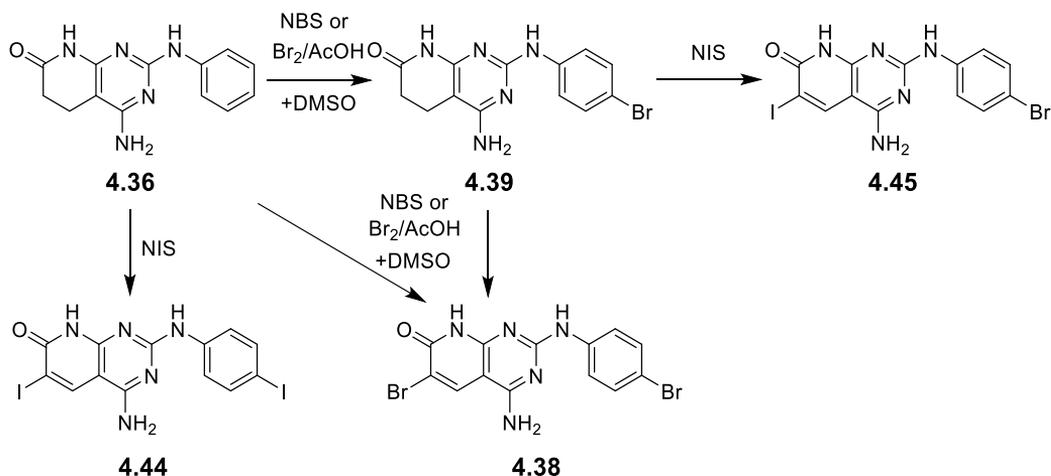


Fig 4.73. Schematic summary of all synthesized structures from 4.36.

Focusing on 4.40, the 6-bromo substituted and 6-iodo substituted pyridopyrimidines 4.42 and 4.43 were obtained (Fig 4.74). Several transformations at the halogenated position were

tested but only the introduction of a cyano group was possible starting from the bromo substituted compound **4.42**. However, the transformation of the 6-cyano substituted compound **4.47** in the corresponding carboxylic acid failed and we discarded this approach for synthesizing anti-DM1 candidates.

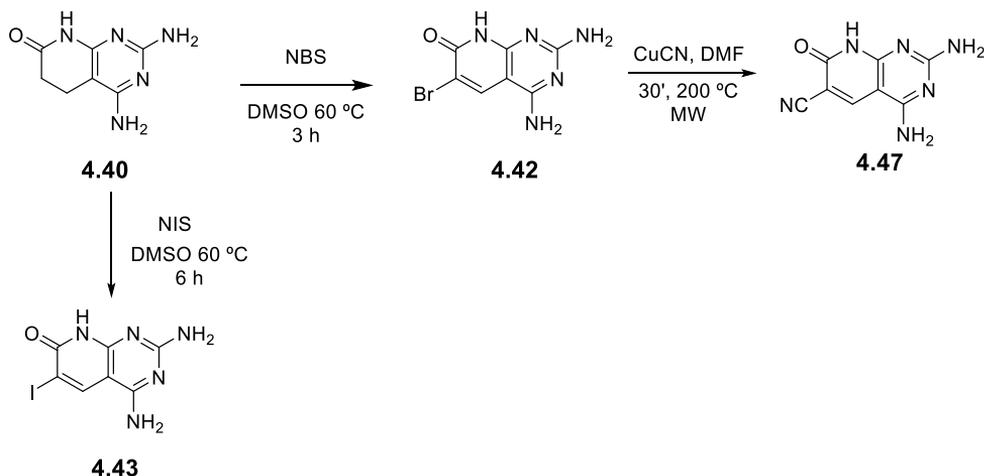


Fig 4.74. Schematic representation of the synthetic products obtained from 4.40.

Finally, we decided to introduce a carboxymethyl group at position C6 of the pyrido[2,3-*d*]pyrimidine system and, consequently, structure **4.55** was synthesized. **4.54** was initially obtained using the acyclic strategy achieving the desired product in four steps and a 5.3% global yield. Subsequently, we synthesized **4.55** using a multicomponent reaction, reducing the number of steps, and increasing the overall yield to 15.4%. **4.54** was transformed to the corresponding carboxylic acid **4.55** that in preliminary assays shows reactivity against primary amines under Stechlich coupling conditions.



Fig 4.74.

Contrary to structure **4.24** or its potential derivatives which have a binding capacity against only one uracil, the possible structures derived from **4.55** have potential interaction with two uracils, being possible their Janus-Wedge interaction with two uracils in dsCUG systems. Therefore, **4.55** was selected to build the next generation of anti-DM1 candidates.

References

4.5 References

1. Hamby, J. M. *et al.* Structure-activity relationships for a novel series of pyrido[2,3-d]pyrimidine tyrosine kinase inhibitors. *J. Med. Chem.* **40**, 2296–2303 (1997).
2. Klutchko, S. R. *et al.* 2-Substituted aminopyrido[2,3-d]pyrimidin-7(8H)-ones. Structure-activity relationships against selected tyrosine kinases and in vitro and in vivo anticancer activity. *J. Med. Chem.* **41**, 3276–3292 (1998).
3. Boschelli, D. H. *et al.* Synthesis and tyrosine kinase inhibitory activity of a series of 2-amino-8H-pyrido[2,3-d]pyrimidines: Identification of potent, selective platelet-derived growth factor receptor tyrosine kinase inhibitors. *J. Med. Chem.* **41**, 4365–4377 (1998).
4. Dorsey, J. F., Jove, R., Kraker, A. J. & Wu, J. The pyrido[2,3-d]pyrimidine derivative PD180970 inhibits p210(Bcr-Abl) tyrosine kinase and induces apoptosis of K562 leukemic cells. *Cancer Res.* **60**, 3127–3131 (2000).
5. Wisniewski, D. *et al.* Characterization of potent inhibitors of the Bcr-Abl and the c-kit receptor tyrosine kinases. *Cancer Res.* **62**, 4244–45 (2002).
6. Huang, M. *et al.* Inhibition of Bcr-Abl kinase activity by PD180970 blocks constitutive activation of Stat5 and growth of CML cells. *Oncogene* **21**, 8804–8816 (2002).
7. Huron, D. R. *et al.* A novel pyridopyrimidine inhibitor of Abl kinase is a picomolar inhibitor of Bcr-abl-driven K562 cells and is effective against STI571-resistant Bcr-abl mutants. *Clin. Cancer Res.* **9**, 1267–1273 (2003).
8. Wolff, N. C. *et al.* PD166326, a novel tyrosine kinase inhibitor, has greater antileukemic activity than imatinib mesylate in a murine model of chronic myeloid leukemia. *Blood* **105**, 3995–4003 (2005).
9. Martinez-Teipel, B. *et al.* 2-Methoxy-6-oxo-1,4,5,6-tetrahydropyridine-3-carbonitriles: Versatile Starting Materials for the Synthesis of Libraries with Diverse Heterocyclic Scaffolds. *J. Comb. Chem.* **7**, 436–448 (2005).
10. Perez-Pi, I., Berzosa, X., Galve, I., Teixido, J. & Borrell, J. I. Dehydrogenation of 5,6-dihydropyrido[2,3-d]pyrimidin-7(8H)-ones: a convenient last step for a synthesis of pyrido[2,3-d]pyrimidin-7(8H)-ones. *Heterocycles* **82**, 581–591 (2010).
11. Galve, I. *et al.* Synthesis of 2-arylamino substituted 5,6-dihydropyrido[2,3-d]pyrimidine-7(8H)-ones from arylguanidines. *Mol. Divers.* **16**, 639–649 (2012).
12. Camarasa, M., Barnils, C., Puig Raimon, de la B., Teixido, J. & Borrell, J. I. A new and practical method for the synthesis of 6-aryl-5,6-dihydropyrido[2,3-d]pyrimidine-4,7(3H,8H)-diones. *Mol. Divers.* **17**, 525–536 (2013).
13. Borrell, J. I. *et al.* Synthesis and biological activity of 7-oxo substituted analogues of 5-deaza-5,6,7,8-tetrahydrofolic acid (5-DATHF) and 5,10-dideaza-5,6,7,8-tetrahydrofolic acid (DDATHF). *J. Med. Chem.* **44**, 2366–2369 (2001).
14. Borrell, J. I. *et al.* An Unequivocal Synthesis of 4-Amino-1,5,6,8-tetrahydropyrido[2,3-d]pyrimidine-2,7-diones and 2-Amino-3,5,6,8-tetrahydropyrido[2,3-d]pyrimidine-4,7-diones. *Collect. Czechoslov. Chem. Commun.* **61**, 901–909 (1996).
15. Mont, N., Teixidó, J., Borrell, J. I. & Kappe, C. O. A three-component synthesis of pyrido[2,3-d]pyrimidines. *Tetrahedron Lett.* **44**, 5385–5387 (2003).
16. Mont, N., Teixidó, J., Kappe, C. O. & Borrell, J. I. A one-pot microwave-assisted

- synthesis of pyrido[2,3-d]pyrimidines. *Mol. Divers.* **7**, 153–159 (2003).
17. I. Borrell, J., Pérez-Pi, I., Berzosa, X., Galve, I. & Teixidó, J. Dehydrogenation of 5,6-Dihydropyrido[3,2-d]pyrimidin-7(8H)-ones: A Convenient Last Step for a Synthesis of Pyrido[2,3-d]pyrimidin-7(8H)-ones. *Heterocycles* **82**, 581 (2010).
 18. Ros Blanco, L. Disseny i síntesi de noves quimioteques de potencials inhibidors d'entrada del VIH. (Universitat Ramon Llull, 2011).
 19. Camarasa, M. *et al.* Design, synthesis and biological evaluation of pyrido[2,3-d]pyrimidin-7-(8H)-ones as HCV inhibitors. *Eur. J. Med. Chem.* **115**, 463–483 (2016).
 20. González, À. L. *et al.* In silico discovery of substituted pyrido[2,3-d]pyrimidines and pentamidine-like compounds with biological activity in myotonic dystrophy models. *PLoS One* **12**, e0178931 (2017).
 21. Victory, P., Cirujeda, J. & Vidal-Ferran, A. Two step synthesis of pyrido[2,3-d]pyrimidines from acyclic precursors. Cyclization of 2-cyanamino-4,6-diphenylpyridine-3-carbonitrile by Hydrogen Halides. *Tetrahedron* **51**, 10253–10258 (1995).
 22. Galve, I. *et al.* Synthesis of 2-Arylamino substituted 5,6-dihydropyrido[2,3-d]pyrimidine- 7(8H)-ones from arylguanidines. *Mol. Divers.* **16**, 639–649 (2012).
 23. Mont, N., Fernández-Megido, L., Teixidó, J., Kappe, C. O. & Borrell, J. I. A diversity-oriented, microwave-assisted synthesis of 4-oxo and 4-chloropyrido[2,3-d]pyrimidin-7(8H)-ones. *QSAR Comb. Sci.* **23**, 836–849 (2004).
 24. Yu, C., Liu, B. & Hu, L. Efficient Baylis-Hillman reaction using stoichiometric base catalyst and an aqueous medium. *J. Org. Chem.* **66**, 5413–5418 (2001).
 25. Vloon, W. J., van den Bos, J. C., Koomen, G.-J. & Pandit, U. K. Synthesis of ring C modified sesbanimide analogues. *Tetrahedron* **48**, 8317–8328 (1992).
 26. Murillo, G. Síntesis de pirido[2,3-d]pirimidin-7(8H)-onas 2-arilamino sustituidas y derivados. (IQS School of engineering, 2012).
 27. Mont, N., Teixidó, J. & Borrell, J. I. A diversity oriented, microwave assisted synthesis of N-substituted 2-hydro-4-amino-pyrido[2,3-d]pyrimidin-7(8H)-ones. *Mol. Divers.* **13**, 39–45 (2009).
 28. WANG, S. Y. Chemistry of Pyrimidines. I. The Reaction of Bromine with Uracils 1-3. *J. Org. Chem.* **24**, 11–13 (1959).
 29. Fauth, F., Peral, I., Popescu, C. & Knapp, M. The new Material Science Powder Diffraction beamline at ALBA Synchrotron. *Powder Diffr.* **28**, 360–370 (2013).
 30. Boultif, A., Louër, D. & IUCr. Powder pattern indexing with the dichotomy method. *J. Appl. Crystallogr.* **37**, 724–731 (2004).
 31. Galve, I. *et al.* A captured room temperature stable Wheland intermediate as a key structure for the orthogonal decoration of 4-amino-pyrido[2,3- d]pyrimidin-7(8 H)-ones. *Org. Biomol. Chem.* **18**, 9810–9815 (2020).
 32. Schoffstall, A. M. Synthesis of 5,6-dihydropyrido[2,3-d]pyrimidine derivatives directly from acyclic precursors. *J. Org. Chem.* **36**, 2385–2387 (1971).
 33. Li, J. *et al.* A Ligand That Targets CUG Trinucleotide Repeats. *Chem. - A Eur. J.* **22**, 14761 (2016).

References

34. Li, J. *et al.* A Dimeric 2,9-Diamino-1,10-phenanthroline Derivative Improves Alternative Splicing in Myotonic Dystrophy Type 1 Cell and Mouse Models. *Chem. - A Eur. J.* **24**, 18115–18122 (2018).

5 Base Recognizers

In this chapter, we present pyrido[2,3-*d*]pyrimidin-7-(8*H*)-ones combined with a flexible linker able to stabilize DM1 toxic transcripts, enabling the release of MBNL1. The novel-designed family of compounds fits in the RNA using the Janus-Wedge interaction model using MD structure-based drug design techniques. Compounds are tested both in the FRET assay and in primary patient-derived myoblasts.

Introduction

5.1 Introduction

5.1.1 Overview

In this last chapter of this thesis, gained knowledge from the rest of the chapters is used to obtain a new family of DM1 candidates capable of achieving relevant biological activity. For this purpose, the design of structures is based on candidate **4.55** (see section 4.3.4.5), the most versatile system to form structures capable of interacting with CUG RNA.

5.1.2 The base interactor

In chapter 3 of the present thesis, the interaction mode of groove binders, compounds capable of recognizing the helical secondary structure of nucleotide chains, has been discussed. On the other hand, two different types of reversible interaction with nucleotides are described for small molecules: base intercalation and nucleobase recognition.

5.1.2.1 Base intercalators

Intercalation between nucleotides is a common phenomenon. Many known bioactive structures have this activity. Compounds such as benzo[*a*]pyrenes (Fig 5.2)¹ or ethidium bromide² are known to have this type of activity. Thus, when a compound has to be embedded between two levels of a nucleotide secondary structure (i.e. dsDNA or dsRNA) is often associated with cytotoxicity or carcinogenesis. Indeed, some known structures provide this feature as an asset for pharmaceutical treatment, such as antibiotics (i.e. actinomycin D)³ or anti-cancer drugs (i.e. topotecan).⁴

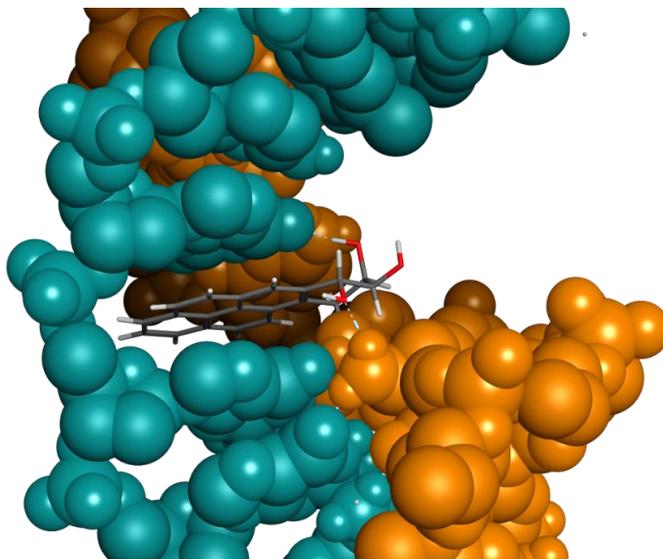


Fig.5.1. Intercalation of a benzo[*a*]pyrene in dsDNA is displayed (adapted PDB id:1BPS).⁵

Introduction

For this reason, it is a relatively little-studied field for the treatment of cells that do not want to be eliminated, although there are many molecules with this type of activity, even among the hormones themselves (i.e. estradiol).⁶ It is therefore essential to differentiate cytotoxicity from activity via nucleotide intercalation. Unfortunately, the problem with these structures is their poor selectivity since the basis of their interaction lies in pi-stacking. This source of intermolecular attraction can attain remarkable energy since nucleotides are mostly electronically deficient and a planar structure with high electron density in a pi-expanded orbital system can lead to a strong interaction. However, most structures designed exclusively towards this type of interaction are poorly selective and, adding the binding strength, unequivocally leads to cytotoxicity.

5.1.2.2 Base recognizer

The second mode of interaction with nucleotides presented in this chapter is base recognition. Lehn et al. proposed a new model for DNA nitrogenated base-pair recognition. According to their hypothesis, the interaction between molecules and nucleotides is possible by falling them in base pairs to maximize the number of Watson and Crick interactions, thus establishing up to 6 hydrogen bonds. This model, called the Janus-Wedge (JW) model,⁷ was initially intended for use in DNA but is adaptable to any double-stranded nucleotide secondary structure (including RNA).

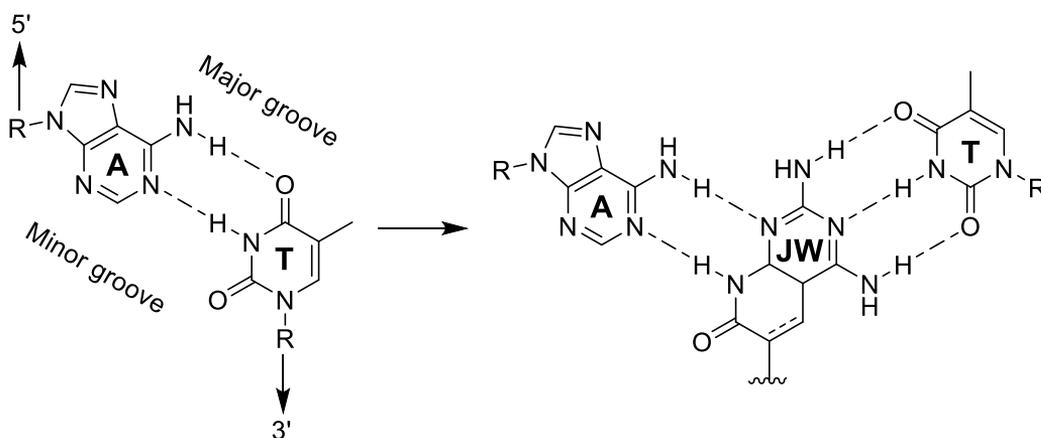


Fig 5.2. General Janus-Wedge interaction scheme (left) and JW interaction mechanism adapted to a pyrido[2,3-d]pyrimidine to fall into a A-T base-pair (right).⁸

To the best of our knowledge, a limited number of compounds have been described using this strategy, and its range of applicability is scarce. It is recognized that Janus-Wedge structures have their major defect in DNA recognition. This is because DNA has a very stable secondary structure and opening the strands to fit a molecule requires a hybridization process, which exogenous molecules can rarely achieve without the help of a thermal process. Moreover, there is the problem that a single structure of this style cannot distinguish between major and

minor grooves. Consequently, it cannot distinguish between an inverted pair of nucleotides (i.e. A-T or T-A), thus drastically reducing its DNA selectivity.⁹ For this reason, this strategy is often used along with one of the already described in this thesis such as intercalation or groove binding (or both, see Fig 5.3).

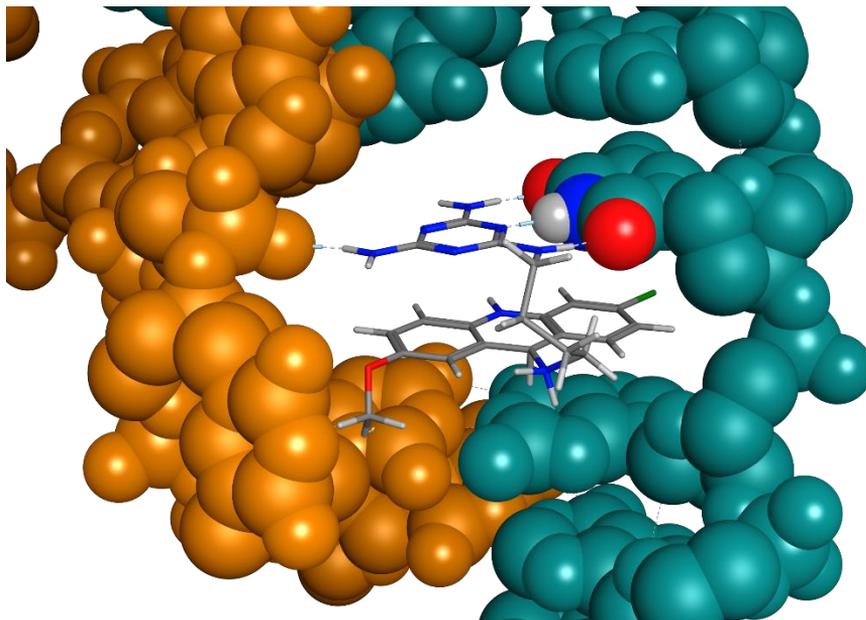


Fig 5.3. Base recognition in dsDNA is displayed paired with an intercalation acridine moiety (adapted PDB id:6M4T).¹⁰

Conversely, the same features that preclude base-recognizers from reaching such structured targets as dsDNA are an opportunity for ncRNA. Given the dynamic nature of many RNA-based secondary structures, the hybridization energy is much lower. Assuming that a ligand with sufficient affinity can find its free target and stabilize a regular conformation with the energy of its interaction with the receptor. An example of this type of ncRNAs is TREDs toxic transcripts. These ncRNAs, which are of high therapeutic interest, form repetition-based sequences. Precisely, transcripts of *DMPK* expansions form CUG chains, which, as described above, can form helical secondary structures. For JW-type ligands, they are an ideal target, as they can selectively bind to these unstructured RNAs, facilitating the stabilization of the dsRNA secondary structure, one of the hallmarks of DM1.

5.1.3 Candidate selection

Some research groups have made exciting progress in the *de novo* design of compounds with either or both modes of interaction. Given the intrinsic nature of *DMPK* transcripts, we have focused our interest on those developments relevant to nucleotide recognition.

Introduction

5.1.3.1 Zimmerman

They are the first group to create *de novo* compounds using the Janus-Wedge strategy in DM1.¹¹ Their contribution is of great value in demonstrating that this is a valid strategy in this field. This group has published numerous candidates over the last decade.^{12–15} They have used all the nucleotide interaction methods described in this thesis, maintaining uracil recognition as the primary therapeutic strategy. They use triaminotriazine, a simple and commercially available core.

Early studies conjugated this structure to acridine, similar to the one depicted in Fig 5.3. The product obtained synergistically utilized the binding potency of an intercalator such as acridine with the selectivity to uracil of triaminotriazine.¹¹ Despite its high potency, the compound stalls in one of the common drawbacks of bearing a non-specific intercalator in its structure: it exhibited high cytotoxicity. In addition, the compound had very low solubility in the culture media and failed to cross membranes. This is one of the most common problems in drug design, which worsens in DM1 since modular compounds are usually even heavier and more hydrophobic than classic pharmacological hits. In addition, their target is located inside the nucleus, behind a second barrier. In a subsequent generation of candidates, they managed to relatively improve the properties of the candidate by coupling the structure to a spacer exhibiting groove binder activity.^{12,13}

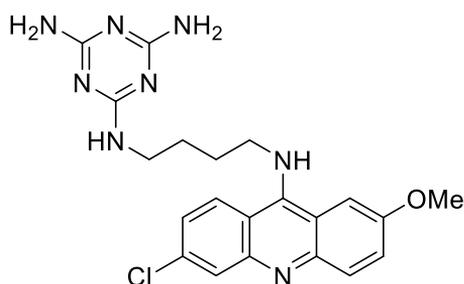


Fig 5.4. Acridine- Triaminotriazine conjugate structure.¹¹

Another variant of the triaminotriazine-based structures is designed initially as a dimer, with a distance between the recognizing units optimal for binding to two uracils in the CUG sequence. They place an aliphatic spacer between them with a *p*-phenyl bisamidinium unit, a group that provides two essential characteristics to the candidate. On the one hand, groove binding potency, its positive charge confers binding capacity to any nucleotide chain since these are essentially anionic due to the phosphate units in the nucleotide backbone. In addition, forming an ionic compound confers to this structure solubility and the ability to reach the nucleus much better than the previous generation.

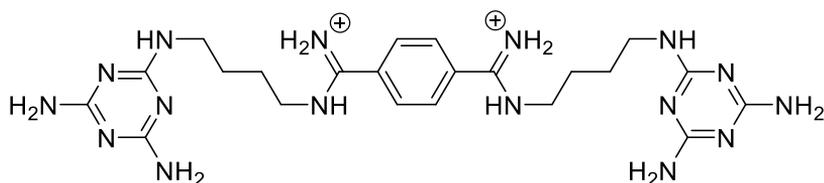


Fig 5.5. Bisamidinium triaminotriazine structure.¹⁴

On the other hand, the structure has a lower biological activity than the acridine-conjugated product. The transformation of the dimer into an oligomer compensates this defect. Thus, generating a compound that, in addition to the advantages already mentioned, improves the binding capacity and selectivity since more equispaced triaminotriazine units reduce the number of possible off-targets for binding. The oligomeric mixture product shows activity in the submicromolar range and no toxicity.¹⁵

5.1.3.2 Nakatani

Nakatani's group is another relevant group in the study of Janus-Wedge recognition. According to the studies performed by this group, a way to improve the binding potency against uracil. Their most promising ligand against DM1 has a differential feature. It does not pair the hydrogen bridges canonically but slightly increases the space between hydrogen bond groups, building a structure that seems to surround the target nucleotide. According to their hypothesis, cross-repulsion between hydrogen bond acceptors and donors is diminished (see Fig 5.6).¹⁶ Furthermore, their naphthyridine structures can form potent interactions by base intercalation.

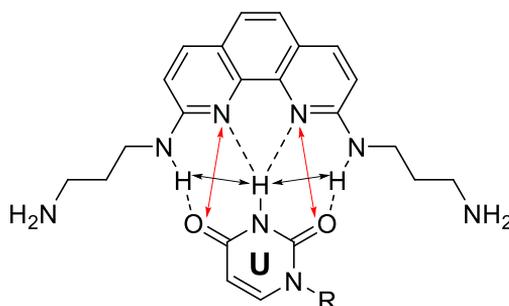


Fig 5.6. Highlighted hydrogen bond repulsions on a typical Nakatani structure. Adapted from Li *et al.*¹⁷

On the other hand, these structures target RNA bulges on CUG-disrupted structures. Unlike the structures of Zimmerman *et al.*, they excel in interacting with only one nucleotide. In the same way as other candidates' evolution process, they reported a product with the same uracil-recognizing scaffold in dimer form bearing an aliphatic spacer.¹⁸ They have reported another exciting type of structure with DM1 as a target in recent times. It consists of a diaminopyridinic structure, complying with the original Janus-Wedge canon. However, they add 1,3-diaminoisoquinoline on the opposite side of the Janus-Wedge front to seek selective

Introduction

interaction with ribose (see Fig 5.7 left).¹⁹ Moreover, taking into account the characteristic somatic instability of DM1, they reported specific ligands for slipped-DNA bases. Despite the mechanism that generates the nucleotide repetition to increase is unknown, sources point to this phenomenon as its origin.²⁰ Finally, although it is not a study directly related to DM1, they introduce a concept for this type of heterocyclic structure dependent on the arrangement of their hydrogen bridge donors and acceptors. The pKa of these heterocyclic candidates is an essential variable in drug design. They design a ligand that only binds to its target in a given pH range when one of the nitrogen atoms of its heterocycles is protonated (see Fig 5.7 right).²¹

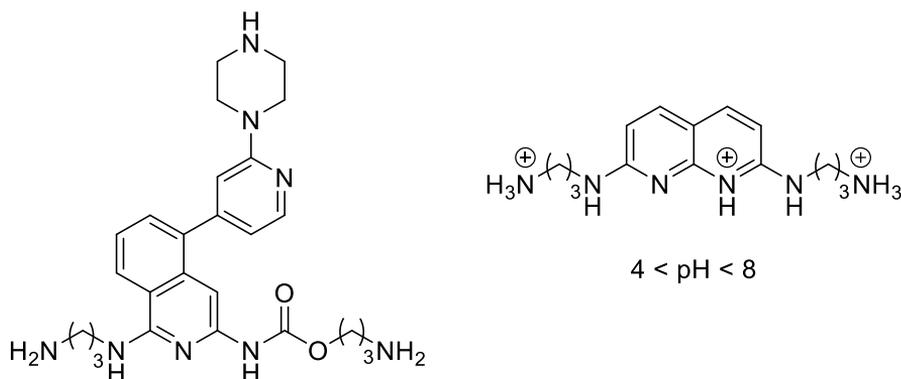


Fig 5.7. Left side: pyridinylpiperazine unit in the backbone introduced for ribose interaction.¹⁹ Right side: pKa specific protonation after the hydrogen bond profile in this specific case of naphthridine.²¹

5.1.3.3 This work

To develop new anti-DM1 ligands, we decided to adapt structures used in the group in the past years. We harnessed the experience in the field of CUG molecular design, the development of RNA-specific scoring functions, or the study of structural parameters in molecular dynamics previously developed,²² in combination with the synthetic development described in the previous chapter to obtain a new set of candidates that take advantage of the concept of JW recognition. To fulfill the model, the recognizers must fit in the target nucleobases. Specifically, in the DM1 model, the target is the U•U mismatches.²³ By stabilizing the irregular structure of the foci, the toxic r(CUG)^{EXP} may return to its soluble and non-pathogenic form, relieving MBNL1. (Fig.5.8)

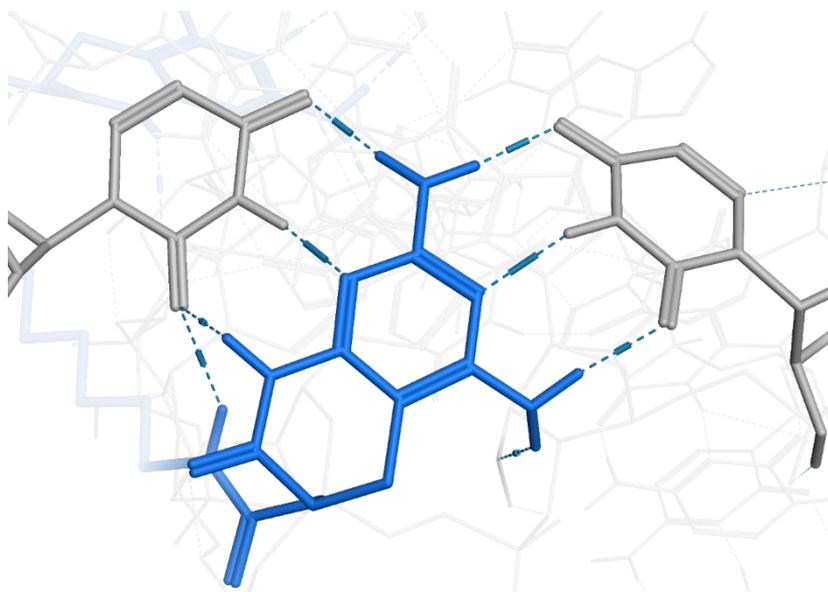


Fig. 5.8. Janus-Wedge interaction mechanism suggested between a U•U mismatch and the pyrido[2,3-*d*]pyrimidine scaffold, allowing base recognition of singular nucleic acids such as dsCUG.

In the following pages, we describe a *de novo* computer-aided drug design procedure to predict the ability of novel compounds to bind to $r(\text{CUG})^{\text{EXP}}$. The consideration of Janus-Wedge interactions as the base mechanism for CUG recognition led us to identify a novel active core based on pyrido[2,3-*d*]pyrimidine derivatives. Pyrido[2,3-*d*]pyrimidine is a privileged scaffold with notable biological activity in fields such as cancer or viral diseases.^{24,25} We adapted such structures to obtain a set of new multivalent compounds composed of two pyrido[2,3-*d*]pyrimidin-7-(8*H*)-ones units linked by an aliphatic spacer, able to target two U•U mismatches in $r(\text{CUG})^{\text{EXP}}$ structures via Janus-Wedge interactions.

Objectives

5.2 Objectives

Develop an *in silico* procedure able to evaluate the potency of novel base recognizers.

Design synthetic methodologies to join recognizing moieties to aliphatic spacers.

Synthesis of a novel base recognizer family with potential activity against DM1.

Evaluate the activity of obtained compounds *in vitro* using our previously adapted AID2675 test.

Characterize the activity of such novel compounds in patient-derived DM1 muscular cells.

5.3 Rational design

5.3.1 Strategy design

We design a whole new family of compounds based on compound **4.55**. Nucleobase recognition is achieved by the interaction shown in Fig. 5.8. This structure with a terminal carboxylic acid functionality allows coupling to spacers with nucleophilic functional groups by acyl substitution reactions such as the Steglich reaction. Since the intended interaction mechanism for these structures is entirely different from Chapter 3, so must be the rational design pipeline.

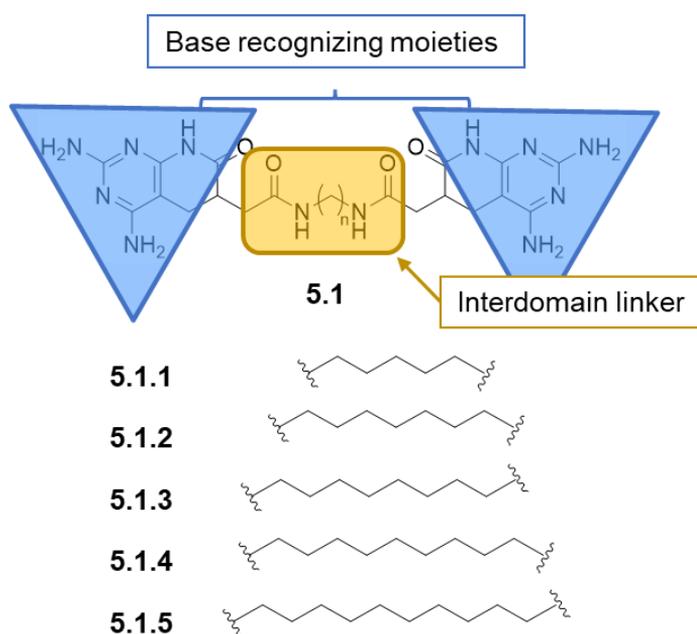


Fig 5.9. The general structure of ligands proposed in the current work. All structures are characterized by bearing a pyrido[2,3-*d*]pyrimidine scaffold at both terminal ends, separated by aliphatic spacers of different lengths. Both terminal scaffolds are bound to the central spacer with an amide bond.

For the first generation of structures based on JW interaction, we intend to obtain the U•U selectivity of these recognizers by designing dimer-type compounds. By adjusting the distance between the two recognizers, a selective activity against DM1 can be obtained. We have selected a series of spacers to connect the two recognizer structures with different lengths. In this first experiment, we chose aliphatic diamines of different chain lengths. These compounds provide a strong nucleophilicity, which allows better reactivity. Moreover, the aliphatic chains do not incorporate, a priori, any non-specific interaction, thus focusing the interaction's attention on the potency of the pyrido[2,3-*d*]pyrimidinic scaffold. Finally, the use of flexible structures can improve the interaction of these ligands by allowing a dynamic

Rational design

adjustment since, unlike groove binders, they must interact within the CUG basepair structures.

We assessed the binding mechanism of the compound family **5.1** by combining molecular docking and molecular dynamics simulations, using a short hairpin CUG₁₆ (shCUG) model as the receptor in a three-step pipeline.²² Although allowing moderate motion of the side chains, flexible docking does not guarantee the generation of Janus-Wedge arrangement. Given the difficulty in dealing with static structures, we applied molecular dynamics simulations to generate an ensemble of shCUG conformations.

5.3.2 Receptor preparation

We performed this first MD to capture conformations suitable to use in the following two stages: the ligands' docking and a second MD simulation to evaluate the potency of the interaction.

5.3.2.1 Molecular Dynamics

We heat the system to 300 K within 500 ps, restraining the RNA's position with a constant 2.0 kcal/mol-Å² force. We progressively remove the restraints along the system's equilibration stage, performing this stage at constant temperature and pressure (1.0 bar, 300 K correspondingly). We conduct a final NPT production stage during 100 ns without restraints. Moreover, we select The Particle Mesh Ewald method for electrostatic interactions under periodic boundary conditions and SHAKE²⁶ algorithm for hydrogen atoms. We conduct the simulations using AMBER18 software, fixing the time step to 2 fs (University of California, San Francisco, CA). A set of frames exhibiting opening values similar to those expected from alternate U•U mismatches was selected using the cpptraj²⁷ module.

The length of the hairpin model ensures a sufficiently large double-stranded conformation for bivalent ligands to bind consecutive or non-consecutive uridine units. It is also short enough to allow long production times in MDs, which is critical in this study. Therefore, the dynamics of non-canonical U•U pairs were analyzed to capture the preferred conformation to stabilize JW interactions.

5.3.2.2 Induced-fit docking

After that, the second step is to fit **5.1** compounds in a selected conformation. Docking was pharmacophore-guided, considering that base recognizers fall in base pairs via major groove (Fig 5.10.A).²⁸ The pyrido[2,3-*d*]pyrimidine unit fulfills the requirements to act as a uracil recognizer, establishing the desired hydrogen bond (Fig. 5.8).

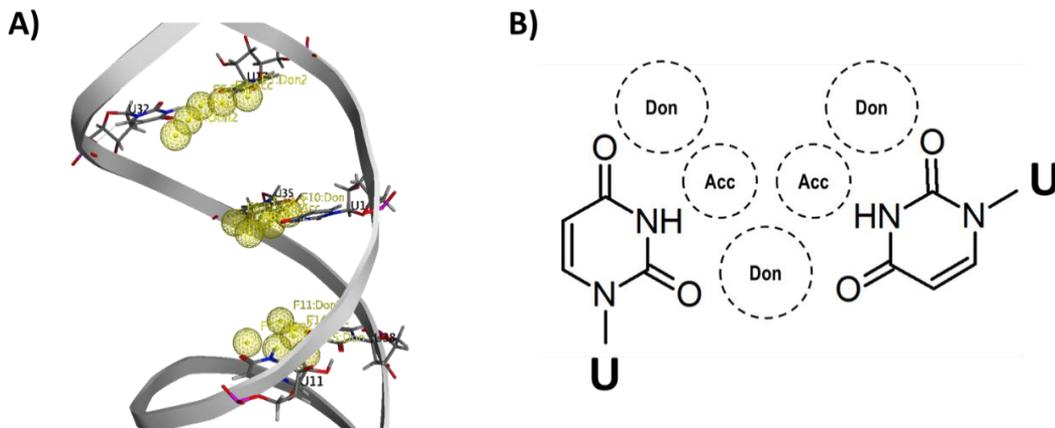


Fig. 5.10. a) Graphical representation of the pharmacophore features within the RNA structure. b) Schematic view of the pharmacophore model generated for each U•U mismatch.

We developed the pharmacophore model with Molecular Operating Environment (MOE v.2019.06) software (Chemical Computing Group, Montreal, QC) to evaluate the ability of the ligands to establish a Janus-Wedge interaction pattern with two U•U mismatches.

The characteristics required of the conformations extracted from the first MD promote the opening of the uracil base pairs toward the major hairpin groove. The generated model involved three consecutive U•U mismatches, leading to 5 pharmacophore features (3 H-bond donors and 2 H-bond acceptors) (Fig. 5.10). Thus, we can explore different interaction modes and systems that may interact with consecutive (short side chain) or alternating mismatches (longer side chain).

Thereafter, we perform pharmacophore-guided dockings using MOE under induced-fit conditions. For the docking study, we considered two different score functions (London dG for the placement and GBVI/WSA dG in the refinement). We select the best conformation for each ligand for the subsequent calculations.

5.3.3 Ligand evaluation

5.3.3.1 Molecular dynamics and MMPBSA

To further evaluate the interaction stability of docked compounds, we perform molecular dynamics (MD, Fig.5.11). The spacer length determines the binding mechanism of pyrido[2,3-*d*]pyrimidine derivatives and their ability to act as bivalent ligands (targeting two U•U pairs); As previously depicted, shorter linkers would not be long enough to reach both mismatches, while longer ones could trigger swapping the interaction to a more distant base pair. The stability of all complexes was assessed by studying the structural parameters and quantified

Rational design

calculating the free energy of binding, through molecular mechanics Poisson-Boltzmann surface area continuum solvation analysis (MMPBSA).

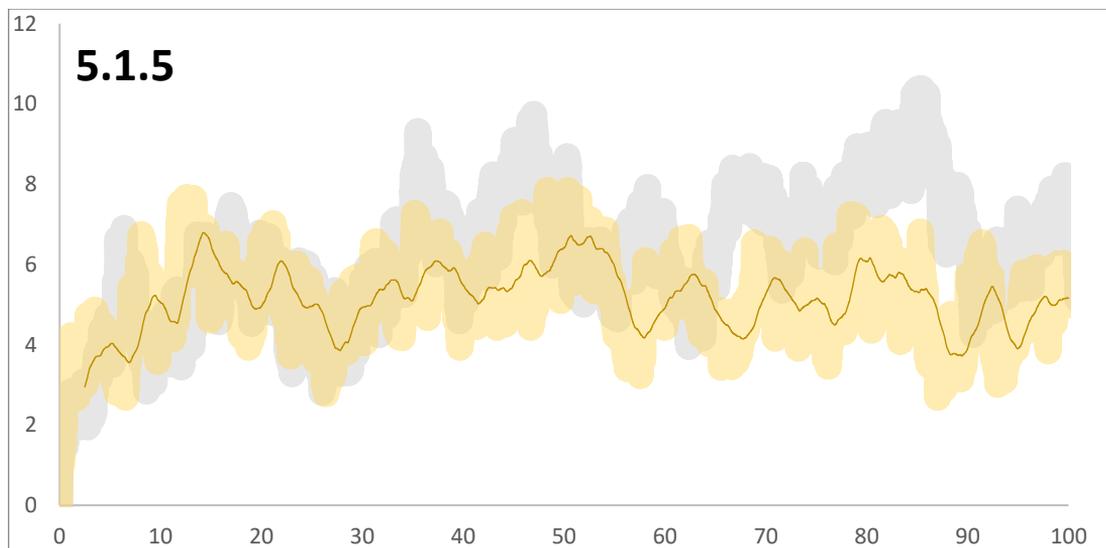


Fig. 5.11. RMSD profile of the RNA structure along MD calculations performed considering the RNA structure only (grey) and in complex with the compound 5.1.5 (yellow). RMSD is represented in Armstrong (Å) and the simulation time in ns.

Although the root mean square deviation (RMSD) values of the center of mass of the system in the ligand dynamics exemplified in Fig 5.11 show a minor drift of these values, we attempted to quantitatively measure the stabilization of the system by initially using the MMPBSA. Variations in the binding mechanism would explain the difference in Poisson-Boltzmann Gibbs free energy of binding ($\Delta^\circ G_{PB}$), predicted by MMPBSA (Fig. 5.12).

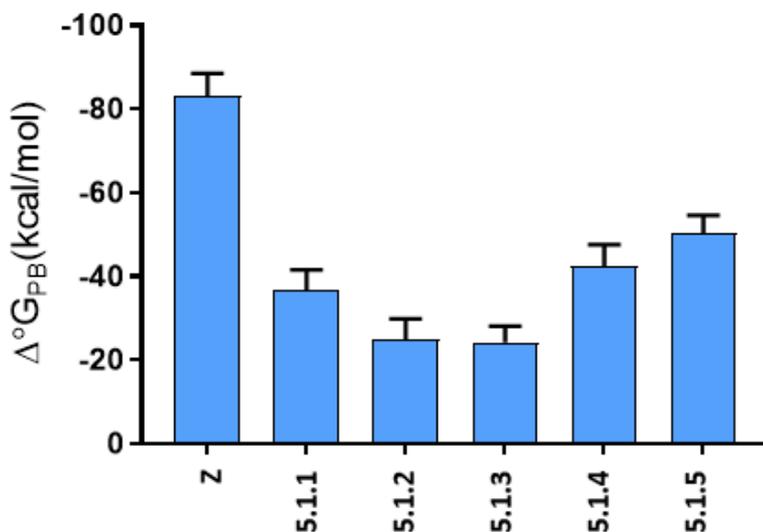


Fig. 5.12. MMPBSA values of free Gibbs energy binding calculated over 20 ns of each ligand in their respective MD. Z compound stands for the Zimmerman compound (Figure 5.5) considered as positive control.

Results show a clear dependence of the spacer length, highlighting **5.1.1** and **5.1.5** (-37 ± 5 kcal/mol and -50 ± 4 kcal/mol, respectively). The pharmacophore allows both consecutive and non-consecutive binding for the ligands. Short ligands (**5.1.1**) can achieve higher binding energies via consecutive mismatch interaction, whereas longer ones reach alternate mismatches (**5.1.4**, **5.1.5**). The nature of the spacer would account for the significant energy gap between pyrido[2,3-*d*]pyrimidine derivatives (**5.1**) and the positive control (**Z**). Zimmerman's bisamidinium linker gives the desired distance between active sites of the molecule and a boost in unspecific binding energy, while linker moieties used in the current work lack any functional groups. As this was the first time that this set of structures has been evaluated against DM1, it is desirable to assess whether the primary scaffolds are active or not without any further interactions from the moiety core. For this reason, linkers were maintained as simple as possible, although it is a drawback in terms of total binding energy.

5.3.3.2 Structural parameters analysis

One way to measure nucleotide conformational stability is through structural parameters. This geometrical data describes the geometric orientation and distances of pairs and sequences of nucleotides. The arrangements of paired bases are defined with respect to the orientation and displacement of an ideal, planar Watson–Crick base pair, where the rigid-body parameters (buckle, propeller, opening, shear, stretch, and stagger) are zero. Therefore, the less variable a set of structural parameter data, the more stable a bond between base-pair turns.

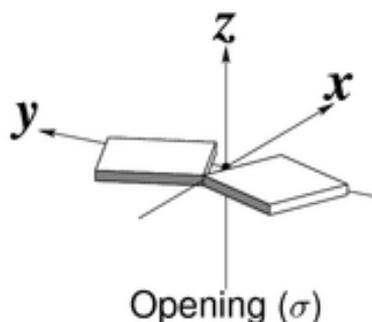


Fig. 5.13. a) Graphical representation of opening structural parameters.²⁹

Structural parameters are not the only statistical data available in nucleotide structural data (helical parameters, sugar pucker). Two-base structural parameters are selected as are the more simple way of identifying base-pair small molecule recognition. The opening angle between a base pair is the preferred option among all the given parameters because it is the most affected in recognition via hydrogen bonding.

Rational design

To perform the structural parameter analysis, x3dna is used on the full-length molecular dynamics to understand the statistical variations upon the relaxation of the CUG hairpin receptor and a given ligand.

Even though this software allows automatic recognition of pair bases, we find issues in the recognition of U•U mismatches. Even noting them as uncanonical pairs, the distance among them surpassed the preset cut-off for any base pair. As the mismatches are non-canonically paired bases, their distance starts at a minimum of 6 Å, while conventional base pairing is considered nonbonding upon 4 Å. After tuning the inputs inducing base-pair to force the recognition of U•U non-canonically pairs throughout the analysis, all structural parameters can be calculated even in the presence of the ligands.

Once obtained, the results of every step of the MD simulation are statistically analyzed using R. The plot options are tuned to show the statistical variation of the structural parameter data in a boxplot.

As we design this analysis *ad hoc*, the comprehension of the obtained results may not be trivial. The expected output is a reduction in the variance of the selected structural parameters. The presented hypothesis relies on the fact that CUG RNA destabilization leads to MBNL1 sequestration phenomenon. Therefore, structures able to stabilize the hairpin conformation may lead to MBNL1 recovery.

Shear, stretch, stagger, buckle, propeller, opening, shift, slide, rise, tilt, roll, and twist are monitored for all de MD simulations. However, as explained previously, the plotted final results comprehend only the opening values.

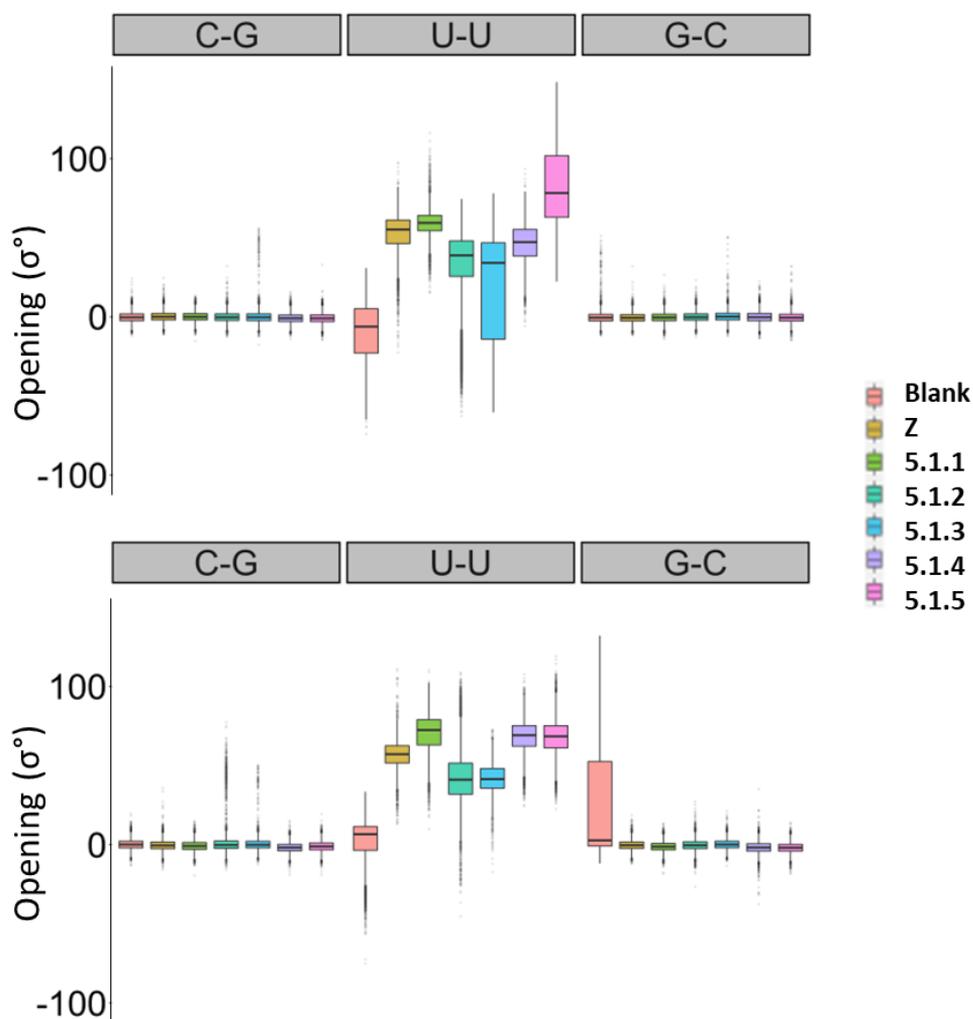


Fig. 5.14. Structural parameter analysis of opening values of two alternate CUG mismatches. The graph displays the full-length MD analysis in boxplot format.

The analysis of structural parameters in the hairpin base pairs evidenced that most of the compounds under study increase the system's stability by reducing the dispersion of structural parameters in the U•U mismatch and its vicinities (Fig. 5.14), particularly of the opening value. From all the structural parameters considered, opening explicitly represents the instability of a generated U•U mismatch, and it gives valuable data when a Janus-Wedge interaction mechanism is established (inducing an opening value shift of about 60°). As a result, we observe a reduction in the data dispersion for some of the compounds. Surprisingly, we observe a reduction in the primary U•U mismatches and the vicinity of these base pairs, representing how the introduced ligands' stability improves stability in the local mismatch and the surrounding nucleotides.

Synthesis

5.4 Synthesis

After synthesizing compound **4.55** as described in the previous chapter, we have carried out the functionalization to obtain the symmetric ligands. For this purpose, we perform a Steglich-type reaction for the amidation between the **4.55** acid and the diamine-type spacers, as shown in Fig 5.15.

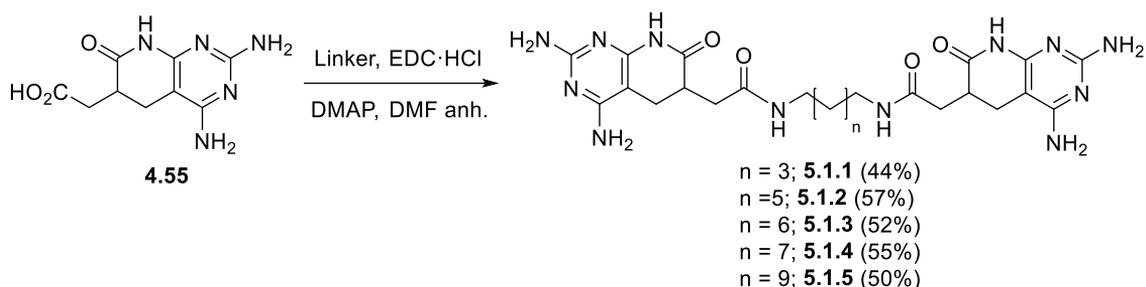


Fig 5.15. Coupling reaction between acid **4.55** and different diamine-type spacers to obtain ligands **5.1**.

First, we dissolve the acid **4.55** with EDC-HCl and DMAP base in anhydrous DMF. This mixture is stirred for 15 minutes at room temperature. Next, the diamine spacer is added and allowed to react for 24 hours at room temperature. After this time, the DMF is concentrated, and the solid is washed with 0.5 M HCl to remove possible residues of the reagents used, obtaining the ligands in the form of a gray solid.

We base the characterization of these compounds on complete NMR spectroscopy. In the following, we discuss the assignment of the signals of the candidates, taking ligand **5.1.4** as an example.

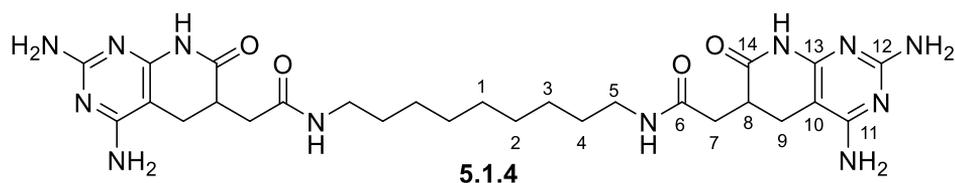


Fig 5.16. Structure of the 9-carbon spacer ligand **5.1.4** with numerical assignment for each carbon in the structure.

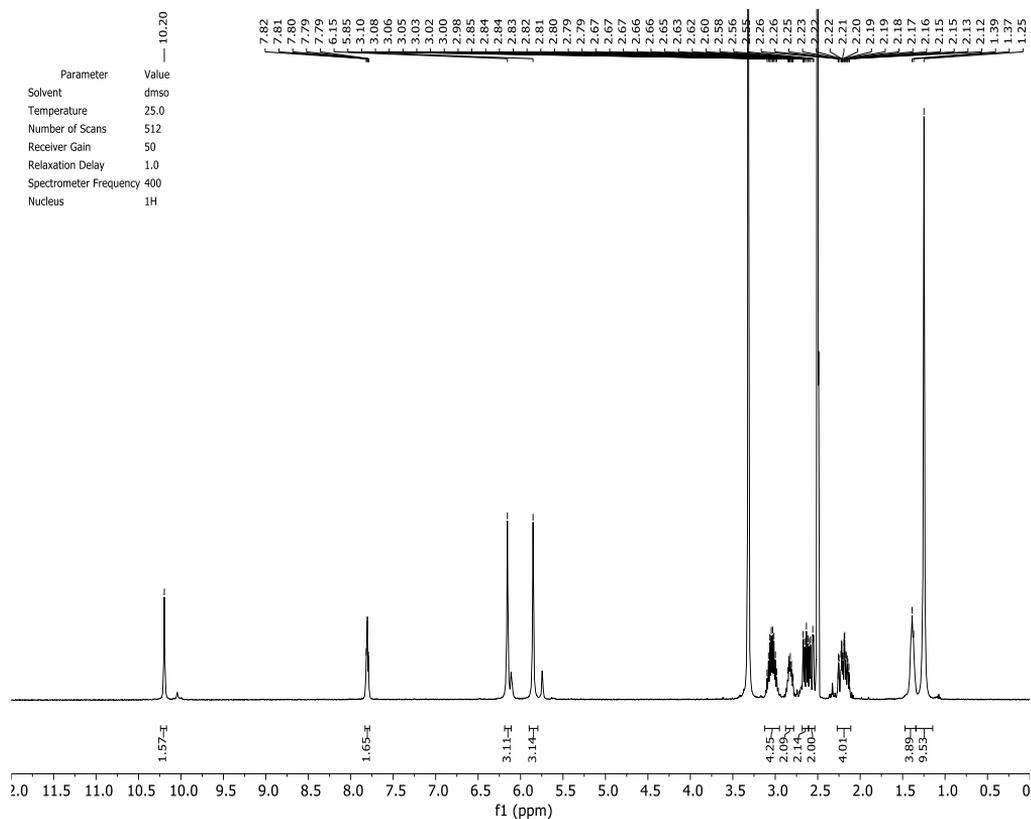


Fig 5.17. ¹H-NMR of the 9-carbon spacer ligand 5.14.

The signals observed between 1.0 and 1.5 ppm correspond to the aliphatic protons of the spacer. Due to their chemical similarity, the central C1-H, C2-H, and C3-H fall in the same signal.

The low-field signal at 10.4 ppm corresponds to the lactam protons in the pyridone ring. The triplet observed at 7.6 ppm belongs to the protons of the two amides formed in this reaction and their coupling to the neighbor methylene. The two singlets observed above 5.8 and 8.2 ppm belong to the primary amines of the molecule.

There are five signals between the chemical shifts of 2.0 and 3.2 ppm. Using HSQC, we can determine that the signal at 2.85 ppm corresponds to the C8-H proton because it is the only carbon with only one hydrogen atom and, therefore, it is the only red signal (see Fig 5.18).

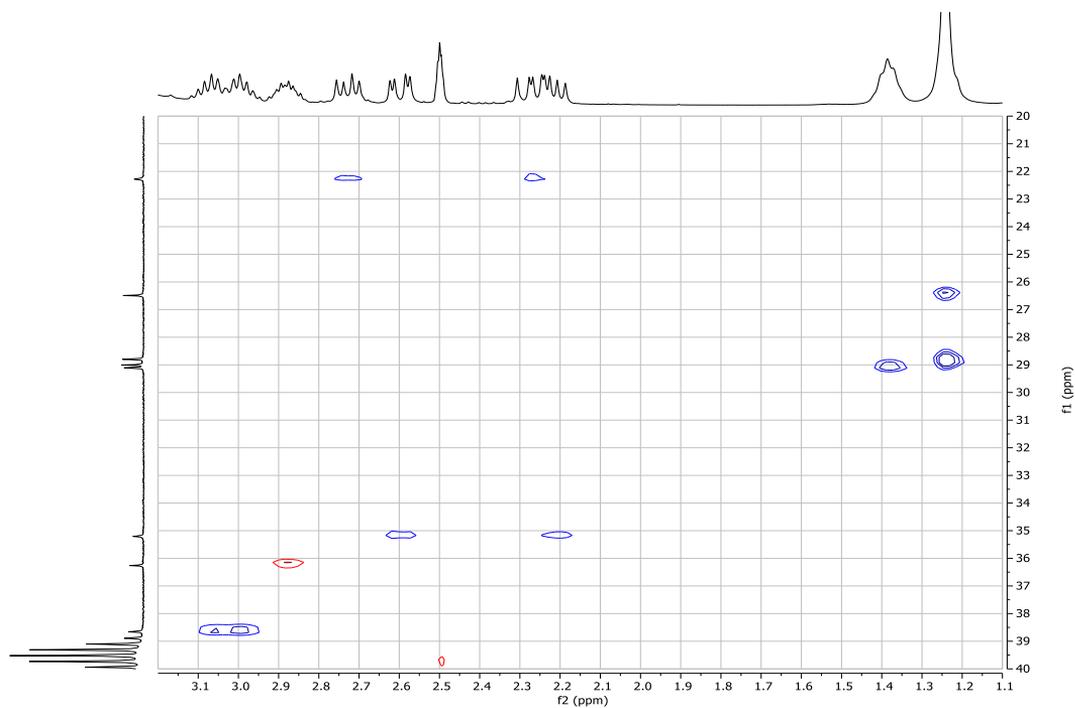


Fig 5.18. Magnification of the HSQC spectrum of ligand 5.1.4.

One common phenomenon in this kind of structure is diastereotopy. One carbon corresponds to two different proton signals, indicating they correspond to the C7-H and C9-H signals as they are in close range with a stereogenic center. Therefore, the 3.0-3.1 ppm proton signal corresponds to the C5-H. To confirm those assignments and assign with the carbon spectrum we performed HMBC.

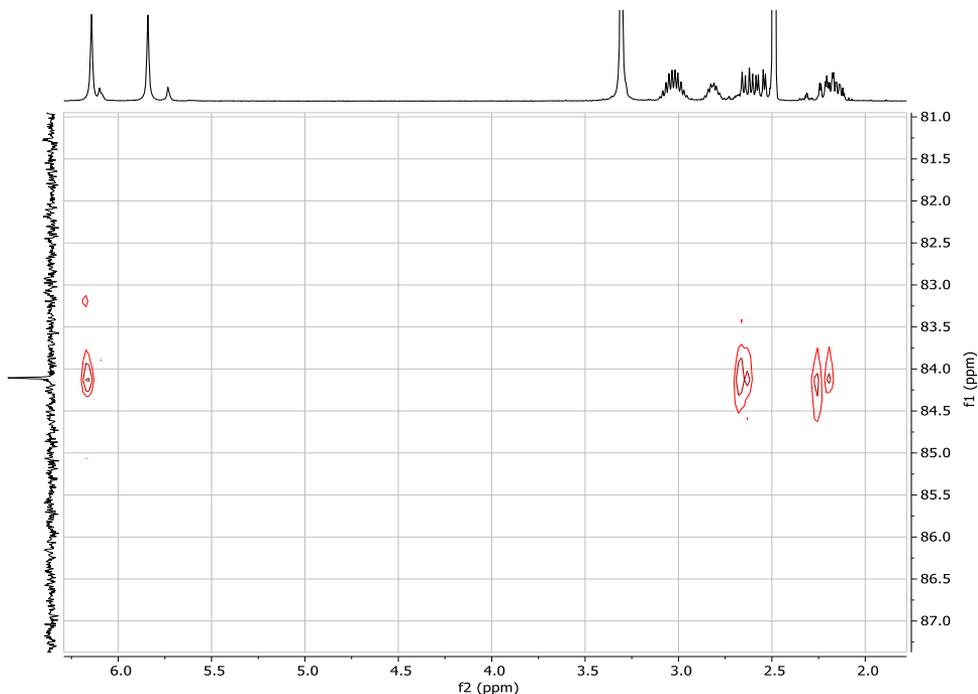


Fig 5.19. Magnification of the HMBC spectrum of ligand 5.1.4.

The carbon signal at 84.1 ppm corresponds to one of the two bridging carbons, C10, since the signal also correlates with one of the primary amines (see Fig 5.19). That confirms that signals at 2.2 and 2.7 ppm correspond to C9-H, and signals at 2.2 and 2.55 ppm correspond to C7-H.

The reaction confirming correlation is the one between the triplet of the proton spectrum corresponding to the amides and the carbonyl carbon of the amide. The carbonyl signal is observed at 169.9 ppm and correlates with the protons of the pyridopyrimidine backbone at 2.6 and 2.2 ppm. Observing the amide proton signal, a correlation is observed with such carbonyl, confirming the bond between the spacer and the heterocyclic moiety.

We isolated the rest of the structures and characterized them in the same way as **5.1.4**. High-resolution mass spectrometry confirmed all the series of novel structures and, after optimizing the work-up, were obtained with yields between 40-60%.

5.4.1 Second generation of recognizers

The poor solubility of these compounds originates from the pyrido[2,3-*d*]pyrimidine structures, a phenomenon that we have been able to verify in chapter 4 and that is part of the polar nature and the ability to form intermolecular hydrogen bonds of the selected heterocyclic moiety.

Synthesis

Based on this, and given that we cannot modify the base-recognizing unit because otherwise, they would not fulfill the JW model, we decided to test other types of spacers that could increase solubility to the systems.

We decided to use spermine, a compound with two terminal amines and two secondary amines, as a proof of concept. The main potential problem with using spermine is setting a competition between the two types of amine in the Steglich-type reaction, leading to different products. Since primary amines are more reactive than secondary amines, we tried to approach the synthesis without using protecting groups for the amines that should not react.

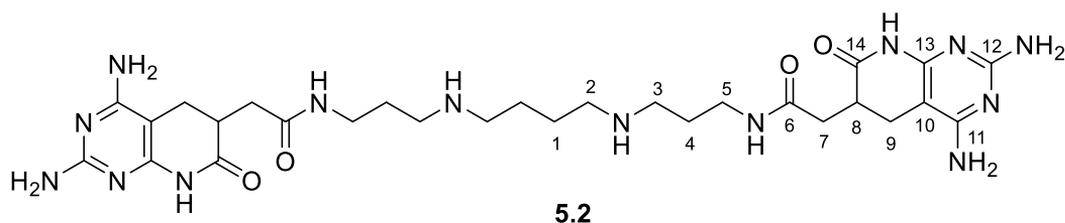


Fig 5.20. Scheme of the spermine-type ligand 5.2 numbering each of the carbons of the molecule.

All reagents were dissolved in anhydrous DMF in a microwave vial under an inert atmosphere. The mixture was stirred for 18 h at 200 °C, reaching a pressure of 14 bar. At the end of the reaction, a black crude is obtained. We evaporate the mixture, and the resulting solid is suspended in diethyl ether. The suspension was then filtered, yielding a brown solid. That product is dissolved in methanol and precipitated in diethyl ether.

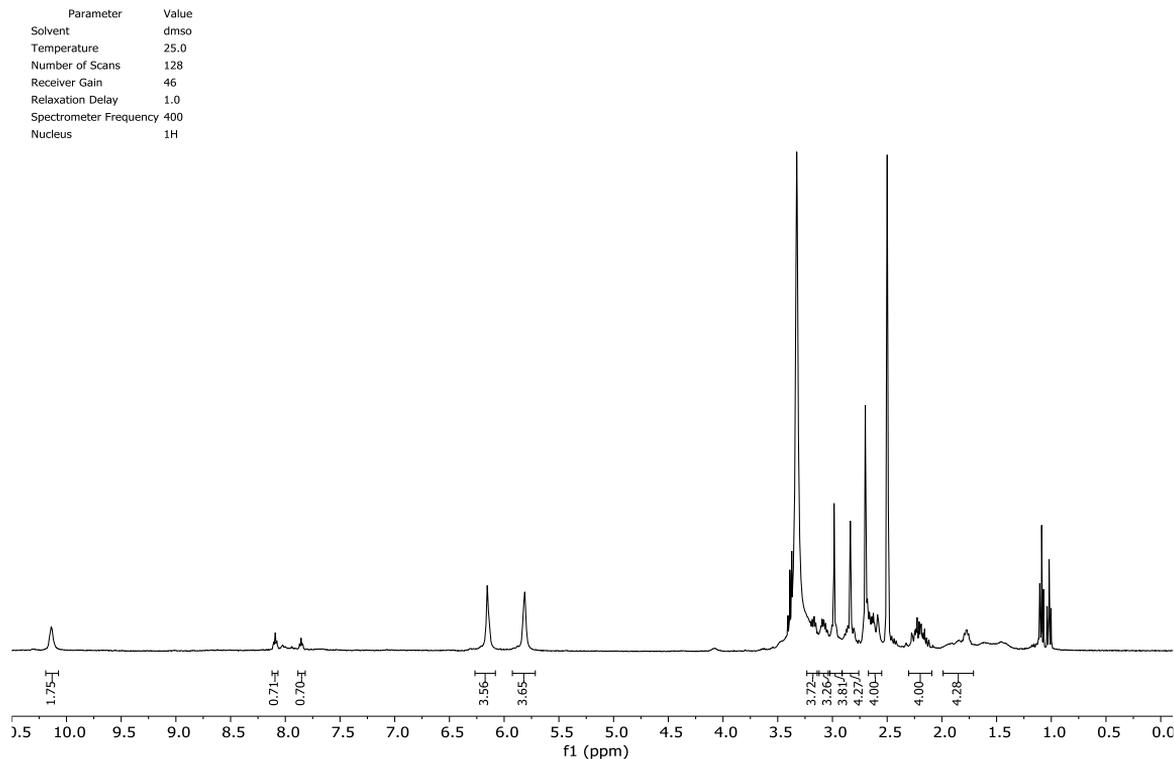


Fig 5.21. ¹H-NMR of the spermine spacer ligand **5.2**.

We proceed to the characterization of the new compound **5.2**. The signal between 2.1 and 2.4 ppm is very similar to the signal of the protons corresponding to adjacent positions of the chiral center appearing in the spectra of the aliphatic ligands (see Fig 5.21). HSQC verifies this since it presents two distinct blue signals for the carbons at 22.9 and 35.4 ppm, corresponding to the C7-H and C9-H protons.

We could not completely isolate the product **5.2**, and DMF and DMAP impurities can be observed in the spectroscopy. Despite that, we will discuss the key signals that have allowed us to confirm that the reaction is fulfilled. For this, we need both HSQC and HMBC.

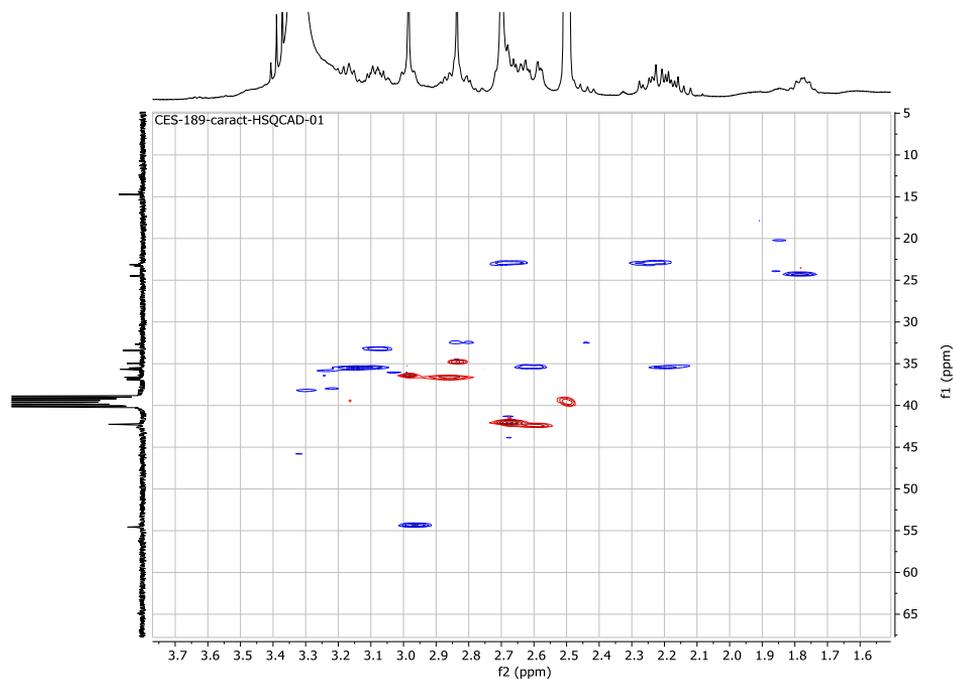


Fig 5.22. HSQC spectrum of the spermine-like ligand 5.2.

HSQC spectrum confirms the assignment of the C7-H and C9-H protons. In addition, the only signal that appears red and does not belong to any of the impurities, being the stereogenic proton C8-H, can also be assigned. Since the three protons coming from pyridopyrimidine are already assigned, the remaining methylene-type signals belong to the spermine. To confirm the formation of the product, we will use the same strategy searching for amide correlation in HMBC.

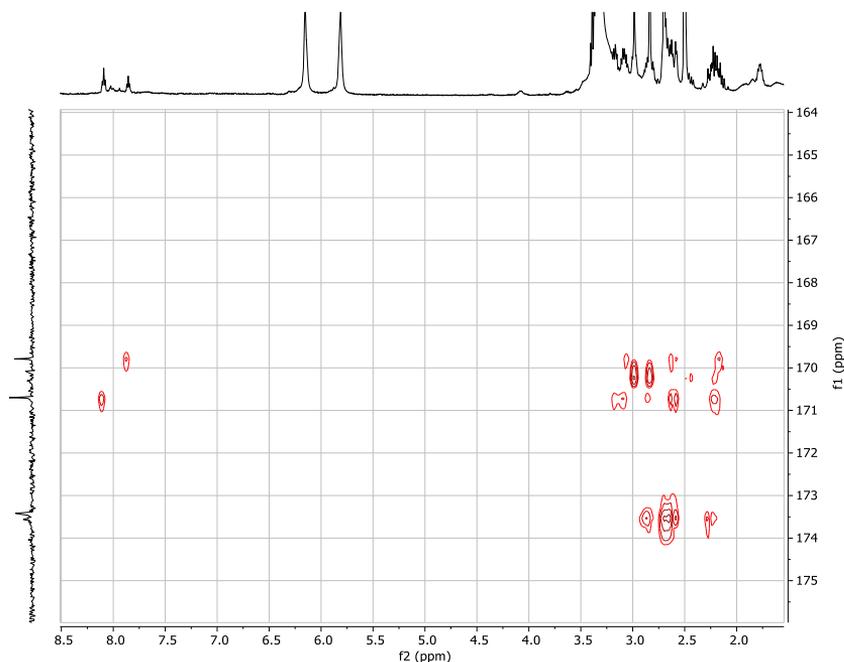


Fig 5.23. HMBC spectrum of the spermine-type ligand 5.2.

It is observed how each of the triplet signals correlates with a different carbonyl carbon signal each. Both carbon signals further correlate with the pyridopyrimidine protons discussed above, confirming the formation of the amide bond. This also suggests that the existing signals may be in duplicate, since those two carbon signals and the two triplet proton signals should be one and the same signal for each case. With all this, it is confirmed that the desired product has indeed been formed, even if it contains impurities.

Given the fact that this product is not optimized and its use in biological systems could generate misleading results, we decided not to include it in the activity study.

5.5 Biologic Evaluation

5.5.1 AID 2675

Once again, the first step in determining anti DM1 activity is the evaluation with the adapted AID2675 test described in chapter 2. This fluorescence-based test can measure the ability of candidates to free MBNL1 of an DM1-like oligo-RNA. The ability of MBNL1 (and the recombinant form used) to recognize and bind to small GC rich structures is the key factor of the AID 2675 molecular basis. Two fluorophores bind to both protein and RNA (via antibody recognition and avidin-biotin interaction respectively) making FRET possible when the complex is formed. If the complex is broken by a candidate, fluorescence intensity falls as the fluorophores are no longer near enough. Thus, FRET does not occur.

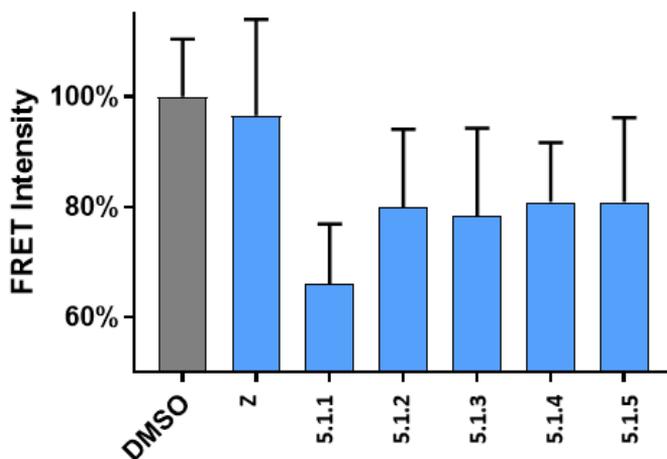


Fig. 5.24. Normalized FRET intensity was measured as a relationship between the fluorescence intensity of the signals of 545 and 665 nm (average of three measures, $n=3$). Final concentrations of studied candidates were set at $0.1 \mu\text{M}$ and of protein-RNA complex at 20 nM . Z is included for comparison.

To obtain the recombinant MBNL1, transfected *E.coli* were grown, and the purified recombinant protein was harvested prior to its use in the test. Minor changes to the original protocol allowed achieving the final recombinant protein with fewer impurities and to avoid aggregation issues (Fig. S4). The whole preparation process was conducted in RNase-free conditions, and all reagents were RNase-free grade to avoid synthetic oligo CUG from degrading before measuring the protein displacement.

Compound concentration was set at a very low range to avoid fluorescent interactions and possible precipitation issues. However, it is interesting to see that even in these conditions, structure **5.1.1** can bind tightly to RNA displacing recombinant MBNL1, reducing the FRET intensity by 33% compared to the negative control (Fig. 5.24). The rest of the candidates also

exhibit some MBNL1 displacing potency. These results encouraged us to study these candidates in patient-derived cells further.

Biologic Evaluation

5.5.2 Cell Biology

In order to determine the drug efficacy and gain understanding of the behavior of compounds **5.1** in a complex system, we performed the biological characterization of the candidates. Dermal fibroblasts of both healthy and DM1 affected individuals were used to run the experiments. Both cell lines were previously immortalized by infection with a retroviral vector carrying the human telomerase (hTERT) gene. Moreover, cells were also retrovirally infected with a vector producing an inducible myogenic differentiation 1 (MYOD1) transcription factor fused to the estrogen receptor (ER) hormone-binding domain (MYOD1-ER).³⁰ On the one hand, TERT ensures a homogeneous proliferation over cycles of duplication with no evidence of genetic affection related upon infection. On the other hand, MYOD1 allows, via estradiol addition to the culture media, the expression of myogenic factors leading to differentiation into myoblasts and finally forming mature myotubes. All the characteristic events of DM1 such as CUG nuclear foci, splicing factor sequestration and missplicing events were maintained in the DM1 patient-derived cells in culture and absent in control cells derived from healthy individuals.³¹ Original fibroblasts from DM1 patient had 290 repeats. Nonetheless, further expansion occurs in culture. Thus, and despite the cell population's heterogeneity, the most abundant length in PCR amplification is larger than 1000 CTG repeats. The biological activity of the compounds on these DM1 features was tested on these cell models. Being skeletal muscle the most affected tissue in DM1 patients, the use of myogenic cell models offers the advantage to test the effect of the designed drugs on both DM1 hallmarks and differentiation capacity of these cells, in the perspective of a human therapy. To evaluate **5.1** and **Z** in cell culture some considerations were made. Some of these candidates underwent crystallization in the culture media at concentrations over 50 μM in previous studies. For this reason, and to make all candidates comparable, their concentration was set to 10 μM . These test concentrations are lower than the ones used in previously reported studies.²⁸ Nevertheless, it is compatible with a possible in vivo use of the candidates.

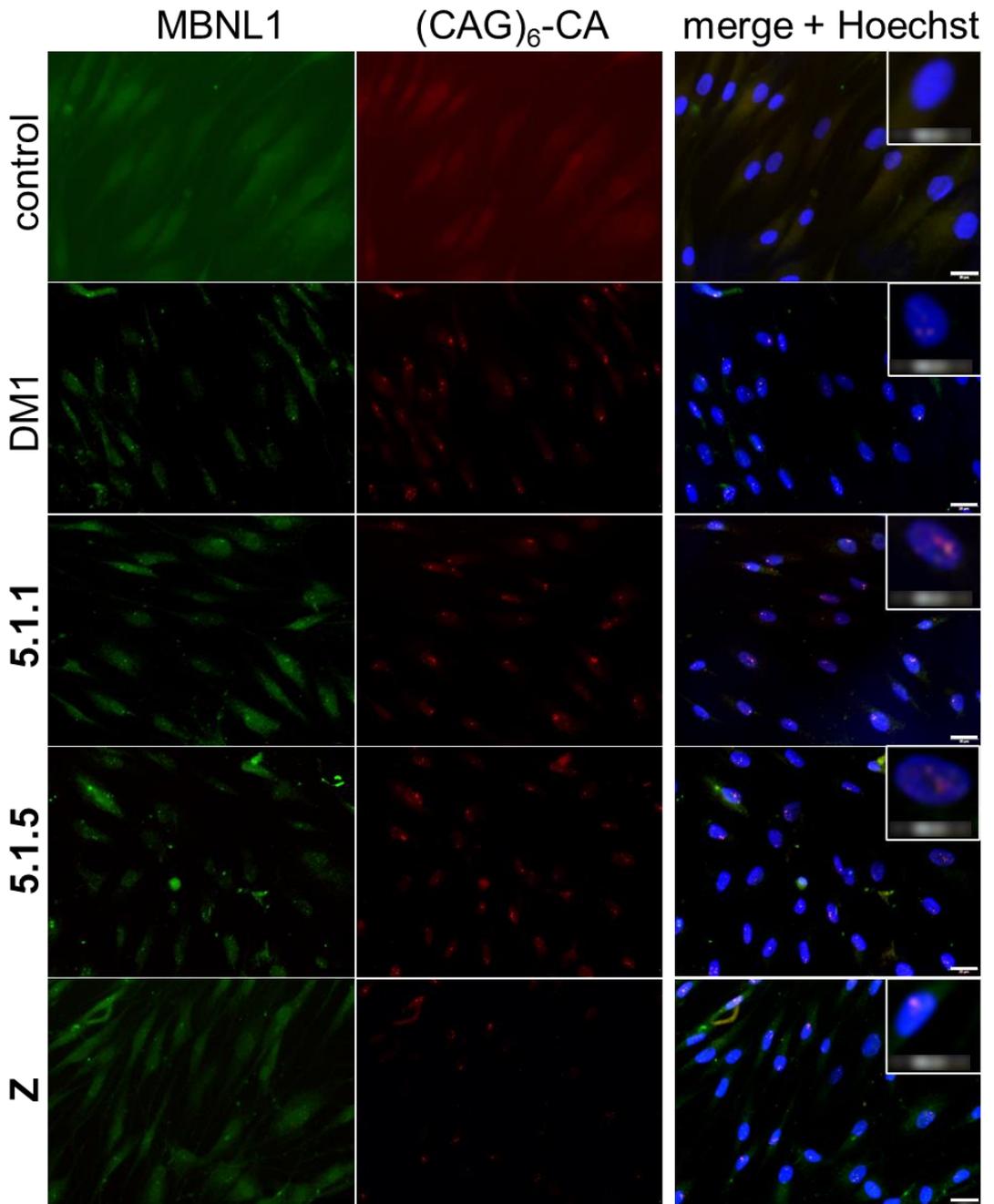


Fig. 5.25. FISH images using a (CAG)₆ probe labelled with Texas Red at the 5' end in combination with immunofluorescence staining. To verify the co-localization of MBNL1 in ribonuclear inclusions, following the last post-hybridization wash, cells were stained sequentially with antibodies to MBNL1 and with goat anti-mouse antibody conjugated with Alexa Fluor 488. Nuclei were visualized with Hoechst 33258 dye. Insets show zoomed-in images (scale bars 20 μ m).

In order to directly see whether MBNL1 was sequestered in CUG foci, RNA Fluorescent In Situ Hybridization (RNA-FISH) analysis combined to immunofluorescent staining (IF) of MBNL1 protein with a specific antibody were performed as described previously (Fig. 5.25,

Biologic Evaluation

analysis in Fig. 5.26).³² Ligands **5.1** were designed to specifically bind to CUG motifs and obtained FISH results are consistent with this prediction. Nevertheless, a reduction in the total number of nuclear foci was not expected as that means transcription inhibition. The staining with the (CAG)₆ RNA-FISH probe (red) showed no significant difference between untreated and treated cells in the foci count per cell ratio for **5.1** compounds. This is consistent with the prediction since the specific binding of these ligands to CUG motifs is not expected to affect the expression level of the mutant transcripts. However, some compounds were able to reduce the amount of colocalized MBNL1 in those nuclei. That was observed via IF using anti-MBNL1 mAb (green). Interestingly, **5.1.1** induces an important depletion of MBNL1 sequestered in nuclear foci of above 23%, which means that this compound can selectively attach to CUG and competitively displace MBNL1, eventually reducing the number of nuclear foci containing the protein in a significant manner (Fig. 5.26).

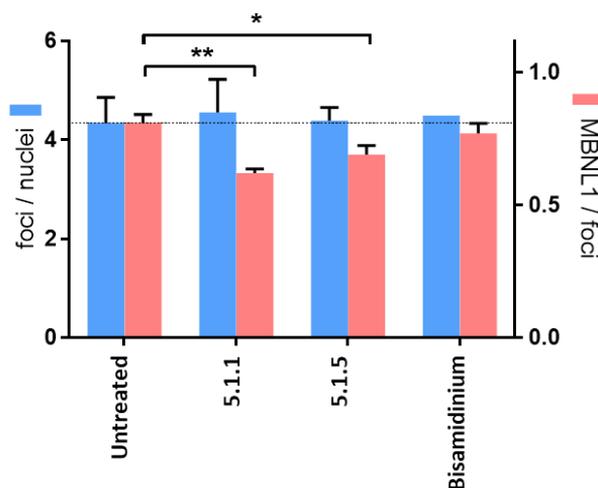


Fig. 5.26. Quantitative FISH analysis of total foci per cell and colocalized MBNL1 in foci. Results are presented as mean (\pm SD) of $n=3$ experiments. Statistical significance was assessed from a two-tailed Student's *t*-test (** $p<0.01$, * $p<0.05$).

Another important event in DM1 is missplicing deficiencies of a large number of gene transcripts. To study possible recoveries in the splicing pattern we chose two gene transcripts known to be affected in DM1: the sarcoplasmic/endoplasmic reticulum calcium ATPase 1 (SERCA1) and the insulin receptor (INSR). The inclusion or not of exon 22 of SERCA1 and exon 11 of INSR was monitored via semiquantitative RT-PCR to check whether missplicing events were modulated by test compounds present in the media. The results showed that at the studied conditions (see Supporting information of Ondono *et al.*)³³ the missplicing of these genes was not significantly affected by the addition to the culture media of either ligands **5.1** or Bisamidinium **Z** (Fig. 5.27).

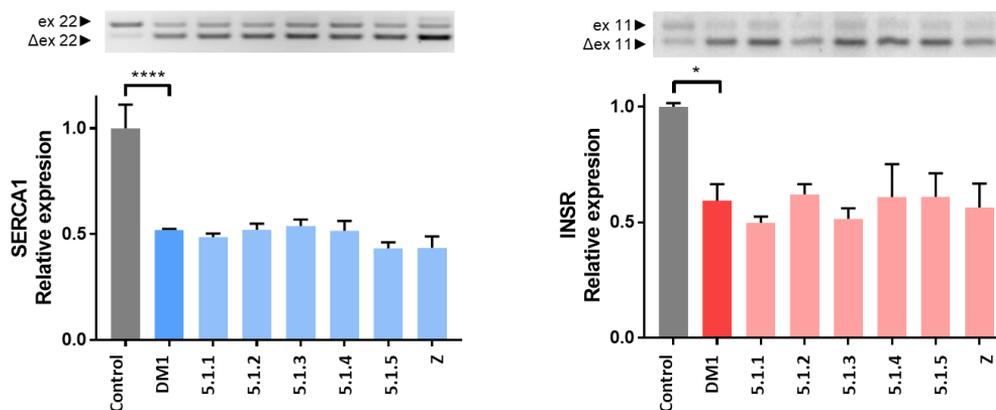


Fig. 5.27. PCR study of SERCA1 and INSR alternative splicing. The results are presented the relative intensity of the upper band divided to the sum of intensities of both bands and normalised to control cells. All results are presented as the mean \pm SD of three experiments. P values were assessed by two-tailed t test. Strong significance was observed between control and diseased cells. On the other hand, no significance was observed between treated cells and DM1 cells. (**** $p < 0.0001$, * $p < 0.05$).

To check if structures **5.1** affected the expression levels of DM1-related splicing factors or myogenic differentiation markers, Western blot analysis was performed. The effect of compounds was studied on 4 protein levels: MBNL1, CELF1, myosin and myogenin (Fig. S7). Ligands **5.1** were observed not to affect the total amount of MBNL1 or CELF1, which is in fact a positive result, as these proteins are essential and any anti-DM1 drug candidate aims to recover the natural balance between these antagonists. On the other hand, myosin and myogenin are crucial muscle differentiation proteins.^{34,35} A decrease in their total cell concentration may alter cell differentiation. Therefore, looking at the state of differentiation is an important indication of a possible toxic effect of a tested compound. Our findings with Western blot analysis show that only in the case of ligand **5.1.5** a visible decrease of Myogenin levels is observed while no effect at all was observed for the remaining compounds.

Biologic Evaluation

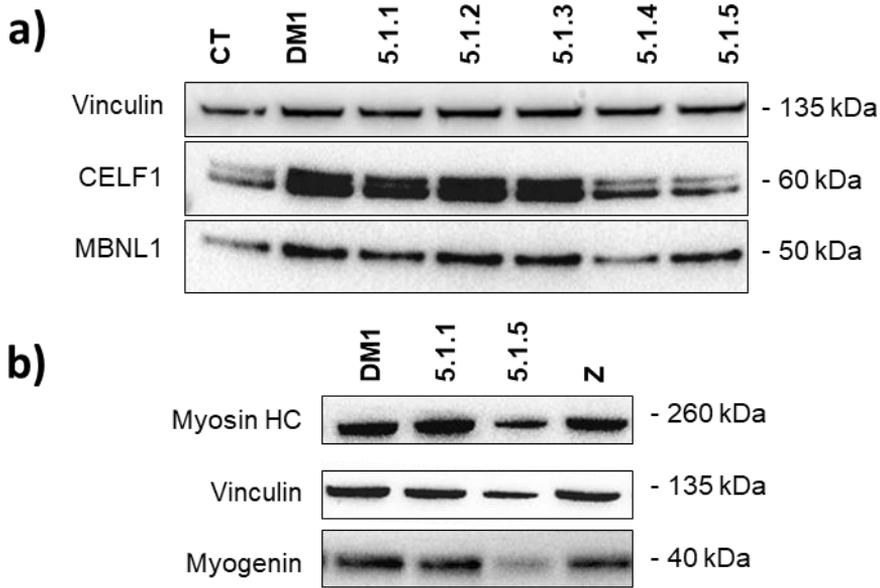


Fig. 5.28. Western blot analysis of: a) CUGBP Elav-like family member 1 (CELF1), muscleblind like protein 1 (MBNL1) and constitutively expressed vinculin (Vinculin) in control and DM1 cells. B) muscle specific factors myogenin (Myogenin), myosin (Myosin HC) and constitutively expressed vinculin (Vinculin).

5.6 Chapter Discussion

Base recognition has been one of the main trends in small molecule drug development in the DM1 field. This is due to the implementation of a comprehensive structure-activity relationship.^{18,36} The triangular shape of pyrido[2,3-*d*]pyrimidin-7(6*H*)-ones, combined with the spatial disposition of hydrogen bond donors and acceptors in such core, resulted in a potential U•U recognizer interaction.

In recent years, computer-aided rational design studies have been described in the literature to study ligand-RNA interaction.^{18,37,38} It is worth noting that studies describing a rational approach leaned on ligand-based drug design strategies are commonly applied to commercially available chemical libraries.³⁹ In this chapter, we develop tools to design novel compounds with structure-based methods Janus-Wedge recognition as a strategy to stabilize CUG in its non-toxic conformation. A set of 5 novel chemical entities based on two pyrido[2,3-*d*]pyrimidin-7(6*H*)-one scaffold separated by an aliphatic linker of variable length was studied (**5.1**). According to their predicted Gibbs free energy of binding, these candidates would exhibit different interaction mechanisms with CUG repeats depending on their length. Simulations point to the 5-carbon length linker for optimal recognition of consecutive U•U mismatches in the 10-carbon length linker and for alternate mismatches the 10-carbon linker. Moreover, the analysis of base-pair opening showed a reduction in data variance for **5.1** stating that the candidate can stabilize the CUG environment and maintain the JW interaction during the whole simulation.

Interestingly, this observation was validated in biological studies that proved the ability of **5.1** for CUG binding and triggering MBNL1 recovery *in vitro* (AID2675 test). It is worth of attention that **5.1** also exhibited these features in DM1 patient-derived myoblasts, reducing the amount of observed MBNL1 in FISH-IF microscopy. These findings provide valuable data about the length needed to reach two consecutive U•U mismatches for the pyrido[2,3-*d*]pyrimidines scaffold suggested in this study. Moreover, results confirm the usefulness of molecular modelling for the rational design of JW base recognizers with potential anti-DM1 activity.

Furthermore, a new structure based on the nucleic acid folding agent spermine has been tested.^{12,13} This product has been previously used in the design of RNA-targeting compounds. We have been able to synthesize a highly soluble product in most solvents and derived from spermine, but not being able to guarantee a sufficient purity level, we discarded it for biological evaluation.

These results allow a better understanding of the biological interaction of pyrido[2,3-*d*]pyrimidine structures and set a starting point in future drug discovery studies and they were published in *Comput Struct Biotechnol J.* 2021; 19: 51–61.³³

References

5.7 References

1. Woodgate, R. *et al.* Crystal structure of a benzo [a] pyrene diol epoxide adduct in a ternary complex with a DNA polymerase. *Proc. Natl. Acad. Sci.* **101**, 2265–2269 (2004).
2. Tsuboi, M., Benevides, J. M. & Jr, G. J. T. The Complex of Ethidium Bromide with Genomic DNA : Structure Analysis by Polarized Raman Spectroscopy Preparation of EtBr single crystals. *Biophys. J.* **92**, 928–934 (2007).
3. Chen, H., Liu, X. & Patel, D. J. DNA Bending and Unwinding Associated with Actinomycin D Antibiotics Bound to Partially Overlapping Sites on DNA. *J. Mol. Biol.* **258**, 457–479 (1996).
4. Staker, B. L. *et al.* Structures of Three Classes of Anticancer Agents Bound to the Human Topoisomerase I–DNA Covalent Complex. *J. Med. Chem.* **48**, 2336–2345 (2005).
5. Schwartz, J. L. *et al.* Solution Structure of the Minor Conformer of a DNA Duplex Containing a dG Mismatch Opposite a Benzo[a]pyrene Diol Epoxide/dA Adduct: Glycosidic Rotation from Syn to Anti at the Modified Deoxyadenosine. *Biochemistry* **36**, 11069–11076 (1997).
6. Wei, H. *et al.* DNA Structural Integrity and Base Composition Affect Ultraviolet Light-Induced Oxidative DNA Damage. *Biochemistry* **37**, 6485–6490 (1998).
7. Branda, N., Kurz, G. & Lehn, J. JANUS WEDGES: a new approach towards nucleobase-pair recognition. *Chem. Commun.* 2443 (1996) doi:10.1039/cc9960002443.
8. Thadke, S. A. *et al.* Shape selective bifacial recognition of double helical DNA. *Commun. Chem.* **1**, 1–10 (2019).
9. Hendry, L. B., Mahesh, V. B., Bransome, E. D. & Ewing, D. E. Small molecule intercalation with double stranded DNA: Implications for normal gene regulation and for predicting the biological efficacy and genotoxicity of drugs and other chemicals. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **623**, 53–71 (2007).
10. Chien, C. *et al.* Structural basis for targeting T : T mismatch with triaminotriazine-acridine conjugate induces a U-shaped head-to-head four-way junction in CTG repeat DNA Structural basis for targeting T : T mismatch with triaminotriazine-acridine conjugate induces a U-s. (2020) doi:10.1021/jacs.0c03591.
11. Arambula, J. F., Ramisetty, S. R., Baranger, A. M. & Zimmerman, S. C. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc. Natl. Acad. Sci.* **106**, 16068–16073 (2009).

12. Jahromi, A. H., Honda, M., Zimmerman, S. C. & Spies, M. Single-molecule study of the CUG repeat-MBNL1 interaction and its inhibition by small molecules. *Nucleic Acids Res.* **41**, 6687–6697 (2013).
13. Jahromi, A. H. *et al.* A novel CUGexp-MBNL1 inhibitor with therapeutic potential for myotonic dystrophy type 1. *ACS Chem. Biol.* **8**, 1037–1043 (2013).
14. Wong, C. H. *et al.* Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. *J. Am. Chem. Soc.* **136**, 6355–6361 (2014).
15. Lee, J. Y. *et al.* Intrinsically cell-penetrating multivalent and multitargeting ligands for myotonic dystrophy type 1. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8709–8714 (2019).
16. Li, J. *et al.* A Ligand That Targets CUG Trinucleotide Repeats. *Chem. - A Eur. J.* **22**, 14761 (2016).
17. Li, J. *et al.* A Ligand That Targets CUG Trinucleotide Repeats. *Chemistry - A European Journal* vol. 22 14761 (2016).
18. Li, J. *et al.* A Dimeric 2,9-Diamino-1,10-phenanthroline Derivative Improves Alternative Splicing in Myotonic Dystrophy Type 1 Cell and Mouse Models. *Chem. - A Eur. J.* **24**, 18115–18122 (2018).
19. Matsumoto, J. *et al.* The Dimeric Form of 1,3-Diaminoisoquinoline Derivative Rescued the Mis-splicing of Atp2a1 and Clcn1 Genes in Myotonic Dystrophy Type 1 Mouse Model. *Chem. – A Eur. J.* **26**, 14305–14309 (2020).
20. Nakamori, M. *et al.* A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.* **52**, 146–159 (2020).
21. Aikawa, H., Yano, A. & Nakatani, K. A 2,7-diamino-1,4,8-triazanaphthalene derivative selectively binds to cytosine bulge DNA only at a weakly acidic pH. *Org. Biomol. Chem.* **15**, 1313–1316 (2017).
22. González, À. L., Teixidó, J., Borrell, J. I. & Estrada-Tejedor, R. On the applicability of elastic network models for the study of RNA CUG trinucleotide repeat overexpansion. *PLoS One* **11**, 1–20 (2016).
23. Yuan, Y. *et al.* Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **35**, 5474–5486 (2007).
24. Camarasa, M. *et al.* Design, synthesis and biological evaluation of pyrido[2,3-d]pyrimidin-7-(8H)-ones as HCV inhibitors. *Eur. J. Med. Chem.* **115**, 463–483 (2016).
25. Recasens-Zorzo, C. *et al.* Pharmacological modulation of CXCR4 cooperates with BET bromodomain inhibition in diffuse large B-cell lymphoma. *Haematologica* **104**, 778–788 (2019).

References

26. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
27. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
28. Wong, C. H. *et al.* Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. *J. Am. Chem. Soc.* **136**, 6355–6361 (2014).
29. Lu, X. J. & Olson, W. K. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).
30. Hollenberg, S. M., Cheng, P. E. I. F. & Weintraub, H. Trans-Activation and Muscle Determination. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8028–8032 (1993).
31. Provenzano, C. *et al.* CRISPR/Cas9-Mediated Deletion of CTG Expansions Recovers Normal Phenotype in Myogenic Cells Derived from Myotonic Dystrophy 1 Patients. *Mol. Ther. - Nucleic Acids* **9**, 337–348 (2017).
32. Cardani, R., Mancinelli, E., Sansone, V., Rotondo, G. & Meola, G. Biomolecular identification of (CCTG)_n mutation in myotonic dystrophy type 2 (DM2) by FISH on muscle biopsy. *Eur. J. Histochem.* **48**, 437–442 (2004).
33. Ondono, R. *et al.* Design of novel small molecule base-pair recognizers of toxic CUG RNA transcripts characteristics of DM1. *Comput. Struct. Biotechnol. J.* **19**, 51–61 (2021).
34. Sharma, A. *et al.* Myosin Heavy Chain-embryonic is a crucial regulator of skeletal muscle development and differentiation. *bioRxiv* 261685 (2018) doi:10.1101/261685.
35. Flynn, J. M., Meadows, E., Fiorotto, M. & Klein, W. H. Myogenin regulates exercise capacity and skeletal muscle metabolism in the adult mouse. *PLoS One* **5**, (2010).
36. Lee, J. Y. *et al.* Intrinsically cell-penetrating multivalent and multitargeting ligands for myotonic dystrophy type 1. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8709–8714 (2019).
37. Velagapudi, S. P., Gallo, S. M. & Disney, M. D. Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat. Chem. Biol.* **10**, 291–297 (2014).
38. Wong, C. H. *et al.* Investigating the Binding Mode of an Inhibitor of the MBNL1{dot operator}RNA Complex in Myotonic Dystrophy Type 1 (DM1) Leads to the Unexpected Discovery of a DNA-Selective Binder. *ChemBioChem* **13**, 2505–2509 (2012).

39. González, À. L. *et al.* In silico discovery of substituted pyrido[2,3-d]pyrimidines and pentamidine-like compounds with biological activity in myotonic dystrophy models. *PLoS One* **12**, e0178931 (2017).

References

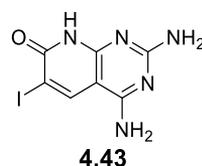
6 Conclusions

Conclusions

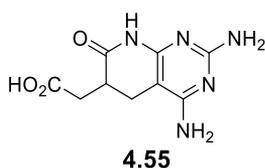
- The plasmid pGEX-6P-MBNL1- Δ 105-H has been designed and constructed from scratch on the basis of the poor information available in the literature to contain a recombinant form of MBNL1 that has successfully been transformed in two bacterial cell lines: BL21(DE3)pLysS and DH5 α . The tuning of the production and isolation conditions allowed us to obtain GST-MBNL1- Δ 105-H.

- The tuning of the AID2675 TR-FRET fluorescent test has been carried out. It allows rapid screening of candidates, increasing the efficiency of the overall drug discovery process.

- A family of groove-binding candidates has been designed and developed in analogy to pentamidine. Ligands were designed using SBDD techniques and synthesized using the click chemistry approach.

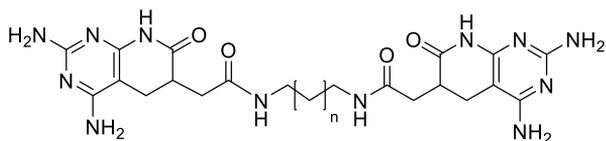


- New synthetic methodologies have been developed for the introduction of halo and carboxymethyl substituents at position C6 of pyrido[2,3-*d*]pyrimidin-7(8*H*)-ones. In particular, the 6-bromo, 6-iodo, and 6-carboxymethyl compounds (**4.42**, **4.43**, and **4.55**) have proved to be potential sources of a novel family of anti-DM1 candidates



- Structure **4.55** has been selected as a potential binding scaffold to build a family of candidates based on amide bond formation.

- Consequently, a new family of base recognizer small molecules has been developed using **4.55** as starting material and the Janus-Wedge strategy. Such uracil mismatch binders have been designed using MD-based SBDD, and their structural parameters show the stabilization of the complex between CUG and compounds **5.1**.



5.1

- The potency of family **5.1** has been tested using the AID2675 TR-FRET fluorescent test and patient-derived human myoblasts. FISH-IF results show a significant reduction in the amount of MBNL1 colocalized inside the foci when the cells are treated with candidate **5.1.1** (n=3).

Conclusions

7. Experimental

This chapter describes all the experimental procedures carried out during this thesis

7.1 Biochemistry

7.1.1 Protein purification.

Competent BL21(DE3)pLysS *E. coli* (Invitrogen, Thermo Scientific, Carlsbad, CA) were transformed with the pGEX-6P-MBNL1- Δ 105-His plasmid (GE Healthcare & Invitrogen).¹ A single colony of the transfected BL21 was picked into 10 mL of LB with 100 μ g/mL ampicillin and grown for 16 hours at 37 °C and 200 rpm. 1 mL of the resulting culture mixture was inoculated to 1 L of LB with 100 μ g/mL ampicillin and incubated at 37 °C and 150 rpm in a shaking incubator until OD_{600nm} reached 0.5-0.6. The cell culture was then induced with 500 μ M IPTG for 3 hours at 30 °C. Thereafter, the culture media was centrifuged at 4000 xg and 4 °C for 30 minutes*. The supernatant was discarded and the pellet was resuspended in 50 mL of cell lysis buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.00), 5 mM imidazole, 0.1 % IGEPAL CA-630, 5 % glycerol and freshly added: 0.1 M PMSF, 5 mM benzamidine, 1 EDTA-free complete protease inhibitors tablet, 1 mM DNase and 1 mg/mL lysozyme). The cell culture was lysed by alternating 15 cycles of sonication (1 min sonication – 1 min icing). To remove cell debris, the sample was centrifuged at 10000 xg and 4 °C for 50 min and the supernatant was collected for purification. The supernatant was added to 2 mL of a previously equilibrated TALON® Metal Affinity Resin (Takara Bio, Kusatsu, JP) and incubated at 4 °C for 1 hour. The resin was washed three times with 10 column volumes (30 mL x 3 times) of washing buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.00), 25 mM imidazole, 0.1% IGEPAL CA-630, 5% glycerol). The protein was eluted with 5mL of elution buffer (150 mM NaCl, 50 mM Tris-HCl (pH 8.00), 250 mM imidazole, 0.1% IGEPAL CA-630, 5% glycerol). The eluates were added to PD-10 Desalting Columns (GE Healthcare, Chicago, IL) that was previously preequilibrated using fluorescence test buffer (25 mM HEPES pH 7.4, 110 mM KCL, 10 mM NaCl, 1 mM MgCl₂, 15 μ M ZnCl₂, 0.02%Tween-20). Protein purity was assessed via Comassie staining of SDS-PAGE electrophoresis 10% polyacrylamide gel (Figure S4). Finally, GST-MBNL1-His₆ was quantified via direct absorbance spectroscopy at 280 nm using the ExpASy Protparam (protein parameter) tool to assess extinction coefficient (57300 M⁻¹.cm⁻¹).² GST-MBNL1-His₆ samples were stored at -80 °C with 50% glycerol. Prior to each use, samples were

Rational design

ultracentrifuged at 150000 xg, 4 °C for 1 h and quantified to avoid aggregation issues.

7.1.2 AID 2675.

The previously described assay was adapted to RNase pretreated 96-well plates (Greiner, Kremsmünster, AT). Reagents were premixed and added to the wells containing compounds solution (20 nM Biot-(CUG)₆, 20 nM MBNL1-His₆, 0,11 ng/mL anti-His-Tb, 10 µM SA-XL665 and 0,1 µM tested compound). All reagents were diluted in binding buffer composed of 25 mM HEPES pH 7.4, 110 mM KCl, 10 mM NaCl, 1 mM MgCl₂, 15 µM ZnCl₂, 0.02% Tween-20, 0.1% BSA and 5 mM freshly added DTT. After 60 min incubation at room temperature, the fluorescence was measured using a BMG Labtech Fluostar Optima microplate reader. Samples were irradiated at 340 nm, and fluorescence was read at 545 and FRET at 665 nm with a time gap of 30 µs and an integration time of 1500 µs.

7.2 Rational design

7.2.1 Base recognizers

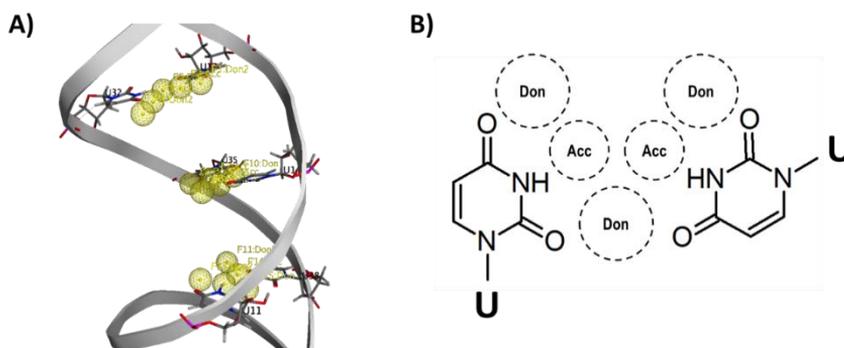
7.2.1.1 Receptor preparation.

The receptor used in the current study is a CUG₁₆ RNA model previously developed.³ The flexible behavior of the system was assessed by molecular dynamics simulations. The system was initially heated to 300 K within 500 ps restraining the position of the RNA with a 2.0 kcal/mol·Å² force constant. These restraints were gradually reduced along the system's equilibration stage, performed at constant temperature and pressure (1.0 bar, 300 K correspondingly). A final NPT production stage was conducted during 100 ns without restraints. The Particle Mesh Ewald method was used for electrostatics interactions under periodic boundary conditions and SHAKE⁴ algorithm was applied for hydrogen atoms. The time step was fixed to 2 fs and simulations were conducted using AMBER18 software (University of California, San Francisco, CA). Upon that, a set of frames exhibiting opening values similar to

the ones expected from alternate U•U mismatches were selected using the cpptraj⁵ module.

7.2.1.2 Pharmacophore model and docking.

A pharmacophore model was developed using Molecular Operating Environment (MOE) software (Chemical Computing Group, Montreal, QC), to evaluate the ability of the ligands to establish a Janus-Wedge interaction pattern with two U•U mismatches. The generated model involved three consecutive U•U mismatches, each of them led to 5 pharmacophore features (3 H-bond donors and 2 H-bond acceptors). Thereafter, pharmacophore guided dockings were performed using MOE under induced fit conditions. We considered two different score functions (London dG and GBVI/WSA dG) for the docking study. The best conformation for each ligand was selected for the subsequent calculations.



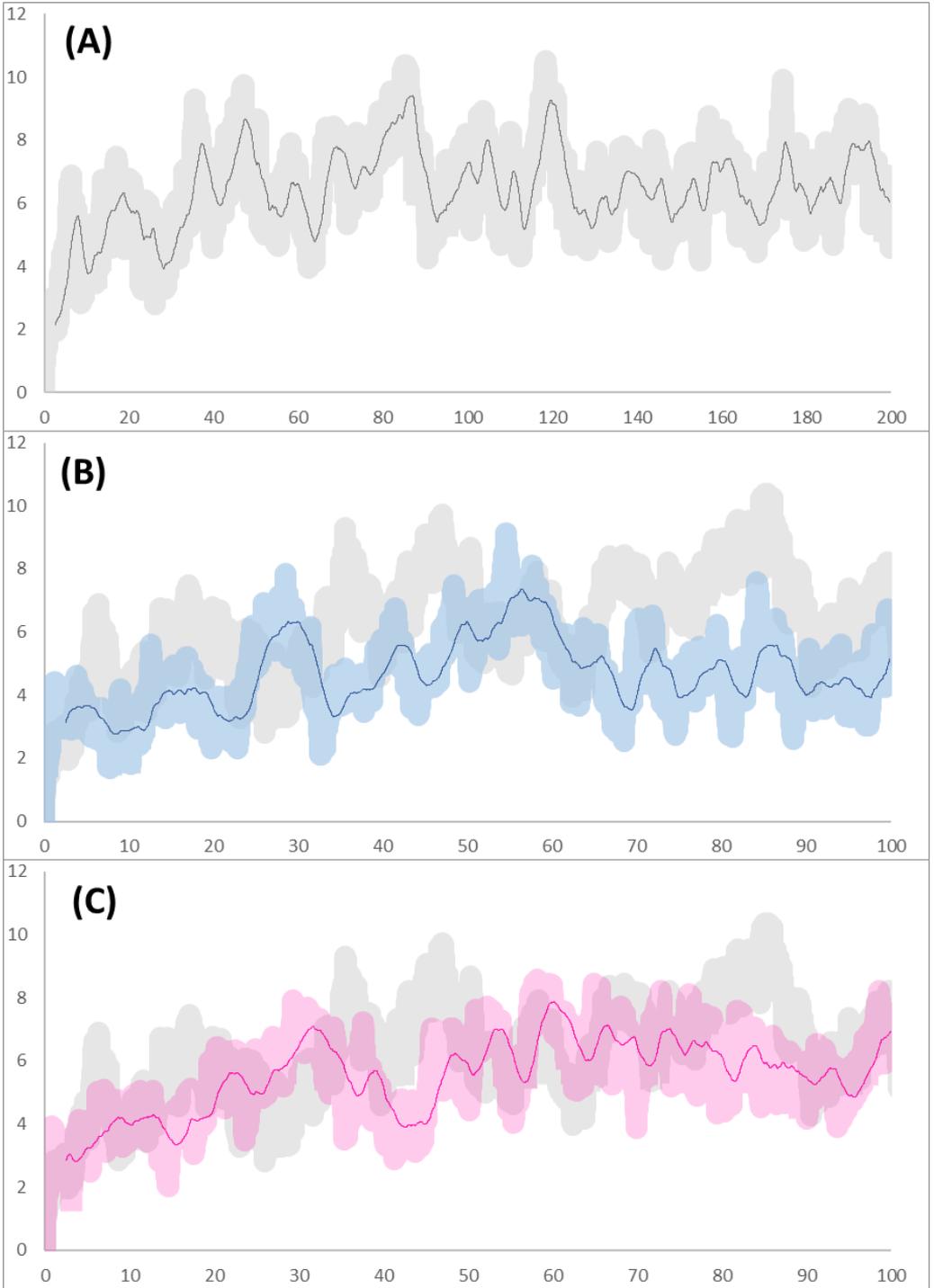
7.2.1.3 Molecular dynamics and data analysis

Ligands were prepared in UCSF-Chimera⁶, calculating AM1-BCC atomic charges and using the GAFF⁷ forcefield. Complex structures were generated using tleap, defining the Amber14 and Rochester torsions forcefields⁸ for the RNA. Complexes were neutralized and solvated using the OPC water model in a truncated octahedral box.⁹ All molecular dynamics simulations were conducted using pmemd.cuda module in Amber18.¹⁰⁻¹² and run through the following steps. The system was initially heated to 300 K within 500 ps restraining the position of the RNA with a 2.0 kcal/mol·Å² force constant. These restraints were gradually reduced along the system's equilibration stage, performed at constant temperature and

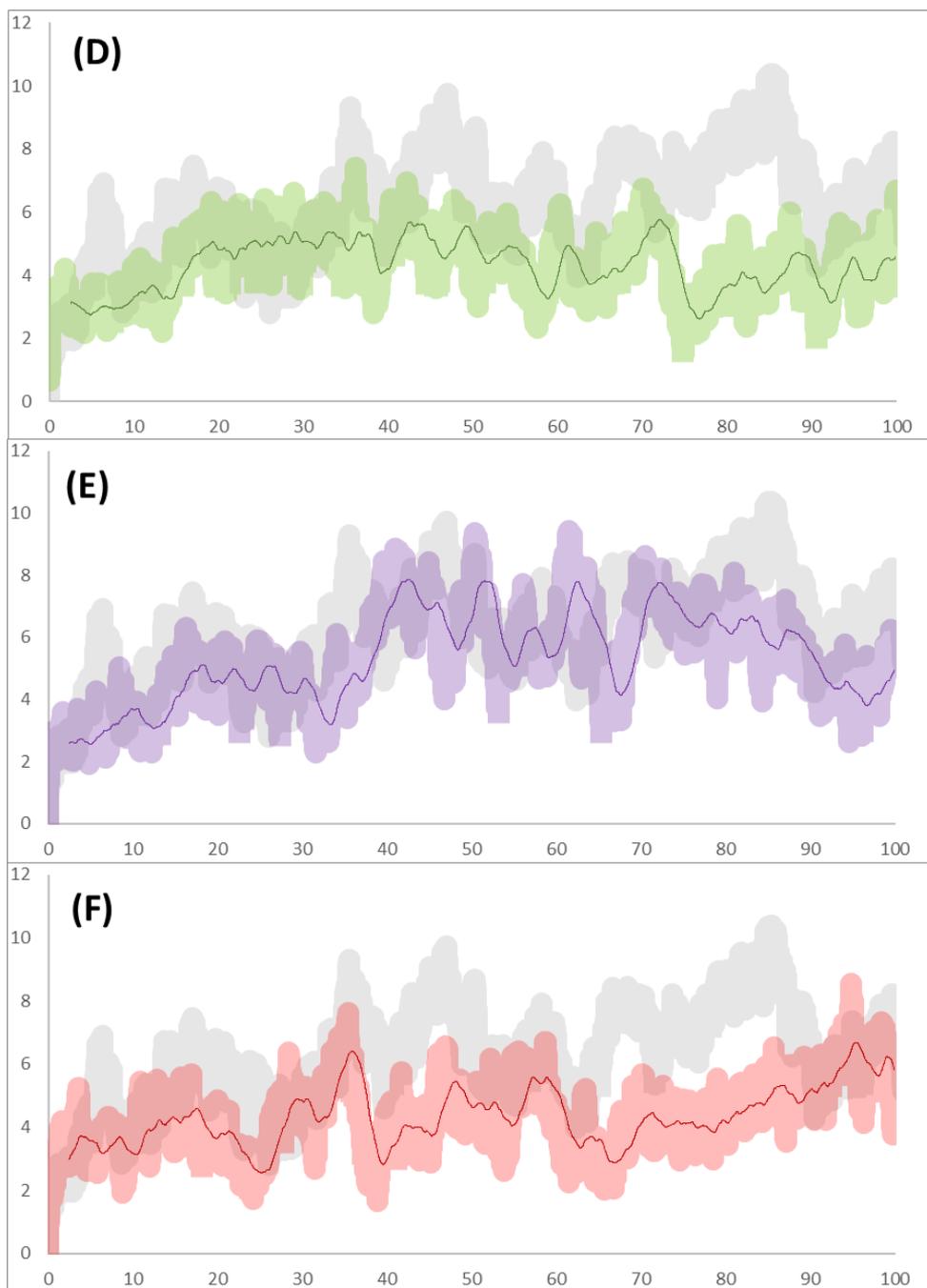
Rational design

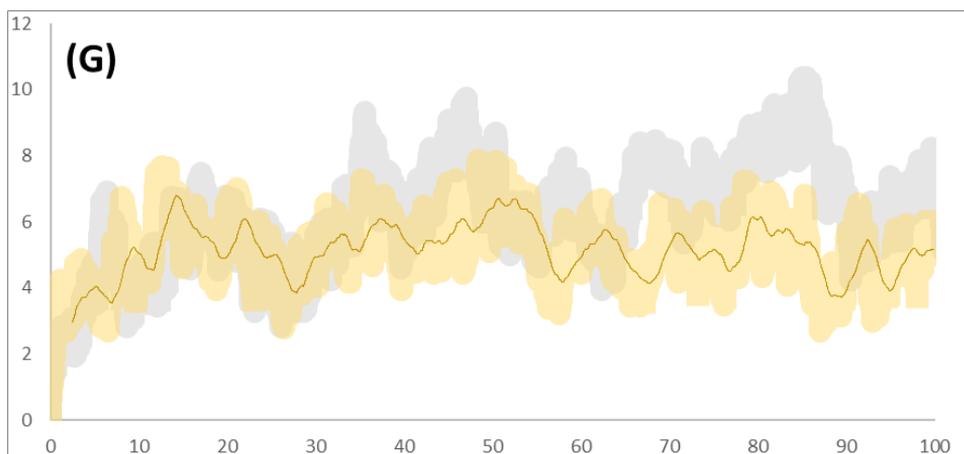
pressure (1.0 bar, 300 K correspondingly). A final NPT production stage was conducted during 100 ns without restraints. The Particle Mesh Ewald method was used for electrostatics interactions under periodic boundary conditions and SHAKE⁴ algorithm was applied for hydrogen atoms. The time step was set to 2 fs. RMSD analysis for the full-length trajectories was conducted using ccptraj module and MMPBSA binding free energies were calculated for the last 20 ns of simulation. RNA structural parameters analysis was performed using the X3DNA software.¹³ Data analysis and plotting were performed using R (R Foundation for Statistical Computing, Vienna, AT).

Experimental



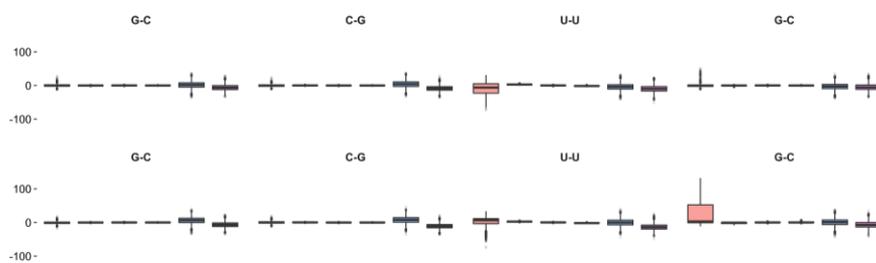
Rational design



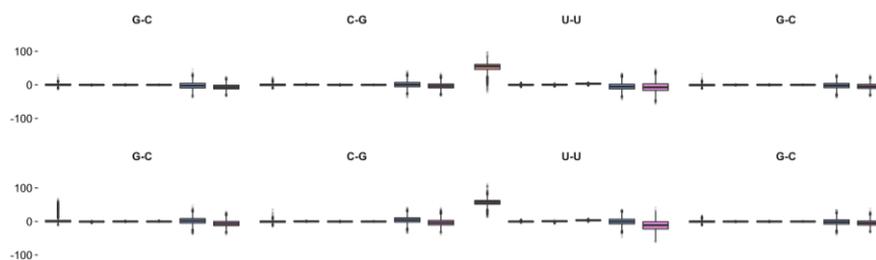


RMSD profile of the RNA structure along MD calculations performed considering the RNA structure only (A) and in complex with the compound of reference Z (B) and 1a-1e candidates (C-G). RMSD is represented in Armstrongs (Å) and the simulation time in ns.

Blank:

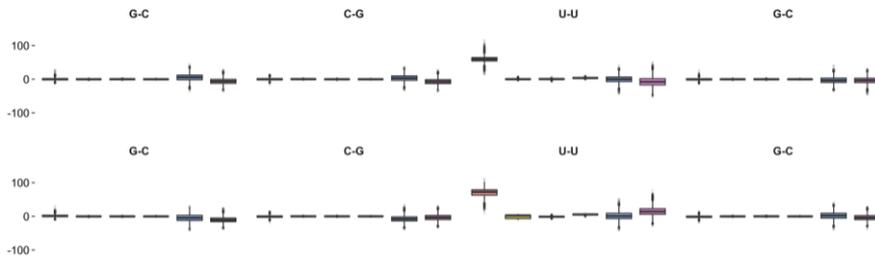


Z:

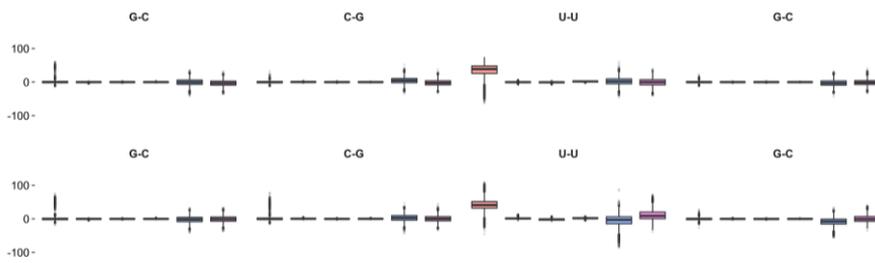


Rational design

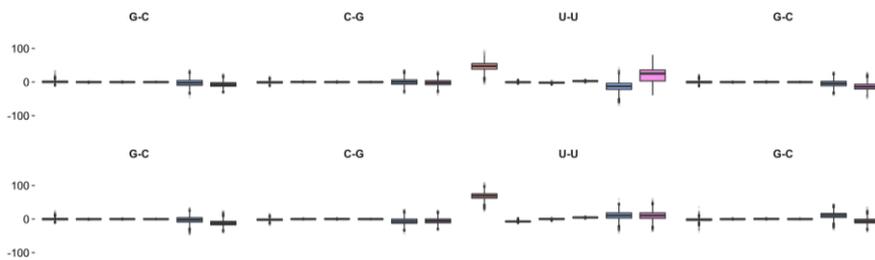
5.1.1



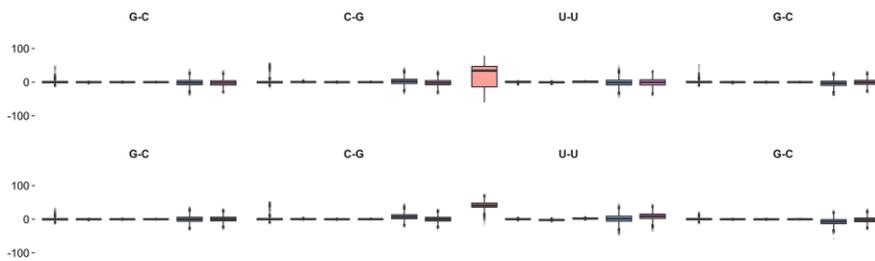
5.1.2



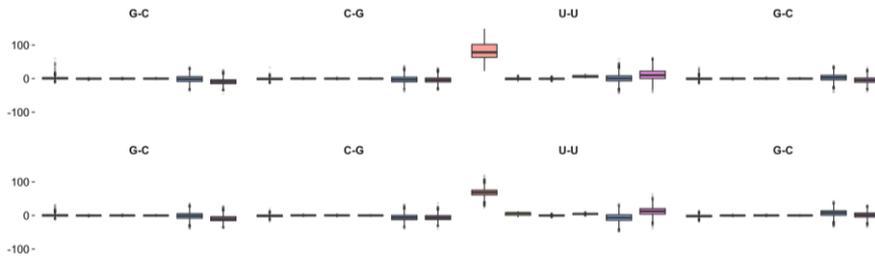
5.1.3



5.1.4



5.1.5



analysis of single base-pair structural parameters (Opening, Shear, Stagger, Stretch, Buckle, Propeller). Full-length MD is displayed in boxplot graph format.

7.3 Cell culture

7.3.1 **Groove Binders activity determination**

7.3.1.1 Polarization fluorescence assays

Carboxyfluorescein-labeled CUG RNA (23 CUG repeats; 6-FAM-CUG23) at 6 nM was incubated with compounds at different concentrations in binding buffer (50 mM Tris-HCl pH 7.0, 250 mM NaCl, 50 μ M ZnCl₂, 10% glycerol, 1 mM DTT) on ice for 20 min in the dark. Polarization was measured in a EnVision® Multilabel Reader using as excitation filter FP480 and as emission filter FP535. The readings for each concentration or all compounds were quadruplicated.

7.3.1.2 Cell culture conditions

Cell model of the disease (provided by the D. Furling's laboratory, Institute of Myologie, Paris) consisted of normal and DM1 (1300 CTG repeats) immortalized (hTERT) skin fibroblasts expressing conditional MyoD. Fibroblast cells were grown in Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g/L of glucose, 1 % of penicillin and streptomycin (P/S) and 10% foetal bovine serum (FBS) (Sigma). Fibroblasts were transdifferentiated in myoblasts by inducing expression of MyoD. Cells were plated in muscle differentiation medium (MDM) made of DMEM 4.5 g/L glucose with 1% P/S, 2% horse serum, 1% apo-transferrin (10 mg/ml), 0,1 % insulin (10 mg/ml) and 0,02 % doxycyclin (10 mg/ml) for 48 h. For compound testing fibroblasts were aliquoted in 24-well plate with 3,5 x 10⁴ cells per well and were differentiated as before. Caffeine (Sigma-Aldrich) was added to a final concentration of 125 μ M and 250 μ M in MDM medium and cells were incubated for 48 h.

7.3.1.3 Immunodetection of MBNL1

Cells were fixed in 4% paraformaldehyde (PFA) for 10 min at room temperature followed by several washes in PBS 1x. Cells were then permeabilized with 0.3 % Triton in PBS (PBT), blocked (PBT 1% donkey serum) for 30 min at room temperature, and incubated with primary antibody (mouse anti-MBNL1 1:500; Sigma) at 4°C overnight. After several washes with PBT cells were incubated for

45 min with biotin-conjugated secondary antibody (Sigma) at a 1:200 dilution. Cells were then incubated with ABC solution (ABC kit, VECTASTAIN) for 30 min at room temperature, followed by PBT washes and incubation with streptavidin-FITC (1:1000) for 20 min. Samples were finally mounted in Vectashield (Vector) with 2 µg/ml DAPI.

7.3.1.4 Foci detection

In situ hybridization with CUG repeat RNA. Cells were fixed in 4% paraformaldehyde (PFA) for 10 min at room temperature followed by several washes in PBS 1x. Fixed cells were incubated in pre-hybridization buffer (SSC 2x, 30% deionized formamide) for 10 min at room temperature, hybridized with Cy3-(CAG)₇-Cy3 labelled probe diluted 1:100 in hybridization buffer (40% formamide, 2x SSC, 0.2% BSA, 10% dextran sulfate, 2 mM vanadyl complex, 10% tRNA (10 mg/ml), 10% herring sperm) for 2 h at 37°C, washed twice in pre-hybridization buffer for 15 min at 45°C, washed in PBS 1x for 15 min at room temperature and mounted in Vectashield (Vector) with 2 µg/ml DAPI. Images were taken using a Leica DM2500 fluorescence microscope and foci were manually counted from at least 50 cells per compound.

7.3.1.5 Toxicity studies

Fibroblast cells were grown in DMEM with 4.5g/L of glucose, 1% of penicillin and streptomycin (P/S) and 10% foetal bovine serum (FBS) (Sigma). Cells were aliquoted in 96-well plate with 1.0×10^4 cells per well and incubated for 24 h. The compound was added to a final concentration of 75 µM, 125 µM, 250 µM and 500 µM in DMEM and cells were incubated for 24 h. To measure cell viability, MTS tetrazolium salt was added to each well and was incubated for 4 h at 37°C in a humidified chamber with 5% CO₂. The conversion of MTS into soluble formazan (accomplished by dehydrogenase enzymes from metabolically active cells) was measured by absorbance at 490 nm (CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay, Promega). Data were transformed to percentage of survival relative to cells not exposed to caffeine, which established 100% viability.

7.3.2 Base recognizers activity determination

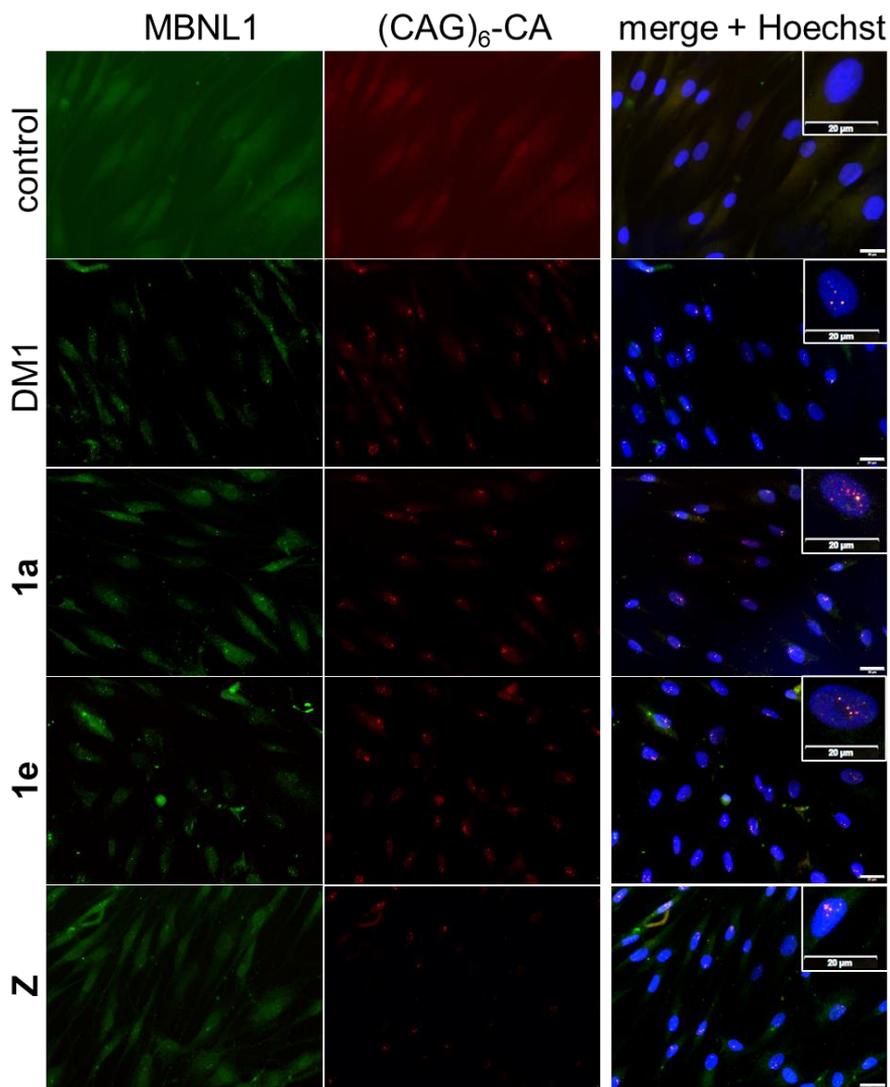
7.3.2.1 Cell culture.

Immortalized human myoblasts control and DM1 derived from primary dermal fibroblasts of control and DM1 affected individuals respectively were used as described in previous work.¹⁴ Fibroblasts were sequentially infected with retroviral vectors carrying TERT and Hygromycin selection (Addgene #1773, Cambridge, MA)¹⁵, and with retroviruses carrying estrogen-inducible mouse Myod1 and puromycin selection (Addgene #13494).¹⁶ Cells were propagated in dishes coated with collagen from rat tail at 5 µg/cm² (Gibco, Thermo Fisher Scientific, Waltham, MA), using DMEM without phenol red (Gibco, Thermo Fisher Scientific, Waltham, MA) supplemented with 15% FBS (Sigma-Aldrich, St. Louis, MO) and incubated for 48 h. Differentiation to myotubes was induced by growing cells to confluency and replacing the proliferation medium with differentiation medium consisting of DMEM without phenol red supplemented with 10⁻⁷ M β-estradiol (Sigma-Aldrich) and 2% horse serum and incubating for 24 h. All cells were incubated under a 5% CO₂ atmosphere at 37°C. For PCR and WB experiments 35 mm dishes were seeded at 1.3·10³ cells/cm² whereas for FISH and IF cells were seeded on glass coverslips at 3.2·10³ cells/cm². DM1 cells were treated with the compounds at 10 µM for 48 h in all experiments.

7.3.2.2 FISH IF.

Treated cells were washed twice in PBS for 5 minutes. After that, cells were fixed with 2% formaldehyde for 30 min at 4 °C. After fixation, coverslips were washed 3 times in PBS, and permeabilized in 0.4% Triton™ X-100 (Sigma-Aldrich) in PBS for 10 min. After washing with PBS 3 times, coverslips were incubated in 40% formamide and 2x SSC for 10 min at room temperature and then hybridized with (CAG)₆ probe labeled with Texas Red at the 5' end (IDT, Coralville, IA) (1 ng/µL for 2 h at 37 °C in 30% formamide, 2x SSC, 0.02% BSA, 67 ng/µL yeast tRNA, 2 mM vanadyl ribonucleoside complex). Cells were then washed first in 40% formamide and 2x SSC at 45 °C for 30 min. After that, coverslips were washed sequentially in 2x SSC and 1x SSC at 45 °C for 15 min each. Then, samples were incubated with anti-MBNL1 3A4 antibody

(Santa Cruz Biotechnology) solution in PBS at room temperature for 1h. After that samples were washed 3 times with PBS and incubated using goat anti-mouse IgG-Alexa Fluor 488 (Invitrogen, Thermo Scientific, Carlsbad, CA). After that, samples were incubated with Hoechst 33258 (Invitrogen) in PBS at room temperature for 10 min. Finally, coverslips were washed 3 times in PBS, water and mounted on glass slides with ProLong™ Gold Antifade Mountant (Invitrogen). Samples were examined with an Olympus AX70 immunofluorescence microscope. Images were recorded on an Olympus XM10 camera and processed using the Olympus CellSens Standard 1.8.1 software. Cell quantification was conducted using ImageJ software.



7.3.2.3 Western Blot.

Treated cells were washed twice with PBS for 5 minutes. After that, cells were ice cooled and lysed by adding RIPA buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 0.5% Sodium deoxycholate, 0.1% SDS, 1% NP40, 1 mM EDTA, 5 mM NaF, 5 mM Na₃VO₄ and Protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO). After 5 minutes of incubation at 0 °C, cells were scraped, and the lysate was transferred to an Eppendorf tube. The extract was centrifuged at 14000 xg, 4 °C for 10 minutes to remove cell debris. The supernatant

was transferred to a new Eppendorf tube and total protein concentration was assessed using the Micro BCA Protein Assay Kit (Thermo Scientific, Waltham, MA). Homogeneous protein samples mixed with Laemmli sample buffer and Novex Nupage sample reducing agent (Invitrogen, Thermo Scientific, Carlsbad, CA) were loaded and separated using Bolt 4-20% Tris-glycine SDS-PAGE (Invitrogen, Thermo Scientific, Carlsbad, CA). Gels were run in Novex Nupage MOPS buffer (Invitrogen, Thermo Scientific, Carlsbad, CA) setting voltage to 220 V for 40 min. After gel running, proteins were transferred to PVDF or nitrocellulose membranes using a stack soaked in transfer buffer (25 mM Tris, 190 mM glycine, 10% methanol) and setting current at 350 mA for 90 min at 4 °C. After checking correct transferring using ponceau staining, membranes were blocked for 1 h with 5% non-fat dry milk in Tris-buffered saline containing 0.05% Tween 20 (TBST). After 3 washes with TBST, membranes were then exposed to primary antibodies overnight at 4 °C (MBNL1, CUGBP1, myogenin, myosin). Subsequently, membranes were washed 3 times in TBST and incubated in horseradish-peroxidase-conjugated anti-mouse antibody solution at room temperature for 1h. Finally, Chemiluminescence detection was achieved using ETA C 2.0 and NOVA 2.0 reagents (Cyanagen, Bologna, Italy) in a ChemiDoc XRS+ (Bio-Rad, Hercules, CA). Imaging and quantitation of the bands were carried out by the ChemiDoc XRS Western Blot Imaging System using the ImageLab 4.0 software (Bio-Rad, Hercules, CA). Equal loading of proteins onto the gel was confirmed by immunodetection of vinculin using Monoclonal Ab α Vinculin (Sigma-Aldrich, St. Louis, MO) at room temperature for 1h.

7.3.2.4 PCR.

RNAs from both control and patient cell lines were extracted with TRIzol reagent (Invitrogen, Thermo Scientific, Carlsbad, CA). Cells were washed twice with PBS and frozen at -80°C for 30 min. After that, TRIzol reagent was added and cells were scraped off the plates, transferred into Eppendorf tubes and incubated at room temperature for 5 min. Chloroform was added and the resulting mixture was vigorously shaken and incubated 3 minutes at room temperature. The aqueous phase was separated by centrifugation, and RNA precipitated by adding IPA. The sample was centrifuged at 12000 xg for 10 min at 4 °C. The supernatant was discarded, and

the RNA was washed twice with ethanol 80% prechilled at -20 °C. After drying, samples were dissolved in RNase free water, incubated 5 min at 60 °C, cooled in ice and quantified using Nanodrop ND-1000 (Nanodrop Technologies, Wilmington, DE). After that, RNA samples were retro transcribed with the GoScript Reverse Transcription System (Promega, Fitchburg, WI). Oligo (dT) and random primers were preincubated with RNA for 5 min at 70 °C. The resulting mixture was chilled in ice for 5 min. Next, reactions were prepared by adding 3mM MgCl₂, PCR Nucl. Mix, Rnasin, Reverse Transcriptase and placed in the thermal cycler at 42 °C for 1 h, 70 °C for 15 min and finally chilled. The resulting cDNA samples were amplified with GoTaq Flexi DNA Polymerase (Promega, Fitchburg, WI) for 35 cycles using the specific primers (Table S2). Upon reaction ending, samples were loaded on agarose gel containing Midori staining (Nippon Genetics, Düren, GE). Quantitation of amplified bands on gel images was performed using UVITec1D software. Exon inclusion was quantified as the percentage of the total intensity of both isoforms.

7.3.2.5 Statistics.

All biological experiments were performed at least in triplicate, and pictures represent typical examples. Data are presented as mean ± standard deviation (SD). A two-tailed unpaired t-test was applied for assessment of the statistical significance of differences between two groups. A p-value lower than 0.05 was deemed statistically significant. Statistical analyses were conducted using GraphPad Prism version 7.01 (GraphPad Software, Inc., La Jolla, CA, USA).

7.4 Chemical Synthesis

7.4.1 Instrumental

Nuclear Magnetic Resonance spectra ($^1\text{H-NMR}$ and $^{13}\text{C-NMR}$) were recorded in a Varian 400-MR spectrometer ($^1\text{H-NMR}$ 400 MHz and $^{13}\text{C-NMR}$ 100.6 MHz), in the Organic Chemistry Department at IQS under the leadership of Dr. X. Batllori. Chemical shifts are reported in parts per million (ppm) on the δ scale and referenced to the residual signal of the solvent (DMSO- d_6 : 2.5 ppm in $^1\text{H-NMR}$ and 39.52 ppm in $^{13}\text{C-NMR}$, CDCl_3 : 7.26 ppm in $^1\text{H-NMR}$ and 77.16 ppm in $^{13}\text{C-NMR}$). Coupling constants (J) are reported in Hertz (Hz). Standard and peak multiplicities are designated as follows: s (singlet), d (doublet), dd (doublet of doublets), ddd (doublet of doublets of doublets), t (triplet), q (quartet), qn (quintet), m (multiplet), br (broad signal).

Infrared Spectra (IR) were recorded in a Nicolet iS10 FTIR spectrometer with Smart iTr (Thermo Scientific) by Mrs. N. Ruiz at the Organic and Pharmaceutical Chemistry department at IQS under the leadership of Dr. X. Batllori, using a potassium bromide (KBr) disc. Wave number values are reported in cm^{-1} . The notation used is: st (stretching), b (bending).

Mass Spectrometry (MS) was conducted on an Agilent Technologies 5975 mass spectrometer operating in electron ionization (EI) mode at 70 eV and at 4kV accelerating potential.

High Resolution Mass Spectrometry (HRMS) was conducted on a microTOF (Bunker) high resolution mass spectrometer operating in APCI, ESI or EI mode at the Unidade de espectrometría de masas e proteomica (Universidade de Santiago de Compostela) under the leadership of Dr. Esteban Guitián.

The melting point (MP) was determined with a SMP3 melting point apparatus (Stuart Scientific).

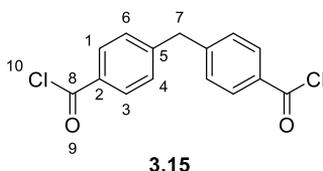
Automatic flash chromatography was performed in a Combiflash®Rf (Teledyne Isco) with RediStep® silica gel columns or basic alumina columns.

Microwave irradiation experiments were carried out in an Initiator™ (Biotage) microwave apparatus, operating at a frequency of 2.45 GHz with continuous irradiation power from 0 to 400 W. Reactions

Chemical Synthesis

were performed in 2.5, 5 and 20 mL glass tubes, sealed with aluminum/Teflon crimp tops, which can be exposed up to 250 °C and 20 bar internal pressure. Temperature was measured with an IR sensor on the outer surface of the process vial. After the irradiation period, the reaction vessel was cooled rapidly to 50 °C by air jet cooling.

7.4.2 Synthesis of 4,4'-methylenedibenzoyl chloride (3.15).



262.4 mg of 4,4'-methylenedibenzoic acid were suspended in 3.5 mL of methylene chloride. 5 mL of thionyl chloride were added and the mixture was heated at reflux temperature under nitrogen atmosphere for 24h. The resulting solution was concentrated *in vacuo* to afford 298,1mg (99%) of 4,4'-methylenedibenzoyl chloride(**3.15**) as a brownish solid.

Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3426.26, 2988.88, 2670.96, 2550.91, 1740.90, 1604.12, 1424.22, 1317.08, 1292.10, 1175.36, 1114.72, 1018.38, 881.93, 744.96, 545.86.

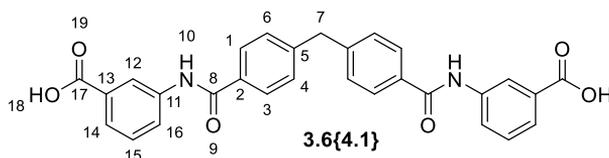
$^1\text{H-NMR}$ (400 MHz, d_4 -MeOH) δ (ppm): 8.07 – 7.84 (m, 4H), 7.42 – 7.22 (m, 4H), 4.11 (s, 2H).

The spectroscopic data found match to the described in bibliography¹⁷.

7.4.3 Method 1: Schotten-Baumann amide bond formation.

To a suspension of the corresponding primary amine in brine was added dropwise a solution containing the dibenzoyl chloride in acetone. A saturated solution of sodium hydrogencarbonate was added and the reaction mixture was stirred for 15 minutes. Upon stirring the solution was acidulated using concentrated hydrochloric acid. The resulting precipitate was isolated by filtration, washed with hydrochloric acid (1M) and cold water and dried to afford the title compound.

7.4.4 Synthesis of 3,3'-((4,4'-methylenebis(benzoyl))bis(azanediyl))dibenzoic acid(3.6{4.1}).



Using method 1, 148,8 mg (88%) of 3,3'-((4,4'-methylenebis(benzoyl))bis(azanediyl)) dibenzoic acid(3.6{4.1}) were obtained as white solid.

Spectroscopic data:

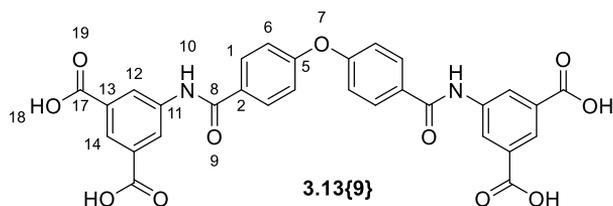
IR (KBr) ν (cm⁻¹): 3325.28, 3043.02, 2576.64, 1689.85, 1593.36, 1541.00, 1434.68, 1304.74, 1263.39, 1118.88, 940.46, 755.38, 678.74

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 12.95 (s, 2H), 10.35 (s, 2H), 8.40 (t, *J* = 1.8 Hz, 2H), 8.03 (ddd, *J* = 8.2, 2.2, 1.1 Hz, 2H), 7.96 – 7.90 (m, 4H), 7.67 (dt, *J* = 7.7, 1.2 Hz, 2H), 7.49 – 7.46 (m, 2H), 7.46 – 7.41 (m, 4H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm): 167.20 (C17), 165.50 (C8), 144.74 (C5), 139.44 (C11), 132.62 (C2), 131.19 (C13), 128.86 (C15), 128.79 (C4-C6), 128.04 (C1-C3), 124.38 (C14), 124.36 (C16), 121.07 (C12), 40.67 (C7).

HRMS (ESI-TOF) *m/z* (%): calculated for C₂₉H₂₃N₂O₆[M+1]⁺: 495,1551; found [M+1]⁺:495,1541.

7.4.5 Synthesis of 5,5'-((4,4'-oxybis(benzoyl))bis(azanediyl))diisophthalic acid(3.13{9}).



Using method 1, 170,3 mg (86%) of 5,5'-((4,4'-oxybis(benzoyl))bis(azanediyl))diisophthalic acid (**3.13{9}**) were obtained as white solid.

Spectroscopic data:

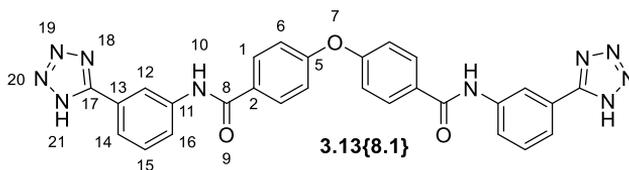
IR (KBr) $\nu(\text{cm}^{-1})$: 3418.01, 3080.98, 1707.64, 1595.73, 1556.21, 1501.53, 1245.84, 1175.07, 759.93, 665.37, 594.40.

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 13.25(s, 4H), 10.75 (s, 2H), 8.94 – 8.58 (m, 4H), 8.21 (q, *J* = 1.4 Hz, 2H), 8.16 (d, *J* = 8.1 Hz, 4H), 7.23 (d, *J* = 8.7 Hz, 4H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm): 166.52(C17), 164.97(C8), 158.90(C5), 139.95(C11), 131.64(C13), 130.38(C1), 129.76(C2), 124.93(C14), 124.86(C12), 118.54(C6).

HRMS (ESI-TOF) *m/z* (%): calculated for C₃₀H₂₁N₂O₁₁ [M+1]⁺: 585.1140; found [M+1]⁺:585.1134.

7.4.6 Synthesis of 4,4'-oxybis(N-(3-(1H-tetrazol-5-yl)phenyl)benzamide) (3.13{8.1}).



Using method 1, 149.5 mg (81%) of 4,4'-oxybis(N-(3-(1H-tetrazol-5-yl)phenyl)benzamide) (**3.13{8.1}**) were obtained as white solid.

Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3425.34, 3301.43, 1644.41, 1596.85, 1568.86, 1534.13, 1501.34, 1450.95, 1261.00, 1171.91, 1085.78, 852.22, 838.10, 685.02.

$^1\text{H-NMR}$ (400 MHz, d_6 -DMSO) δ (ppm): 12.89 (s, 2H), 10.59 (s, 2H), 8.63 (t, $J = 1.9$ Hz, 2H), 8.17 – 8.10 (m, 4H), 8.04 – 7.95 (m, 2H), 7.84 (dt, $J = 7.7, 1.2$ Hz, 2H), 7.58 (t, $J = 8.0$ Hz, 2H), 7.29 – 7.21 (m, 4H).

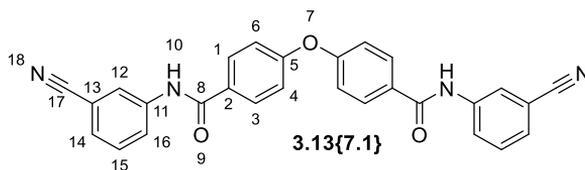
$^{13}\text{C-NMR}$ (100 MHz, d_6 -DMSO) δ (ppm): 164.89(C8), 158.82(C5), 140.09(C11), 130.28(C1), 130.02(C13), 129.70(C15), 122.89(C14), 122.19(C16), 118.93(C2), 118.80(C12), 118.53(C6).

HRMS (ESI-TOF) m/z (%): calculated for $\text{C}_{28}\text{H}_{21}\text{N}_{10}\text{O}_3$ $[\text{M}+1]^+$: 545.1793; found $[\text{M}+1]^+$: 545.1783.

7.4.7 Method 2: Conventional amide bond formation.

A mixture of 4,4'-oxibis(benzoyl chloride) and the corresponding primary amine were dissolved in anhydrous THF, trimethylamine was added and the solution was refluxed for 24h under nitrogen atmosphere. The solvent was evaporated *in vacuo* and the resulting solid was suspended using 1M hydrochloric acid and finally filtered and washed with water to obtain the title compound.

7.4.8 Synthesis of 4,4'-oxybis(*N*-(3-cyanophenyl)benzamide)(**3.13**{**7.1**}).



Using method 2, 227.0 mg (99%) of 4,4'-oxybis(*N*-(3-cyanophenyl)benzamide)(**3.13**{**7.1**}) were obtained as white solid.

Spectroscopic data:

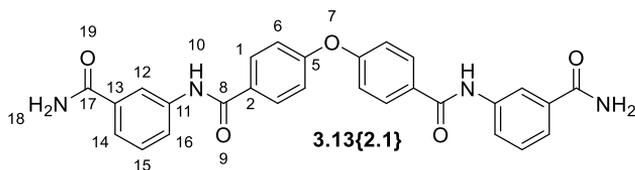
IR (KBr) $\nu(\text{cm}^{-1})$: 3384.95, 3265.65, 2231.56, 1661.32, 1648.40, 1593.94, 1544.85, 1498.73, 1427.97, 1247.56, 1170.80, 793.17, 680.79.

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 10.55 (s, 2H), 8.26 (q, *J* = 1.3 Hz, 2H), 8.09 – 8.02 (m, 6H), 7.62 – 7.55 (m, 4H), 7.28 – 7.22 (m, 4H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm): 165.05(C8), 158.89(C5), 140.01(C11), 130.26(C1-C3), 130.12(C14), 129.76(C2), 127.11(C15), 124.85(C16), 122.99(C12), 118.73(C17), 118.58(C4-C6), 111.45(C13).

HRMS (ESI-TOF) *m/z* (%): calculated for C₂₈H₁₉N₄O₃ [M+1]⁺: 459.1452; found [M+1]⁺: 459.1449.

7.4.9 Synthesis of 4,4'-oxybis(*N*-(3-carbamoylphenyl)benzamide) (**3.13{2.1}**).



Using method 2, 221,7 mg (90%) of 4,4'-oxybis(*N*-(3-carbamoylphenyl)benzamide) (**3.13{2.1}**) were obtained as white solid.

Spectroscopic data:

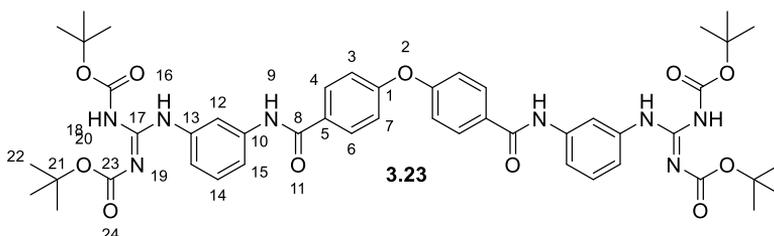
IR (KBr) ν (cm⁻¹): 3309.63, 3178.97, 1658.79, 1586.99, 1550.18, 1501.01, 1446.77, 1391.46, 1238.08, 1177.37, 1113.93, 761.42, 676.51.

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 10.37 (s, 1H), 8.25 (t, *J* = 1.8 Hz, 1H), 8.14 – 8.04 (m, 2H), 7.95 (ddd, *J* = 8.1, 2.1, 0.9 Hz, 2H), 7.64 – 7.56 (m, 1H), 7.42 (t, *J* = 7.9 Hz, 1H), 7.36 (s, 1H), 7.28 – 7.17 (m, 2H).

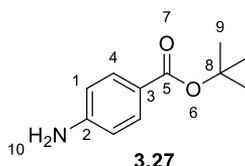
¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm): 167.89(C17), 164.69(C8), 158.74(C5), 139.18(C11), 134.97(C13), 130.14(C1-C2), 128.43(C15), 123.14(C16), 122.41(C14), 120.14(C12), 118.51(C6).

HRMS (ESI-TOF) *m/z* (%): calculated for C₂₈H₂₂N₄O₅Na [M+1]⁺: 517.1482; found [M+1]⁺: 517.1483.

7.4.10 Synthesis of 4,4'-oxybis(N-(N',N''-di-tert-butylloxycarbonyl-3-guanidinophenyl)benzamide) (3.23).



Using method 2, 278.7 mg (89%) of 4,4'-oxybis(N-(N',N''-di-tert-butylloxycarbonyl-3-guanidinophenyl)benzamide) (3.23) were obtained as pale yellowish solid and was used in the next step without further purification.

7.4.11 Synthesis of tert-butyl 4-aminobenzoate (3.27).

A mixture of 4-aminobenzoic acid (**3.26**) and DCC were dissolved in *t*-ButOH and heated under reflux for 18h under anhydrous atmosphere. The reaction mixture was cooled to 0°C, 20 mL of water were added. The resulting mixture was extracted with diethyl ether. The extract was dried over anhydrous magnesium sulphate filtered and the solvent was evaporated in vacuo. The obtained crude product was further purified by column chromatography to give 1,4559g (91%) of *tert*-butyl 4-aminobenzoate (**3.27**).

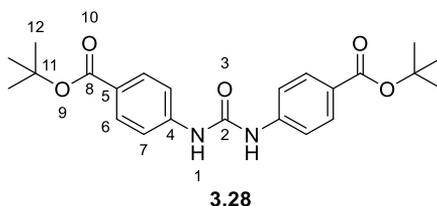
Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): δ 7.69 – 7.50 (m, 2H), 6.60 – 6.47 (m, 2H), 5.86 (s, 2H), 1.49 (s, 9H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm) 165.29(C5), 153.08(C2), 130.88(C4), 117.63(C3), 112.54(C1), 78.87(C8), 28.02(C9).

The spectroscopic data found match to the described in bibliography¹⁸.

7.4.12 Synthesis of 4,4'-ureylene-di-(*tert*-butyl benzoate) (3.28).



A mixture of *tert*-butyl 4-aminobenzoate (**3.27**) and CDI in anhydrous THF was heated under reflux for 24h under nitrogen atmosphere. Upon cooling, solvent was evaporated in vacuo. The residue was dissolved in ethyl acetate and washed with hydrochloric acid and brine. The extract was dried over anhydrous magnesium sulphate, filtered and concentrated in vacuo to give 1,5596g (97%) of 4,4'-Ureylene-di-(*tert*-butyl benzoate)(**3.28**).

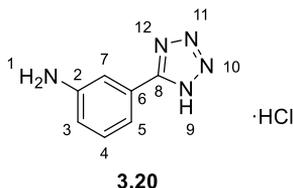
Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 9.23 (s, 2H), 7.83 (d, J = 8.9 Hz, 4H), 7.57 (d, J = 8.9 Hz, 4H), 1.53 (s, 18H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm) 164.66(C8), 151.89(C2), 143.61(C4), 130.23(C6), 124.54(C5), 117.32(C7), 80.11(C11), 27.87(C12).

The spectroscopic data found match to the described in bibliography¹⁸.

7.4.13 Synthesis of 3-(1H-tetrazol-5-yl)aniline hydrochloride (3.20).



A mixture of 3-aminobenzonitrile (**3.19**), thiethylamine hydrochloride and sodium azide in toluene was heated under reflux for 24h under nitrogen atmosphere. Upon cooling the reaction mixture, water was added and two phases were observed. The aqueous phase was separated and acidulated to pH 2. Upon cooling the extract, crystallization occurs. The resulting solid was filtered, washed with cold water and dried to afford 360,1mg (48%) of 3-(1H-tetrazol-5-yl)aniline hydrochloride (**3.20**) as an orange crystalline solid.

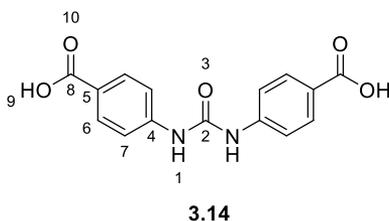
Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3434.35, 3045.67, 2899.17, 2839.55, 2778.63, 2749.40, 2597.14, 1631.24, 1594.99, 1553.83, 1524.54, 1466.63, 1131.44, 1114.69, 1087.60, 830.99, 741.58, 684.61, 482.90.

¹H-NMR (400 MHz, d_6 -DMSO) δ (ppm): 7.97 – 7.85 (m, 2H), 7.66 – 7.57 (m, 1H), 7.46 – 7.38 (m, 1H).

The spectroscopic data found match to the described in bibliography¹⁹.

7.4.14 Synthesis of 4,4'-Ureylene-dibenzoic acid(3.14)



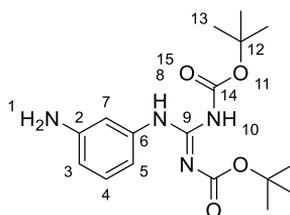
4,4'-Ureylene-di(*tert*-butyl benzoate)(**3.28**) was dissolved in trifluoroacetic acid and stirred for 1h at room temperature. Diethyl ether was added and the reaction mixture was filtered, washed with diethyl ether and dried to give 1,1380g (80%) of 4,4'-Ureylene-dibenzoic acid (**3.14**) as a fine white powder.

Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 12.62 (s, 2H), 9.17 (s, 2H), 7.88 (d, *J* = 8.7 Hz, 4H), 7.57 (d, *J* = 8.8 Hz, 4H).

The spectroscopic data found match to the described in bibliography¹⁸.

7.4.15 Synthesis of *N,N'*-di-*tert*-butyloxycarbonyl-3-guanidinoaniline (3.18).



3.18

A mixture of *m*-phenylenediamine (**3.17**), 1,3-Bis(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea, mercury (II) chloride and triethylamine in anhydrous DMF was stirred at 0°C for 1h and at room temperature for 18h under nitrogen atmosphere. Ethyl acetate was added and the reaction mixture was filtered through cellite pad to remove the mercury salts. The reaction solution was washed with water and brine, dried over anhydrous magnesium sulphate, filtered and concentrated *in vacuo*. The obtained crude product was further purified by column chromatography to give 962,6mg (92%) of *N,N'*-di-*tert*-butyloxycarbonyl-3-aminophenylguanidine (**3.18**) as a white solid.

Spectroscopic data:

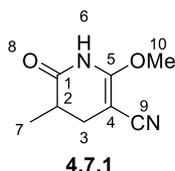
IR (KBr) $\nu(\text{cm}^{-1})$: 3425.88, 3329.36, 2979.63, 293.66, 1718.43, 1642.16, 1624.32, 1412.06, 1337.25, 1156.96, 1109.90, 1060.02, 863.69, 803.82, 779.49, 691.10, 653.63, 597.16.

¹H-NMR (400 MHz, d_6 -DMSO) δ (ppm): 11.42 (s, 1H), 9.85 (s, 1H), 6.97 (t, $J = 8.0$ Hz, 1H), 6.73 (d, $J = 8.0$ Hz, 1H), 6.66 – 6.59 (m, 1H), 6.37 (ddd, $J = 8.0, 2.2, 0.9$ Hz, 1H), 5.18 (s, 2H), 1.66 – 1.22 (m, 18H).

¹³C-NMR (100 MHz, d_6 -DMSO) δ (ppm): 162.75, 152.72, 152.29, 149.04, 137.11, 129.07, 110.95, 110.15, 107.88, 83.33, 78.72, 27.86, 27.67, 26.34.

HRMS (EI) m/z (%): calculated for $\text{C}_{17}\text{H}_{26}\text{N}_4\text{O}_4$ $[\text{M}+1]^+$: 350.1954; found $[\text{M}+1]^+$: 350.1971.

7.4.16 Synthesis of 2-methoxy-5-methyl-6-oxo-1,4,5,6-tetrahydropyridine-3-carbonitrile (4.7.1).



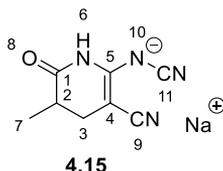
Sodium methoxide is dissolved in anhydrous methanol. Malononitrile is added and the reaction mixture is cooled at room temperature. Upon cooling, methyl methacrylate (**4.5.1**) is added and the reaction mixture is heated to reflux under nitrogen atmosphere for 5h. Solvent is evaporated in vacuo and the residue is redissolved in deionized water. The resulting solution is neutralized by adding 2M hydrochloric acid dropwise, during the addition, massive precipitation of a yellow solid occur. When precipitation was complete, the precipitate was filtered, washed with ice water and dried to afford 2.9329g (49%) of 2-methoxy-5-methyl-6-oxo-1,4,5,6-tetrahydropyridine-3-carbonitrile(**4.7.1**) as a pale yellowish powder.

Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 10.49 (s, 1H), 3.91 (s, 3H), 2.60 – 2.45 (m, 1H), 2.42 (dd, *J* = 15.2, 6.7 Hz, 1H), 2.20 (dd, *J* = 15.2, 12.2 Hz, 1H), 1.08 (d, *J* = 6.8 Hz, 3H).

The spectroscopic data found match to the described in bibliography²⁰.

7.4.17 Synthesis of the sodium salt of *N*-(3-cyano-5-methyl-6-oxo-1,4,5,6-tetrahydropyridin-2-yl)cyanamide (4.15).



A mixture of 2-methoxy-5-methyl-6-oxo-1,4,5,6-tetrahydropyridine-3-carbonitrile (**4.7.1**) and sodium hydrogencyanamide in 1,4-dioxane was heated to reflux for 5h. Upon cooling the reaction mixture was filtered washed with fresh 1,4-dioxane, diethylether and dried to afford 1.3017g (63%) of the sodium salt of *N*-(3-cyano-5-methyl-6-oxo-1,4,5,6-tetrahydropyridin-2-yl)cyanamide (**4.15**) as a pale greenish powder.

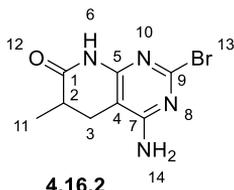
Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 8.53 (s, 1H), 2.38 – 2.27 (m, 1H), 2.24 (dd, *J* = 14.4, 6.2 Hz, 1H), 2.05 (dd, *J* = 14.4, 11.3 Hz, 1H), 1.03 (d, *J* = 6.8 Hz, 3H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm): 173.60(C1), 156.39(C5), 123.19(C9), 119.76(C11), 55.17(C4), 35.19(C2), 29.98(C3), 14.83(C7).

The spectroscopic data found match to the described in bibliography²⁰.

7.4.18 Synthesis of 4-amino-2-bromo-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (4.16.2).



The sodium salt of *N*-(3-cyano-5-methyl-6-oxo-1,4,5,6-tetrahydropyridin-2-yl)cyanamide (**4.15**) was dissolved in concentrated hydrobromic acid and stirred at room temperature for 1h. The reaction mixture is cooled at 0°C and aqueous ammonia is added to the reaction mixture dropwise until neutralization. The resulting mixture is filtered washed with water and dried to afford 313.0mg (84%) of 4-amino-2-bromo-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (**4.16.2**) as a pale reddish powder.

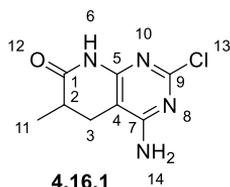
Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 10.48 (s, 1H), 7.10 (s, 2H), 2.77 (dd, *J* = 16.0, 6.9 Hz, 1H), 2.65 – 2.53 (m, 1H), 2.24 (dd, *J* = 16.0, 11.6 Hz, 1H), 1.12 (d, *J* = 6.9 Hz, 3H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm): 173.41(C1), 162.09(C5-C7), 156.90(C5-C7), 148.51(C9), 92.55(C4), 33.75(C2), 24.94(C3), 15.36(C11).

The spectroscopic data found match to the described in bibliography²⁰.

7.4.19 Synthesis of 4-amino-2-chloro-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (4.16.1).



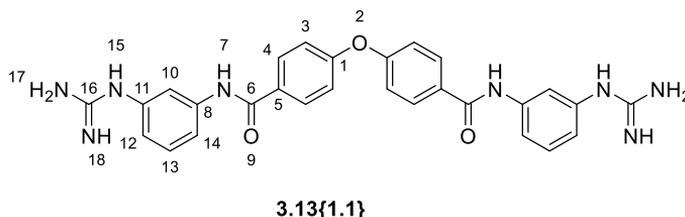
The sodium salt of *N*-(3-cyano-5-methyl-6-oxo-1,4,5,6-tetrahydropyridin-2-yl)cyanamide (**4.15**) was dissolved in 4M hydrogen chloride (in 1,4-dioxane) and stirred for 1 h at room temperature. The reaction mixture is cooled at 0 °C and aqueous ammonia is added to the reaction mixture dropwise until neutralization. The resulting mixture is filtered washed with water and dried to afford 341,5 mg (69%) of 4-amino-2-chloro-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (**4.16.1**) as a pale yellowish powder.

Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 10.48 (s, 1H), 7.09 (s, 2H), 2.79 (dd, *J* = 16.0, 6.9 Hz, 1H), 2.66 – 2.54 (m, 1H), 2.26 (dd, *J* = 16.0, 11.7 Hz, 1H), 1.12 (d, *J* = 6.9 Hz, 3H).

The spectroscopic data found match to the described in bibliography²⁰.

7.4.20 Synthesis of 4,4'-oxybis(N-guanidinophenylbenzamide)(3.13{1.1}).



4,4'-oxybis(*N,N'*-di-*tert*-butyloxycarbonyl-*N*-guanidinophenylbenzamide) **3.23** was dissolved in trifluoroacetic acid and dichloromethane (1:1) and stirred for 3h at room temperature. The reaction mixture was concentrated in vacuo to afford the trifluoroacetate salt of 4,4'-oxybis(*N*-guanidinophenylbenzamide)(**3.13{1.1}**).

Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3418.65, 3274.22, 3143.09, 2983.11, 1686.89, 1652.01, 1601.13, 1554.91, 1500.10, 1461.24, 1371.14, 1252.08, 1088.65, 1064.18, 1043.35, 813.54, 740.86, 602.50.

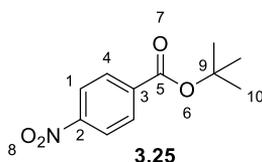
¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 10.41 (s, 2H), 9.97 (s, 2H), 8.13 – 8.02 (m, 4H), 7.81 (t, *J* = 2.0 Hz, 2H), 7.72 – 7.65 (m, 2H), 7.55 (s, 6H), 7.42 (t, *J* = 8.1 Hz, 2H), 7.27 – 7.19 (m, 4H), 7.03 – 6.93 (m, 2H).

¹³C-NMR (100 MHz, *d*₆-DMSO) δ (ppm): 164.84(C6), 158.77(C1), 155.88(C16), 140.40(C11), 135.64(C8), 130.18(C5), 130.16(C4), 129.85(C13), 119.46(C12), 118.53(C3), 118.00(C14), 115.9(C10).

HRMS (EI) *m/z* (%): calculated for C₂₈H₂₇N₈O₃ [M+1]⁺: 523.2201; found [M+1]⁺: 523.2203.

7.4.21 Method 3: Protection of nitrobenzoic acid.

The corresponding isomer of nitrobenzoic acid was dissolved in dichloromethane and dimethylformamide. The reaction mixture was cooled at 0°C and tert-butanol and dimethylaminopyridine were added. Once the resulting solution was cold, dicyclohexylcarbodiimide was added and the reaction mixture was stirred for 4h at room temperature. Upon stirring, the crude was filtered and washed with 1M hydrochloric acid and brine. The extract was dried over anhydrous magnesium sulphate, filtered and concentrated *in vacuo*. The resulting crude solid was further purified by column chromatography to give the title compound.

7.4.22 Synthesis of tert-butyl 4-nitrobenzoate (3.25).

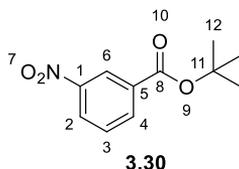
Using method 3, 221,7 mg (90%) of *tert*-butyl 4-nitrobenzoate (**3.25**) were obtained as a white solid.

Spectroscopic data:

¹H-NMR (400 MHz, CDCl₃) δ (ppm): 8.38 – 8.22 (m, 2H), 8.22 – 8.06 (m, 2H), 1.62 (s, 9H).

The spectroscopic data found match to the described in bibliography¹⁸.

7.4.23 Synthesis of *tert*-butyl 3-nitrobenzoate (**3.30**).



Using method 3, 1.8153 g (91%) of *tert*-butyl 3-nitrobenzoate (**3.30**) were obtained as a white solid.

Spectroscopic data:

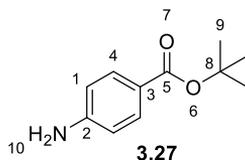
¹H-NMR (400 MHz, CDCl₃) δ (ppm): 8.79 (ddd, J = 2.2, 1.6, 0.4 Hz, 1H), 8.38 (ddd, J = 8.2, 2.4, 1.1 Hz, 1H), 8.32 (ddd, J = 7.7, 1.6, 1.2 Hz, 1H), 7.66 – 7.57 (m, 1H), 1.63 (s, 9H).

The spectroscopic data found match to the described in bibliography¹⁸.

7.4.24 Method4: Catalyzed reduction from nitro to amine group.

The corresponding isomer of *tert*-butylnitrobenzoate was dissolved in anhydrous methanol. Palladium (over carbon 10% wt) was added. Hydrogen was bubbled and the reaction mixture was stirred at room temperature until the reaction was complete. The completion of the reaction was monitored by TLC. The reaction mixture was filtered through celite pad and the solvent was evaporated *In vacuo* to afford title compound.

7.4.25 Synthesis of *tert*-butyl 4-aminobenzoate (**3.27**).

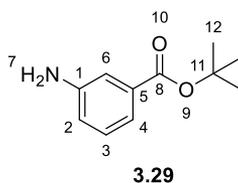


Using method 4, 1.7495 g (98%) of *tert*-butyl 4-aminobenzoate (**3.27**) were obtained as a white solid.

Spectroscopic data:

¹H-NMR (400 MHz, CDCl₃) δ (ppm): 8.06 – 7.72 (m, 2H), 6.72 – 6.51 (m, 2H), 3.98 (s, 2H), 1.56 (s, 9H).

The spectroscopic data found match to the described in bibliography¹⁸.

7.4.26 Synthesis of tert-butyl 3-aminobenzoate (3.29).

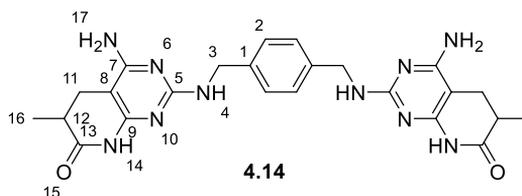
Using method 3, 1.5086 g (96%) of *tert*-butyl 3-aminobenzoate (**3.29**) were obtained as a white solid.

Spectroscopic data:

¹H-NMR (400 MHz, CDCl₃) δ (ppm): 7.49 – 7.44 (m, 1H), 7.43 – 7.39 (m, 1H), 7.23 (t, J = 7.8 Hz, 1H), 6.95 (ddd, J = 7.9, 2.4, 1.0 Hz, 1H), 4.68 (s, 2H), 1.58 (s, 9H).

The spectroscopic data found match to the described in bibliography¹⁸.

7.4.27 Synthesis of 2,2'-((1,4-phenylenebis(methylene))bis(azanediyl))bis(4-amino-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one) (4.14).



A mixture of 4-amino-2-bromo-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one (**4.16.2**), xylylenediamine (**4.17**) and anhydrous sodium carbonate in anhydrous dimethylsulfoxide were heated under microwave radiation at 190°C for 9h. The reaction mixture was poured into deionized water. The formed precipitate was filtered, washed with water and dried to afford 346.4 mg (91%) of 2,2'-((1,4-phenylenebis(methylene))bis(azanediyl))bis(4-amino-6-methyl-5,8-dihydropyrido[2,3-d]pyrimidin-7(6H)-one) (**4.14**) as a yellowish solid.

Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3303.23, 3205.30, 2931.03, 1686.64, 1627.83, 1578.80, 1553.03, 1451.05, 1383.01, 1266.20, 1213.86, 785.12, 594.25.

¹H-NMR (400 MHz, d_6 -DMSO) δ (ppm): 9.83 (s, 2H), 7.20 (s, 4H), 6.65 (t, $J = 6.2$ Hz, 2H), 6.10 (s, 4H), 4.36 (d, $J = 6.4$ Hz, 4H), 2.69 (dd, $J = 15.4, 6.7$ Hz, 2H), 2.49 – 2.42 (m, 2H), 2.14 (dd, $J = 15.3, 11.5$ Hz, 2H), 1.10 (d, $J = 6.8$ Hz, 6H).

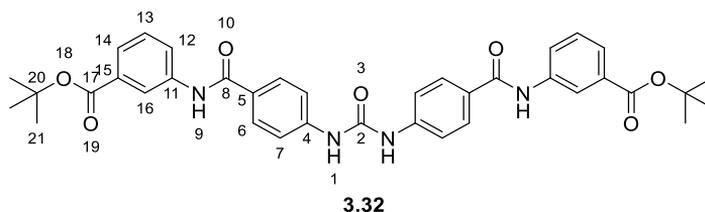
¹³C-NMR (100 MHz, d_6 -DMSO) δ (ppm): 174.17(C13), 161.57(C7), 160.74(C5), 156.45(C9), 139.20(C1), 126.85(C2), 83.84(C8), 43.53(C3), 34.73(C12), 25.37(C11), 15.61(C16).

HRMS (EI) m/z (%): calculated for $\text{C}_{24}\text{H}_{29}\text{N}_{10}\text{O}_2$ $[\text{M}+1]^+$: 489.2469; found $[\text{M}+1]^+$: 489.2468.

7.4.28 Method 5: EDC mediated amide bond formation.

A mixture of the corresponding primary amine, 4,4'-Ureylene-dibenzoic acid, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride, N-hydroxibenzotriazole hydrate and triethylamine in dimethylformamide was stirred for 24h at room temperature under nitrogen atmosphere. The solvent was evaporated *in vacuo*, deionized water was added to the resulting oil and the mixture was sonicated until the crude is completely dissolved. The white solid that precipitated was filtered and washed with water to afford the title compound.

7.4.29 Synthesis of di-tert-butyl 3,3'-((4,4'-(carbonylbis(azanediyl))bis(benzoyl)) bis(azanediyl))dibenzoate (3.32).

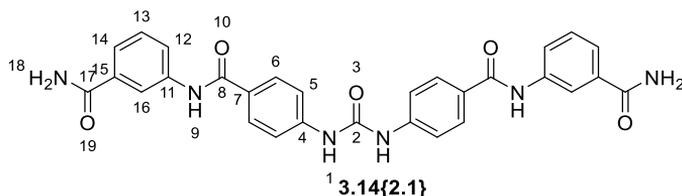


Using method 5, 205.2 mg (95%) of di-tert-butyl 3,3'-((4,4'-(carbonylbis(azanediyl)) bis(benzoyl)) bis(azanediyl))dibenzoate (**3.32**) as a white solid.

Spectroscopic data:

¹H-NMR (400 MHz, *d*₆-DMSO) δ (ppm): 10.30 (s, 2H), 9.17 (s, 2H), 8.35 (t, *J* = 1.9 Hz, 2H), 8.06 (ddd, *J* = 8.2, 2.2, 1.1 Hz, 2H), 7.99 – 7.94 (m, 4H), 7.66 – 7.60 (m, 6H), 7.47 (t, *J* = 7.9 Hz, 2H), 1.57 (s, 18H).

7.4.30 Synthesis of 4,4'-(carbonylbis(azanediyl))bis(*N*-(3-carbamoylphenyl)benzamide) (3.14{2.1}).



Using method 5, 80.7 mg (90%) of 4,4'-(carbonylbis(azanediyl))bis(*N*-(3-carbamoylphenyl)benzamide) (3.14{2.1}) as a white solid.

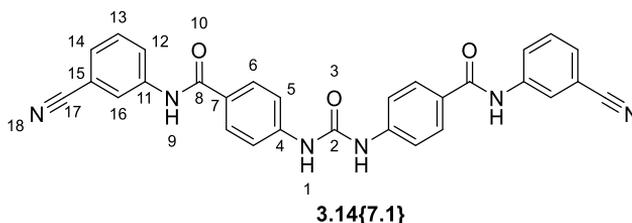
Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3375.78, 3171.25, 1661.25, 1590.77, 1535.65, 1508.87, 1485.78, 1446.50, 1387.23, 1316.67, 1240.02, 1184.50, 847.85, 763.14, 642.48.

$^1\text{H-NMR}$ (400 MHz, d_6 -DMSO) δ (ppm): 10.23 (s, 2H), 9.16 (s, 2H), 8.24 (t, $J = 1.7$ Hz, 2H), 7.98 (d, $J = 8.8$ Hz, 4H), 7.94 (dd, $J = 7.2$, 2.0 Hz, 4H), 7.62 (d, $J = 8.8$ Hz, 4H), 7.58 (d, $J = 7.7$ Hz, 2H), 7.41 (t, $J = 7.9$ Hz, 2H), 7.34 (s, 2H).

$^{13}\text{C-NMR}$ (100 MHz, d_6 -DMSO) δ (ppm): 167.99(C17), 164.99(C8), 152.14(C2), 142.73(C4), 139.35(C11), 134.95(C15), 128.86(C6), 128.41(C13), 127.79(C7), 123.13(C12), 122.26(C14), 120.10(C16), 117.39(C5).

7.4.31 Synthesis of 4,4'-(carbonylbis(azanediyl))bis(N-(3-cyanophenyl)benzamide) (3.14{7.1}).



Using method 5, 76.6 mg (92%) of 4,4'-(carbonylbis(azanediyl))bis(N-(3-cyanophenyl)benzamide) (3.14{7.1}) were obtained as a white solid.

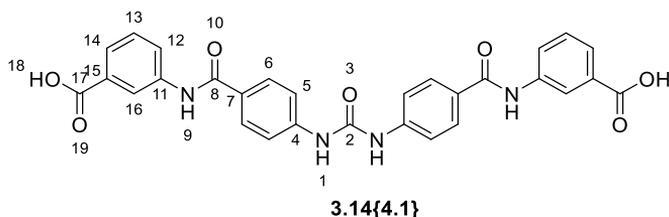
Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3361.94, 2232.53, 1650.23, 1591.38, 1531.21, 1509.52, 1485.51, 1428.19, 1414.32, 1316.34, 1242.25, 1184.30, 850.07, 792.72, 762.15, 679.93, 520.17.

$^1\text{H-NMR}$ (400 MHz, d_6 -DMSO) δ (ppm): 10.41 (s, 1H), 9.19 (s, 1H), 8.29 – 8.24 (m, 1H), 8.05 (dt, $J = 7.5, 2.1$ Hz, 1H), 8.00 – 7.92 (m, 1H), 7.68 – 7.62 (m, 2H), 7.61 – 7.53 (m, 1H).

$^{13}\text{C-NMR}$ (100 MHz, d_6 -DMSO) δ (ppm): 165.35(C8), 152.08(C2), 142.98(C4), 140.20(C11), 130.11(C13), 128.97(C6), 127.37(C7), 126.92(C14), 124.79(C12), 122.93(C16), 118.81(C17), 117.44(C5), 111.44(C15).

7.4.32 Synthesis of 3,3'-((4,4'-(carbonylbis(azanediyl))bis(benzoyl))bis(azanediyl)) dibenzoic acid (3.14{4.1})



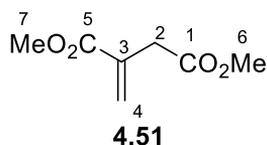
Di-*tert*-butyl 3,3'-((4,4'-(carbonylbis(azanediyl))bis(benzoyl))bis(azanediyl))dibenzoate (**3.32**) was dissolved in trifluoroacetic acid and dichloromethane (1:1) and stirred for 3h at room temperature. The reaction mixture was filtered, washed with dichloromethane and dried to afford 77,1 mg (93%) of 3,3'-((4,4'-(carbonylbis(azanediyl))bis(benzoyl))bis(azanediyl)) dibenzoic acid (**3.14{4.1}**) as a fine white powder.

Spectroscopic data:

IR (KBr) $\nu(\text{cm}^{-1})$: 3429.39, 3277.62, 2575.58, 1692.00, 1665.38, 1645.87, 1592.34, 1533.16, 1508.98, 1438.58, 1413.68, 1311.92, 1239.29, 1127.36, 848.90, 763.45, 754.27, 678.03.

$^1\text{H-NMR}$ (400 MHz, d_6 -DMSO) δ (ppm): 12.97 (s, 2H), 10.29 (s, 2H), 9.18 (s, 2H), 8.54 – 8.36 (m, 2H), 8.08 – 8.02 (m, 2H), 7.98 (d, J = 8.8 Hz, 4H), 7.67 (dt, J = 7.8, 1.2 Hz, 2H), 7.63 (d, J = 8.8 Hz, 4H), 7.47 (t, J = 7.9 Hz, 2H).

$^{13}\text{C-NMR}$ (100 MHz, d_6 -DMSO) δ (ppm): 167.24(C17), 165.07(C8), 152.11(C2), 142.76(C4), 139.60(C11), 131.17(C15), 128.88(C6), 128.84(C13), 127.72(C7), 124.37(C12), 124.23(C14), 121.08(C16), 117.37(C5).

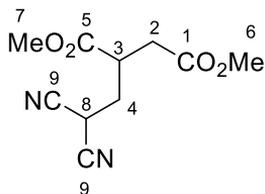
7.4.33 Synthesis of dimethyl 2-methylenesuccinate (4.51)

Itaconic acid (5.02 g, 38.5 mmol) was dissolved in anhydrous MeOH (17.5 mL) and H₂SO₄ (1.5 mL) was added to the solution under Ar atmosphere, and then, the reaction was brought to reflux for 21 h. After that, the crude was diluted with H₂O and extracted with AcOEt (3 x 15 mL), the organic layer was dried over MgSO₄, filtered and evaporated under reduced pressure. The residue was purified by flash chromatographic column (cy:AcOEt, 80:20) to afford dimethyl 2-methylenesuccinate (2.94 g, 18.6 mmol, 48% yield) as a colorless oil.²¹

Spectroscopic data:

¹H NMR (CDCl₃, 400 MHz): □ 6.18 (s, 1H, H-C4), 5.60 (s, 1H, H-C4), 3.66-3.61 (m, 3H, H-C6), 3.59-3.54 (m, 3H, H-C7), 3.21 (s, 2H, H-C2).

7.4.34 Synthesis of dimethyl 2-(2,2-dicyanoethyl)succinate (4.52)



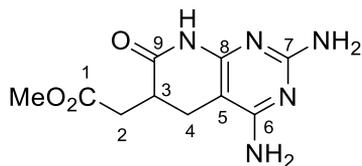
4.52

NaH (0.1320 g, 5.5 mmol) was dissolved in anhydrous THF (5 ml), the mixture was stirred for 10 minutes at room temperature. Then, malononitrile (0.2076 g, 3.14 mmol) was dissolved in anh. THF (5 mL) and added to the mixture that was stirred for 20 minutes at room temperature. After that, diester **4.51** (0.5190 g, 3.28 mmol) was dissolved in anh. THF (5 mL) and added to the mixture that was stirred for 2 hours at 40 °C. After this period, the crude was diluted with H₂O, acidified with conc. HOAc and extracted with DCM (3 x 15 mL). The organic layer was dried over MgSO₄, filtered and evaporated under reduced pressure to afford 2-(2,2-dicyanoethyl)succinate (0.250 g, 1.1 mmol, 34% yield) as a yellowish oil.

Spectroscopic data:

¹H NMR (DMSO-*d*₆, 400 MHz): □ 3.64 (s, 3H, H-C7), 3.60 (s, 3H, H-C6), 2.91 (tt, *J* = 8.1, 6.0 Hz, 1H, H-C3), 2.69 (dd, *J* = 6.9, 3.2 Hz, 2H, H-C2), 2.36-2.24 (m, 2H, H-C4).

7.4.35 Synthesis of methyl 2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetate (**4.53**)

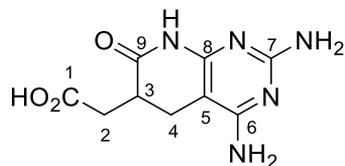
**4.53**

Diester **4.51** (1.62 g, 10.2 mmol), malononitrile (0.34 g, 5.2 mmol) and guanidine carbonate (0.92 g, 5.1 mmol) were added to a microwave vial with 20 mL of anhydrous MeOH. The mixture was heated at 140 °C under microwave radiation for 15 min. The resultant solid was collected by filtration, washed with H₂O, EtOH and Et₂O and oven dried to afford 2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetate (1.32 g, 5.3 mmol, 51% yield) (**4.53**) as a yellowish solid.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 10.25 (s, 1H, N-H lact.), 6.16 (s, 2H, H-NH₂), 5.83 (s, 2H, H-NH₂), 3.62 (s, 3H, H-C10), 2.90 – 2.68 (m, 3H, H-C4, H-C2, H-C3), 2.49 – 2.42 (m, 1H, H-C2), 2.26 (dd, *J* = 15.3, 13.0 Hz, 1H, H-C4).

7.4.36 Synthesis of 2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetic acid (4.55)



4.55

Pirido **4.53** (0.502 g, 2.0 mmol) was dissolved in NaOH 0.5 M (7 mL) and the solution was stirred for 21 hours. The crude was acidified with conc. HCl, then, the mixture was filtered and the resulting solid was washed with H₂O and oven dried to afford 2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetic acid (0.2980 g, 1.26 mmol, 63% yield) (**4.55**) as a white solid.

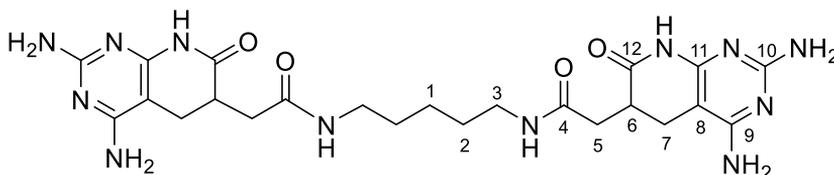
Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 10.74 (s, 1H, N-H lact.), 7.75 (s, 2H, NH₂), 7.52 (s, 2H, NH₂), 2.87 (ddt, *J* = 30.2, 14.9, 7.0 Hz, 3H), 2.71 (dd, *J* = 16.9, 4.9 Hz, 1H, H-C2, H-C4, H-C3), 2.43 (dd, *J* = 17.0, 6.6 Hz, 1H, H-C2), 2.35 (dd, *J* = 14.8, 12.9 Hz, 1H, H-C4).

7.4.37 Method 6: Amide bond procedure for ligands 5.1

Acid **4.55**, diamine, EDC·HCl, DMAP and K₂CO₃ were dissolved in anhydrous DMF (2.5 mL) and the mixture was stirred for 24 hours at room temperature. After that, DMF was evaporated under reduced pressure to obtain a solid which was washed with HCl 0.5 M to afford ligands by filtration.

7.4.38 Synthesis of *N,N'*-(pentane-1,5-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide)



5.1.1

Acid **4.55** (116 mg, 0.5 mmol), cadaverine dihydrochloride (44 mg, were reacted using method 6. *N,N'*-(pentane-1,5-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (60 mg, 0.11 mmol, 44% yield) (**5.1.1**) was obtained by filtration as a white solid.

Spectroscopic data:

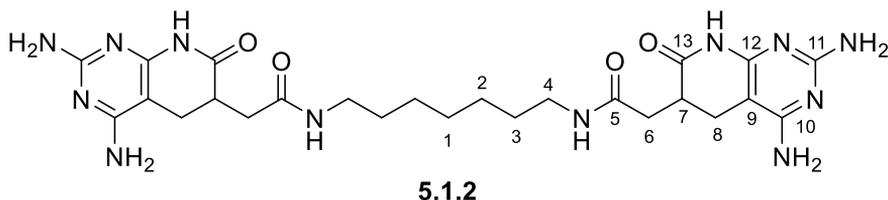
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.59 (s, 2H, N-H lact.), 7.91 (t, *J* = 5.5 Hz, 2H, N-H amide), 7.50 (s, 4H, NH₂), 7.27 (s, 4H, NH₂), 3.13 – 2.95 (m, 4H, H-C3), 2.89 (dtd, *J* = 12.3, 7.6, 4.5 Hz, 2H, H-C6), 2.73 (dd, *J* = 15.6, 7.0 Hz, 2H, H-C7), 2.60 (dd, *J* = 15.4, 4.5 Hz, 2H, H-C5), 2.25 (ddd, *J* = 22.5, 15.5, 9.9 Hz, 4H, H-C5, H-C7), 1.40 (p, *J* = 7.2 Hz, 4H, H-C2), 1.26 (qd, *J* = 8.7, 7.3, 2.5 Hz, 2H, H-C1).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 172.8 (C12), 169.7 (C4), 83.5 (C11), 38.6 (C3), 36.2 (C6), 35.2 (C5), 28.8 (C2), 23.9 (C1), 22.2 (C7).

IR (KBr) ν_{\max} (cm⁻¹): 3361, 2932, 1648, 1569, 1495, 1462, 1395, 1329, 1274, 1226, 1164, 1088, 762, 600.

HRMS (ESI) *m/z* (%): calculated for C₂₃H₃₂N₁₂O₄[M+1]⁺: 540.2669; found [M+1]⁺: 541.2737.

7.4.39 Synthesis of *N,N'*-(heptane-1,7-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (5.1.2)



Acid **4.55** (119 mg, 0.5 mmol), 1,7-heptandiamine (32 mg, 0.25 mmol), were reacted using method 6. *N,N'*-(heptane-1,7-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (80 mg, 0.14 mmol, 57% yield) (**5.1.2**) was obtained by filtration as a white solid.

Spectroscopic data:

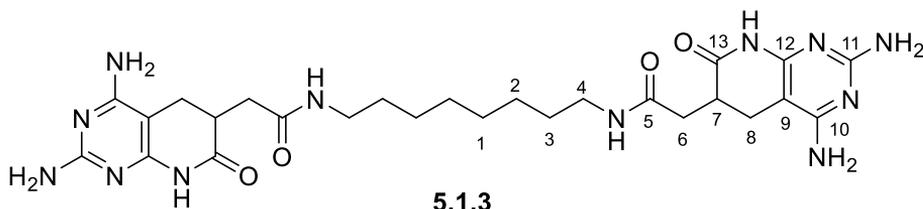
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.73 (s, 2H, N-H lact.), 7.91 (t, *J* = 5.6 Hz, 4H, N-H amide), 7.83 (s, 4H, NH₂), 7.62 (s, 4H, NH₂), 3.04 (m, 4H, H-C4), 2.90 (m, 2H, H-C7), 2.76 (dd, *J* = 15.4, 6.9 Hz, 2H, H-C8), 2.60 (dd, *J* = 15.5, 4.5 Hz, 2H, H-C6), 2.35 – 2.20 (m, 4H, H-C6, H-C8), 1.44 – 1.34 (m, 4H, H-C3), 1.30 – 1.21 (m, 6H, H-C1, H-C2).

¹³C NMR (100 MHz, dmsO) δ 173.3 (C13), 170.0 (C5), 153.9 (C11), 83.9 (C12), 39.1 (C4), 36.5 (C7), 35.5 (C6), 29.5 (C3), 29.0 (C1), 26.9 (C2), 22.4 (C8).

IR (KBr) ν_{\max} (cm⁻¹): 3383, 2927, 2853, 1642, 1497, 1387, 1318, 1270, 1226, 1162, 1086, 818, 761, 534.

HRMS (ESI) *m/z* (%): calculated for C₂₅H₃₆N₁₂O₄[*M*+1]⁺: 568.2982; found [*M*+1]⁺: 569.3058.

7.4.40 Synthesis of *N,N'*-(octane-1,8-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (5.1.3)



Acid **4.55** (119 mg, 0.5 mmol), 1,9-octandiamine (36 mg, 0.25 mmol), were reacted using method 6. *N,N'*-(propane-1,3-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (76 mg, 0.13 mmol, 52% yield) (**5.1.3**) was obtained by filtration as a white solid.

Spectroscopic data:

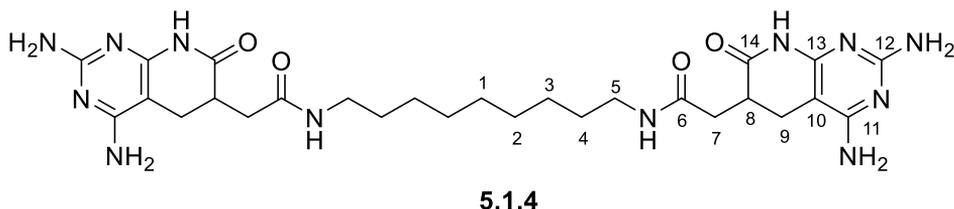
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.13 (s, 2H, NH lact.), 7.81 (dt, *J* = 8.0, 3.8 Hz, 2H, NH amide), 6.13 (s, 4H, NH₂), 5.81 (s, 4H, NH₂), 3.03 (m, *J* = 12.7, 6.6 Hz, 4H, H-C4), 2.81 (m, *J* = 7.1 Hz, 2H, H-C7), 2.71 – 2.52 (m, 4H, H-C6, H-C8), 2.23 – 2.08 (m, 4H, H-C6, H-C8), 1.38 (m, 4H, H-C3), 1.24 (m, 8H, H-C1, H-C2).

¹³C NMR (100 MHz, dmsO) δ 173.8 (C13), 170.4 (C5), 84.1 (C12), 39.1 (C4), 37.3 (C7), 36.1 (C6), 29.5 (C3), 29.2 (C2), 26.9 (C1), 23.5 (C8).

IR (KBr) ν_{\max} (cm⁻¹): 3372, 3192, 2925, 1643, 1598, 1569, 1506, 1458, 1387, 1277, 1225, 1093, 1015, 815, 760, 678, 536, 436.

HRMS (ESI) *m/z* (%): calculated for C₂₆H₃₈N₁₂O₄[M+1]⁺: 582.3139 found [M+1]⁺: 583.3211.

7.4.41 Synthesis of *N,N*-(nonane-1,9-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (5.1.4)



Acid **4.55** (116 mg, 0.5 mmol), 1,9-nondiamine (40 mg, 0.25 mmol), were reacted using method 6. *N,N*-(nonane-1,9-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (82 mg, 0.14 mmol, 55% yield) (**5.1.4**) was obtained by filtration as a white solid.

Spectroscopic data:

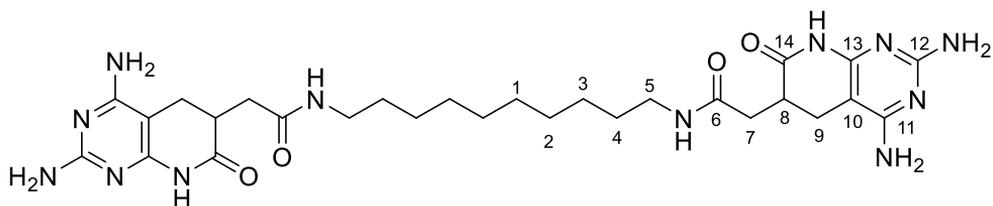
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.61 (s, 2H, NH lact.), 7.87 (t, *J* = 5.5 Hz, 2H, NH amide), 7.52 (s, 4H, NH₂), 7.32 (s, 4H, NH₂), 3.13 – 2.94 (m, 4H, H-C5), 2.94 – 2.83 (m, 2H, H-C8), 2.78 – 2.69 (m, 2H, H-C9), 2.60 (dd, *J* = 15.5, 4.5 Hz, 2H, H-C7), 2.34 – 2.16 (m, 4H, H-C7, H-C9), 1.42 – 1.32 (m, 4H, H-C4), 1.24 (s, 10H, H-C1, H-C2, H-C3).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 173.3 (C14), 170.1 (C6), 83.9 (C13), 39.1 (C5), 36.6 (C8), 35.6 (C7), 29.6 (C4), 29.4 (C3), 29.2 (C2), 26.9 (C1), 22.6 (C9).

IR (KBr) ν_{\max} (cm⁻¹): 3338, 2925, 2852, 1646, 1565, 1494, 1465, 1389, 1320, 1269, 1226, 1165, 1087, 818, 762, 668, 535.

HRMS (ESI) *m/z* (%): calculated for C₂₇H₄₀N₁₂O₄[M+1]⁺: 596.6970; found [M+1]⁺: 597.3365.

7.4.42 Synthesis of *N,N'*-(decane-1,10-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide)



5.1.5

Acid **4.55** (119 mg, 0.5 mmol), 1,9-decandiamine (43 mg, 0.25 mmol), were reacted using method 6. *N,N'*-(nonane-1,9-diyl)bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (75 mg, 0.12 mmol, 50% yield) (**5.1.5**) by filtration as a grey solid.

Spectroscopic data:

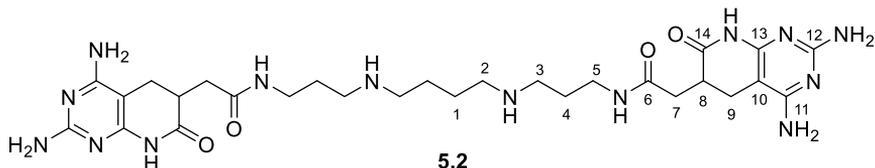
¹H NMR (400 MHz, DMSO-*d*₆) δ 10.53 (s, 2H, NH lact.), 7.86 (t, *J* = 5.6 Hz, 2H, NH amide), 7.32 (s, 4H, NH₂), 7.08 (s, 4H, NH₂), 3.12 – 2.92 (m, 4H, H-C5), 2.86 (ddt, *J* = 11.9, 7.7, 4.4 Hz, 2H, H-C8), 2.71 (dd, *J* = 15.5, 7.1 Hz, 2H, H-C9), 2.59 (dd, *J* = 15.4, 4.5 Hz, 2H, H-C7), 2.23 (ddd, *J* = 21.4, 15.5, 10.0 Hz, 4H, H-C7, H-C9), 1.37 (t, *J* = 6.8 Hz, 4H, H-C4), 1.23 (s, 12H, H-C1, H-C2, H-C3).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 173.0 (C14), 169.7 (C6), 83.5 (C13), 36.3 (C5), 35.2 (C8), 29.1 (C7), 29.0 (C3), 28.8 (C2), 26.5 (C1), 22.3 (C9).

IR (KBr) ν_{\max} (cm⁻¹): 3382, 2923, 2851, 1641, 1566, 1461, 1387, 1320, 1271, 1224, 1161, 1087, 816, 762, 534.

HRMS (ESI) *m/z* (%): calculated for C₂₈H₄₂N₁₂O₄[M+1]⁺: 610.3452; found [M+1]⁺: 611.3520.

7.4.43 Synthesis of *N,N'*-((butane-1,4-diylbis(azanediyl))bis(propane-3,1-diyl))bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (5.2)



In a 5 mL microwave vial, **4.55** (167 mg, 0.7 mmol), spermine (70 mg, 0.35 mmol), EDC·HCl (406 mg, 2.10 mmol) and DMAP (490 mg, 4 mmol) were dissolved in anhydrous DMF (3.5 mL) and the reaction was carried out under microwave irradiation during 16 h at 180 °C. After that, the solvent was removed under reduced pressure, and the crude was resuspended in Et₂O. The resulting precipitate was filtered and the solid was washed with small portions of MeOH and Et₂O and oven dried affording *N,N'*-((butane-1,4-diylbis(azanediyl))bis(propane-3,1-diyl))bis(2-(2,4-diamino-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-6-yl)acetamide) (85 mg, 0.13 mmol, 38% yield) (**5.2**) as a black solid.

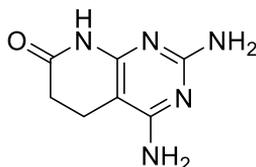
Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 10.11 (s, 2H, NH lact.), 8.06 (t, *J* = 5.8 Hz, 1H, NH amide), 7.83 (t, *J* = 5.4 Hz, 1H, N'H amide), 6.12 (s, 4H, NH₂), 5.79 (s, 4H, NH₂), 3.20 – 3.11 (m, 4H, H3/H2), 3.10 – 3.00 (m, 4H, H1/H2/H3), 2.99 – 2.88 (m, 4H, H5), 2.87 – 2.75 (m, 4H, H8/), 2.62 – 2.53 (m, 4H, H7/H9), 2.27 – 2.07 (m, 4H, H7/H9), 1.96 – 1.68 (m, 4H, H4).

¹³C NMR (100 MHz, DMSO-*d*₆) δ 173.6 (C14), 170.7 (C6), 170.1 (C12), 169.8 (C13), 161.7 (C11), 83.6 (C10), 54.6 (C2), 36.6 (C8), 35.5 (C7), 35.1 (C5), 33.4 (C1), 24.5 (C4), 23.0 (C9).

IR (KBr) ν_{\max} (cm⁻¹): 3412, 3212.9, 2934.6, 1636.3, 1568, 1460, 1387, 761.

7.4.44 Synthesis of 2,4-diamino-5,8-dihydropyrido[2,3-*d*]pyrimidin-7(6*H*)-one (4.40)



4.40

Methyl acrylate (2.02 g, 23.4 mmol), malononitrile (0.44 g, 7.1 mmol) and guanidine carbonate (0.82 g, 4.5 mmol) were added to a microwave vial with 20 mL of anhydrous MeOH. The mixture was heated at 140 °C under microwave radiation for 15 min. The resultant solid was collected by filtration, washed with H₂O, EtOH and Et₂O and oven dried to afford 2,4-diamino-5,8-dihydropyrido[2,3-*d*]pyrimidin-7(6*H*)-one (0.98 g, 5.5 mmol, 25% yield) (**4.40**) as a yellowish solid.²²

Spectroscopic data:

¹H-NMR (400 MHz, DMSO-*d*₆) δ (ppm): 10.30 (s, 1H, H-N7), 6.17 (s, 2H, NH₂-C2), 5.88 (s, 2H, NH₂-C4), 2.63 – 2.29 (m, 4H, H-C5&C6).

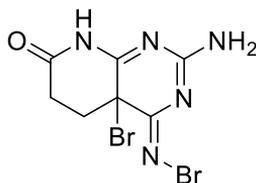
¹³C-NMR (100 MHz, DMSO-*d*₆) δ(ppm): 172.2 (C7), 161.8 (C2), 161.6 (C8a), 156.5 (C4), 83.8 (C4a), 30.5 (C6), 17.0 (C5).

IR (KBr) (cm⁻¹): 3416, 3337, 3185, 1624, 1575, 1442, 1378, 1322, 1289, 1216, 835, 783.

Anal. Calcd. for C₇H₉N₅O: C, 46.92; H, 5.06; N, 39.09. Found: C, 46.81; H, 4.99; N, 39.29.

MS (EI) *m/z*: 180.1 [10%, M+1], 179.1 [100%, M+], 178.1 [59%], 159.9 [10%], 150.1 [23%], 110.1 [11%].

7.4.45 Synthesis of 2-amino-4a-bromo-4-(bromoimino)-4,5,6,8-tetrahydropyrido[2,3-d]pyrimidin-7(4aH)-one (4.41)



4.41

Bromine (610 μL , 11.2mmol) was added to a suspension of **4.40** (1.00 g, 5,6 mmol) in 100 mL of acetic acid. The resulting mixture was stirred 3h at room temperature. Then, water was added, and the formed precipitate was collected by filtration and washed thoroughly with water, ethanol and ether. The product was oven dried providing the title compound as a yellow solid (1.52 g, 81%) (**4.41**).

Spectroscopic data:

$^1\text{H-NMR}$ (400 MHz, $\text{DMSO-}d_6$) δ (ppm): 11.63 (s, 1H, H-N7), 8.29 (s, 1H, $\text{NH}_2\text{-C2}$), 7.89 (s, 1H, $\text{NH}_2\text{-C2}$), 2.91 – 2.75 (m, 1H, H-C6), 2.72 – 2.59 (m, 2H, H-C6&C5), 2.54 – 2.41 (m, 1H, H-C5).

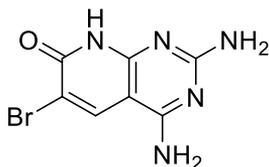
$^{13}\text{C-NMR}$ (100 MHz, $\text{DMSO-}d_6$) δ (ppm): 170.8 (C7), 168.5, 167.4, 161.0, 43.6 (C4a), 30.5 (C6), 28.9 (C5).

IR(KBr) (cm^{-1}): 3311, 3190, 1703, 1644, 1521, 1392, 1348, 1295, 1268, 1220, 1045, 993, 899, 844, 748, 710.

Anal. Calcd. for $\text{C}_7\text{H}_7\text{N}_5\text{OBr}_2$: C, 24.95; H, 2.09; N, 20.78. Found: C, 25.28; H, 2.09; N, 20.72.

HRMS(ESI) m/z (%): calculated for $\text{C}_7\text{H}_8\text{Br}_2\text{N}_5\text{O}[\text{M}]^+$: 335.9090; found $[\text{M}+1]^+$: 335.9091

7.4.46 Synthesis of 2,4-diamino-6-bromopyrido[2,3-d]pyrimidin-7(8H)-one (4.42)



4.42

A solution of **17a** (1.52 g, 4.51 mmol) in DMSO was heated *in vacuo* for 3h at 80°C. Then, water was added, and the formed precipitate was collected by filtration and washed thoroughly with water, ethanol and ether. The product was oven dried providing the title compound as a brownish solid (992.4 mg, 86%, **4.42**).

Spectroscopic data:

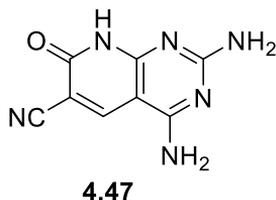
¹H-NMR (400 MHz, DMSO-*d*₆) δ (ppm): 12.21 (s, 1H, H-N7), 8.47 (s, 1H, H-C5), 7.25 (s, 2H, NH₂), 6.75 (s, 2H, NH₂).

¹³C-NMR (100 MHz, DMSO-*d*₆) δ (ppm): 162.9 (C2), 160.8, 159.7, 156.1, 137.1 (C5), 105.3 (C6), 91.3 (C4a).

IR(KBr) (cm⁻¹): 3344, 3184, 2844, 1649, 1589, 1534, 1466, 1274, 1020, 790, 554.

MS (EI) *m/z*: 257.0 [98%, M+2], 255.0 [100%, M+], 179.1 [27%], 178.2 [17%], 176.1 [16%], 133.1 [16%].

HRMS(ESI) *m/z* (%): calculated for C₇H₇BrN₅O[M]⁺: 255.9898; found [M+1]⁺: 255.9898

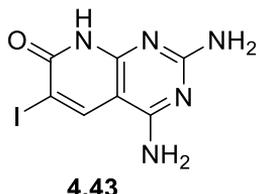
7.4.47 Synthesis of 2,4-diamino-7-oxo-7,8-dihydropyrido[2,3-*d*]pyrimidine-6-carbonitrile (4.47)

Bromo **4.42** (0.21 g, 0.80 mmol) and CuCN (0.32 g, 3.50 mmol) were added to a microwave vial with 4 mL DMF. The mixture was heated at 200 °C under microwave radiation for 12 hours. The solution was filtered and the solid was washed with an EDTA solution. The solid was oven dried to afford 2,4-diamino-7-oxo-7,8-dihydropyrido[2,3-*d*]pyrimidine-6-carbonitrile (**4.47**) (153.6 mg, 0.76 mmol, 95% yield) as a yellowish solid.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 11.89 (s, 1H, NH lact.), 8.59 (s, 1H, H-C2), 7.54 (s, 2H, NH₂), 6.97 (s, 2H, NH₂).

7.4.48 Synthesis of 2,4-diamino-6-iodo-4,8-dihydropyrido[2,3-*d*]pyrimidin-7(4*aH*)-one (4.43)



A solution of compound **4.40** (177 mg, 1.0 mmol) and *N*-iodosuccinimide (467 mg, 2.1 mmol) in DMSO (5 mL) was heated in vacuo for 6 h at 60 °C. Then, water was added, and the formed precipitate was collected by filtration and washed thoroughly with water, ethanol and ether. The product was oven dried providing 2,4-diamino-6-iodopyrido[2,3-*d*]pyrimidin-7(8*H*)-one **4.43** (249 mg, 0.8 mmol, 83% yield) as a yellow solid.

Spectroscopic data:

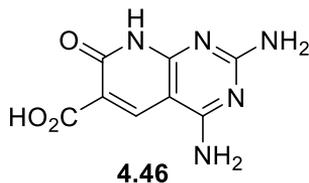
¹H NMR (400 MHz, DMSO-*d*₆) δ 12.07 (s, 1H, H-N7), 8.62 (s, 1H, H-C5), 7.26 (s, 2H, NH₂), 6.76 (s, 2H, NH₂).

¹³C NMR (100 MHz, DMSO-*d*₆) δ (ppm): 163.0 (C2), 160.7, 160.4, 156.9, 143.6 (C5), 92.8 (C4a), 81.2 (C6).

IR(KBr) (cm⁻¹): 3355, 3185, 1625, 1583, 1529, 1460, 1276, 788, 550.

HRMS(ESI) *m/z* (%): calculated for C₇H₇N₅O[M]⁺: 303.9690; found [M+1]⁺: 303.9692

7.4.49 Synthesis of 2,4-diamino-7-oxo-7,8-dihydropyrido[2,3-*d*]pyrimidine-6-carboxylic acid (4.46)

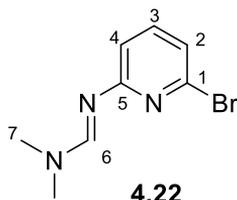


Nitrile **4.47** (158 mg, 0.78 mmol) was dissolved in H₂SO₄ (6.0 mL, 56.4 mmol) and the solution was heated to reflux for 16 h. The resultant crude solution was neutralized with Na₂CO₃ until reach pH ≈ 7-8. The solid was filtered and oven dried to afford 2,4-diamino-7-oxo-7,8-dihydropyrido[2,3-*d*]pyrimidine-6-carboxylic acid as a black solid.

Spectroscopic data:

IR (KBr) ν_{\max} (cm⁻¹): 2915, 1690, 1126, 1017, 874.

7.4.50 Synthesis of *N*-(6-bromopyridin-2-yl)-*N,N*-dimethylformimidamide (4.22)



After sulfonyl chloride (0.83 g, 4.3 mmol) was dissolved in DMF (14 mL) for 5 min, **4.21** (0.50 g, 2.90 mmol) was added at room temperature. The reaction mixture was stirred for the 2 h. After the solvent was removed, K_2CO_3 (4M) solution was added. The mixture was then extracted with AcOEt. The organic layer was dried over $MgSO_4$ and the solvent was removed to afford *N*-(6-bromopyridin-2-yl)-*N,N*-dimethylformimidamide (0.31 g, 1.4 mmol, 47% yield) (**4.22**) as a white solid.

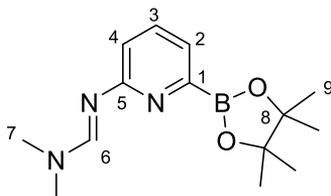
Spectroscopic data:

1H NMR (400 MHz, $DMSO-d_6$) δ 8.36 (s, 1H, NH amidine), 7.49 – 7.44 (m, 1H, H-C3), 7.05 – 7.01 (dd, $J = 7.9, 0.7$ Hz, 1H, H-C2/H-C4), 6.78 (dd, $J = 7.9, 0.7$ Hz, 1H, H-C2/H-C4), 3.08 (s, 3H, H-C7), 2.96 (s, 3H, H-C7).

^{13}C NMR (100 MHz, $DMSO-d_6$) δ 154.9 (C6), 140.3 (C3), 120.0 (C4), 116.7 (C2), 40.2 (C7), 34.3 (C7).

IR (KBr) ν_{max} (cm^{-1}): 2926, 1695, 1631, 1571, 1493, 1443, 1395, 1351, 1153, 1107, 859, 794, 698.

7.4.51 Synthesis of *N,N*-dimethyl-*N*-(6-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)pyridin-2-yl)formimidamide (4.23)



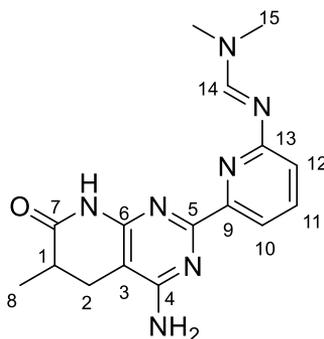
4.23

Pyridine **4.22** (75 mg, 0.2 mmol), B₂Pin₂ (66 mg, 0.26 mmol), PdCl₂(dppf) (15 mg, 0.02 mmol) and KOAc (59 mg, 0.6 mmol) were dissolved in anhydrous THF (2 mL) and the mixture was stirred at 80 °C for 16 h. Upon that, the reaction was allowed to cool at room temperature and EtOAc (10 mL) was added. The resulting mixture was filtered through celite pad and concentrated *in vacuo* affording *N,N*-dimethyl-*N*-(6-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)pyridin-2-yl)formimidamide (**4.23**) as an impure oil and it was used in the next step without further purification.

Spectroscopic data:

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.33 (s, 1H, H-C6), 7.62 – 7.51 (m, 1H, H-C3), 7.23 (dd, *J* = 7.1, 1.2 Hz, 1H), 6.89 (dd, *J* = 7.1, 1.2 Hz, 1H), 3.07 (s, 3H, H-C7), 2.96 (s, 3H, H-C7), 1.21 (s, 12H, H-C9).

7.4.52 Synthesis of *N*-(6-(4-amino-6-methyl-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-2-yl)pyridin-2-yl)-*N,N*-dimethylformimidamide (4.24)



4.24

A mixture of piridopirimidine **4.16.2** (43 mg, 0.17 mmol), pyridinilboronate (94 mg, 0.178 mmol), Na_2CO_3 (60.4 mg, 0.57) dissolved in a minimum amount of water) and glyme was placed in a microwave process vial containing a stirring bar. Then, $\text{Pd}(\text{OAc})_2$ (4 mg, 0.018 mmol) and PPh_3 (7 mg, 0.03 mmol) were added to the mixture, and the vial was sealed and subjected to microwave irradiation for 15 min at 180 °C. After that, the solvent was removed under vacuum and the residue was dissolved in AcOEt and washed with water. The organic layer was dried over MgSO_4 , and the solvent was evaporated under reduced pressure. The residue was purified by flash chromatographic column (cy:AcOEt, 10:90) to afford *N*-(6-(4-amino-6-methyl-7-oxo-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidin-2-yl)pyridin-2-yl)-*N,N*-dimethylformimidamide (13 mg, 0.04 mmol, 24% yield) (**4.24**) as a white solid.

Spectroscopic data:

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 9.74 (s, 1H, NH lact.), 7.66 – 7.52 (m, 4H, H-C14, H-C12, H-C11, H-C10), 2.98 (s, 6H, H-C15), 2.71 (dd, $J = 15.4, 6.7$ Hz, 1H, H-C2), 2.47 (m, $J = 11.3$ Hz, 1H, H-C1), 2.16 (dd, $J = 15.4, 11.1$ Hz, 1H, H-C2), 1.10 (d, $J = 6.9$ Hz, 3H, H-C8).

^{13}C NMR (100 MHz, $\text{DMSO-}d_6$) δ 174.4 (C7), 161.8 (C14), 161.1 (C6/C4), 156.7 (C6/C4), 133.5, 132.5, 132.0 (C10/C11/C12), 131.9

Chemical Synthesis

(C10/C11/C12), 129.3, 129.1 (C10/C11/C12), 83.4 (C3), 36.9 (C15), 35.2 (C1), 25.7 (C2), 16.1 (C8).

IR (KBr) ν_{\max} (cm^{-1}): 3474, 3381, 3345, 3185, 3000, 2951, 2890, 2850, 1717, 1683, 1635, 1620, 1570, 1493, 1263, 849, 749.

7.5 References

1. Chen, C. Z. *et al.* Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* **402**, 1889–1898 (2012).
2. Artimo, P. *et al.* EXPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, 597–603 (2012).
3. González, À. L., Teixidó, J., Borrell, J. I. & Estrada-Tejedor, R. On the applicability of elastic network models for the study of RNA CUG trinucleotide repeat overexpansion. *PLoS One* **11**, 1–20 (2016).
4. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
5. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
6. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
7. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
8. Aytenfisu, A. H., Spasic, A., Grossfield, A., Stern, H. A. & Mathews, D. H. Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J. Chem. Theory Comput.* **13**, 900–915 (2017).
9. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building water models: A different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
10. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
11. Le Grand, S., Götz, A. W. & Walker, R. C. SPFP: Speed

References

- without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380 (2013).
12. Götz, A. W. *et al.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).
 13. Lu, X. J. & Olson, W. K. 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3**, 1213–1227 (2008).
 14. Provenzano, C. *et al.* CRISPR/Cas9-Mediated Deletion of CTG Expansions Recovers Normal Phenotype in Myogenic Cells Derived from Myotonic Dystrophy 1 Patients. *Mol. Ther. - Nucleic Acids* **9**, 337–348 (2017).
 15. Counter, C. M. *et al.* Telomerase activity is restored in human cells by ectopic expression of hTERT (hEST2), the catalytic subunit of telomerase. *Oncogene* **16**, 1217–1222 (1998).
 16. Hollenberg, S. M., Cheng, P. E. I. F. & Weintraub, H. Trans-Activation and Muscle Determination. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8028–8032 (1993).
 17. Hu, D. & Kluger, R. Functional cross-linked hemoglobin bis-tetramers: Geometry and cooperativity. *Biochemistry* **47**, 12551–12561 (2008).
 18. Drewe, W. C. *et al.* Rational design of substituted diarylureas: A scaffold for binding to G-quadruplex motifs. *J. Med. Chem.* **51**, 7751–7767 (2008).
 19. Tagad, H. D. *et al.* Design of pentapeptidic BACE1 inhibitors with carboxylic acid bioisosteres at P1' and P4 positions. *Bioorganic Med. Chem.* **18**, 3175–3186 (2010).
 20. Borrell, J. I. *et al.* An Unequivocal Synthesis of 4-Amino-1,5,6,8-tetrahydropyrido[2,3-d]pyrimidine-2,7-diones and 2-Amino-3,5,6,8-tetrahydropyrido[2,3-d]pyrimidine-4,7-diones. *Collection of Czechoslovak Chemical Communications* vol. 61 901–909 (1996).
 21. Gualandi, A. *et al.* Photocatalytic Radical Alkylation of Electrophilic Olefins by Benzylic and Alkyl Zinc-Sulfinates. *ACS Catal.* **7**, 5357–5362 (2017).

22. Mont, N., Teixidó, J., Borrell, J. I. & Kappe, C. O. A three-component synthesis of pyrido[2,3-d]pyrimidines. *Tetrahedron Lett.* **44**, 5385–5387 (2003).

References