

Development of modified nucleosides for targeting the DNA repair enzyme **SNM1A**



*A thesis submitted to the School of Chemistry, Trinity College Dublin, for the
degree of Doctor of Philosophy*

by Mark Berney B.A. (Mod.).

Based on research carried out under the supervision of Prof. Joanna
McGouran.

May 2022

Trinity College Dublin, The University of Dublin

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work, except where the contributions of others are specifically noted.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiners retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Mark Berney

Mark Berney

Abstract

Certain cancers exhibit upregulation of DNA interstrand crosslink repair pathways, which contributes to resistance to crosslinking chemotherapy drugs and poor prognoses. Inhibition of enzymes implicated in interstrand crosslink repair is therefore a promising strategy for improving the efficacy of cancer treatment. One such target enzyme is SNM1A, a zinc-dependent 5'-3' exonuclease. Human cells deficient in SNM1A, as well as whole mice deficient in SNM1A, show increased sensitivity to crosslinking chemotherapy agents. Previous studies have demonstrated the feasibility of inhibiting SNM1A using modified nucleosides appended with zinc-binding groups. There remains a need however to develop improved inhibitors for this enzyme.

Furthermore, due to the complexity of interstrand crosslink repair, and the additional functions of SNM1A in cell cycle checkpoint pathways, the biological role of this enzyme has yet to be fully elucidated. The development of chemical probes to study SNM1A in a cellular environment would facilitate further research in this area, allowing the full potential of this enzyme as a drug target to be explored.

The work presented herein aimed to further the development of modified nucleosides for binding to SNM1A. Zinc-binding groups previously untried against SNM1A, including sulfonamide, 2-thiophene carboxamide, oxime, and hydroxylamine moieties, were incorporated at the 3'-position of nucleoside derivatives. Biochemical testing of these compounds showed that sulfonamide and hydroxylamine zinc-binding groups are effective. Modified nucleosides were also prepared bearing a hydroxamic acid and a hydrazide at the 3'-position, zinc-binding groups previously trialled at the 5'-position of SNM1A inhibitors.

The synthesis of a family of squaramide- and thiosquaramide-bearing nucleoside derivatives and their evaluation as SNM1A inhibitors is also described in this work. Initial screening identified *N*-hydroxysquaramide, squaric acid, and thiosquaramide moieties as promising zinc-binding groups for targeting SNM1A. Nucleoside derivatives bearing squaramides at the 3'-position were found to be more effective inhibitors than those bearing squaramides at the 5'-position. The reverse trend was observed in the case of thiosquaramides. Quantitative IC₅₀ determination showed that a thymidine derivative bearing a 5'-thiosquaramide was the most potent inhibitor, followed by a thymidine derivative bearing a 3'-squaric acid. UV-vis titrations were performed to study the binding interactions of the (thio)squaramides with zinc ions, allowing the order of potency of the SNM1A inhibitors to be rationalised. The membrane permeability of the active inhibitors was investigated, with several compounds showing promise for future *in vivo* applications.

The development of more effective SNM1A inhibitors through exploiting interactions with the phosphate-binding pocket adjacent to the enzyme's active site, in addition to targeting the catalytic zinc ions, was also investigated in this work. A series of nucleoside derivatives bearing phosphate moieties at the 5'-position, as well as zinc-binding groups at the 3'-position, were prepared and tested. This showed that incorporation of a 5'-phosphate dramatically increased the potency of the inhibitors.

Overall, the results presented in this work represent significant progress in the development of nucleoside-based SNM1A inhibitors. Insights generated will inform the design of next-generation SNM1A inhibitors, and elaboration of inhibitor structures will also enable the development of chemical probes to study SNM1A *in vivo*.

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Joanna McGouran, for giving me the opportunity to undertake a PhD in her lab, and for her guidance throughout the past four years. I'd also like to acknowledge the Irish Research Council for their financial support.

I owe a debt of gratitude to the technical staff in the School of Chemistry for their help. I'm grateful to Dr. John O'Brien for your huge knowledge of NMR and for loving what you do, and to Dr. Manuel R  ther for support with NMR and for knowing how to fix everything. Thank you as well to Dr. Gary Hessman and Dr. Martin Feeney for support with mass spectrometry.

I have been lucky to work with some wonderful people in the McGouran Group, and I want to express my appreciation for all current and former members. William, thank you for being a patient teacher and the person I went to with my stupid questions. Eva, thanks for being a great colleague, lab alpha, friend, and sounding board for science ideas, and thank you Neil for your humour and for being a constant calm presence in the lab. Susie and Ellen, thank you for your friendship and for always bringing fun and laughter to the lab even during difficult moments. Fergus, thanks for all your advice on UV-vis titrations, and for sharing your curiosity and enthusiasm. Thank you Werner for being a friend in the lab, and for inviting us all to Switzerland and being a great host and tour guide. I'd also like to thank my undergraduate students Manav and David - I learnt as much from you as you did from me. Daryl, Sean, Connor, and Valerio, although our time together in the lab was relatively brief or was interrupted by lockdowns and pod systems, thanks for being part of the journey of the past few years, and best of luck in the future.

I'm also grateful to the wider School of Chemistry community, particularly everyone in the Scanlan Group with whom we shared our lab space and many fun times, and my good friend Gary. Thank you as well to Oxana and Louise, my mentors during my first undergraduate research experiences who helped put me on the path to doing a PhD.

I'd also like to thank the Fulbright Commission for their financial support of my visit to the University of Chicago, and Prof. Bryan Dickinson for hosting me there. I'm grateful to all the members of the Dickinson Group for welcoming me and being great colleagues during my stay. Thank you especially to Yang for being excellent at everything you do and teaching me how to do some of it.

Finally, I'm grateful to all the people outside the world of chemistry who supported me throughout my PhD. In particular Alex, Peter, Gillian, Liagh, Billy, Sergio, and my family.

Publications

Berney, M. and McGouran, J. "Methods for detection of cytosine and thymine modifications in DNA", *Nature Reviews Chemistry* **2**, 332-348 (2018).

Berney, M., Doherty, W., Jauslin, W. T., T Manoj, M., Dürr, E.-M., and McGouran, J. "Synthesis and evaluation of squaramide and thiosquaramide inhibitors of the DNA repair enzyme SNM1A", *Bioorganic and Medicinal Chemistry* **46**, 116369 (2021).

- This work is presented in Chapter 3.

Berney M., T Manoj, M. and McGouran, J. "5'-Phosphorylation increases the efficacy of nucleoside inhibitors of the DNA repair enzyme SNM1A" *ChemMedChem* **17**, e202100603 (2022)

- This work is presented in Chapters 2 and 4.

Abbreviations

[H]	reduction
[P]	protecting group
53BP1	p53 binding protein 1
ABP	activity-based probe
ACE	angiotensin-1 converting enzyme
AfBP	affinity-based probe
AMP	adenosine monophosphate
AP1	activator protein 1
APC	anaphase promoting complex
APCI	atmospheric pressure chemical ionisation
APE1	apurinic/aprimidinic (AP) endonuclease 1
app	apparent
Ar	aromatic
ATM	ataxia-telangiectasia mutated protein kinase
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and Rad3-related protein
AZT	azidothymidine
BER	base excision repair
BHQ	black hole quencher
br s	broad singlet
BSA	bovine serum albumin
C	Celsius
<i>ca.</i>	<i>circa</i>
calc.	calculated
cAMP	cyclic adenosine monophosphate
cGAMP	2',3'-cyclic guanosine monophosphate-adenosine monophosphate
cGMP	cyclic guanosine monophosphate
CHCA	α -cyano-4-hydroxycinnamic acid
Chfr	checkpoint with forkhead and ring finger domains
CHK1	checkpoint kinase 1

cm	centimetres
Conc	concentration
CPSF73	cleavage and polyadenylation specificity factor 73
CSB	Cockayne syndrome B
Cy3	cyanine3
D	aspartate
d	doublet
DCM	dichloromethane
DCTB	<i>trans</i> -2-[3-(4- <i>tert</i> -butylphenyl)-2-methyl-2-propenylidene]malononitrile
dd	doublet of doublet
ddd	doublet of doublet of doublets
decomp.	decomposed
DHB	2,5-dihydroxybenzoic acid
DIAD	diisopropyl azodicarboxylate
DIPEA	<i>N,N</i> -diisopropylethylamine
DMF	<i>N,N</i> -dimethylformamide
DMSO	dimethyl sulfoxide
DMT	4,4'-dimethoxytrityl
DMTCl	4,4'-dimethoxytrityl chloride
DNA	deoxyribonucleic acid
dq	doublet of quartets
Dr.	Doctor
dt	doublet of triplets
DTT	1,4-dithiothreitol
<i>E. coli</i>	Escherichia coli
E1cb	elimination, unimolecular, conjugate base
EDCI	<i>N</i> -(3-dimethylaminopropyl)- <i>N'</i> -ethylcarbodiimide
EDTA	Ethylenediaminetetraacetic acid
eIF4E	eukaryotic translation initiation factor 4E
ELAC2	elaC ribonuclease Z 2
ESI	electrospray ionisation

EXO1	exonuclease 1
FAM	fluorescein
FAN1	Fanconi-associated nuclease 1
FANCD1	Fanconi anaemia complementation group 1
FANCD2	Fanconi anaemia complementation group D2
FDA	United States Food and Drug Administration
FemX _{WV}	UDP-N-acetylmuramoylpentapeptide-lysine N(6)-alanyltransferase from <i>Weissella viridescens</i>
FEN1	flap endonuclease 1
Fmoc	fluorenylmethyloxycarbonyl
fmol	femtomoles
g	grams
G0 phase	resting phase
G1 phase	growth 1 phase <i>or</i> gap 1 phase
GDP	guanosine diphosphate
h	hour
HDAC	histone deacetylase
HEPES	<i>N</i> -(2-hydroxyethyl)piperazine- <i>N'</i> -(2-ethanesulfonic acid)
HIV-1	human immunodeficiency virus type 1
HOAt	1-hydroxy-7-azabenzotriazole
HPA	3-hydroxypicolinic acid
HR	homologous recombination
HRMS	high resolution mass spectrometry
Hz	hertz
IC ₅₀	half-maximal inhibitory concentration
ICL	interstrand crosslink
IR	infrared
IRES	internal ribosomal entry site
J	coupling constant
K	binding constant
K	lysine

LC-MS	liquid chromatography–mass spectrometry
LG	leaving group
m	multiplet
M	molar
<i>m/z</i>	mass-to-charge ratio
MALDI	matrix-assisted laser desorption ionisation
MBL	metallo- β -lactamase
<i>m</i> CPBA	<i>m</i> -chloroperoxybenzoic acid
MDO1	macro domain protein 1
mg	milligrams
MHz	megahertz
mL	millilitres
mM	millimolar
mmol	millimoles
MMP	matrix metalloproteinase
MMR	mismatch repair
mp	melting point
mRNA	messenger RNA
MUS81-EME1	MUS81 structure-specific endonuclease subunit-essential meiotic structure-specific endonuclease 1
MutL α	DNA mismatch repair protein MutL α
n.d.	not determined
nA	nanoamps
NDM	New Delhi metallo- β -lactamase
NER	nucleotide excision repair
NF κ B	nuclear factor κ B
NHEJ	non-homologous end-joining
NIH	National Institutes of Health
nm	nanometres
nM	nanomolar
NMR	nuclear magnetic resonance

nt	nucleotides
p.	page
p21	cyclin-dependent kinase inhibitor 1
p53	tumour protein P53
PAGE	polyacrylamide gel electrophoresis
PAMPA	parallel artificial membrane permeability assay
PBS	phosphate-buffered saline
PBZ	poly(ADP-ribose)-binding zinc finger
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PDB	Protein Data Bank
P_e	effective permeability
PEG	polyethylene glycol
PET	positron emission tomography
PIAS1	protein inhibitor of activated STAT 1
PIP box	PCNA-interacting protein box
pp.	pages
ppm	parts per million
Prof.	Professor
PSO2	sensitive to psoralen 2
q	quartet
qC	quaternary carbon
qPCR	quantitative polymerase chain reaction
R	arginine
Rad18	E3 ubiquitin-protein ligase RAD18
RAD51	DNA repair protein RAD51
R_f	retention factor
RNA	ribonucleic acid
RNAi	RNA interference
RPA	replication protein A
rt	room temperature

RT-PCR	reverse transcription-polymerase chain reaction
s	singlet
S	serine
S phase	synthesis phase
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
siRNA	small interfering ribonucleic acid
SLX1-SLX4	SLX1-SLX4 structure-specific endonuclease complex
SNM1	sensitive to nitrogen mustard 1
Sq	squaryl
ssDNA	single-strand deoxyribonucleic acid
SUMO	small ubiquitin-like modifier
t	triplet
T	thymine or thymidine
T	threonine
T4 PNK	T4 polynucleotide kinase
TBAF	tetrabutylammonium fluoride
TBE	tris-borate-EDTA
TBS	<i>tert</i> -butyldimethylsilyl
TBSCl	<i>tert</i> -butyldimethylsilyl chloride
td	triplet of doublets
TFA	trifluoroacetic acid
Th	thiophene
THF	tetrahydrofuran
TLC	thin-layer chromatography
TLS	translesion synthesis
TOF	time-of-flight
tRNA	transfer RNA
U	uracil
UBZ	ubiquitin-binding zinc finger
UV	ultraviolet

UV-vis	ultraviolet-visible
V	Volts
v/v	volume per volume
w/v	weight per volume
WEE1	Wee1-like protein kinase 1
wt%	weight percent
XPF-ERCC1	XPF-excision repair cross-complementation group 1
Y	tyrosine
ZBG	zinc-binding group
β -CASP	metallo- β -lactamase-associated CPSF Artemis SNM1/PSO2
γ -H2AX	phosphorylated H2A histone family member X
μ L	microlitres
μ m	micrometres
μ M	micromolar
μ mol	micromoles
v_{\max}	maximum absorption

Contents

Declaration	ii
Abstract.....	iii
Acknowledgements	v
Publications	vii
Abbreviations.....	viii
Contents.....	xv
1. Introduction	1
1.1. DNA damage and repair.....	1
1.1.1. Repair of interstrand crosslinks.....	2
1.1.2. The role of SNM1A in ICL repair	2
1.1.2.1 ICL repair after replication fork stalling	4
1.1.2.2. ICL repair during transcription	6
1.1.2.3. ICL repair independent of transcription or replication.....	7
1.2. The role of SNM1A in cell cycle checkpoints	7
1.2.1. G1 checkpoint.....	8
1.2.2. Entry to metaphase checkpoint	9
1.3. SNM1A structure and mechanism of action.....	10
1.4. Inhibition of DNA repair enzymes.....	13
1.4.1. SNM1A as a therapeutic target.....	14
1.4.2. Previously reported inhibitors of SNM1A	16
1.5. Molecular probes for studying enzyme activity	20
1.5.1. The utility of developing probes for SNM1A.....	20
1.5.2. Activity-based protein profiling	20
1.5.3. Probes for targeting metallo- β -lactamases and other metallohydrolases.....	23
1.5.4. Probes for nucleic-acid-processing enzymes	25
1.6. Zinc-binding groups.....	29
1.6.1. Hydroxamic acids and related ZBGs.....	29
1.6.2. Sulfur-containing ZBGs.....	31
1.6.3. Phosphorous-based ZBGs.....	33
1.6.4. Other ZBGs	35
1.7. Squaramides.....	36
1.7.1. General properties and cation binding	36

1.7.2. Squaramides in chemical biology	38
1.7.3. Squaramides as phosphate bioisosteres	40
1.7.4. Thiosquaramides	43
1.8. Objectives of this work – the design of novel inhibitors and probes for SNM1A.....	44
2. Nucleosides bearing ZBGs at the 3'-position inhibit SNM1A.....	49
2.1. Introduction.....	49
2.2. Synthesis.....	51
2.2.1. Protecting group strategies for derivatisation of the uridine 3'-position.....	51
2.2.2. Uridine derivatives bearing hydroxamic acid and hydrazide ZBGs	53
2.2.3. A uridine derivative bearing a 3'-amino group	55
2.2.4. Uridine derivatives bearing oxime and hydroxylamine ZBGs	58
2.2.5. Uridine derivatives bearing sulfonamide and 2-thiophene carboxamide ZBGs	60
2.3. Biochemical testing	61
2.3.1. Gel electrophoresis assay	61
2.4. Conclusion	65
3. Squaramide and thiosquaramide inhibitors of SNM1A.....	68
3.1. Introduction.....	68
3.2. Synthesis.....	71
3.2.1. Thymidine derivatives bearing squaramide ZBGs at the 5'-position	71
3.2.2. Thymidine derivatives bearing squaramide ZBGs at the 3'-position	72
3.2.3. Uridine derivatives bearing squaramide ZBGs at the 3'-position	73
3.2.4. Thymidine derivatives bearing thiosquaramide ZBGs.....	74
3.3. Biochemical testing of squaramides and thiosquaramides as SNM1A inhibitors	76
3.3.1. Gel-electrophoresis-based assays	76
3.3.2. IC ₅₀ determination	81
3.3.3. UV-vis titrations	84
3.3.4. Membrane permeability assay	86
3.4. Conclusion	87
4. 5'-Phosphorylation increases the efficacy of nucleoside inhibitors of SNM1A.....	91
4.1. Introduction.....	91
4.2. Synthesis of phosphorylated uridine derivatives.....	94
4.3. Biochemical testing of phosphorylated uridine derivatives	98
4.4. Synthesis of phosphorylated thymidine derivatives.....	103

4.5. Biochemical testing of thymidine derivatives.....	110
4.6. Conclusion.....	114
5. Conclusion and Future Work.....	118
6. Experimental.....	123
6.1. General experimental.....	123
6.2. Synthetic methods.....	124
6.2.1. Synthetic methods for Chapter 2.....	124
6.2.2. Synthetic methods for Chapter 3.....	151
6.2.3. Synthetic methods for Chapter 4.....	172
6.3. Biochemical evaluation.....	202
6.3.1. Gel electrophoresis assay.....	202
6.3.2. Real-time fluorescence assay.....	203
6.3.3. UV-vis titrations.....	207
6.3.4. Parallel artificial membrane permeability assay (PAMPA).....	209
References.....	211

1. Introduction

1.1. DNA damage and repair

Damage to cellular DNA occurs continuously as a result of both endogenous and exogenous factors. Endogenous DNA lesions may result from hydrolytic depurination,¹ deamination of cytosine² or 5-methylcytosine,³ alkylation by *S*-adenosylmethionine,⁴ damage by reactive oxygen species produced in oxidative respiration, damage by products of lipid peroxidation,⁵ and mismatches introduced during DNA replication,⁶ among other processes. Examples of exogenous causes of DNA lesions include ultraviolet light,⁷ other forms of ionising radiation,⁷ and carcinogenic chemicals such as those found in tobacco smoke and well-done meat.⁸

DNA damage can lead to harmful mutations, and block transcription and replication of DNA. Therefore, sophisticated mechanisms of DNA damage repair have evolved. Due to the diverse range of DNA lesions that can form in cells, multiple distinct repair mechanisms are required. Some forms of DNA damage can be directly repaired by a single protein. For example *O*⁶-methylguanine, a highly carcinogenic lesion, can be demethylated to guanine by *O*⁶-methylguanine methyltransferase.⁹ Most DNA damage repair mechanisms however are complex pathways involving multiple proteins. The main DNA repair pathways in mammalian cells are nucleotide excision repair (NER), base excision repair (BER), non-homologous end-joining (NHEJ), homologous recombination (HR), and mismatch repair (MMR).¹⁰

Progression through the cell cycle is regulated as a function of DNA damage.¹¹ Multiple checkpoints arrest the cell cycle in response to DNA damage, in order to allow time for repair to occur before the cell cycle resumes, or to prevent extensively damaged cells from

replicating. A number of DNA repair enzymes are involved in signal transduction between pathways involved in DNA repair and those comprising cell cycle checkpoints. Mutations in DNA repair enzymes and other checkpoint proteins are often associated with cancer.^{12,13}

1.1.1. Repair of interstrand crosslinks

Interstrand crosslinks (ICLs) are a type of DNA lesion. ICLs covalently link the two opposing strands of DNA to each other and thus prevent the double helix from unwinding, inhibiting the two basic functions of DNA; replication and transcription. Although ICLs are a relatively infrequent form of DNA damage, they are highly cytotoxic. Reactive α,β -unsaturated aldehydes such as acrolein, crotonaldehyde, and malondialdehyde, produced endogenously during lipid peroxidation or found in exogenous sources such as automobile exhaust fumes, are a common cause of ICLs.^{14,15} Reactive ionic or radical species formed upon exposure to radiation can also lead to ICLs.^{16,17} A large number of widely used cancer drugs such as cisplatin, mitomycin C, and melphalan, induce ICLs.^{18,19}

Repair of ICLs is less well understood than the repair of other types of DNA damage, due to the variety of structures that can result from different crosslinking agents, and the difficulty associated with generating model duplex oligonucleotides containing site-specific ICLs.²⁰ It is also challenging to detect and measure ICLs in biological samples.²¹ ICL repair in mammalian cells is a complex process, and elements of several biochemical pathways are involved. These include proteins from nucleotide excision repair (NER), homologous recombination (HR), translesion synthesis (TLS) and the Fanconi anaemia pathway.

1.1.2. The role of SNM1A in ICL repair

SNM1 (sensitive to nitrogen mustard 1), also known as PSO2 (sensitive to psoralen 2), is a nuclease that was first discovered in *S. cerevisiae* in the 1980s.^{22,23} Yeast cells deficient in this gene display hypersensitivity to crosslinking agents, but retain wild-type sensitivity to

other sources of DNA damage such as monofunctional alkylating agents, UV light, or ionising radiation.²⁴ There are five homologues of SNM1 in human cells. These include SNM1A, SNM1B/Apollo, SNM1C/Artemis, CPSF73 and ELAC2.²⁵ CPSF73 and ELAC2 are RNA processing enzymes involved in maturation of mRNAs and mitochondrial tRNAs, respectively.^{26,27} The human SNM1 family, comprising SNM1A, SNM1B and SNM1C, are DNA processing nucleases. SNM1A, and to a lesser extent SNM1B, are implicated in ICL repair pathways. Human SNM1A is the closest functional homologue of yeast SNM1, and ectopic expression of human SNM1A in yeast cells rescues DNA repair deficiencies caused by knockout of the SNM1 gene.²⁸

The human SNM1A gene was first identified in 1995 on chromosome 10 by Nagase *et al.*²⁹ The 5'-untranscribed region of the SNM1A gene contains binding sequences for several known transcription factors, some of which, such as NFκB and AP1, are expressed in response to DNA damage.³⁰ The first characterisation of human SNM1A protein, and identification of the SNM1B and SNM1C genes, was carried out in 2000 by Dronkert *et al.*, who reported that mice deficient in SNM1A showed increased sensitivity to the crosslinking agent mitomycin C.³¹

Human ICL repair consists of a number of overlapping biochemical pathways involving many proteins. In addition to SNM1A and SNM1B, at least four other nucleases; XPF-ERCC1, MUS81-EME1, SLX1-SLX4, and FAN1, have been implicated.³² Unravelling the individual roles of these nucleases, which in some cases may be redundant with each other, is a topic of ongoing research.³³ There are three distinct scenarios in which ICLs can be detected in mammalian cells, which can be used to broadly categorise ICL repair pathways.³⁴ These include recognition of an ICL when it blocks a replication fork, recognition upon stalling of

the transcription apparatus at the site of an ICL, and recognition of an ICL in otherwise normal DNA in the absence of replication or transcription.

1.1.2.1 ICL repair after replication fork stalling

The main ICL repair pathway (Figure 1.1) occurs after replication fork stalling.³⁵ Although the details of this process have not been fully elucidated, it is known that the enzyme SNM1A plays an important role. SNM1A is a 5'-3' exonuclease – an enzyme that digests DNA by hydrolysing the phosphodiester backbone from the 5'-end.^{28,36} The pathway, which most likely occurs after convergence of two replication forks at the site of an ICL,^{32,37} is first switched on as the FANCD2-FANCD1 complex is ubiquitinated and recruits Slx4, which in turn recruits XPF-ERCC1.³⁸ Rad18, an E3 ubiquitin ligase, monoubiquitinates PCNA, which then recruits SNM1A.³⁹ SNM1A contains a PIP box (PCNA-interacting protein box) and a UBZ domain (ubiquitin-binding zinc finger) that mediate interaction with ubiquitinated PCNA. Interactions with PIAS1,^{40,41} a SUMO E3 ligase, and γ -H2AX,^{42,43} a phosphorylated histone, are also implicated in recruitment of SNM1A to sites of DNA damage.

It has been proposed that XPF-ERCC1 makes an incision 5' to the ICL, possibly being stimulated to do so by interaction with RPA.⁴⁴ This forms an entry point for SNM1A, which then hydrolyses one strand of the DNA past the point of the ICL, leaving a single mononucleotide covalently linked by the ICL to the opposing strand.³⁸ The resulting gapped intermediate with a tethered mononucleotide is a substrate for TLS, which synthesises past the point of the lesion using specialised DNA polymerases.^{45,46} The tethered mononucleotide is subsequently removed by the NER pathway (Figure 1.1).³⁸

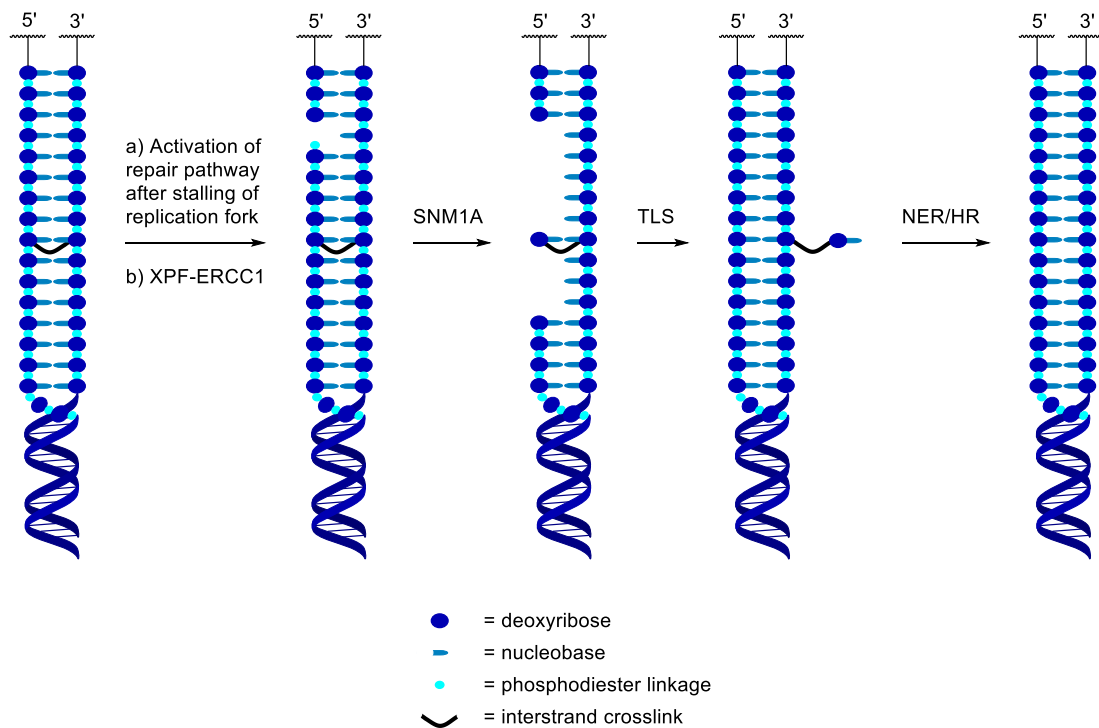


Figure 1.1. Key steps in replication-dependent ICL repair in mammalian cells. Note that for clarity the diagram omits a replication fork or two converging replication forks stalling at the site of an ICL. The TLS and NER pathways repair one chromatid in the final step as shown in the diagram, while the second chromatid is left with a double strand break that is repaired by the HR pathway.^{44,47}

The identity of the endonuclease that generates the incision 5' to the ICL has been called into question however, due to the observed preference of XPF for nicking in the 3'-flap of a stalled replication fork.⁴⁸ Junop and co-workers reported that SNM1A has endonuclease activity *in vitro* and when expressed in yeast cells, suggesting it may play a broader role in ICL repair than previously believed.⁴⁹ This is consistent with the observation that yeast SNM1 also has endonuclease activity,⁵⁰ as does human SNM1C.⁵¹ Although prior studies indicated that SNM1A did not show endonuclease activity *in vitro*,^{36,52} this discrepancy may be due to differences in the design of the oligonucleotide substrates used in these studies, and differences in experimental conditions, as the assays carried out by Junop and co-workers used a substantially higher concentration of SNM1A compared to other works.

Exonuclease activity of SNM1A is significantly higher than endonuclease activity, and the use of oligonucleotide substrates labelled with a fluorophore at the 3'-end therefore may fail to detect endonuclease activity. Junop and co-workers therefore propose that SNM1A could itself form an incision 5' to the ICL through its endonuclease activity, allowing for SNM1A exonuclease activity to then take place in the 5'-3' direction past the ICL. FAN1 is another candidate that may form the incision 5' to the ICL. Like SNM1A, this enzyme is reported to possess endonuclease activity as well as phosphate-dependent 5'-3' exonuclease activity,⁵³⁻⁵⁵ and is recruited to sites of ICL damage by PCNA in a pathway activated by FANCD2.⁵⁶ SNM1A is epistatic with XPF-ERCC1 in response to ICL damage however, suggesting that these enzymes do indeed act in the same pathway.³⁸ Given the variety of different ICL structures, and the importance of repairing these lesions, it is plausible that there could be some redundancy between several enzymes in "unhooking" of ICLs. Studies in yeast cells⁵⁷ and in mouse embryonic fibroblasts⁵⁸ support the hypothesis of redundancy between SNM1A and FAN1.

1.1.2.2. ICL repair during transcription

While replication-dependent ICL repair is the dominant repair pathway during S phase of the cell cycle, ICL repair can also occur during G0/G1 phase due to blockage of transcription. This pathway has been less well studied than replication-dependent repair, which may be due to the lower sensitivity of cells in G0/G1 phase to ICLs. An ICL may be tolerated in G0/G1 phase if it does not prevent transcription of an essential gene, whereas in S phase an ICL in any part of the genome will block replication.⁵⁹ Defects in replication-dependent ICL repair and transcription-dependent ICL repair are additive in terms of the sensitivity of cells to crosslinking agents, indicating that these repair pathways operate independently of each other.⁵⁹

SNM1A is implicated in transcription-dependent ICL repair as well as replication-dependent repair however.³⁴ After stalling at the site of an ICL, RNA polymerase II is strongly bound by CSB, which coordinates transcription-dependent repair.⁶⁰ CSB has been shown to recruit SNM1A to the site of ICLs in human cells and stimulates its exonuclease activity.³⁴ Accumulation of SNM1A at ICL sites is significantly reduced in CSB-deficient cells. Another study found that CSB is also associated with SNM1A in mouse cells, indicating this interaction is conserved across mammalian species.⁶¹

1.1.2.3. ICL repair independent of transcription or replication

ICLs may be recognised independently of transcription or replication due to the structural distortion of the DNA duplex they induce.⁶² This repair is mediated by proteins of the MMR pathway, with the nucleases MutL α and EXO1 proposed to perform ICL unhooking. ICLs which cause the least structural distortion fail to be recognised and repaired by this pathway.⁶² SNM1A-depleted cells show increased sensitivity to crosslinking agents mitomycin C and SJG-136, which target the DNA minor groove and form minimally distorting ICLs. In contrast these cells show only marginally increased sensitivity to nitrogen mustard, which reacts in the major groove to form highly distorting ICLs.³⁸ This indicates that SNM1A likely does not play a role in ICL repair independent of transcription or replication.

1.2. The role of SNM1A in cell cycle checkpoints

Transitions between different phases of the cell cycle are regulated by a series of checkpoints. These checkpoints function to prevent damaged cells from replicating, or to delay replication to allow time for repair to take place. Upon checkpoint activation, cell cycle arrest or apoptosis may occur. SNM1A is known to play a role in these processes. For example, induction of apoptosis upon treatment with the DNA-damaging agent etoposide

was found to be reduced in chicken DT40 cells deficient in SNM1A.⁶³ As well as functioning in ICL repair, SNM1A is involved in at least two cell cycle checkpoints,^{64,65} highlighted in Figure 1.2.

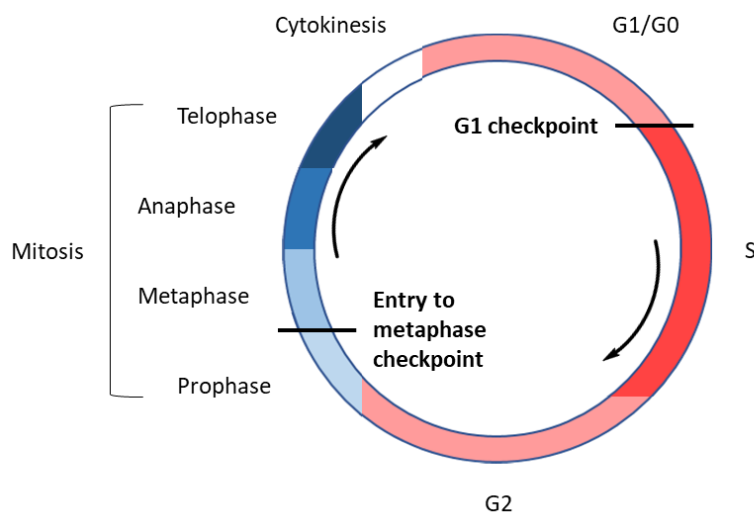


Figure 1.2. A summary of the phases of the cell cycle, including the G1 checkpoint and entry to metaphase checkpoint in which SNM1A plays a role.

1.2.1. G1 checkpoint

The G1 checkpoint controls entry to S phase, preventing initiation of DNA replication in damaged cells. It is the point at which cells commit to entering the cell cycle, or, upon checkpoint activation, enter G0 phase. In short, p53 is phosphorylated by a number of kinases in response to DNA damage, and induces expression of p21.⁶⁶ In turn, p21 inhibits cyclin-dependent kinases that drive progression through the cell cycle.

In human and mouse cells deficient in SNM1A, there is an inadequate G1 checkpoint response upon exposure to ionising radiation.⁶⁴ SNM1A thus seems to play a role in activation of the G1 checkpoint in response to ionising radiation, through interaction with ATM. ATM is a kinases that phosphorylates a number of DNA repair enzymes and checkpoint proteins, including p53, in response to DNA damage. SNM1A has been observed

to localise to the site of double strand breaks induced by ionising radiation, in a manner dependent on interaction with ATM.⁴² SNM1A has also been shown to be a substrate of ATM *in vitro*.⁶⁴ Depletion of SNM1A reduces phosphorylation of p53 by ATM, which in turn reduces transcriptional activation of p21.⁶⁴

1.2.2. Entry to metaphase checkpoint

SNM1A is also involved in a checkpoint that occurs during mitosis.⁶⁵ Human SNM1A mRNA has an unusually long 5'-untranslated region that is predicted to fold into a complex secondary structure. This likely inhibits interaction of eukaryotic initiation factors and ribosomes with a chemically modified region of the mRNA known as the 5'-cap, which is the normal mechanism of translational initiation. The SNM1A mRNA instead contains an internal ribosomal entry site (IRES), allowing ribosomes to begin translating independently of the 5'-cap. This arrangement normally causes SNM1A expression to be quite minimal. During mitosis however, 5'-cap dependent translation is downregulated, allowing more ribosomes to be recruited to mRNAs that contain IRESs, and SNM1A expression is increased.^{67,68}

The Chfr protein plays a central role in a checkpoint that delays progression from prophase to metaphase in response to mitotic stress.⁶⁹ Chfr has been found to be under-expressed or mutated in several human cancer cell lines, which is likely a factor in the sensitivity of some cancers to antimetabolic agents such as taxol.⁶⁹ Cells deficient in SNM1A show similar characteristics to those of cells deficient in Chfr. Mouse cells lacking SNM1A, and HeLa cells in which SNM1A expression has been silenced using an siRNA, show increased sensitivity to the spindle poisons taxol and colcemid.⁶⁵ This suggests SNM1A and Chfr may function in the same pathway. Interestingly, SNM1A contains a poly(ADP-ribose)-binding zinc finger

(PBZ) motif. The role of the PBZ motif in SNM1A is not known, but similar PBZ domains are found in only two other human proteins, one of which is Chfr.⁷⁰⁻⁷²

Although direct interaction of SNM1A with Chfr has not been observed, SNM1A has been found to interact with 53BP1,⁴² and both SNM1A and 53BP1 interact with the anaphase promoting complex (APC).⁶⁵ Both 53BP1 and the APC are implicated in mitotic stress checkpoints, and bioinformatics evidence indicates that the APC may interact with Chfr.⁷³ It has been suggested that SNM1A may act in a pathway that negatively regulates the APC, preventing degradation of Chfr by the APC, and thus contributing to activation of the entry to metaphase checkpoint.⁷⁴ Further evidence for the involvement of SNM1A in responding to mitotic stress is provided by the observation that SNM1A interacts with Astrin, a spindle-associated protein essential for progression through mitosis,⁷⁵ however the role of this interaction *in vivo* has not been expanded upon.

Although responding to spindle disruption is believed to be the main function of the entry to metaphase checkpoint, it has been suggested that DNA damage may also be a trigger for this checkpoint, and an additional function could be to allow time for DNA repair to occur prior to chromatin condensation.^{65,76,77} Further research is needed to fully elucidate the biological role of the entry to metaphase checkpoint, and an improved understanding of SNM1A may be key to unravelling the complexities of this pathway.

1.3. SNM1A structure and mechanism of action

Human SNM1A is a 1040 amino acid protein.²⁹ The identified structural domains are summarised in Figure 1.3. The SNM1 family of enzymes are characterised by the presence of a metallo- β -lactamase (MBL) fold domain and a β -CASP domain, which together are referred to as the SNM1 domain and form the core catalytic part of the enzymes.^{78,79} While

MBLs are often associated with bacterial antibiotic resistance, they are a diverse family of enzymes, and other MBLs have also been shown to have nuclease activity.⁸⁰

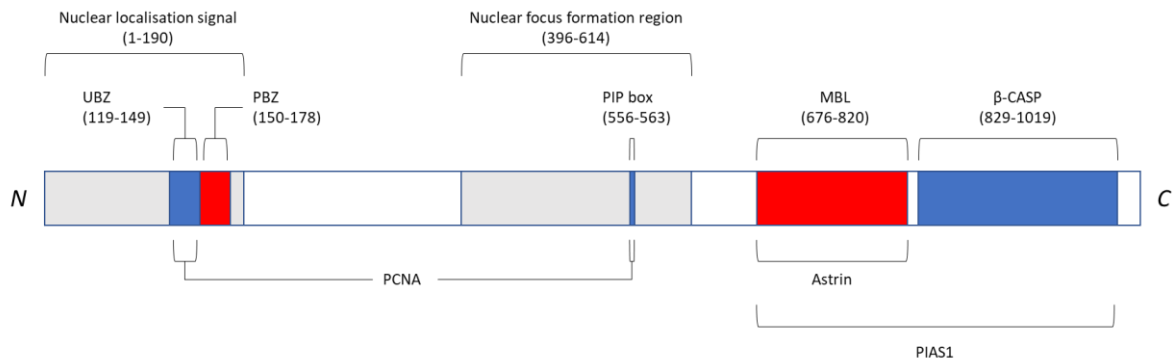


Figure 1.3. Structural domains of human SNM1A. Domains are labelled above while identified regions of interaction with other proteins are labelled below.

A crystal structure (PDB: 5AHR) of a truncated version of SNM1A containing the MBL fold and β -CASP domains has been reported, and shows a single zinc(II) ion coordinated by three histidine residues and one aspartate in the active site of the enzyme.⁷⁹ Another crystal structure of truncated SNM1A (PDB: 4B87) also shows one zinc ion in the active site, while a series of crystal structures obtained through analysis of low occupancy crystal states (PDB group deposition: G_1002036), as well as crystal structures of SNM1A in complex with an inhibitor (PDB: 5NZW and PDB: 5NZX), show one nickel(II) bound to the SNM1A active site. However, a crystal structure of the closely related enzyme SNM1B (PDB: 5AHO) shows two zinc ions in the active site (Figure 1.4).⁷⁹ All the residues involved in zinc coordination in SNM1B are also present in SNM1A, and all are in the same conformation except one aspartate residue, D736, which corresponds to D35 in SNM1B. D35 in SNM1B is hydrogen bonded to an asparagine residue that is absent in SNM1A, which may account for the difference in conformation. Nuclease activity of SNM1A is abolished by mutation of D736 however.^{36,38,52} This suggests that there is a second zinc ion present in the active form

of SNM1A, but that it is more loosely bound and therefore not observed in the crystal structure. It is likely that these two zinc ions activate a water molecule or hydroxide ion for nucleophilic attack on a phosphodiester, and also stabilise a build-up of negative charge on the phosphodiester as it undergoes hydrolysis. Another crystal structure of SNM1B has been reported with two adenosine 5'-monophosphate (AMP) molecules in the active site (PDB: 7A1F), corresponding to the products of a phosphodiester cleavage reaction. One AMP molecule was coordinated to an active site metal ion through two phosphate oxygen atoms, while the deoxyribose ring and nucleobase portions of the molecule had minimal interactions with the enzyme. The other AMP molecule was found to have extensive hydrogen bonding interactions through its phosphate moiety with a phosphate-binding pocket adjacent to the active site. Mutation of residues in the phosphate binding pocket was found to reduce SNM1B nuclease activity.⁸¹

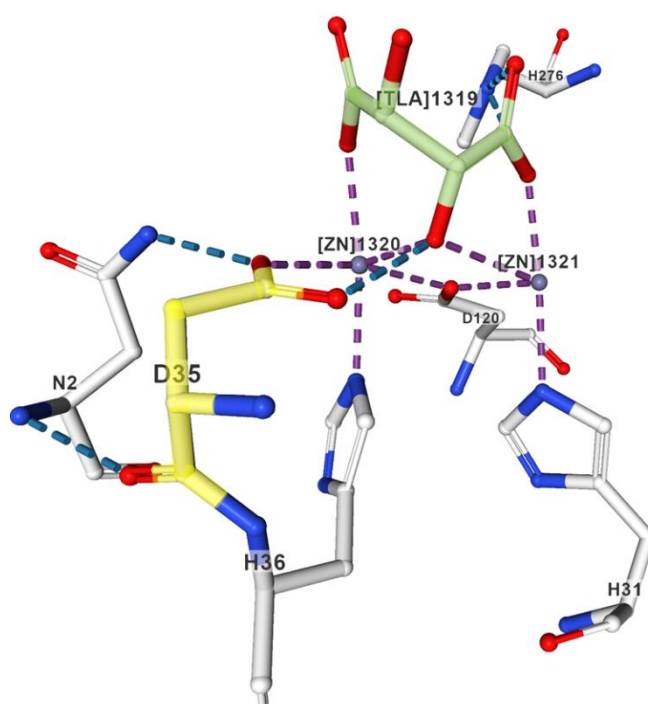


Figure 1.4. Crystal structure of the active site of SNM1B, with a tartrate molecule (green) co-crystallised.⁷⁹ One of the two zinc ions is coordinated by aspartate D35 (yellow). In the crystal

structure of SNM1A the corresponding aspartate residue is in a different conformation, and a second zinc ion is not observed.⁷⁹

Similarly to SNM1B, the active site of SNM1A also contains a binding pocket which recognises a terminal phosphate at the 5'-end of a DNA strand. This terminal phosphate moiety in the substrate is required for SNM1A activity.⁵² The presence of a phosphate binding pocket in SNM1A and SNM1B is suggested to determine the preference of these enzymes for exonuclease activity,⁸¹ while SNM1C, which acts as an endonuclease, lacks a phosphate binding pocket.⁸² The active site of SNM1A is surrounded by a region of positive charge, which may help to prevent dissociation of negatively charged DNA.⁷⁹ A wide DNA-binding groove allows SNM1A to accommodate large DNA substrates containing structure-distorting ICLs. Mutations of residues in this groove reduce the processivity of the enzyme without affecting the kinetics of a single phosphodiester cleavage event.⁷⁹

1.4. Inhibition of DNA repair enzymes

Compounds which damage DNA, including crosslinking agents, are commonly used to treat cancer.^{18,83} Certain cancers however can develop resistance to crosslinking agents through overexpression of the enzymes involved in ICL repair.⁸⁴⁻⁸⁷ Even cancers naïve to crosslinking chemotherapy may also overexpress ICL repair enzymes. For example, upregulation of DNA repair pathways, including overexpression of SNM1A, is associated with the propensity of melanoma cells to metastasise, as the ability to remove lesions that inhibit DNA replication is essential for rapid proliferation.⁸⁸ Similarly, colon cancer cells not previously exposed to crosslinking agents showed a greater than normal ability to repair ICLs induced by oxaliplatin.⁸⁹

The development of inhibitors for DNA repair enzymes, including those involved in ICL repair, is therefore a promising strategy for improving the efficacy of cancer

chemotherapy. Inhibitors for DNA repair enzymes have shown efficacy in clinical trials,⁹⁰⁻⁹² and some, including inhibitors for enzymes such as *O*⁶-methylguanine methyltransferase and poly(ADP-ribose) polymerases, have already been approved for use in cancer treatment.⁹⁰ Drug discovery programmes have also targeted other DNA repair enzymes such as FEN1 (NER and BER),^{93,94} APE1 (BER),^{95,96} and RAD51 (HR).⁹⁷

1.4.1. SNM1A as a therapeutic target

A study by Kohno *et al.* found that a single nucleotide polymorphism in SNM1A is associated with small cell lung carcinoma.⁹⁸ There is also evidence linking SNM1A expression levels with the prognosis of patients with ovarian cancer,⁹⁹ and mutations in the SNM1A gene have been linked with predisposition to breast cancer.¹⁰⁰ Expression of SNM1A was found to be upregulated in colorectal cancer tissue.¹⁰¹ The NIH National Cancer Institute's Genomic Data Commons database contains records of 157 cases of human cancer reported to have mutations in SNM1A.¹⁰² Most notably, point mutations were observed in 8.49% of cancers of the *corpus uteri* in a sample of 530 cases, while 27.18% of 585 patients with ovarian cancer had copy number variations in the SNM1A gene. A detailed investigation of the significance of SNM1A mutations in these cancers has not been reported. Due to the role of SNM1A in cell cycle checkpoints that control replication, and in repairing ICLs which could be oncogenic if they inhibit transcription of a tumour suppressor gene, SNM1A can itself be considered a tumour suppressor. Loss of function mutations in SNM1A may therefore play a role in cancer development. Alternatively, increased expression levels or gain of function mutations in SNM1A may contribute to acquired resistance to crosslinking chemotherapeutics or facilitate rapid proliferation through removal of replication-blocking ICLs even in cancers naïve to crosslinking agents. Although further elucidation of the role of SNM1A in human cancer is needed, it is likely

that SNM1A inhibitors used in the context of crosslinking chemotherapy would increase the efficacy of cancer treatment. Inhibitors which target the SNM1A active site may be able to impair its function in ICL repair without disrupting interactions with other proteins necessary for cell cycle checkpoints. Alternatively, inhibitors that do disrupt the function of SNM1A in cell cycle checkpoints could also be useful in cancer therapy. Many human cancers display dysregulation of cell cycle checkpoints due to mutations in key proteins such as p53.¹³ While this dysregulation may be a contributing factor in tumourigenesis,¹⁰³ it also represents a potentially exploitable vulnerability in fully transformed cells, as further dysregulation beyond a certain point may lead to catastrophic cell death, especially in highly proliferative tissue.^{13,104} Inhibitors of checkpoint proteins such as ATR, CHK1, and WEE1 are already in clinical trials for cancer treatment.¹⁰⁵

Reports of the phenotypes that result from SNM1A deficiency demonstrate the potential effects of inhibiting this enzyme in order to sensitise cancer cells to crosslinking chemotherapy agents, or to target cancer cells through disruption of cell cycle checkpoints. Human cells deficient in SNM1A show increased sensitivity to crosslinking agents such as SJG-136 and the anti-cancer drug mitomycin C.³⁸ Another study found that human cells deficient in SNM1A also exhibit increased sensitivity to the anti-cancer drugs cisplatin and chlorambucil.¹⁰⁶ Mouse embryonic stem cells deficient in SNM1A showed increased sensitivity to mitomycin C, but not to several other crosslinking agents, likely due to redundancy in mammalian ICL repair pathways.³¹ It was also reported in this study that whole mice lacking SNM1A showed no major abnormalities, but were sensitive to mitomycin C.³¹ Two other mouse SNM1A knockouts have been reported. Ahkter *et al.* observed that SNM1A deficient mice showed sensitivity to crosslinking agents, and otherwise normal development, but had shortened lifespans due to increased

tumourigenesis and susceptibility to bacterial infection.¹⁰⁷ This suggests SNM1A acts as a tumour suppressor. Hemphill *et al.* reported however that SNM1A deficient mice showed normal lifespans with no accelerated tumourigenesis.¹⁰⁸ A cross of these animals with FANCD2 deficient mice resulted in perinatal lethality however, suggesting that SNM1A and FANCD2 are each essential in the other's absence.¹⁰⁸ The different mouse phenotypes observed in these studies may be due to other genetic or epigenetic differences between the different strains used in these studies.

XPF-ERCC1, which is believed to act in the same pathway as SNM1A during replication-dependent ICL repair, has been proposed as a drug target,¹⁰⁹ as high levels of XPF-ERCC1 have been associated with poor response to platinum-based chemotherapy in cancer patients.¹¹⁰⁻¹¹³ A number of inhibitors for XPF-ERCC1 have been reported.¹¹⁴⁻¹¹⁸ However, selective targeting of SNM1A in place of XPF-ERCC1 may allow ICL repair to be inhibited with fewer unwanted side effects due to the roles of XPF-ERCC1 in a number of other biochemical pathways.¹¹⁹⁻¹²¹ SNM1A is thus a promising drug target, and inhibitors of this enzyme could be used to sensitise cancer cells to crosslinking chemotherapeutics, and perhaps to spindle poisons such as taxol.⁶⁵

1.4.2. Previously reported inhibitors of SNM1A

A glimpse of the potential utility of targeting SNM1A for cancer treatment was reported by Wu *et al.*, who showed that the steroidal natural product bufalin induces apoptosis in human lung cancer cells.¹²² Bufalin alters expression levels of numerous genes involved in DNA repair and cell cycle regulation, including SNM1A.

A number of compounds that directly inhibit SNM1A have also been reported. As SNM1A is a member of the metallo- β -lactamase superfamily of enzymes, it is inhibited by the β -

lactam antibiotics cephalosporins. A panel of β -lactam antibiotics were screened for inhibitory activity against SNM1A and SNM1B and four cephalosporins: 7-aminocephalosporinic acid, cephalosporin C, cefotaxime, and ceftriaxone, proved to be effective inhibitors of SNM1A (Figure 1.5A).¹²³ The strongest inhibitors, cefotaxime and ceftriaxone, have IC_{50} values in the 4-7 μ M range and appear to act as competitive reversible inhibitors. A crystal structure of SNM1A in complex with ceftriaxone (PDB: 5NZW) shows the inhibitor coordinating to an active site metal ion through its the α,β -dicarbonyl moiety, while another crystal structure (PDB: 5NZX) shows ceftriaxone bound to SNM1A at an alternative allosteric site. A crystal structure (PDB: 5NZY) of SNM1A in complex with cefotaxime has also been reported. Two molecules of cefotaxime are bound to the enzyme in this structure, but do not coordinate the active site metal ion. Other classes of β -lactams including penicillins, carbapenems, and monobactams did not show activity against SNM1A or SNM1B. While the cephalosporins are potent inhibitors of SNM1A *in vitro*, they have low membrane permeability, as they are designed to inhibit bacteria through an extracellular mechanism, and thus do not inhibit SNM1A *in vivo*. Another study reported that a high throughput screening of a library of bioactive compounds identified several other SNM1A inhibitors (Figure 1.5B), which have an unclear mechanism of action.¹²⁴

The development of substrate-mimic inhibitors for SNM1A facilitates the use of rational design to achieve improvements in efficacy, and previous research in the McGouran laboratory has shown promising initial results in this area. Modified DNA oligonucleotides bearing a squaramide at the 5'-end (Figure 1.5C) were shown to inhibit SNM1A.¹²⁵ The squaramide moiety may function as a zinc-binding group (ZBG), targeting the zinc ion(s) in the SNM1A active site. Further work in the McGouran laboratory showed that monomeric

modified nucleosides bearing a ZBG at the 5'-position (Figure 1.5D) could provide a scaffold for development of improved SNM1A inhibitors, as a thymidine derivative containing a hydroxamic acid ZBG inhibited SNM1A *in vitro*.¹²⁶ A number of nucleoside derivatives bearing malonate-based ZBGs were also shown to inhibit SNM1A (Figure 1.5E).¹²⁷ Malonate ZBGs were linked to a thymidine scaffold either through a carbon-carbon bond (C-linked malonates), or through formation of an amide (N-linked malonates). N-Linked malonates were found to inhibit SNM1A more effectively when the ZBG was placed at the 3'-position rather than the 5'-position. 5'-C-linked malonates were the most potent inhibitors of this series, but 3'-C-linked malonates were not successfully synthesised for comparison. Several dinucleosides containing a malonate internucleotide linkage were also reported in this study. These are structurally similar to the 3'-N-linked and 5'-C-linked mononucleosides. The dinucleosides inhibit SNM1A more effectively than the corresponding 3'-N-linked mononucleosides, showing that the presence of a second nucleoside residue increases favourable non-covalent interactions with SNM1A. The dinucleosides are less effective inhibitors than the 5'-C-linked mononucleosides however, likely because attachment of a second nucleoside residue through an amide bond precludes the inclusion of a second carboxylate or hydroxamic acid moiety as part of the ZBG, or because formation of a dinucleoside restricts the flexibility of the malonate, preventing it from adopting the optimum conformation for binding to the active-site zinc ion(s).

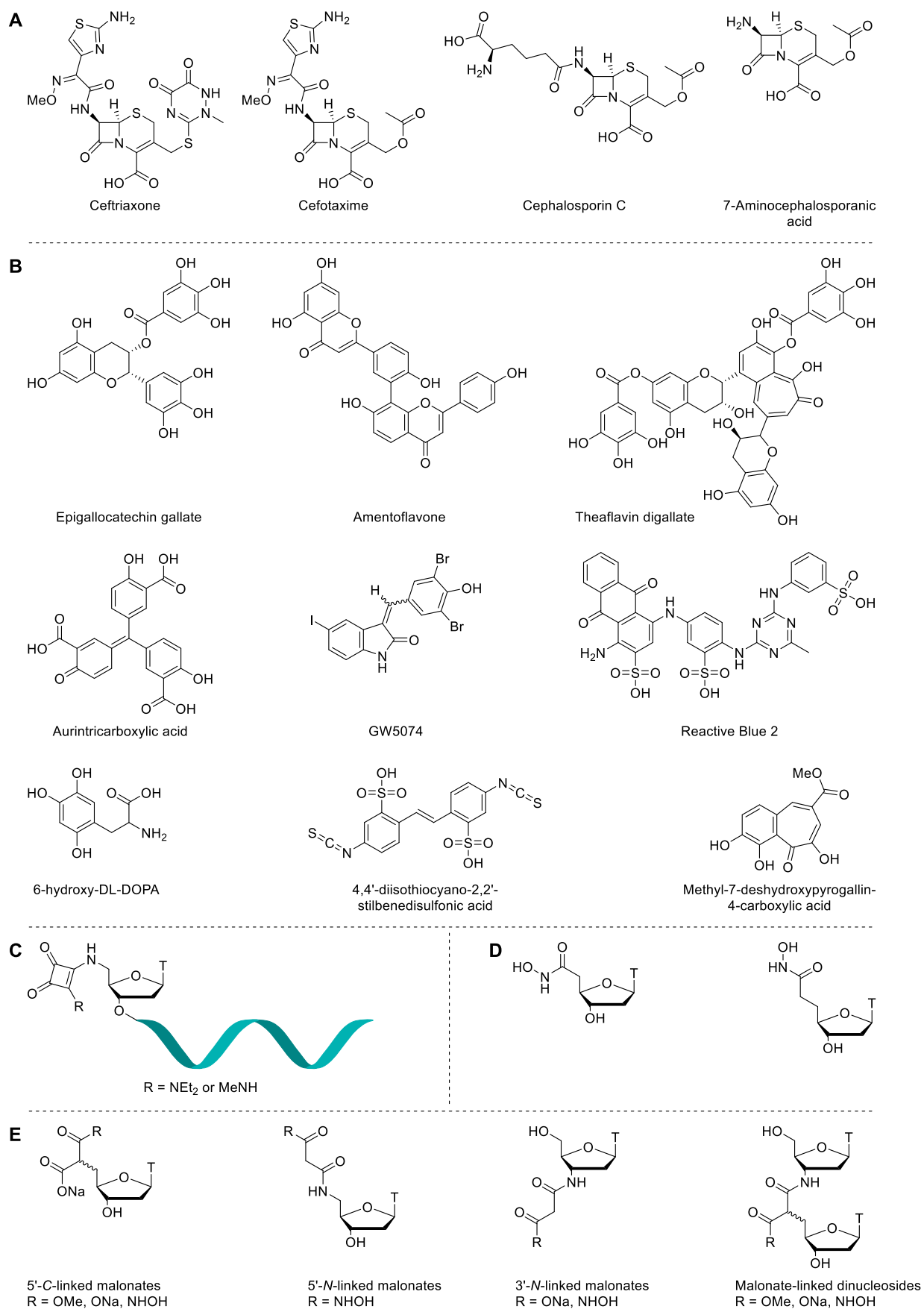


Figure 1.5. Previously reported inhibitors of SNM1A.

1.5. Molecular probes for studying enzyme activity

1.5.1. The utility of developing probes for SNM1A

The clinical relevance of SNM1A and other DNA repair enzymes makes them important targets for further investigation. As well as their role in cancer, ICLs and other forms of DNA damage contribute to the ageing process, and mutations of DNA damage repair enzymes can lead to premature ageing disorders. For example, mutation of CSB, an interactor of SNM1A, causes Cockayne syndrome.^{34,128} Research in this area would be accelerated by the availability of improved tools for the study of these enzymes.

SNM1A cannot be stably overproduced from mammalian cells.^{42,52} *In vitro* biochemical studies and crystallographic studies of SNM1A have therefore utilised recombinant human SNM1A produced from insect cells^{52,79} or yeast,³⁸ or truncated SNM1A expressed in *E. coli*.³⁸ Truncation of SNM1A may alter its activity and interaction with other proteins, as reported in the case of other truncated enzymes.^{129–131} Furthermore, SNM1A expressed in other organisms may lack post-translational modifications acquired by the enzyme in human cells which may modulate its activity. Also, *in vitro* studies may not account for other interacting molecules which may be present in cells. It is therefore desirable to develop methods to facilitate the study of SNM1A within human cells.

1.5.2. Activity-based protein profiling

Activity-based probes (ABPs) are a type of molecular probe which target only the active form of an enzyme, allowing enzymatic activity to be examined in a cellular context. ABPs typically consist of three main parts: a recognition element that mimics the natural substrate of the enzyme, a functional group, sometimes referred to as a warhead, which allows the probe to bind strongly to the active site of the enzyme, and a detection element such as a fluorophore or biotin moiety (Figure 1.6).¹³²

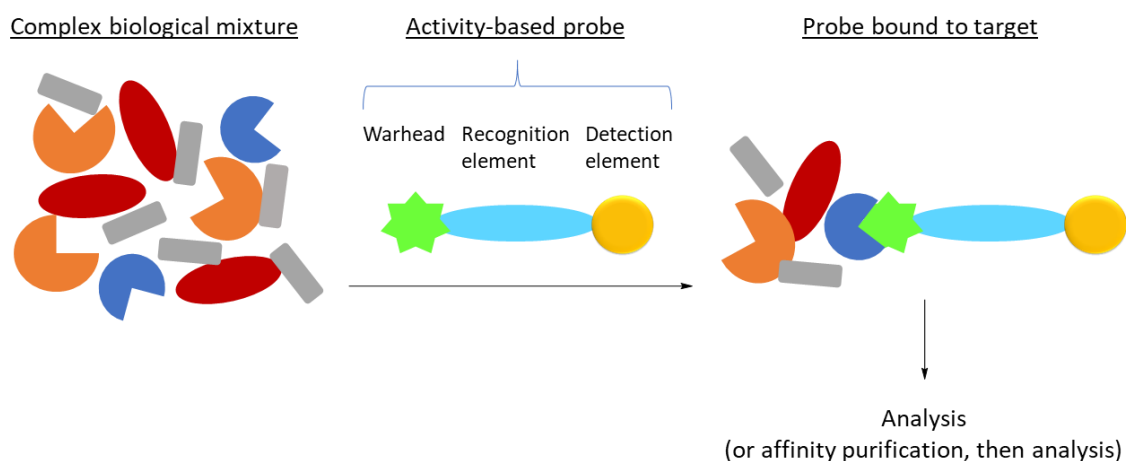


Figure 1.6. Activity-based protein profiling.

In some cases, the presence of the detection element may interfere with the membrane permeability of the probe or cause steric hindrance that blocks interaction with the target enzyme. Click chemistry can then be used to add the detection element to the probe subsequent to binding to the target enzyme and cell lysis.¹³³ As many enzymes contain catalytic nucleophilic residues such as cysteine, serine, or lysine in the active site, electrophilic functional groups such as Michael acceptors or epoxides are commonly used as warheads to covalently trap these nucleophilic residues.¹³⁴ Enzymes that don't contain a catalytic nucleophile are often metalloenzymes, and so are targeted using probes containing metal-binding groups, sometimes termed affinity-based probes (AfBPs). In this case the binding of the probe to the metalloenzyme may be reversible, so a photocrosslinking group is often included in the structure of an AfBP (Figure 1.7). After initial reversible binding of the probe to the target, the sample is irradiated to form a covalent adduct.^{135,136} Diazirines,¹³⁷ often with a trifluoromethyl-substituent,¹³⁸ are the most commonly used photocrosslinking group due to their small size, relatively long excitation wavelengths and high crosslinking efficiency, and can be incorporated into oligonucleotides through attachment at the 5-position of modified pyrimidine

nucleobases.^{139,140} Upon irradiation they form a carbene that inserts into a C-H bond in the target enzyme. Benzophenones,¹⁴¹ which form diradicals upon irradiation, and aryl azides,¹⁴² which form nitrenes, are also used as photocrosslinkers.



Figure 1.7. Diazirine, benzophenone and aryl azide photocrosslinking groups commonly used in molecular probes.

Following treatment of cells with an ABP and cell lysis, gel-electrophoresis-based methods can be used to identify enzymes bound by the ABP.¹⁴³ For biotin-labelled ABPs, a streptavidin pull-down can be performed followed by LC-MS/MS analysis.^{143,144} Similarly, other affinity handles such as HaloTag can be used in place of biotin-streptavidin.¹⁴⁵ Alternatively, whole cells treated with fluorophore-labelled probes, or even tissues or organisms, can be analysed using fluorescence microscopy and other forms of *in vivo* imaging.^{146–149} For example, positron emission tomography (PET) has been used in combination with isotope-labelled probes.¹⁵⁰

Unlike other proteomics methods, ABPs target only the active form of an enzyme. Many enzymes are initially expressed as inactive zymogens which are later activated through cleavage of a peptide from the protein. The activity of many other enzymes is dependent on factors such as the presence of cofactors, the presence of interacting proteins or other molecules, post-translational modifications, and cellular localisation.¹³² Expression levels, investigated by other proteomics techniques, therefore may not be well correlated with

activity in biological processes. ABPs which target only the active form of an enzyme are a useful tool for investigating factors other than abundance that regulate enzyme activity. ABPs can also be used in drug discovery to detect changes in enzyme activity in the presence of candidate inhibitors.¹⁵¹

Broad spectrum ABPs target an entire class or many members of a class of enzymes. These can prove useful for examining differences in activity levels of classes of enzymes between healthy tissue and diseased samples, and thus identify enzymes that are dysregulated in disease states.^{152,153} Broad spectrum ABPs also allow candidate inhibitors to be screened against multiple enzymes simultaneously *in vivo*, and can be used to evaluate selectivity across a range of similar enzymes and identify off-target effects.¹⁵⁴ Additionally, broad spectrum ABPs which screen for a particular activity can be used to annotate enzymes whose functions were previously unknown.¹⁵⁵

ABPs with greater specificity have been developed to target a small number of closely related enzymes, or even a single enzyme.¹⁵⁶ These are particularly useful for cellular localisation studies,¹⁵⁷ and are especially suited to enzymes, such as SNM1A, that are expressed at a low level, and thus might be obscured by more abundant enzymes in studies using broad spectrum ABPs.

1.5.3. Probes for targeting metallo- β -lactamases and other metallohydrolases

ABPs and AfBPs have been used successfully to study diverse types of enzymes, such as proteases, kinases, phosphatases, glycosidases, and deubiquitylating enzymes,¹⁵⁸ among others.¹⁵¹ No probes targeting SNM1A or any of the SNM1 family of enzymes have been reported to date. However, the successful development of molecular probes for a number

of related classes of enzymes demonstrates the potential for SNM1A to be studied using this approach.

Many metallohydrolases, like SNM1A, use catalytic zinc ions to activate a water molecule to hydrolyse their substrates. Two particularly important groups of metallohydrolases that have been studied extensively using molecular probes are matrix metalloproteases (MMPs) and histone deacetylases (HDACs). Many substrate-mimic inhibitors for MMPs have been reported. Typically, they contain a ZBG to enhance binding to the active site. Probes have been developed through elaboration of these inhibitors. For example, the MMP inhibitor GM6001 was modified to include a benzophenone photocrosslinker and a rhodamine fluorophore, to produce a broad spectrum AfBP for MMPs (Figure 1.8).¹³⁵ Later versions of this probe included an alkyne moiety in place of the rhodamine fluorophore.¹⁵⁹ Similarly, probes for HDACs have been prepared through modification of inhibitors for these enzymes. The HDAC inhibitor vorinostat for example, which contains a hydroxamic acid ZBG, was modified to include a benzophenone photocrosslinker and alkyne handle (Figure 1.8).¹⁶⁰

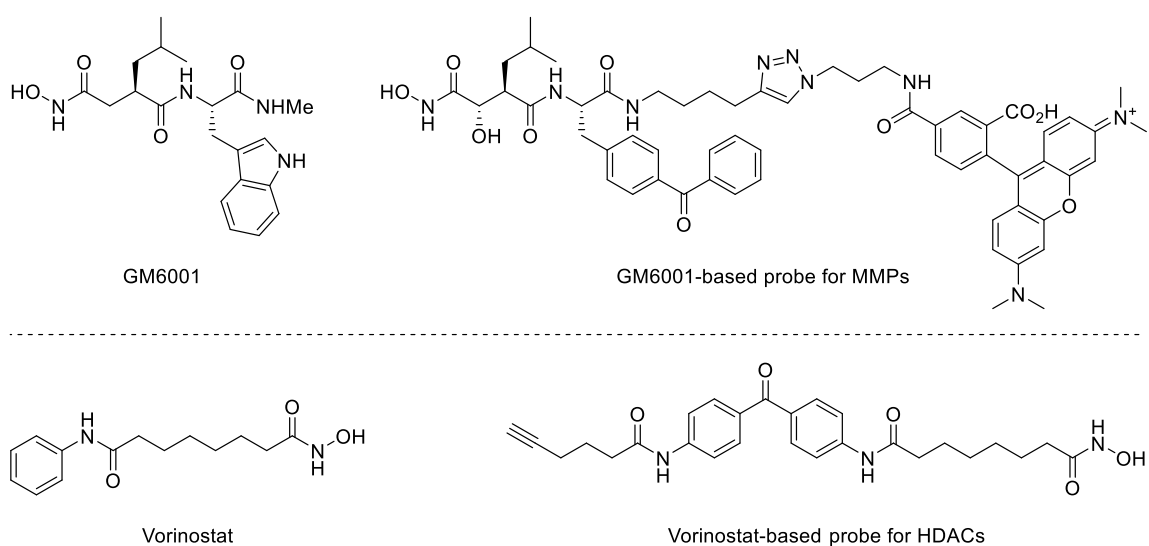


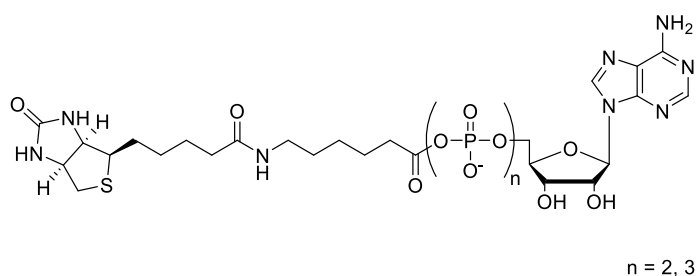
Figure 1.8. AfBPs for metallohydrolases based on elaboration of inhibitor structures.

MBLs, a subclass of metallohydrolases that includes the SNM1 family, have also been targeted using molecular probes. MBLs have primarily been studied using molecular probes in the context of their role in bacterial antibiotic resistance. Thus, substrate mimic probes based on synthetic β -lactams resistant to hydrolysis and appended with an alkyne handle have been developed.¹⁵⁹ Comparative profiling of antibiotic-sensitive *S. aureus* and multidrug resistant *S. aureus* using these probes led to the identification of two novel enzymes with β -lactamase activity.¹⁶¹ In another study, a fluorescent probe for New Delhi metallo- β -lactamase-1 (NDM-1) was developed, which unusually, does not interact directly with the catalytic zinc ions, but instead forms a disulfide bond with a cysteine residue that normally coordinates to one of the zinc ions.¹⁶² Zinc-chelating AfBPs for NDM-5 and NDM-8 have also been reported.¹⁶³ These probes feature an azidonaphthalimide moiety that functions as both a fluorophore and an aryl azide-type photocrosslinker. Another strategy involves the use of fluorogenic β -lactam probes, enzymatic hydrolysis of which releases a fluorescent product.^{164,165}

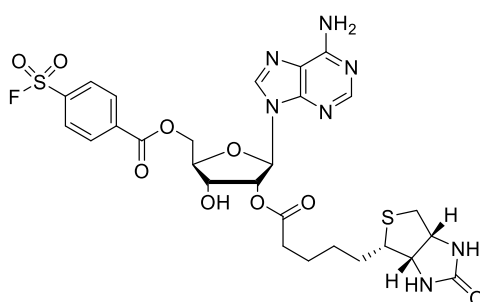
1.5.4. Probes for nucleic-acid-processing enzymes

A number of modified mononucleotides have been developed as molecular probes. Adenosine triphosphate (ATP) derivatives containing a biotin moiety appended to the 5'-position *via* a reactive acyl phosphate (Figure 1.9) have been used to study kinases and other ATP-dependent enzymes through labelling of catalytic lysine residues.^{166–168} Similar adenosine-based ABPs with an acyl phosphate warhead at the 5'-position have also been reported with an alkyne group to enable later conjugation of a detection element.^{169,170} Nucleotide-based ABPs for studying kinases and other proteins have also been developed with different electrophilic groups at the 5'-position, in particular the *p*-fluorosulfonylbenzoyl group (Figure 1.9).^{171–174} These modified-mononucleotide ABPs

are typically broad spectrum probes due to the large number of enzymes that interact with mononucleotides.



Adenosine-based probes bearing a reactive acyl phosphate



An adenosine-based probe bearing a *p*-fluorosulfonylbenzoyl group

Figure 1.9. Modified nucleotides as molecular probes.

There remains an unmet need to develop molecular probes targeting DNA-processing enzymes with greater specificity. This has been achieved in the case of HIV-1 integrase, which like SNM1A has phosphodiesterase activity. This enzyme is inhibited by a class of diketo acid compounds which chelate to Mg^{2+} ions in the enzyme active site. These inhibitors have been appended with a photocrosslinker and biotin moiety to construct AfBPs (Figure 1.10).^{175–177}

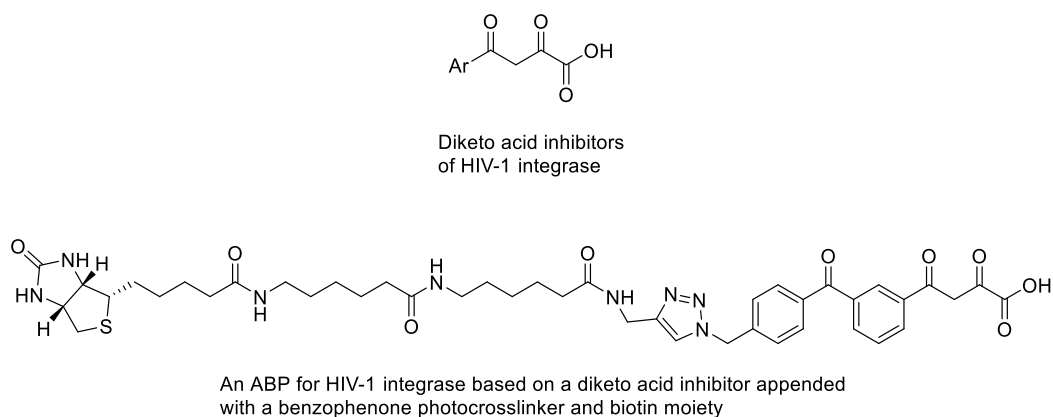


Figure 1.10. Inhibitors and an ABP for HIV-1 integrase based on a diketo acid scaffold.

To develop probes targeting other DNA-processing enzymes with specificity, a logical approach is to increase the size and distinctiveness of the recognition element through the use of an oligonucleotide in place of a mononucleotide or other small molecule scaffold. Furthermore, an oligonucleotide scaffold could potentially serve a dual function as both a recognition element and detection element in molecular probes for DNA-targeting enzymes. An oligonucleotide has been used as a detection element in probes for other enzymes, allowing quantitation through qPCR.¹⁷⁸

Oligonucleotide probes for DNA repair enzymes that have been reported thus far typically consist of an oligonucleotide containing a lesion, such as an ICL,¹⁷⁹ thymine dimer,¹³⁹ or 8-oxoguanine,^{139,180} a photocrosslinker, and a detection element such as biotin. These probes have facilitated the identification of proteins that are recruited to the site of a lesion during DNA repair,¹⁸¹ although they do not allow for the identification and study of the active enzymes that carry out the repair process. A similar strategy has been used to identify proteins that recognise *N*⁶-methyladenosine, a modification common in RNA, through the development of a synthetic oligonucleotide containing *N*⁶-methyladenosine and a diazirine photocrosslinker.¹⁸² AfBPs have also been used to study phosphodiesterase

4,^{183,184} an enzyme that hydrolyses the important second messenger cAMP. AfBPs specific for the related enzyme phosphodiesterase 10 have also been reported.¹⁸⁵

Oligonucleotides containing a fluorophore and a quencher have been used to study nucleases and other DNA repair enzymes. Typically, upon processing of the oligonucleotide by a repair enzyme, the fluorophore is separated from a quencher resulting in an increase in fluorescence.^{186–189} Similar more elaborate luminescence switch-on strategies have also been used.^{190,191} Nuclease activity has also been measured using electrochemical methods.¹⁹² Oligonucleotides containing an electrochemically active ferrocene moiety diffuse faster in solution after truncation by a nuclease, leading to a measurable increase in current. While these luminescence switch-on or electrochemical strategies allow for nuclease activity to be studied using an enzyme kinetics approach, these oligonucleotides function as enzyme substrates and do not covalently capture the active enzyme allowing for downstream analysis in the manner of an ABP.

ABPs combine the advantages of photocrosslinking probes and luminescence switch-on or electrochemical oligonucleotide probes, as they can both covalently capture enzymes in biological samples, and be used to study enzyme activity. However, no substrate-mimic ABPs for DNA-processing enzymes based on an oligonucleotide as a recognition element have been developed to date. One report by Kasai *et al.* described modified electrophilic nucleobases which were incorporated into oligonucleotides and could then undergo nucleophilic aromatic substitution reactions with a cysteine residue in the active site of DNA cytosine 5-methyltransferase.¹⁹³ These oligonucleotides were tested only as inhibitors, but have the potential to form the basis of ABPs for this enzyme through incorporation of a fluorophore or biotin moiety. Such development of oligonucleotide-

based ABPs is a promising but understudied strategy for the investigation of DNA-processing enzymes.

1.6. Zinc-binding groups

A metal-binding group is an important element in many inhibitors and molecular probes for metalloenzymes. For targeting zinc-dependent enzymes, a number of zinc-binding groups (ZBGs) have been utilised. The role of the ZBG is to anchor the inhibitor or probe in the target enzyme's active site, and potentially also to displace a bound water molecule or hydroxide ion from the catalytic zinc ion(s). In particular, ZBGs have been incorporated into inhibitors of two classes of zinc metalloenzymes already discussed: MMPs¹⁹⁴ and HDACs.¹⁹⁵ The results of these studies may be used to inform the design of molecules for targeting the zinc metalloenzyme SNM1A.

1.6.1. Hydroxamic acids and related ZBGs

Hydroxamic acids are one of the most commonly used ZBGs in medicinal chemistry.¹⁹⁶ Numerous X-ray crystallographic structures of enzyme-inhibitor complexes have shown hydroxamic acids coordinating to zinc ions in a bidentate fashion through the carbonyl oxygen and the hydroxyl group.¹⁹⁷ Three hydroxamate-bearing molecules; vorinostat, belinostat, and panobinostat, are HDAC inhibitors approved for clinical use by the US FDA.¹⁹⁵ Hydroxamic acids have also been used in many other experimental inhibitors for HDACs and other enzymes,^{196,198} including a thymidine derivative with a hydroxamic acid at the 5'-position that has been reported as an inhibitor of SNM1A by the McGouran laboratory,¹²⁶ as discussed in section 1.4.2. Hydroxamic acids have pK_a values in the range of 8-10 for both the terminal OH and NH protons in aqueous solution.¹⁹⁷ They are thus not deprotonated under physiological conditions, but upon coordination to a zinc ion their acidity is increased.¹⁹⁹ NMR studies with a ¹⁵N labelled hydroxamic acid showed that it

bound to MMP-3 as the conjugate base of an *O*-acid.²⁰⁰ Other studies have also found that hydroxamic acids were deprotonated subsequent to zinc-binding.^{201,202} Deprotonated hydroxamates bind to Zn²⁺ ions with very high affinity, which has led to the widespread application of hydroxamic acid ZBGs. However, as well as issues with their pharmacodynamic properties,²⁰³ a limitation of hydroxamic acids that has curtailed further clinical use is poor selectivity, as they often bind to other off-target metalloenzymes,²⁰⁴ and may bind to other metals, particularly iron, as well as zinc.²⁰⁵

To obtain better selectivity for a target enzyme, some inhibitors have used ZBGs of intermediate strength, allowing non-covalent interactions between other parts the inhibitor and amino acid residues in the target enzyme to play a larger role. Examples of less potent ZBGs with similar structures to hydroxamic acids include hydrazides, carboxylic acids, and *N*-hydroxyureas, among others (Figure 1.11).

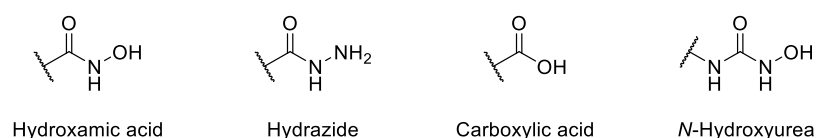


Figure 1.11. Hydroxamic acids and related ZBGs.

Hydrazides are isosteres of hydroxamic acids in which the terminal OH group is replaced by an NH₂. They can chelate to zinc in a similar manner to hydroxamic acids, although with lower affinity. Hydrazides have been used as zinc binding groups in inhibitors of HDACs and MMPs.^{194,206,207}

Many MMP inhibitors containing hydroxamic acid ZBGs have been prepared from carboxylic acid precursors, and in many reports the inhibitory activity of these precursors has also been investigated.²⁰⁸ Carboxylate ZBGs have much lower affinity for zinc ions than

hydroxamates, and may bind in either a unidentate or bidentate manner.²⁰⁹ This lower affinity can however allow for fewer off-target effects. For example, an MMP inhibitor bearing a carboxylic acid ZBG was shown to give excellent selectivity for MMP-12.²¹⁰ Carboxylate ZBGs have also proven effective against other enzymes. For example, valproic acid is a HDAC inhibitor with a carboxylate ZBG that has been tested in clinical trials for treatment of glioblastoma,²¹¹ and lisinopril is a clinically approved inhibitor of angiotensin-1 converting enzyme (ACE) that contains a carboxylate ZBG.²¹² A primary amide has been reported as a carboxylic acid isostere and proved effective as a ZBG in a HDAC inhibitor.²¹³ Malonates are another type of carboxylic-acid based moiety that have shown efficacy as ZBGs in SNM1A inhibitors,¹²⁷ and can be combined with the hydroxamic acid functionality.^{127,214} Malonates have also been reported as phosphate mimics, as both of these functional groups are of similar size and are dianionic under physiological pH.^{215,216} The *N*-hydroxyurea functional group is structurally similar to the hydroxamate moiety and can function as a ZBG. *N*-Hydroxyureas tend to have improved oral bioavailability over analogous hydroxamates, but bind to zinc with lower affinity.²¹⁷ A crystal structure of an inhibitor bearing an *N*-hydroxyurea ZBG in complex with MMP-8 showed unidentate rather than bidentate binding to the active site zinc ion.²¹⁸ Another X-ray crystallography study showed an *N*-hydroxyurea bound to a zinc ion in a bidentate fashion in the carbonic anhydrase active site.²¹⁹ A thymidine derivative bearing an *N*-hydroxyurea at the 3'-position was shown to bind to zinc ions, and was used to inhibit a ribonuclease.²²⁰

1.6.2. Sulfur-containing ZBGs

Thiols bind strongly to zinc ions due to the high affinity of sulfur for zinc, and are therefore commonly employed by Nature as ZBGs; many metalloproteins contain cysteine residues that bind to structural or active site zinc ions.^{221,222} Thiol ZBGs are also used in synthetic

drug molecules, for example in captopril, an inhibitor of ACE (Figure 1.12).²²³ Other thiol-containing ACE inhibitors have been reported,²²⁴ and thiols have also been used as ZBGs in inhibitors of other enzymes, such as MMPs,²⁰⁸ and MBLs.²²⁵ A striking example of the efficacy of thiol ZBGs was reported by Suzuki *et al.*, who prepared an analogue of the HDAC inhibitor vorinostat in which the hydroxamic acid ZBG is replaced by a thiol (Figure 1.12). The IC₅₀ of this inhibitor is comparable to that of vorinostat, despite the fact that the thiol ZBG binds to zinc in a monodentate manner while vorinostat contains a chelating ZBG.²²⁶ As an elaboration on the structure of thiol ZBGs, chelating mercaptosulfide²²⁷ and mercaptoamide²²⁸ ZBGs have been reported (Figure 1.13).

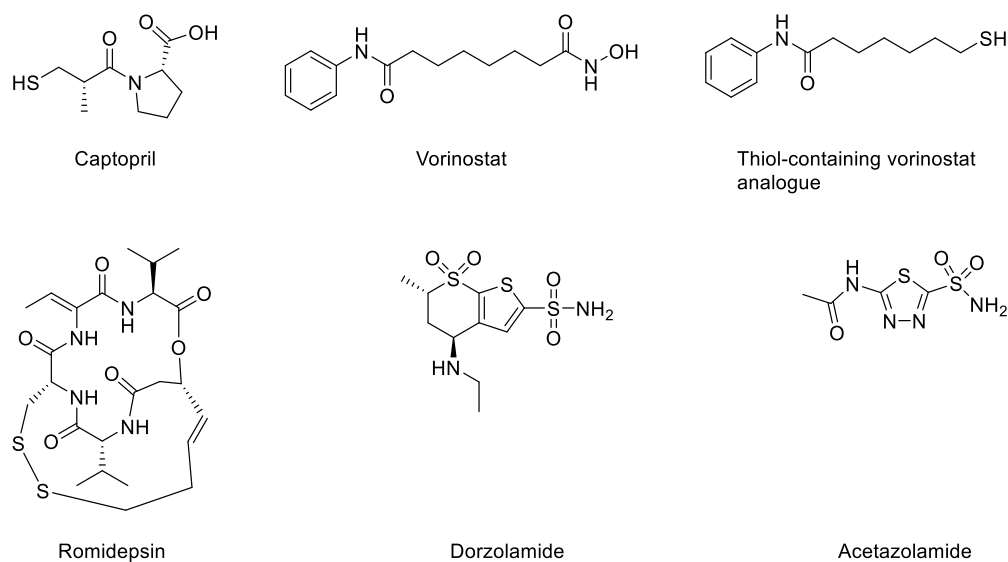


Figure 1.12. Metalloenzyme inhibitors containing sulfur-based ZBGs.

Due to the reactivity of thiol groups, a prodrug strategy is often used to mask this ZBG.²²⁹ This can be achieved using a disulfide. Romidepsin is a bacterial natural product HDAC inhibitor used for cancer treatment (Figure 1.12).²³⁰ It is thought to act as a prodrug, as it contains a disulfide which is reduced to a thiol *in vivo*. An alternative prodrug strategy

involves the use of thioesters. For example thiol ZBGs in HDAC inhibitors have been masked using a thioacetate²³¹ or *S*-isobutyryl²³² group.

Sulfonamides have been investigated as ZBGs in inhibitors of various enzymes, including MMPs.²³³ In particular however, sulfonamide ZBGs have proven effective in inhibitors of carbonic anhydrases.²³⁴ Several sulfonamide-based carbonic anhydrase inhibitors, such as dorzolamide and acetazolamide (Figure 1.12), are in clinical use.²³⁵ Sulfonamides have been shown in X-ray crystallographic studies to bind to Zn(II) in the carbonic anhydrase active site in a unidentate manner through the nitrogen atom, after deprotonation of one of the NH protons.²³⁶ An analysis of a large number of co-crystal structures in the Protein Data Bank showed that sulfonamides primarily bind to zinc ions in a unidentate fashion, however several examples of bidentate binding were also reported.²⁰⁹ In addition to their use as ZBGs, sulfonamides have been reported as phosphate isosteres, for example in tyrosine phosphatase inhibitors.²³⁷ Isosteres of sulfonamides such as sulfamates,^{234,238} sulfamides,^{234,238,239} and sulfodiimines,^{240,241} have also been investigated as ZBGs (Figure 1.13).

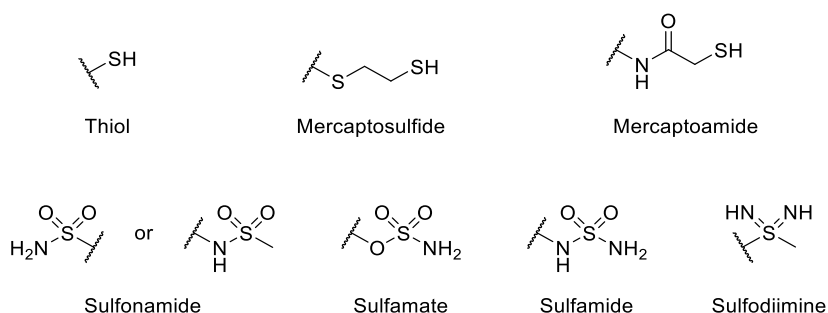


Figure 1.13. Sulfur-based ZBGs.

1.6.3. Phosphorous-based ZBGs

Several phosphorous-based functional groups have been used as ZBGs (Figure 1.14).

Phosphinates for example have been used in inhibitors of MMP-11.^{242,243} Notably,

fosinopril is an FDA-approved ACE inhibitor that contains a phosphinate ZBG.²⁴⁴ Similarly, phosphonates have also been used as ZBGs in MMP inhibitors.^{245,246} Foscarnet is an antiviral drug in clinical use that has been shown to undergo off-target binding to the zinc ion in the active site of a carbonic anhydrase through its phosphonate group.²⁴⁷

Phosphoramidates are structurally similar to sulfonamides, and have thus been used as ZBGs in inhibitors of carbonic anhydrase,^{248,249} in addition to other enzymes.²⁵⁰ Phosphoramidate-based inhibitors of the zinc-dependent enzyme thermolysin were found to be more potent than phosphonate analogues due to the ability of the NH protons of the phosphoramidates to engage in hydrogen bonding with surrounding amino acids.²⁵¹ However other reports have shown that the relative efficacy of phosphinate, phosphonate, and phosphoramidate ZBGs varies across different target enzymes.²⁰⁸ The P-N bond in phosphoramidates is reportedly unstable to hydrolysis under a range of conditions, and this may limit the application of this functional group as a ZBG.²⁵²

The carbamoylphosphonate functional group has been effectively used as a ZBG in MMP inhibitors.^{253–255} Carbamoylphosphonates can chelate to zinc through the carbonyl oxygen and one of the phosphonate oxygens to form a 5-membered ring.²⁵⁶ Carbamoylphosphonates have proven to be more effective ZBGs than other acylphosphonates, likely because of greater electron density on the carbonyl oxygen due to delocalisation of one the nitrogen lone pairs.²⁵³ The amide NH proton(s) may also contribute to hydrogen bonding with amino acids in the active site of the target enzyme. Indeed *N,N*-dialkylcarbamoylphosphonates were found to have little or no inhibitory activity against MMPs, demonstrating the importance of fine-tuning the ZBG for the target enzyme in question.²⁵³

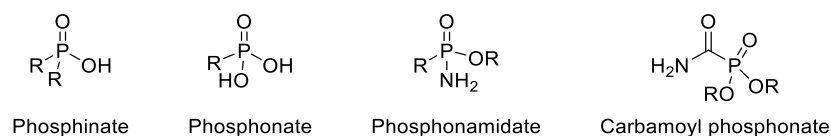


Figure 1.14. Phosphorous-based ZBGs

1.6.4. Other ZBGs

In addition to hydroxamic acids and their isosteres, sulfur-based ZBGs, and phosphorous-based ZBGs, a large number of other ZBGs have been reported. Many of these are elaborations and combinations of the ZBGs discussed above, such as sulfonylhydrazides,²⁰⁶ thiocarbamoyl phosphinates,²⁵⁷ succinyl hydroxamates,²⁰⁸ and hydroxamates adjacent to sulfonamides.²⁰⁸

Other classes of ZBGs have also been developed however. One such class of ZBGs is based on the ketone functional group. Ketones function as ZBGs of intermediate strength,²⁵⁸ as they may be hydrated in solution to form geminal diols which chelate zinc through the two oxygen atoms.²⁵⁹ Trifluoromethyl ketones are more likely to exist as a hydrate in solution, and function as stronger ZBGs.²⁶⁰ Silane diols have also been used to chelate zinc in a similar way.^{260,261}

Benzamides have also been utilised as ZBGs. In particular 2-aminobenzamides have proven effective in HDAC inhibitors.^{262,263} Chidamide is a HDAC inhibitor approved for clinical use in China which contains a benzamide ZBG.²⁶⁴ A potential drawback of the 2-aminobenzamide ZBG is that it contains an aniline moiety, which can potentially be metabolised to form toxic products.²⁶⁵ To avoid this issue, Li and Woster reported an oxazole moiety bearing an amide substituent as an effective isostere of the 2-aminobenzamide group.²⁶⁶ Other heterocycles containing an α -hydroxy carbonyl or

α -hydroxy thiocarbonyl motif have also been used as ZBGs. These have been reviewed by Jacobsen *et al.*¹⁹⁴

A number of macrocyclic ZBGs have been reported. Cyclic peptides, sometimes appended with additional ZBGs, have been used as inhibitors of zinc metalloenzymes. These include natural products such as romidepsin,²³⁰ as well as synthetic peptides.^{267,268} Non-peptidic nitrogen-containing macrocycles have also been proposed as ZBGs.²⁶⁹ Crisaborole is a clinically approved inhibitor of phosphodiesterase 4 that uses a cyclic boronate ZBG.^{270,271}

Selection of an appropriate ZBG for binding to a particular target enzyme is often not amenable to standard structure-activity relationship approaches, and this has driven the development of a diverse range of novel ZBGs. Developments in the area of novel ZBGs for biological applications are ongoing, and there remains a need for improved ZBGs with greater selectivity and more favourable pharmacokinetic properties.

1.7. Squaramides

1.7.1. General properties and cation binding

Squaramides are a family of cyclobutene derivatives that have found extensive application in diverse areas of research. Their stability, conformational rigidity, synthetic versatility, and excellent hydrogen bonding properties make squaramides a useful tool in areas such as supramolecular chemistry, organocatalysis, and bioconjugation.²⁷² Squaramides with various substituents also have the potential to function as novel ZBGs and phosphate mimics, and thus show promise for applications targeting zinc-dependent phosphodiesterases such as SNM1A.

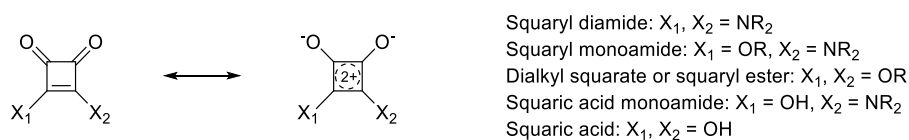


Figure 1.15. Aromaticity of squaramides.

Due to a mesomeric effect, squaric acid, squaryl esters, and squaramides can satisfy the Hückel rule, as they have $4n+2$ ($n = 0$) π electrons delocalised around the cyclobutene ring, and thus have aromatic character (Figure 1.15). Donation of a lone pair of electrons from a nitrogen atom into the cyclobutene ring increases the aromaticity of squaramides relative to squaryl esters.²⁷³ Involvement of NH protons of squaramides in hydrogen bonding likely enhances this effect by increasing the electron density on the nitrogen atoms. Squaramides therefore exhibit strong hydrogen bond donor capabilities. This has facilitated their use in organocatalysis,^{274,275} with many organocatalysts consisting of a squaramide core flanked by a chiral moiety to one side, and a basic functional group such as a tertiary amine on the other side.²⁷² Squaramides can interact with anions through donation of hydrogen bonds and this has led to their use in molecular recognition.^{276,277} Donation of the nitrogen lone pairs into the cyclobutene ring also results in restricted rotation around the nitrogen-carbon bonds, allowing squaramides to exist as a mixture of *syn/anti* rotamers.^{278–280} This property, as well as the fact that squaramides are rigid and planar, allows these moieties to be used to obtain a degree of conformational control of molecules into which they are incorporated.

Protonation of the squaramide oxygen atoms, or acceptance of hydrogen bonds, also enhances the aromaticity of squaramides, and this effect accounts for their excellent hydrogen-bond acceptor capabilities, and likely contributes to their interaction with cations.^{273,281} Squaramides can bind to cations through coordination of the carbonyl

groups, or possibly also through π -cation interactions, with the cation positioned above the plane of the cyclobutene ring.^{279,282,283} *N*-Hydroxysquaramides have the ability to chelate zinc ions (Figure 1.16), and have shown promise as ZBGs in inhibitors for metalloproteases.^{284,285} Squaric acid has been found to form salts with various metals, but in particular reacted rapidly with zinc or with magnesium.²⁸⁶ As squaramides are both strong donors and acceptors of hydrogen bonds, they can self-assemble through formation of extended hydrogen bonded networks, leading to their application in supramolecular chemistry and materials science.²⁷²

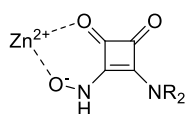


Figure 1.16. Chelation of zinc ions by *N*-hydroxysquaramides in their deprotonated form.

Wide application of squaramides has been facilitated by the ready accessibility of their synthetic precursors, dialkyl squarates.²⁸⁷ Substitution of dialkyl squarates with one amine to form a squaryl monoamide increases the aromatic stabilisation and reduces the electrophilic character of these molecules,²⁷³ allowing them to be sequentially substituted with two different amines without formation of unwanted symmetrical products.

1.7.2. Squaramides in chemical biology

Due to the generally high tolerance for these functional groups in biological systems,^{288,289} squaramides have found application in the field of chemical biology. In particular they have proven useful for bioconjugation due to ease of sequentially substituting dialkyl squarates with two different amines. Squaramides have thus been used as a linker for labelling proteins with monosaccharides and other small molecules,²⁹⁰ and for conjugating oligosaccharides to proteins.^{291–293} Squaramides have also been used as a linker for

attaching oligosaccharides to DNA²⁹⁴ and RNA²⁹⁵ oligonucleotides to improve cellular uptake of nucleic acids, and for linking peptides to DNA²⁹⁶ and RNA.²⁹⁷ A squaramide linkage has been used to conjugate the bacterial natural product desferrioxamine B to trastuzumab, an antibody that targets the HER2 receptor that is commonly overexpressed in breast cancer.²⁹⁸ Desferrioxamine B binds to Zr(IV) ions, and the use of the squaramide linkage is believed to strengthen this binding due to coordination of the squaramide carbonyl groups to zirconium. This conjugate can be used to deliver zirconium-89, an isotope with properties suitable for application in PET imaging, to breast cancer tissue.

Another property of squaramides, conformational rigidity, has led to the incorporation of squaramide linkages in peptidomimetics to afford conformational control of the folding properties of these molecules.^{299–301}

Squaramides bearing fluorophore substituents have been used for imaging cells in fluorescence microscopy. They enter the cell through receptor-mediated endocytosis due to interaction of the squaramide with anionic moieties present on the external surface of the cell membrane.³⁰² The squaramides accumulate in acidic vesicles, particularly late endosomes, and can thus be used for staining of these subcellular compartments.^{302,303}

Squaryl monoamides tend to react slowly with surface-exposed lysine residues in proteins unless the squaryl monoamide is present in large excess, or held in proximity to the lysine due to non-covalent interactions between the protein and the squaramide-bearing molecule. The reaction of squaryl monoamides with lysine residues is tolerant of other nucleophilic functional groups such as thiols that may be present in biological samples. This allows substrate mimics bearing squaryl monoamides to be used for selective covalent labelling of specific lysine residues in enzyme active sites.^{292,297} This strategy has been

applied in the study of nucleic acids and nucleic acid-processing enzymes, as an RNA strand bearing a squaryl monoamide has been used to selectively covalently label a lysine residue in the active site of the tRNA-binding enzyme FemX_{Wv}.²⁹⁷

As well as targeting enzyme active sites, squaryl monoamides can also be used for specific labelling of lysine residues at sites of interaction between biomolecules. A cytidine triphosphate derivative bearing a squaryl monoamide attached to the nucleobase was successfully used as a substrate for DNA-polymerase catalysed synthesis of DNA oligonucleotides.²⁹⁶ The resulting squaryl monoamide-bearing oligonucleotides were able to crosslink with lysine residues in histones. Lysine-DNA interactions are common and important for histone binding to DNA. No crosslinking of these oligonucleotides was observed with non-DNA binding proteins that contain multiple lysine residues, showing the importance of proximity effects. This method could similarly be used to identify other proteins that interact with DNA.

1.7.3. Squaramides as phosphate bioisosteres

Squaramide moieties have been used broadly in medicinal chemistry. Dibutyl squarate is an immunomodulator applied topically for treatment of warts³⁰⁴ and alopecia.³⁰⁵ More complex squaramide-based small molecules have entered clinical trials for treatment of diabetic neuropathy,^{306,307} chronic obstructive pulmonary disease,³⁰⁸ and metastatic solid tumours.³⁰⁹ In recent years there has been growing interest in the development of squaramides for treating a wide range of other diseases.²⁷²

Squaramides and squaric acids can function as bioisosteres for phosphates, as well as a number of other functional groups. Squaryl diamides have been used as a guanidine isostere,³¹⁰ and squaric acids have found application as isosteres for carboxylic acids.³¹¹ For

example, α -amino squaric acids, amino acid analogues in which the carboxylic acid moiety is replaced by a squarate, have been synthesised³¹² and incorporated into enkaphalin analogues.³¹³ α -Amino squaric acids have also been incorporated into other peptides using solid-phase synthesis, and reportedly confer resistance to peptidases.³¹⁴ A squaric acid monoamide has also been used as a carboxylate bioisostere in a mimic of the neurotransmitter L-glutamate.³¹⁵

In addition to the use of squaramides as phosphate mimicking ZBGs in nucleoside and oligonucleotide inhibitors of SNM1A discussed in section 1.4.2,^{125,126} there have been several other reports of the use of squaramides as phosphate bioisosteres. Squaramides are neutral but highly polarised, allowing them to mimic the negatively charged phosphate group. Furthermore, squaric acid monoamides have been reported to have pK_a values around 2.3,³¹⁶ and so exist as monoanions at physiological pH, similar to a phosphodiester linkage in DNA. As squaramides are not easily hydrolysed they are more stable than phosphates or phosphodiesteres. Seto and co-workers reported squaric acids as phosphate mimics in inhibitors of protein tyrosine phosphatases,³¹⁷ and phospholipid analogues bearing squaramides as phosphate isosteres have been prepared.³¹⁸ Most examples of squaramide derivatives as phosphate bioisosteres, however, have focused on the use of these moieties in analogues of nucleotides and oligonucleotides.

Glüsenkamp *et al.* first reported an adenosine derivative bearing a squaryl monoamide at the 5'-position in 1991, although the biological properties of this compound were not investigated.³¹⁹ A similar adenosine derivative bearing a squaryl diamide as a phosphate mimic at the 5'-position was later shown to inhibit the human O-acetyl-ADP-ribose deacetylase MDO1.³²⁰ Guanosine derivatives bearing a sulfonamide or squaric acid

monoamide as a phosphate mimic at the 5'-position have been shown to inhibit the eukaryotic translation initiation factor eIF4E.³²¹

Sugar-nucleotides often bind to glycosyl transferases through interaction of a diphosphate on the nucleotide with a divalent metal cofactor in the enzyme. Nucleosides modified with a squaryl diamide in place of the diphosphate at the 5'-position inhibited a trypanosomal GDP-mannose dependent mannosyl transferase, albeit with limited efficacy.²⁷⁸ These compounds were shown to bind to Mg^{2+} cations through the squaramide moiety in NMR titrations.

The squaramide functional group has been used as a replacement for a phosphodiester linkage in the backbone of a DNA oligonucleotide. Sekine and co-workers reported the synthesis of a dinucleoside consisting of two thymidine residues connected through a squaramide linkage, and incorporation of this molecule into oligonucleotides through solid-phase synthesis.³²² An NMR titration experiment in which aliquots of Mg^{2+} ions were added to a solution of the squaramide-linked dinucleoside indicated that the squaramide internucleotide linkage coordinated to Mg^{2+} through the carbonyl oxygens. The squaramide internucleotide linkage was resistant to cleavage by snake venom phosphodiesterase. Although the squaramide is highly polarised, incorporation of this linkage nonetheless reduces electrostatic repulsion with a phosphodiester linkage in the opposing strand of a duplex, relative to the repulsion observed between two native phosphodiester linkages. Base-pairing with an opposing strand is therefore not disrupted at the site of the squaramide linkage, however the conformational properties of the squaramide internucleotide linkage confer a bent structure to the oligonucleotide and

causes a degree of overall destabilisation of the duplex. Sekine and co-workers have also reported oligonucleotides containing a squaramide as a 2'-5' internucleotide linkage.³²³

Another report showed that two oligonucleotides containing amino groups at the 3'- or 5'-position could be ligated through reaction with a squaryl ester to form a squaramide internucleotide linkage.³²⁴ The rate of this reaction was greatly increased through templating by a DNA or RNA strand complementary to the two amine-modified oligonucleotides. Also, the resulting squaramide-linked oligonucleotide product could be read-through by a DNA polymerase, and thus amplified by PCR. This allowed templated squaramide-ligation of two DNA oligonucleotides complementary to a target RNA sequence to be used as a non-enzymatic alternative to the reverse transcription step of RT-PCR.

Sekine and co-workers prepared mononucleoside derivatives bearing squaric acid monoamides at the 5' and 3' positions, which they proposed could function as phosphate mimics. They also reported a cytosine and a guanosine derivative containing 5'-3' endocyclic squaramides as mimics of cAMP and cGMP. Biological evaluation of these compounds was not reported however.³¹⁶

In addition to their use as phosphate bioisosteres, because of their aromaticity squaramides have also been investigated as a nucleobase replacement in nucleoside derivatives. These compounds were shown to have anticancer activity.³²⁵

1.7.4. Thiosquaramides

Thiosquaramides, in which the squaramide carbonyl groups are replaced by thiocarbonyl moieties, have been less well studied than their oxo analogues due to difficulties in synthesising these compounds arising from the instability of many thiosquaryl ester

intermediates. However, Rombola *et al.* recently developed a method for preparing thiosquaramides by treatment of dicyclopentyl squarate with Lawesson's reagent, followed by substitution with amines,³²⁶ facilitating further research into applications of the thiosquaramide functional group.

Thiosquaramides have significantly more acidic NH protons than squaramides, with typical pK_a values ranging between 4 and 9.³²⁷ One study found that a series of thiosquaramides were more acidic than the analogous squaramides by an average of 6.4 pK_a units.³²⁸ Thiosquaramides are thus even stronger hydrogen bond donors than squaramides,³²⁹ and therefore have been used in organocatalysis.^{330–332} Thiosquaramides are also more lipophilic than squaramides, and these properties have allowed them to be used as pH-controllable transporters of anions across phospholipid membranes, demonstrating the potential for biological applications of these compounds.^{327,333,334}

Due to the affinity of sulfur for zinc, thiosquaramides could prove to be more potent ZBGs than their oxo analogues. In one report, replacement of one of the squaramide carbonyl groups with a thiocarbonyl moiety in squaramide-based metalloprotease inhibitors improved the inhibitory effect, however analogues containing two thiocarbonyls were not investigated.²⁸⁴ The potential of thiosquaramides as ZBGs has yet to be fully explored.

1.8. Objectives of this work – the design of novel inhibitors and probes for SNM1A

Inhibition of interstrand crosslink repair enzymes such as SNM1A is a promising strategy for improving the efficacy of crosslinking chemotherapy drugs for cancer treatment. Initial studies have demonstrated the feasibility of developing SNM1A inhibitors,^{123–127} but the full potential of this enzyme as a drug target has yet to be explored. Furthermore, the biological role of SNM1A in DNA repair pathways and in cell cycle checkpoints has yet to

be fully elucidated, and there is a need for improved tools to facilitate the study of this enzyme. The work presented in this thesis details the development of modified nucleosides that bind to SNM1A. These compounds have been tested as inhibitors of SNM1A in *in vitro* assays, with several showing promising efficacy. It is anticipated that future incorporation of the modified nucleosides into oligonucleotides will allow for the development of probes to study SNM1A in a cellular environment.

Several modified nucleosides were previously tested as SNM1A inhibitors in the McGouran laboratory.^{125–127} These nucleosides were appended with ZBGs to target the zinc ion(s) in the SNM1A active site. ZBGs were incorporated at the 5'-position of several compounds,^{125–127} while only two nucleosides bearing ZBGs at the 3'-position were investigated.¹²⁷ The work presented in this thesis aimed to develop nucleoside-based SNM1A inhibitors of improved potency, through screening a broader range of ZBGs, and further investigating the effect of placing the ZBG at the 3'-position of the nucleoside scaffold. Nucleoside derivatives functionalised at the 3'-position with ZBGs previously untried against SNM1A, including sulfonamide, 2-thiophene carboxamide, oxime and hydroxylamine moieties, were prepared and tested as SNM1A inhibitors, as discussed in Chapter 2.

Another focus of this work has been the potential of squaramides to function as ZBGs in SNM1A inhibitors, as described in Chapter 3. Due to their versatility, squaramides have been employed in diverse areas such as supramolecular chemistry, organocatalysis, bioconjugation, and medicinal chemistry.²⁷² However, the use of squaramides as metal chelators in biological applications has not been extensively explored. An advantage of squaramides is that they are readily functionalisable, either through replacement of the

carbonyl groups with thiocarbonyls, or through derivatisation of squaryl esters *via* reaction with various different amines to produce squaramides appended with different substituents. Squaramides bearing substituents containing an additional oxygen atom, designed to provide an additional site for zinc-coordination and thereby improve binding to SNM1A were prepared, as well as a series of thiosquaramides. These compounds were tested in SNM1A inhibition assays, and their interaction with zinc ions in solution was also studied using UV-vis spectroscopy.

SNM1A contains a phosphate-binding pocket adjacent to the active site,⁷⁹ which has not been targeted by substrate-mimic inhibitors reported previously. This work therefore also sought to investigate the possibility of developing more potent inhibitors that utilise non-covalent interactions with amino acid residues in SNM1A's phosphate-binding pocket, in addition to binding to the catalytic metal ion(s), as reported in Chapter 4.

Investigations of nucleoside-based inhibitors for SNM1A are expected to inform the future development of molecular probes to facilitate study of this enzyme. Probes are envisaged based on a DNA oligonucleotide as a recognition element mimicking the natural substrate of SNM1A, bearing a ZBG appended to the 5'-terminal residue, and a biotin moiety or fluorophore as a detection element at the 3'-end (Figure 1.17). Attachment of the terminal residue at the 5'-end through a 2'-5' linkage will allow space for a ZBG to be incorporated at the 3'-position of this residue, locating the ZBG in the position normally occupied by the first phosphodiester to be hydrolysed by the enzyme.

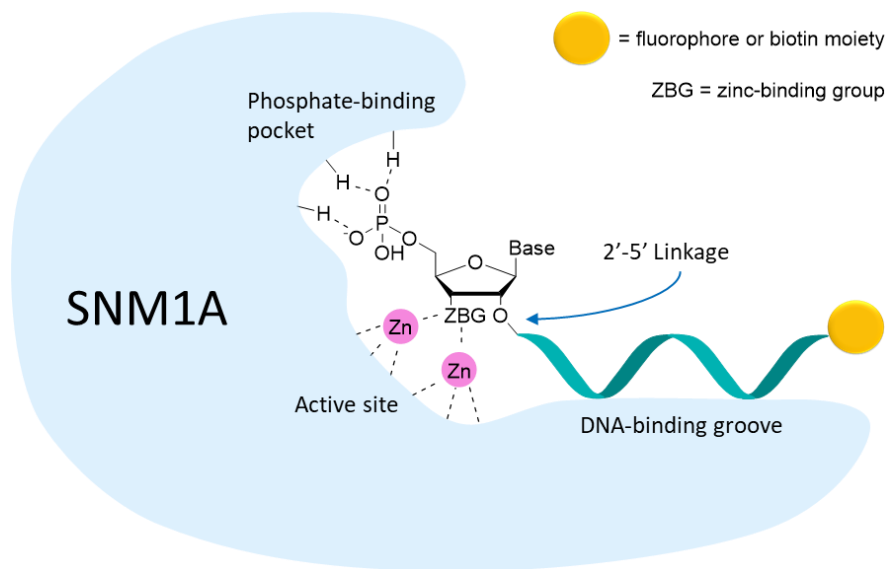


Figure 1.17. Proposed design of oligonucleotide probes for SNM1A.

2. Nucleosides bearing ZBGs at the 3'-position inhibit SNM1A

2.1. Introduction

Previously reported inhibitors of SNM1A include cephalosporins,¹²³ which show strong inhibition *in vitro* but have low membrane permeability. Several compounds identified through high-throughput screening of a library of bioactive molecules appear to inhibit SNM1A through an unclear mechanism of action.¹²⁴ A number of substrate-mimic inhibitors of SNM1A have been reported by the McGouran group. These are based on a nucleoside or oligonucleotide scaffold, and utilise squaramide,¹²⁵ hydroxamic acid,¹²⁶ and malonate¹²⁷ moieties as ZBGs. A broad range of potential ZBGs has yet to be explored for targeting this enzyme however, and the need remains to develop SNM1A inhibitors with greater efficacy.

Furthermore, the optimum placement of a ZBG in an inhibitor molecule has yet to be fully investigated. The use of a nucleoside scaffold offers the possibility of incorporating a ZBG at either the 5'- or the 3'-position of the compound (Figure 2.1). Substrate-mimic inhibitors previously reported feature 5'-ZBGs, other than two thymidine derivatives bearing malonate-based ZBGs at the 3'-position, which showed promising inhibitory activity.¹²⁷ Placement of a ZBG at the 3'-position may prove improve binding to SNM1A, and also allows for future optimisation of lead compounds through functionalisation of the 5'-position to target the phosphate-binding pocket adjacent to the SNM1A active site.⁷⁹

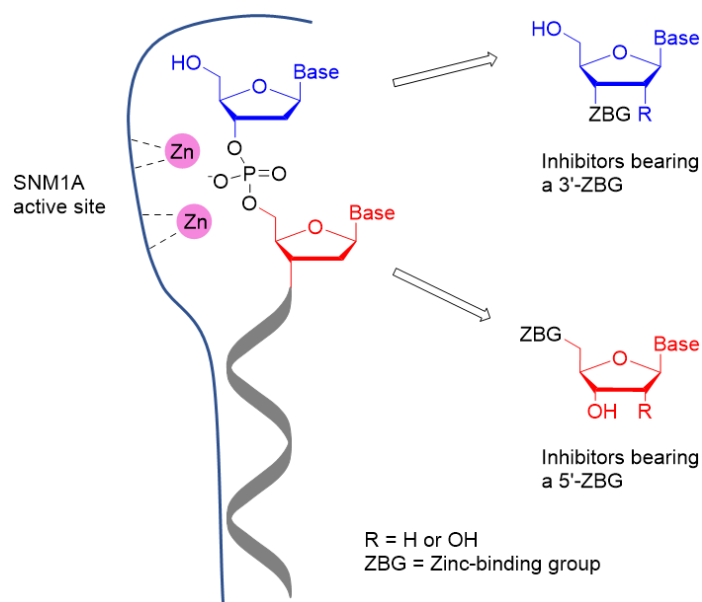


Figure 2.1. A schematic view of a DNA strand binding to the SNM1A active site showing the potential for development of substrate-mimic inhibitors bearing at ZBG at the either the 3'- or 5'-position.

In this chapter, the synthesis and *in vitro* testing of a series of novel substrate-mimic SNM1A inhibitors bearing ZBGs at the 3'-position is reported (Figure 2.2). Nucleoside derivatives bearing ZBGs previously untried against SNM1A, such as sulfonamide, 2-thiophene carboxamide, oxime and hydroxylamine moieties, have been prepared, with some showing promising inhibitory activity in *in vitro* assays. Hydroxamic acid and hydrazide ZBGs, previously tested at the 5'-position of SNM1A inhibitors, have been incorporated at the 3'-position of nucleoside derivatives.

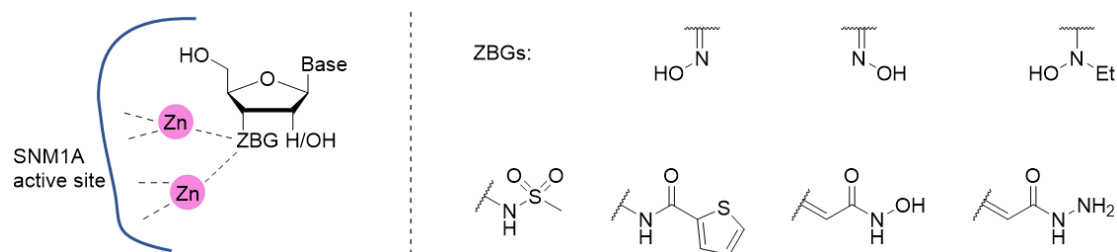


Figure 2.2. Schematic view of substrate-mimic inhibitors bearing ZBGs binding to the SNM1A active site.

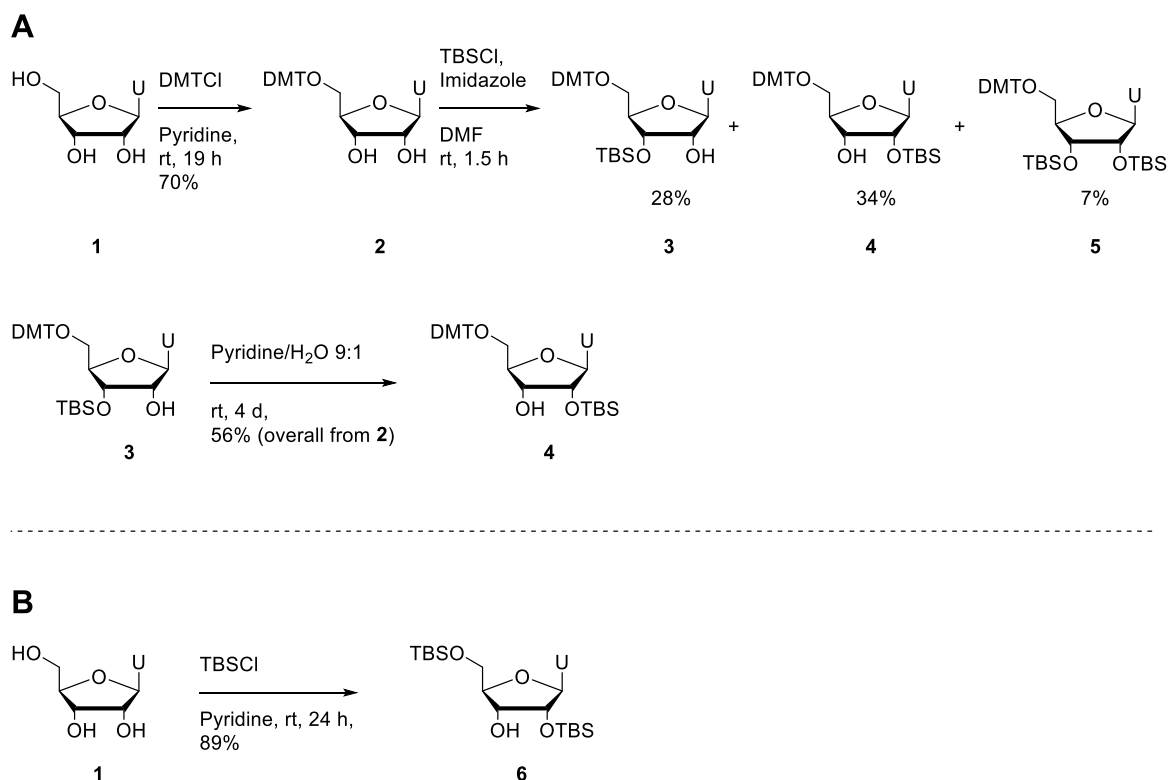
2.2. Synthesis

Candidate SNM1A inhibitors were prepared through derivatisation of uridine. Although SNM1A behaves as a DNA-processing enzyme *in vivo*, it has been shown that this enzyme can also hydrolyse RNA *in vitro*.^{36,52} It was therefore anticipated that the use of a ribonucleoside scaffold as the basis of the inhibitors would not preclude binding to SNM1A. The presence of a ribose in place of a deoxyribose ring in the candidate inhibitors increases the potential for optimisation of lead compounds through derivatisation at the 2'-position, or incorporation of successful inhibitors into oligonucleotides *via* 2'-5' linkages. Uridine was chosen over other nucleosides as a scaffold for the development of inhibitors to minimise the need for protecting group manipulations, as unlike other canonical RNA nucleobases, uracil does not contain an exocyclic amino group.

2.2.1. Protecting group strategies for derivatisation of the uridine 3'-position

To enable functionalisation of the 3'-position of uridine with ZBGs, uridine derivatives bearing protecting groups at the 2'- and 5'-positions were first prepared (Scheme 2.1). The dimethoxytrityl (DMT) group is commonly used to protect the 5'-position of nucleoside derivatives, and has the advantage that it can facilitate the future use of these compounds for applications involving solid-phase oligonucleotide synthesis. A uridine derivative **4** bearing a 5'-DMT and a 2'-*tert*-butyldimethylsilyl (TBS) protecting group was therefore

prepared. The DMT protecting group is unsuitable for some synthetic routes, however, due to its high sensitivity to acid, and a uridine **6** derivative bearing TBS protecting groups at both the 5'- and 2'-positions was therefore also prepared.



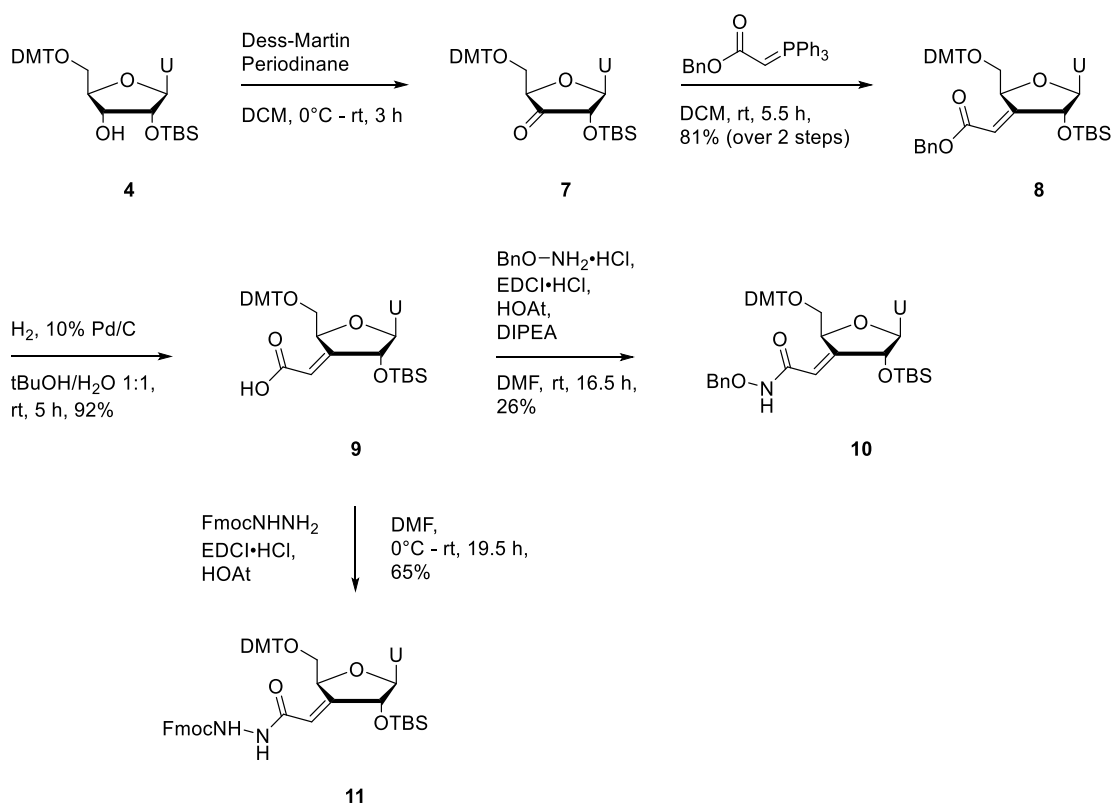
Scheme 2.1. A) Synthesis of uridine derivative **4** bearing a 5'-DMT and 2'-TBS protecting groups. B) Synthesis of uridine derivative **6** bearing TBS protecting groups at the 2'- and 5'-positions.

The 5'-hydroxyl group of uridine (**1**) was protected by reaction with DMTCl to afford the product **2** in 70% yield (Scheme 2.1A).³³⁵ Further reaction of DMT protected uridine **2** with TBSCl produced a mixture of mono-TBS protected uridine isomers **3** and **4** in 28% and 34% yield, respectively, as the steric bulk of a DMT group at the 5'-position appears to be insufficient to adequately differentiate the 3'-hydroxyl group from the 2'-hydroxyl moiety.³³⁶ A bis-TBS protected uridine derivative **5** was also obtained in 7% yield. Isomerisation of the undesired 3'-protected product **3** to the desired 2'-protected isomer **4** was achieved by stirring compound **3** in a pyridine/water mixture. Under these conditions

there is a dynamic equilibrium between isomers **3** and **4**. Following purification from the resulting mixture, the desired isomer **4** bearing 5'-DMT and 2'-TBS protecting groups was obtained in an overall yield of 56%. Uridine derivative **6** bearing TBS groups at both the 2'- and 5'-positions was prepared through reaction of uridine (**1**) with TBSCl (Scheme 2.1B).³³⁷ The 2',5'-protected isomer **6** was obtained as the major product in 89% yield.

2.2.2. Uridine derivatives bearing hydroxamic acid and hydrazide ZBGs

Hydroxamic acids are established zinc-binding groups which have been used in a range of metalloprotease inhibitors.³³⁸ A thymidine derivative containing a hydroxamic acid ZBG at the 5'-position has been reported to inhibit SNM1A *in vitro* with an IC₅₀ of 139 μM.¹²⁶ A homologue of this compound in which the hydroxamic acid ZBG was attached at the 5'-position *via* a longer alkyl chain showed greatly reduced inhibition of SNM1A, demonstrating the importance of the positioning of the ZBG relative to the nucleoside core of the inhibitor. Hydrazides are isosteres of hydroxamic acids, and have also been used as ZBGs.^{194,206} A thymidine derivative bearing a hydrazide at the 5'-position did not show inhibition of SNM1A, unlike the analogous hydroxamic-acid-bearing compound.¹²⁶ Hydrazides are known to be weaker ZBGs than hydroxamic acids, and thus hydrazide-based metalloenzyme inhibitors often have lower potency than analogous hydroxamic acids. However, the use of the weaker hydrazide ZBG in an inhibitor can confer greater selectivity, as non-covalent interactions between other parts of the inhibitor molecule and the enzyme play a greater role.^{206,207} It was hypothesised that placement of a hydroxamic acid or hydrazide ZBG at the 3'-position could lead to improved potency relative to previously reported inhibitors bearing these ZBGs at the 5'-position.¹²⁶ Hydroxamic acid **10** and hydrazide **11** were therefore prepared (Scheme 2.2).



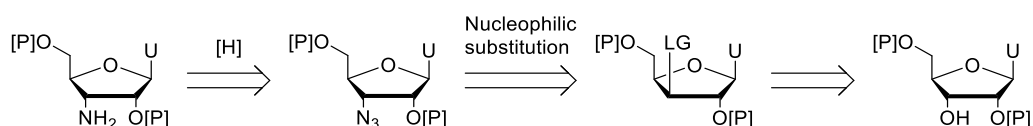
Scheme 2.2. Synthesis of uridine derivatives bearing hydroxamic acid and hydrazide ZBGs.

To facilitate the introduction of a hydroxamic acid or hydrazide moiety, a uridine derivative **9** containing a carboxylic acid appended to the 3' position was synthesised (Scheme 2.2). This is a useful intermediate, as it allows for a range of functionalities to be introduced through the formation of amide bonds. In a sequence analogous to one previously reported for the synthesis of similar nucleosides,³³⁹ protected uridine derivative **4** was oxidised to generate ketone **7**, which was further treated with a Wittig reagent to provide benzyl ester **8** in 81% yield over 2 steps. Hydrogenation of ester **8** yielded carboxylic acid **9** in 92% yield. Intermediate **9** was reacted with *o*-benzylhydroxylamine to afford protected hydroxamic acid **10**. Separately, carboxylic acid **9** was reacted with Fmoc-hydrazine to provide hydrazide **11**. Hydroxamic acid **10** and hydrazide **11** bear protecting groups appropriate for solid-phase oligonucleotide synthesis. Following deprotection of the 2'-TBS groups and phosphoramidite formation, it is anticipated that compounds **10** and **11** will be successfully

incorporated into oligonucleotide probes for SNM1A *via* 2'-5' linkages. Global deprotection of hydroxamic acid **10** and hydrazide **11** to test these monomers as inhibitors of SNM1A was foregone in order to conserve material for future oligonucleotide synthesis.

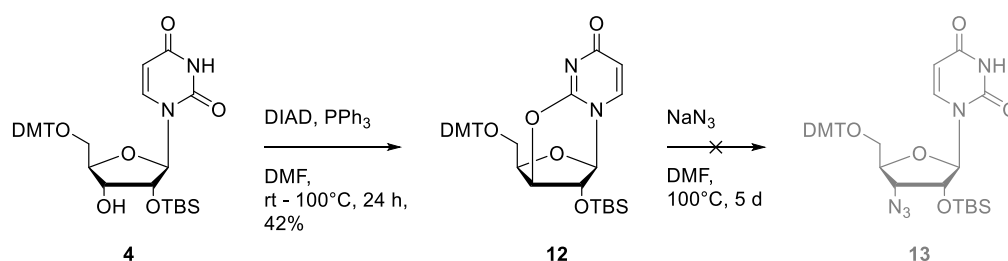
2.2.3. A uridine derivative bearing a 3'-amino group

While carboxylic acid **9** is a useful intermediate for the synthesis of nucleoside derivatives bearing hydroxamic acid and hydrazide ZBGs, the installation of other ZBGs, such as sulfonamide or squaramide moieties, requires a uridine derivative bearing a nucleophilic amino group at the 3'-position. To obtain a 3'-amino-3'-deoxyuridine derivative, a strategy of preparing an intermediate bearing a leaving group with inverted stereochemistry at the 3'-position, followed by displacement of the leaving group by an azide nucleophile, was initially pursued (Scheme 2.3).



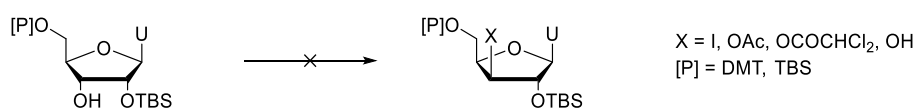
Scheme 2.3. Synthetic strategy for the preparation of a 3'-amino-3'-deoxyuridine derivative through displacement of a leaving group by an azide nucleophile. LG = leaving group.

An intramolecular Mitsunobu reaction of protected uridine derivative **4** was carried out to produce anhydrosugar **12** (Scheme 2.4). This anhydrosugar proved unreactive however, as ring-opening of **12** to generate azide **13** was unsuccessful despite the use of elevated temperature and extended reaction time.



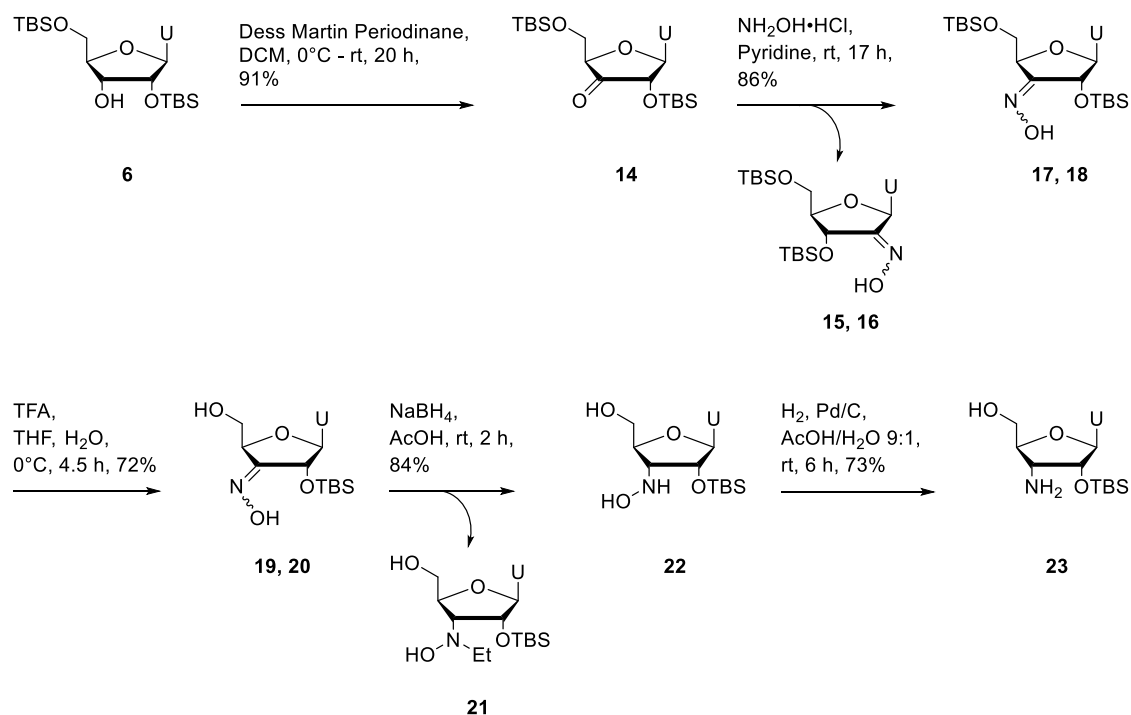
Scheme 2.4. Attempted synthesis of azide **13** through nucleophilic ring-opening of an anhydrosugar.

Attempts to introduce a more reactive leaving group with inverted stereochemistry at the 3'-position of uridine (Scheme 2.5) through an Appel reaction ($X = \text{I}$), Mitsunobu reaction ($X = \text{OAc}$, $X = \text{OCOCHCl}_2$) and through mesylation of the 3'-OH group followed by displacement of the mesylate in an $\text{S}_{\text{N}}2$ reaction with sodium hydroxide ($X = \text{OH}$) were also unsuccessful.



Scheme 2.5. Attempted synthesis of a uridine derivative bearing a leaving group with inverted stereochemistry at the 3'-position.

An alternative synthetic strategy for the synthesis of a 3'-amino-3'-deoxyuridine derivative through formation of a 3'-oxime moiety followed by stereoselective reduction was carried out (Scheme 2.6).



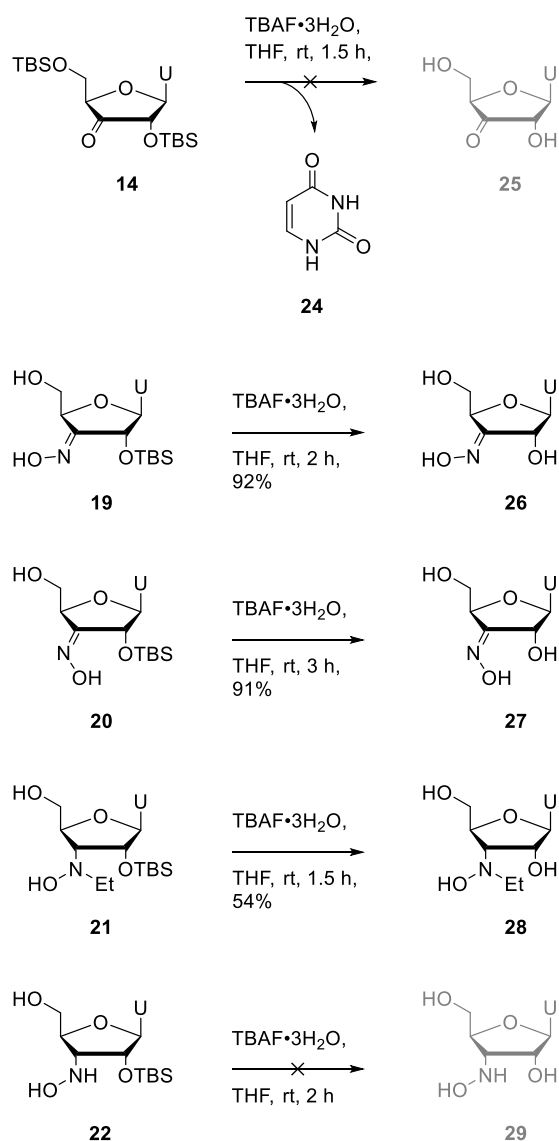
Scheme 2.6. Synthesis of 3'-aminouridine derivative **23**.

3'-Amino uridine derivative **23** was prepared from **6** through modification of a previously reported synthetic route (Scheme 2.6).^{340,341} Oxidation of protected uridine derivative **6** to ketone **14**, followed by condensation with hydroxylamine produced an oxime-bearing uridine derivative isolated as a mixture of E/Z isomers **17** and **18**. Formation of previously unidentified side products **15** and **16**, was also observed. These were presumably formed *via* silyl migration from the 2'- to the 3'-position in the enolate form of ketone **14**. Reducing the concentration of the reaction by approximately one third (from 0.88 M of hydroxylamine and 0.17 M of compound **14** to 0.56 M of hydroxylamine and 0.11 M of **14**) reduced the formation of side products **15** and **16** from 13% yield to a trace amount, and increased the yield of the desired products **17** and **18** from 67% to 86%. Attempts to reduce oxime isomers **17/18** were unproductive. However, after selective deprotection of the 5'-TBS group, E/Z oxime isomers **19** and **20** were successfully reduced using sodium triacetoxyborohydride formed *in situ* from sodium borohydride and acetic acid. Formation

of trace amounts of alkylated side product **21** was observed, likely due to acetylation of hydroxylamine product **22** by boron triacetate followed by reduction of the resulting amide. Hydroxylamine product **22** was formed stereoselectively, which can be attributed to coordination of the 5'-hydroxyl group of oxime **19/20** to boron.³⁴² Hydrogenation of hydroxylamine **22** afforded amine **23**, a key intermediate allowing further functionalisation of the 3'-position of the uridine scaffold with ZBGs.

2.2.4. Uridine derivatives bearing oxime and hydroxylamine ZBGs

Ketones have been used as ZBGs, as upon hydration in solution they form geminal diols which can chelate zinc ions through the two oxygen atoms.^{258,259} Oxime and hydroxylamine moieties contain lone pairs and thus can also potentially coordinate to a metal ion. Silyl deprotection of ketone **14**, oximes **19** and **20**, and hydroxylamine **21** was carried out in order to test these compounds as inhibitors of SNM1A (Scheme 2.7).



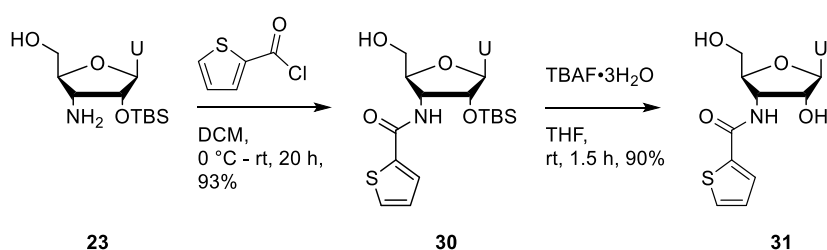
Scheme 2.7. Synthesis of uridine derivatives **26** and **27** bearing oxime ZBGs, and uridine derivative **28** bearing a hydroxylamine moiety. Attempted synthesis of uridine derivative **25** bearing a ketone ZBG and uridine derivative **29** bearing a hydroxylamine group.

Treatment of ketone **14** with tetrabutylammonium fluoride (TBAF) did not provide the desired product **25**, but instead lead to formation of uracil (**24**) in quantitative yield, likely due to deprotonation of ketone **14** at the 2'-position and formation of an enolate, followed by E1cb elimination. Oxime **19** however was successfully reacted with TBAF to provide the deprotected oxime **26** in 92% yield. Similarly, oxime **20** was deprotected to generate

uridine derivative **27** in 91% yield. Hydroxylamine **21** was deprotected to provide the product **28** in 54% yield. Treatment of hydroxylamine **22** with TBAF led to the formation of a product which could not be identified.

2.2.5. Uridine derivatives bearing sulfonamide and 2-thiophene carboxamide ZBGs

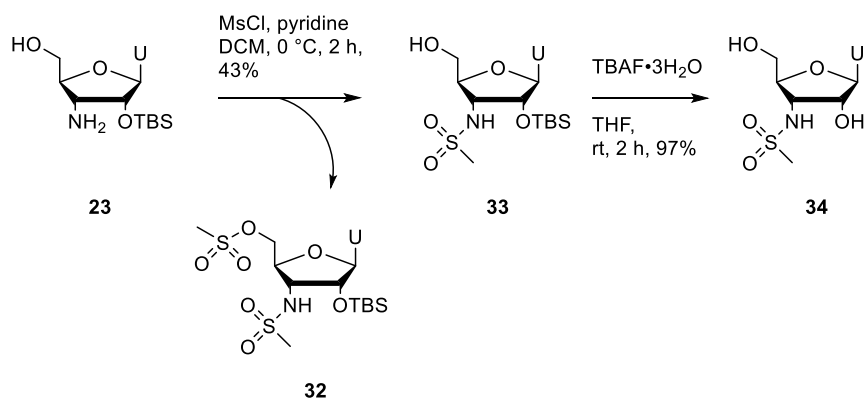
Zinc complexes of 2-thiophene carboxamides have been reported,³⁴³ and 2-thiophene carboxamides bearing a trifluoroacetyl substituent have been used as ZBGs in HDAC inhibitors.^{344,345} A uridine derivative bearing a 2-thiophene carboxamide was therefore prepared to test the efficacy of this ZBG for inhibition of SNM1A (Scheme 2.8).



Scheme 2.8. Synthesis of uridine derivative **31** bearing a 2-thiophene carboxamide ZBG.

3'-Amino uridine derivative **23** was reacted 2-thiophenecarbonyl chloride to prepare compound **30** in 93% yield (Scheme 2.8). Intermediate **30** was treated with TBAF to provide the target thiophene-bearing uridine derivative **31** in 90% yield.

Sulfonamides have been investigated as ZBGs in inhibitors of various enzymes,²³³ and in particular have proven effective for inhibition of carbonic anhydrases.^{234,235} In addition to their use as ZBGs, sulfonamides have been reported as phosphate isosteres, for example in tyrosine phosphatase inhibitors.²³⁷ This potential dual function as a ZBG and a phosphate isostere is desirable for a substrate-mimic SNM1A inhibitor, and a sulfonamide-bearing uridine derivative **34** was therefore prepared (Scheme 2.9).



Scheme 2.9. Synthesis of uridine derivative **34** bearing a sulfonamide ZBG.

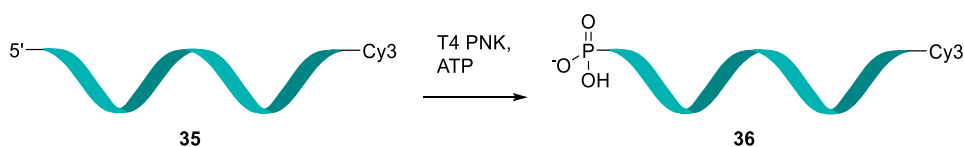
Reaction of amine **23** with mesyl chloride in the presence of a catalytic amount of pyridine provided sulfonamide-containing uridine derivative **33** in 43% yield (Scheme 2.9).

Compound **32** bearing a mesylate at the 5'-position in addition to a 3'-sulfonamide was also obtained from this reaction. The desired sulfonamide **33** was deprotected with TBAF to provide the target sulfonamide-bearing uridine derivative **34** in 97% yield.

2.3. Biochemical testing

2.3.1. Gel electrophoresis assay

Exonuclease activity of SNM1A can be evaluated using a polyacrylamide gel electrophoresis (PAGE) based assay.^{123,125} This assay was performed in the presence of candidate SNM1A inhibitors to determine their efficacy. A 21mer ssDNA oligonucleotide **35** labelled with a Cy3 fluorophore at the 3'-end, provided by the research group of Prof. Tom Brown, was phosphorylated at the 5'-end using bacteriophage T4 polynucleotide kinase (T4 PNK) to produce oligonucleotide **36** (Scheme 2.10). The 5'-phosphorylated oligonucleotide **36** can function as a substrate of SNM1A *in vitro*.



Scheme 2.10. Enzymatic phosphorylation of fluorophore-labelled oligonucleotide **35** to produce SNM1A substrate **36**.

The candidate SNM1A inhibitors **26**, **27**, **28**, **31**, and **34** were each incubated at a concentration of 1 mM in solution with SNM1A at 37 °C for 5 minutes. Oligonucleotide substrate **36** was added and the reaction mixtures were further incubated for 1 hour. Upon incubation with SNM1A oligonucleotide **36** is digested from the 5'-end, producing a shorter oligonucleotide, which in turn can be further digested by the enzyme. While SNM1A is known to hydrolyse higher molecular weight substrates processively, short oligonucleotides such as **36** are digested with low processivity; the enzyme dissociates from the substrate after the hydrolysis of a phosphodiester linkage and another binding event must take place before further digestion.³⁶ The length of the oligonucleotide products is therefore directly related to the efficacy of the inhibitor present. The extent of digestion of the oligonucleotide substrate was visualised using PAGE (Figure 2.3). SNM1A does not show significant nuclease activity on DNA strands of 8 nucleotides or less,⁵² and digestion of the oligonucleotide substrate to this length therefore indicates full activity of the enzyme (Figure 2.3A). The presence of longer oligonucleotides in the assay mixture is indicative of effective inhibition of SNM1A (Figure 2.3B). Thymidine was included in this assay as a control to verify that inhibitory effects result from the inclusion of ZBGs.

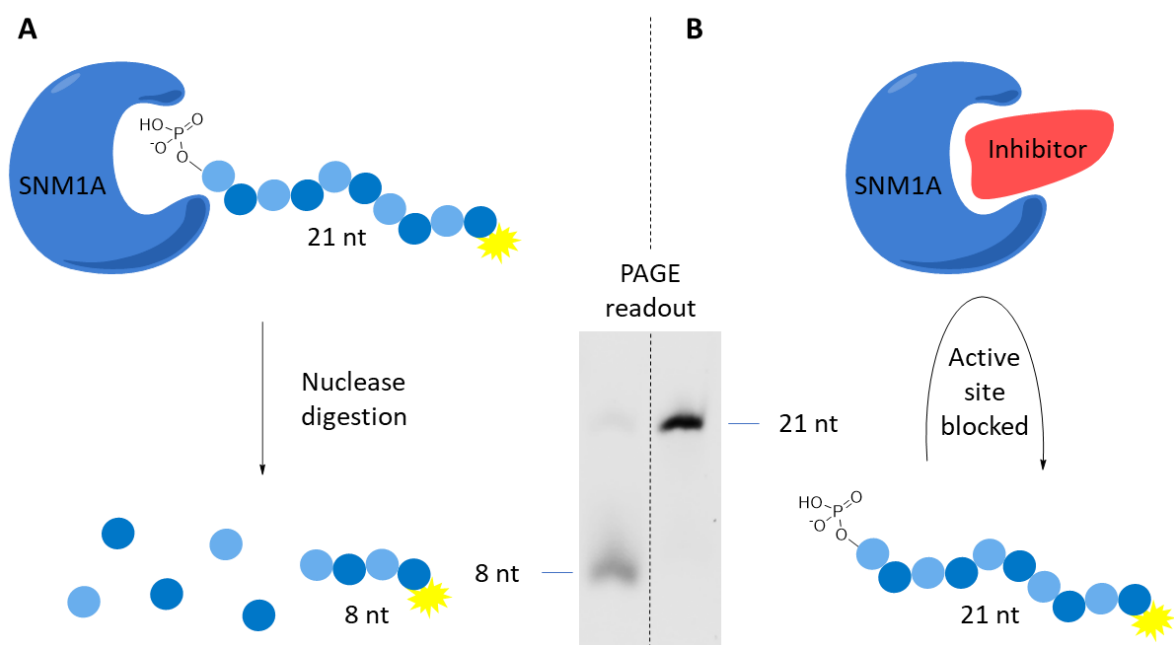


Figure 2.3. A PAGE-based assay to determine the efficacy of SNM1A inhibitors. **A)** In the absence of an effective inhibitor a 21mer fluorescent oligonucleotide substrate is digested to produce shorter fluorescent products. Full activity of the enzyme results in a product 8 nucleotides in length. **B)** In the presence of an effective inhibitor the substrate is not digested and a 21mer fluorescent oligonucleotide is visible in the PAGE readout. nt = nucleotides.

This initial gel electrophoresis-based screening showed moderate inhibition of the enzyme occurred in the presence of sulfonamide **34** and hydroxylamine **28** (Figure 2.4). Interestingly, despite their structural similarity to hydroxylamine **28**, oximes **26** and **27** do not inhibit SNM1A. 2-Thiophene carboxamide **31** also does not appear to inhibit the enzyme.

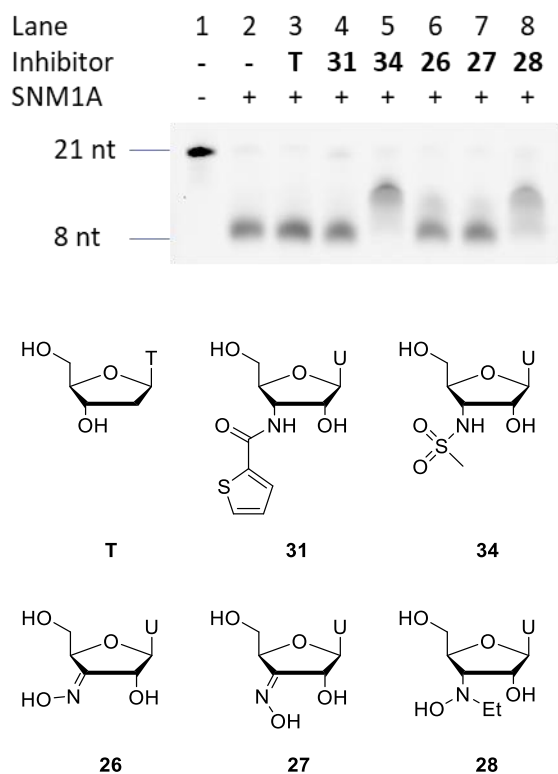


Figure 2.4. Evaluation of uridine derivatives **26-28**, **31**, and **34** through visualisation of the extent of digestion of a 21mer fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides (1 mM) for 5 minutes before the oligonucleotide substrate (80 nM) was added, and then incubated for a further 1 hour. Thymidine (T) was included in this assay as a control. nt = nucleotides.

Following their identification as candidate inhibitors, hydroxylamine **28** and sulfonamide **34** were tested individually in a similar gel-electrophoresis-based assay at concentrations ranging from 1 mM to 0.3 μ M (Figure 2.5). The results of this assay showed that while sulfonamide **34** exhibited more promising inhibition than hydroxylamine **28**, neither compound showed appreciable inhibition of SNM1A at concentrations lower than 333 μ M.

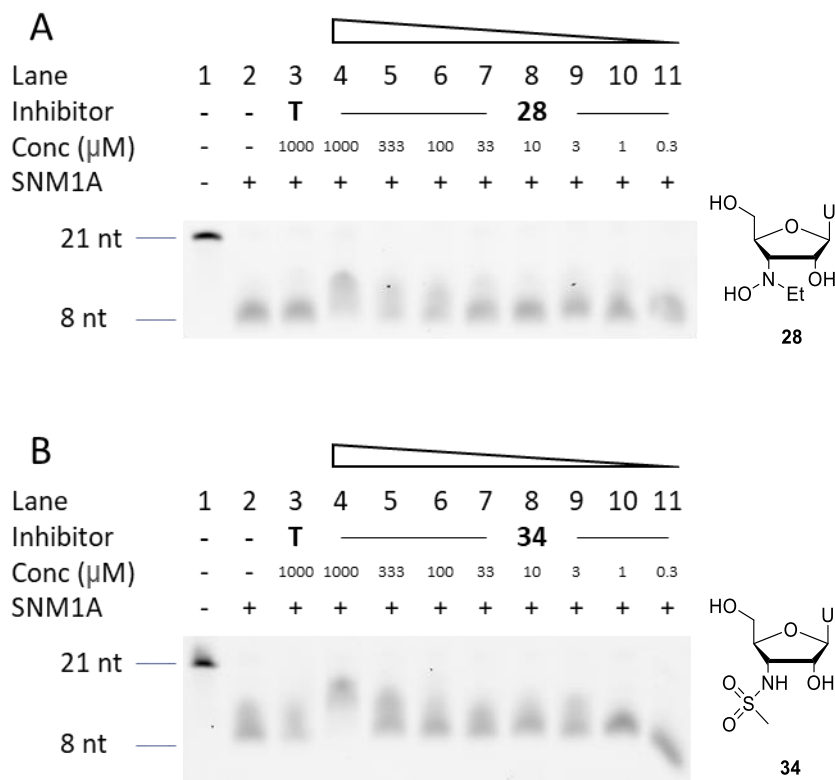


Figure 2.5. Evaluation of SNM1A inhibitors **28** and **34** at a range of concentrations from 1 mM to 0.3 μM through visualisation of the extent of digestion of a fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides for 5 min before the 21mer oligonucleotide substrate (80 nM) was added, and the mixture was incubated for a further 60 min. nt = nucleotides.

2.4. Conclusion

In summary, a series of uridine derivatives bearing ZBGs at the 3'-position have been synthesised. Protecting group strategies to enable selective functionalisation of the 3'-position of uridine were developed, and a uridine derivative bearing a nucleophilic amino group was prepared, allowing for the installation of ZBGs. The modified nucleosides bearing ZBGs were tested as inhibitors of the DNA repair enzyme SNM1A in gel-electrophoresis-based assays. Compound **34** bearing a sulfonamide, a ZBG not previously tested against this enzyme, and compound **28** bearing a hydroxylamine moiety, which is

not well established as a ZBG, demonstrated some inhibitory activity. Nucleoside derivatives bearing hydroxamic acid and hydrazide ZBGs were also prepared for future use in solid-phase oligonucleotide synthesis to develop probes to study SNM1A. Optimisation of the candidate SNM1A inhibitors through incorporation of a 5'-phosphate is described in Chapter 4.

3. Squaramide and thiosquaramide inhibitors of SNM1A

3.1. Introduction

A crystal structure (PDB: 4B87) of a truncated version of SNM1A shows a single zinc ion in the active site.⁷⁹ However a crystal structure (PDB: 5AHO) of the closely related enzyme SNM1B shows two zinc ions in the active site.⁷⁹ It has been suggested that the active form of SNM1A contains a second zinc ion, or perhaps another metal cation, which is more loosely bound and therefore not observed in the crystal structure.³⁶ These two zinc ions are predicted to activate a water molecule for nucleophilic attack on a phosphodiester in the DNA backbone, and stabilise a build-up of negative charge on the phosphodiester as it undergoes hydrolysis.

Many substrate-mimic inhibitors of SNM1A previously developed have utilised ZBGs such as a sulfonamide or hydroxylamine (Chapter 2), or hydroxamic acid¹²⁶ group, designed to interact with one zinc ion. One report however described several thymidine derivatives bearing bifunctional malonate groups designed to chelate two zinc ions.¹²⁷ Although precise IC₅₀ values for these malonate-based inhibitors were not reported, the most potent compound showed a degree of SNM1A inhibition at a concentration as low as 33 μM in a gel-based assay, a marked improvement compared to a structurally similar inhibitor bearing a hydroxamic acid ZBG. While this result indicates the potential benefit of targeting the proposed two zinc ions in the SNM1A active site using novel ZBGs, there remains a need for further development of substrate-mimic SNM1A inhibitors to fully investigate this under-studied enzyme as a drug target.

In this chapter, the synthesis and testing of a series of substrate-mimic SNM1A inhibitors bearing squaramide or thiosquaramide ZBGs is described. Squaramides are synthetically versatile and can be derivatised through substitution of squaryl esters with amines bearing

a variety of substituents. It was therefore envisaged that a squaramide bearing an appropriate substituent could be developed as a ZBG capable of interacting with two zinc ions. Squaramides are known to chelate cations through their two carbonyl oxygen atoms,²⁸³ and *N*-hydroxysquaramides, which bear an additional hydroxy group that can contribute to metal coordination, have shown promise as ZBGs in inhibitors for metalloproteases.^{284,285} Furthermore, squaramides can also function as phosphate bioisosteres,^{318,322} which is desirable for mimicking the natural DNA substrate of SNM1A. An oligonucleotide bearing a squaramide at the 5'-terminus has been shown to bind to SNM1A.¹²⁵ It was therefore hypothesised that squaramides bearing substituents containing an additional oxygen atom could show improved binding to SNM1A due to chelation of the proposed second zinc ion in the SNM1A active site (Figure 3.1A).

Thioquaramides have also been investigated in this work as ZBGs for use in substrate-mimic SNM1A inhibitors. Due to the affinity of sulfur for zinc, and the greater polarisation of thiosquaramides, it was hypothesised that they could prove to be more potent ZBGs than their oxo analogues. For some squaramide-based metalloprotease inhibitors, replacement of one of the squaramide carbonyl groups with a thiocarbonyl moiety improved the inhibitory effect, however analogues containing two thiocarbonyls have not been reported to date.²⁸⁴

A range of nucleoside derivatives bearing differently substituted squaramides and thiosquaramides at the 3'- or 5'-position, as well as a dinucleoside linked through a thiosquaramide in place of a phosphodiester, were designed and synthesised (Figure 3.1B). Several modified ribonucleosides were prepared for comparison with deoxyribonucleoside inhibitors, as SNM1A can hydrolyse RNA as well as DNA *in vitro*.^{36,52} These compounds have

been tested *in vitro* as inhibitors of SNM1A, with several showing inhibitory activity. The interaction of these compounds with zinc ions has been studied in UV-vis titrations, and these results allow the relative potency observed in SNM1A inhibition assays to be rationalised. The membrane permeability of the candidate SNM1A inhibitors has been quantified using a parallel artificial membrane permeability assay (PAMPA) to further evaluate their potential for *in vivo* applications.

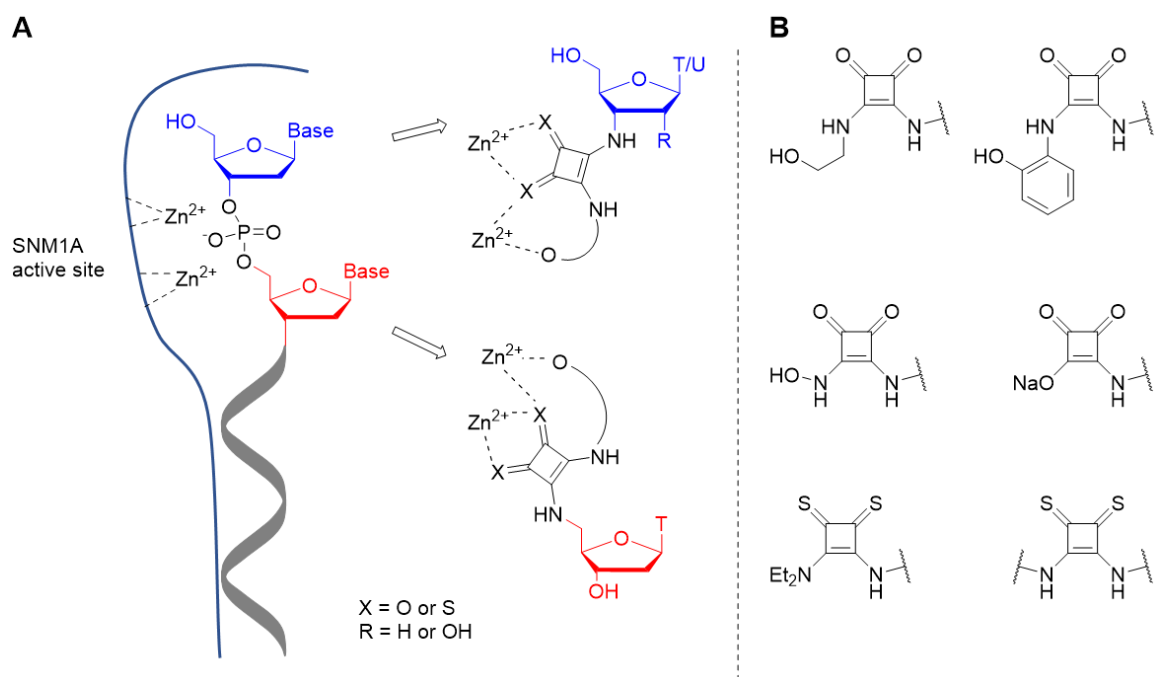
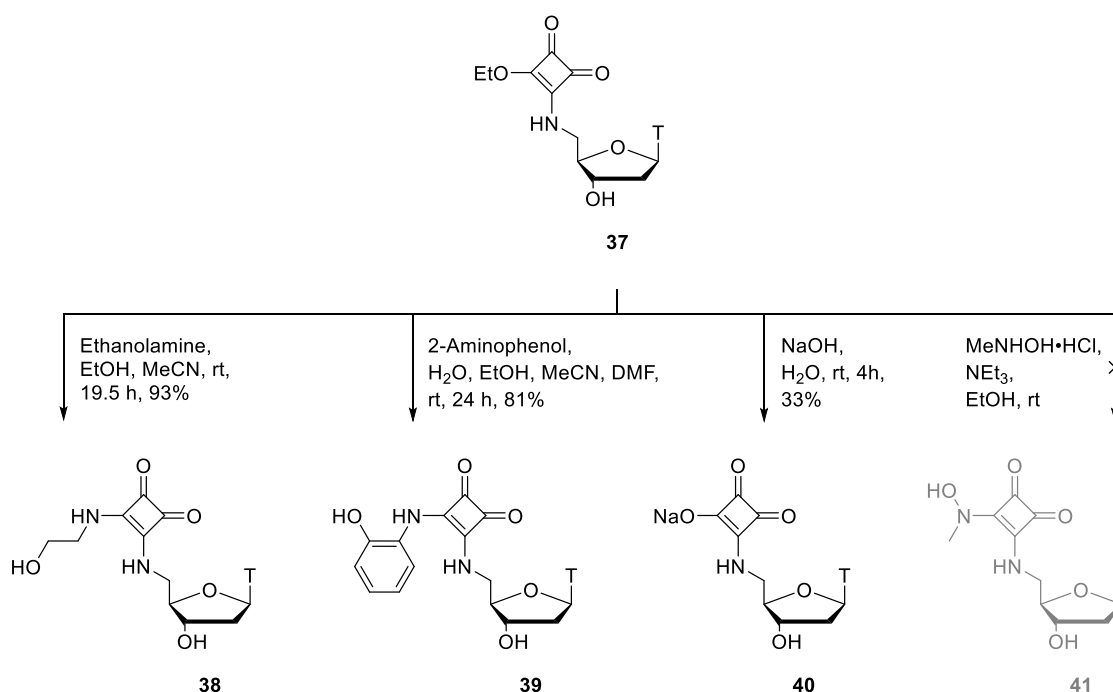


Figure 3.1. A) Design of squaramide-bearing substrate mimic inhibitors of SNM1A based on a schematic view of substrate binding to the enzyme active site. B) Structures of squaramide and thiosquaramide motifs used as zinc-binding groups in SNM1A inhibitors.

3.2. Synthesis

3.2.1. Thymidine derivatives bearing squaramide ZBGs at the 5'-position

A series of thymidine derivatives **38-40** bearing a squaramide group at the 5'-position were synthesised (Scheme 3.1). Squaramides **38** and **39** bear substituents containing an additional hydroxyl group, while compound **40** contains a squaric acid moiety, aimed at improving binding to active site zinc ions in SNM1A.



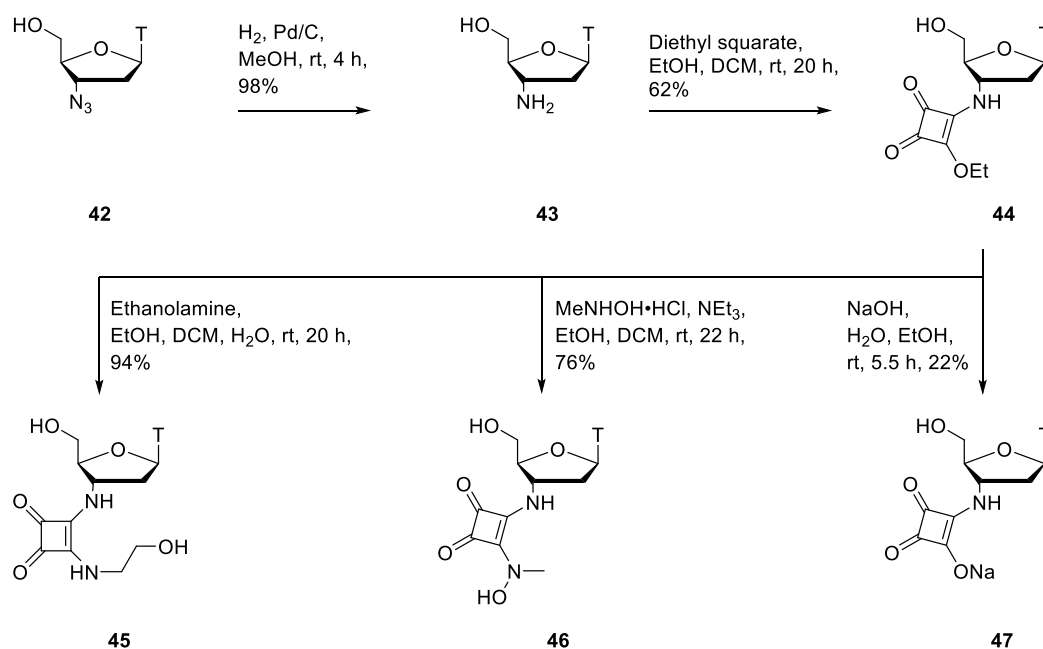
Scheme 3.1. Synthesis of thymidine derivatives **38-40** bearing squaramides at the 5'-position.

Intermediate **37**, provided by Dr. Eva-Maria Dürr,^{125,126} was used for the generation of further functionalised squaramides **38-40**. Intermediate **37** was reacted with ethanolamine to prepare squaramide **38** in 93% yield. Separately, reaction of squaryl monoamide **37** with 2-aminophenol produced squaramide **39**, which is analogous to squaramide **38** in the positioning of the side-chain hydroxy group, but has less conformational flexibility. It was hypothesised that this could improve binding to zinc due to a lower entropic penalty. Intermediate **37** was also hydrolysed under basic conditions to provide squaric acid **40** in

33% yield.³¹⁶ Reaction of intermediate **37** with *N*-methylhydroxylamine hydrochloride did not result in formation of the desired *N*-hydroxysquaramide product **41**.

3.2.2. Thymidine derivatives bearing squaramide ZBGs at the 3'-position

Thymidine derivatives bearing squaramides at the 3'-position (**44-47**) were synthesised for comparison in biochemical assays with compounds **37-40** bearing the squaramide moiety at the 5'-position (Scheme 3.2).

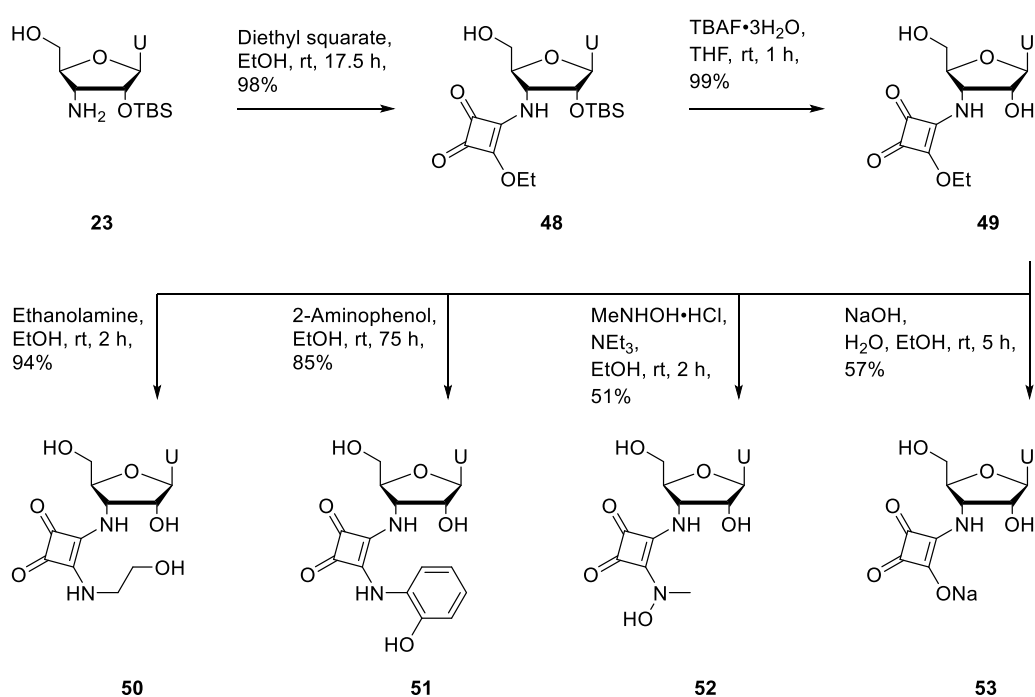


Scheme 3.2. Synthesis of thymidine derivatives **44-47** bearing squaramides at the 3'-position.

3'-Amino-3'-deoxythymidine (**43**) was generated in 98% yield *via* reduction of commercially available AZT (**42**).¹²⁷ Amine **43** was reacted with diethyl squarate to prepare squaryl monoamide **44**, a key intermediate for preparation of further functionalised squaramides **45-47**. Intermediate **44** was reacted with ethanolamine, and separately with *N*-methylhydroxylamine hydrochloride, to furnish squaramide **45** and *N*-hydroxysquaramide **46**, respectively. Hydrolysis of squaryl monoamide **44** under basic conditions yielded squaric acid derivative **47** in a modest 22% yield.

3.2.3. Uridine derivatives bearing squaramide ZBGs at the 3'-position

In vitro studies have shown that SNM1A can hydrolyse RNA as well as DNA,^{36,52} indicating that modified ribonucleosides could potentially function as SNM1A inhibitors. Furthermore, use of a ribonucleoside scaffold in place of a deoxyribonucleoside provides the potential for further functionalisation at the 2'-position for future optimisation of lead compounds. Several uridine derivatives bearing squaramides at the 3'-position were therefore synthesised (Scheme 3.3).



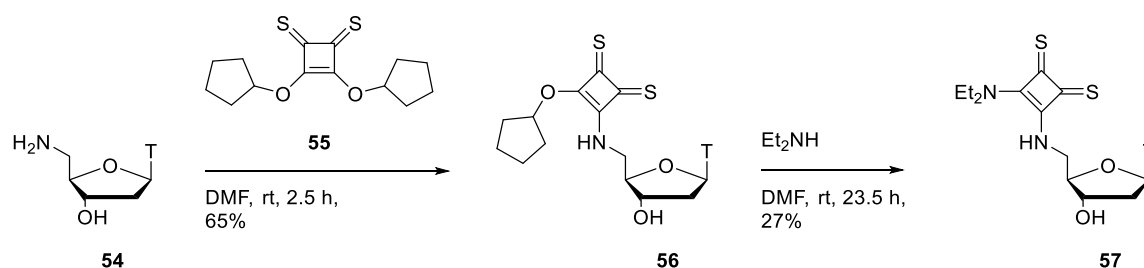
Scheme 3.3. Synthesis of uridine derivatives **49-53** bearing squaramides at the 3'-position.

Uridine derivative **48**, containing a squaramide at the 3'-position, was prepared in 98% yield by reaction of amine **23** with diethyl squarate. The silyl protecting group of squaramide **48** was removed with TBAF to provide the product **49** in 99% yield. Squaryl monoamide **49** acted as a key intermediate in the preparation of squaramides **50-53**. Squaryl monoamide **49** was treated with ethanolamine, and separately with 2-aminophenol to produce squaramides **50** and **51**, respectively. *N*-Hydroxysquaramide **52**

was prepared in 51% yield by reaction of squaryl monoamide **49** with *N*-methylhydroxylamine hydrochloride. Hydrolysis of squaryl monoamide **49** under basic conditions provided squaric acid derivative **53** in 57% yield.

3.2.4. Thymidine derivatives bearing thiosquaramide ZBGs

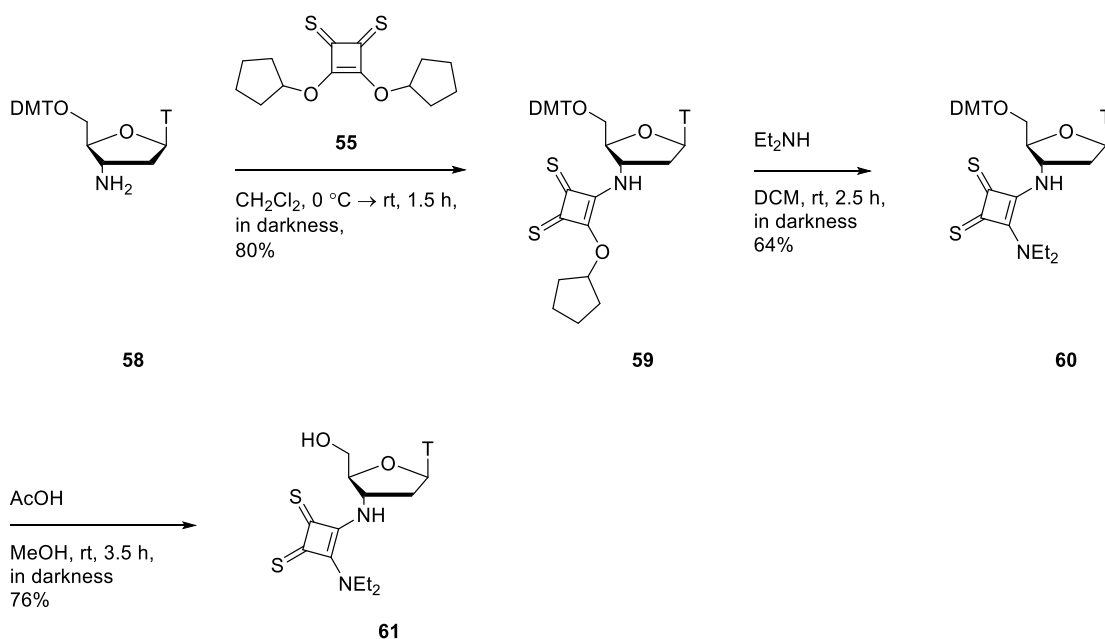
A series of thiosquaramide-bearing nucleosides **57**, **61** and **63** were prepared to test whether these sulfur-containing functional groups would prove to be more effective inhibitors of the zinc metalloenzyme SNM1A than their oxo analogues. 5'-Aminothymidine **54**, provided by Dr. Eva-Maria Dürr, was reacted with dicyclopentyl dithiosquarate³²⁶ **55** to prepare thiosquaryl monoamide **56** in 65% yield (Scheme 3.4). Thiosquaryl monoamide **56** was reacted with diethylamine to provide thiosquaramide **57** in 27% yield.



Scheme 3.4. Synthesis of a thymidine derivative **57** bearing a 5'-thiosquaramide.

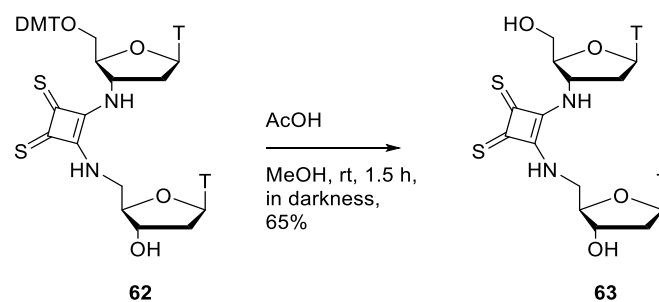
To prepare a thymidine derivative bearing a thiosquaramide at the 3'-position, amine **58**, prepared by Dr. William Doherty, was reacted with dicyclopentyl dithiosquarate **55** to provide thiosquaryl monoamide **59** (Scheme 3.5). Initially a low yield of compound **59** was obtained, however when the reaction was carried out in the dark, the product **59** was generated in 80% yield. It was observed that a number of the other thiosquaramides decomposed after exposure to light for several hours. This is consistent with the reported photoactivity of thiocarbonyls, which can be excited by visible light.^{346,347} Following this observation care was taken to shield thiosquaramides from light during subsequent

reactions. Further reaction of thiosquaryl monoamide **59** with diethylamine provided thiosquaryl diamide **60** in 64% yield. Deprotection of thiosquaramide **60** furnished the target compound **61** in 76% yield.



Scheme 3.5. Synthesis of a thymidine derivative **61** bearing a 3'-thiosquaramide.

It was hypothesised that binding to SNM1A could be improved by using a dinucleoside inhibitor rather than a mononucleoside, more closely mimicking the natural DNA substrate and increasing the number of possible stabilising non-covalent interactions with the enzyme. This strategy was previously applied in the development of SNM1A inhibitors bearing malonate/malonamide ZBGs; dinucleosides linked through a malonamide were slightly more effective inhibitors than their mononucleoside analogues bearing a malonamide at the 3'-position.¹²⁷ A dinucleoside **63** containing a thiosquaramide linkage was therefore prepared (Scheme 3.6). 5'-DMT protected intermediate **62**, provided by Dr. William Doherty, was deprotected under acidic conditions to provide the target dinucleoside **63** in 65% yield.



Scheme 3.6. Synthesis of a dinucleoside **63** containing a bridging thiosquaramide in place of a phosphodiester linkage.

3.3. Biochemical testing of squaramides and thiosquaramides as SNM1A inhibitors

3.3.1. Gel-electrophoresis-based assays

The ability of the squaramide- and thiosquaramide-bearing nucleoside derivatives to inhibit SNM1A was first evaluated using a gel electrophoresis-based assay (Figure 3.2), as previously described (Chapter 2, pp. 61-63). Thymidine was included in this assay as a control to verify that inhibitory effects result from the squaramide or thiosquaramide modifications.

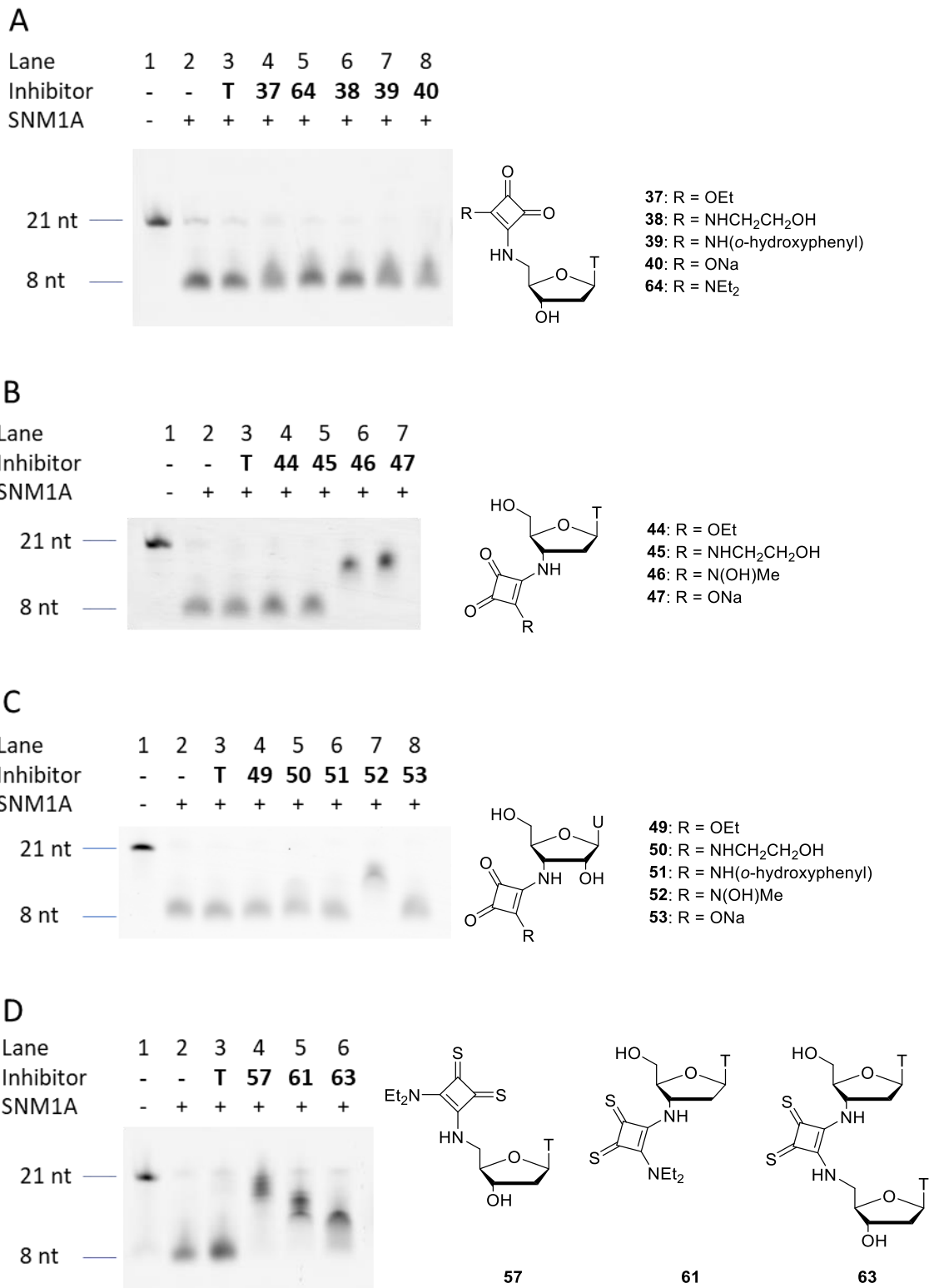


Figure 3.2. Evaluation of SNM1A inhibitors through visualisation of the extent of digestion of a 21mer fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides (1 mM) for 5 minutes before the oligonucleotide

substrate (80 nM) was added, and then incubated for a further 1 hour. nt = nucleotides. A) Evaluation of thymidine derivatives **37-40** bearing a squaramide at the 5'-position. B) Evaluation of thymidine derivatives **44-47** bearing a squaramide at the 3'-position. C) Evaluation of uridine derivatives **49-53** bearing a squaramide at the 3'-position. D) Evaluation of thiosquaramides **57, 61** and **63**.

The results of the inhibitor screen show that thymidine derivatives **37-40** bearing squaramide groups at the 5'-position do not inhibit SNM1A to any significant extent (Figure 3.2A). Compound **64**, a squaramide without a substituent that could contribute to chelation of zinc, was prepared as previously described¹²⁶ and tested as an additional control in this experiment, exhibiting no inhibition of SNM1A. This is consistent with previous results that showed compounds **37** and **64** do not inhibit SNM1A.¹²⁶ Squaramides **38-40** did not show improved binding to SNM1A despite their additional functionalisation. Testing of thymidine derivatives **44-47** bearing squaramides at the 3'-position however revealed that compounds **46** and **47** inhibit SNM1A (Figure 3.2B), as does uridine derivative **52** (Figure 3.2C), an analogue of thymidine derivative **46**. Squaramide-bearing nucleoside derivatives thus appear to inhibit SNM1A more effectively when the squaramide moiety is placed at the 3'-position rather than the 5'-position of the inhibitor. Compound **47** bearing a 3'-squaric acid inhibits SNM1A, while compound **40** bearing a 5'-squaric acid does not. The squaric acid and *N*-hydroxysquaramide moieties appeared to be more effective than the other squaramides tested.

Thiosquaramides **57, 61**, and **63** were also screened in the gel-electrophoresis based assay. Contrary to the trend observed for squaramides, thymidine derivative **57** bearing a thiosquaramide at the 5'-position showed greater inhibition of SNM1A than thymidine derivative **61** bearing a thiosquaramide at the 3'-position (Figure 3.2D). This may indicate

a different binding mode to SNM1A due to the larger steric bulk of thiosquaramides relative to their oxo analogues. Thiosquaramide-linked dinucleoside **63** inhibited SNM1A less effectively than either of the thiosquaramide-bearing mononucleosides **57** and **61**. This was likely due to the conformational rigidity of the thiosquaramide linkage of dinucleoside **63**, restricting the molecule to a less favourable binding conformation, different to that of the natural DNA substrate. Inclusion of a squaramide internucleotide linkage in oligonucleotides has previously been found to cause distortion of normal duplex structure.³²²

Following their identification as candidate SNM1A inhibitors, compounds **46**, **47**, **52**, and **57** were tested individually at concentrations ranging from 1 mM to 3 μ M (Figure 3.3). This showed that thymidine *N*-hydroxysquaramide **46** is a slightly stronger inhibitor, showing inhibition at 333 μ M, (Figure 3.3A) than its uridine analogue **52**, which only shows significant inhibition at 1 mM (Figure 3.3C). When tested simultaneously in the same assay (Figure 3.2B), thymidine derivative **47** bearing a squaric acid moiety appears to show slightly more inhibition of SNM1A at 1 mM than thymidine derivative **46** bearing an *N*-hydroxysquaramide, although it is difficult to confirm this difference in inhibitory effect based on individual testing of compounds **46** and **47** (Figures 3.3A and 3.3B) as both show a degree of inhibition at 333 μ M. *N*-Hydroxysquaramides such as **46** and **52** can chelate to zinc through formation of a 6-membered ring, while chelation of squaric acid **47** to zinc through formation of a 5-membered ring is considered less favourable due to the larger bite angle.³⁴⁸ However, these results indicate that increased electrostatic interaction with zinc due to the negative charge of compound **47** in its deprotonated form may be a more important factor in determining inhibitory potency. Thymidine derivative **57** bearing a

thiosquaramide at the 5'-position appears to be the most potent of all the SNM1A inhibitors tested, showing inhibition at a concentration of 100 μM (Figure 3.3D).

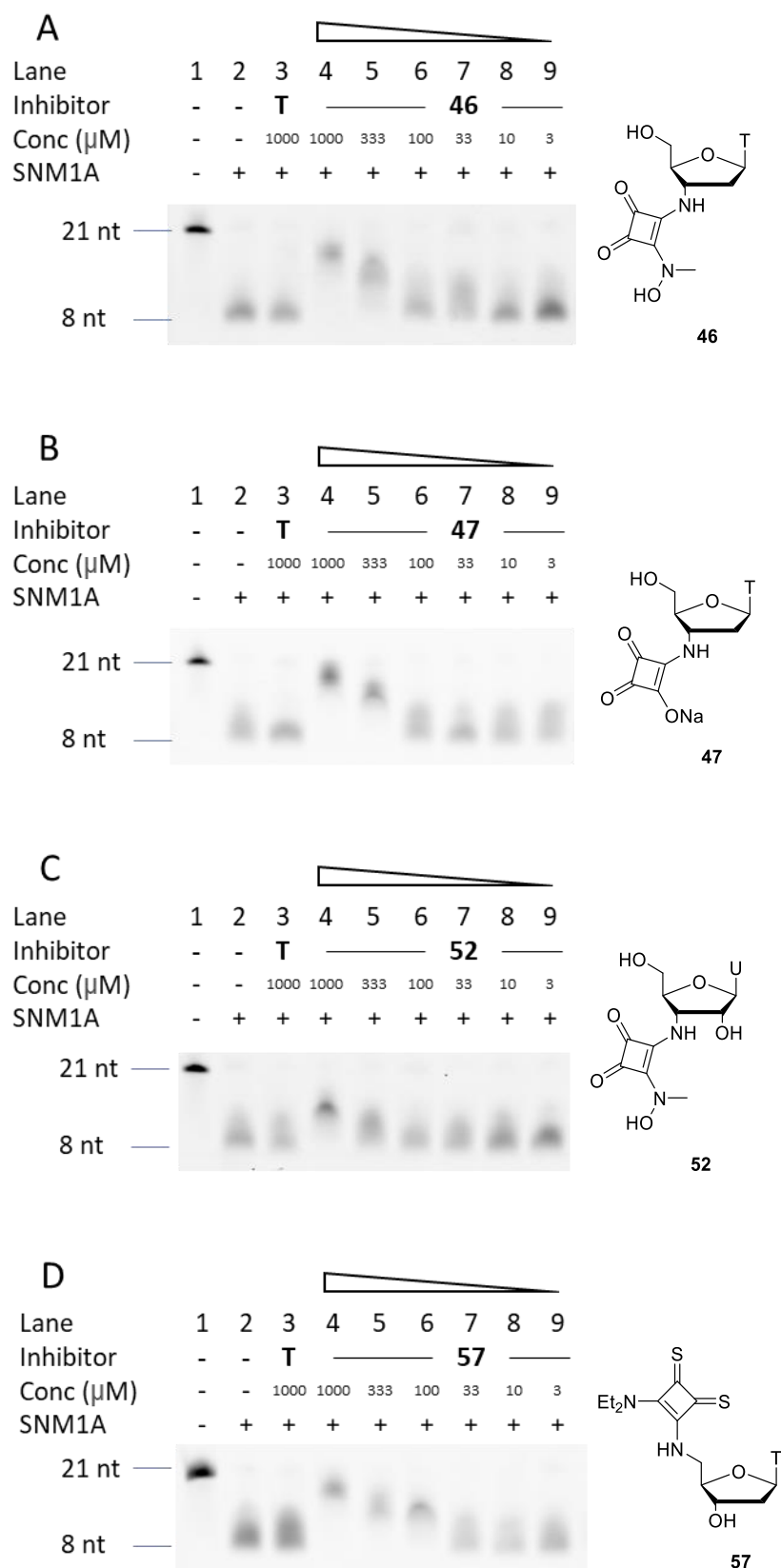
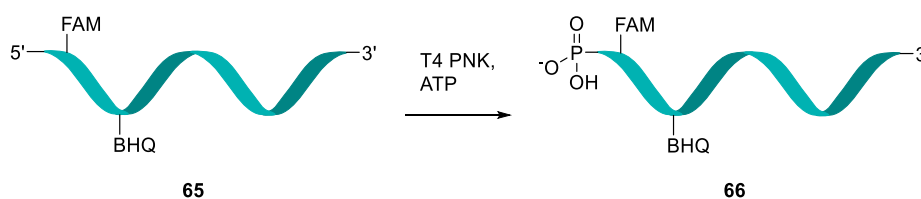


Figure 3.3. Evaluation of SNM1A inhibitors at a range of concentrations from 1 mM to 3 μ M through visualisation of the extent of digestion of a fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides for 5 min before the 21mer oligonucleotide substrate (80 nM) was added and a further 60 min incubation. nt = nucleotides. A) Evaluation of thymidine 3'-*N*-hydroxysquaramide **46**. B) Evaluation of thymidine 3'-squaric acid **47**. C) Evaluation of uridine 3'-*N*-hydroxysquaramide **52**. D) Evaluation of thymidine 5'-thiosquaramide **57**.

3.3.2. IC₅₀ determination

Following identification of the most promising compounds **47** and **57**, a real-time fluorescence assay was carried out,^{123,126} to determine the IC₅₀ values for inhibition of SNM1A. Oligonucleotide **65** bearing a fluorescein (FAM) fluorophore and a black hole quencher (BHQ) moiety, obtained commercially, was phosphorylated at the 5'-end using T4 polynucleotide kinase to produce oligonucleotide **66** (Scheme 3.7), which can function as a substrate for SNM1A.



Scheme 3.7. Enzymatic phosphorylation of oligonucleotide **66**. FAM = fluorescein, BHQ = black hole quencher, T4 PNK = T4 polynucleotide kinase.

Digestion of substrate oligonucleotide **66** by SNM1A led to separation of FAM from BHQ and a consequent increase in fluorescence. Measurement of increasing fluorescence signal over time allowed the rate of the nuclease reaction to be quantified. Carrying out this experiment in the presence of inhibitors at a range of concentrations allowed for IC₅₀ values to be determined (Figure 3.4).

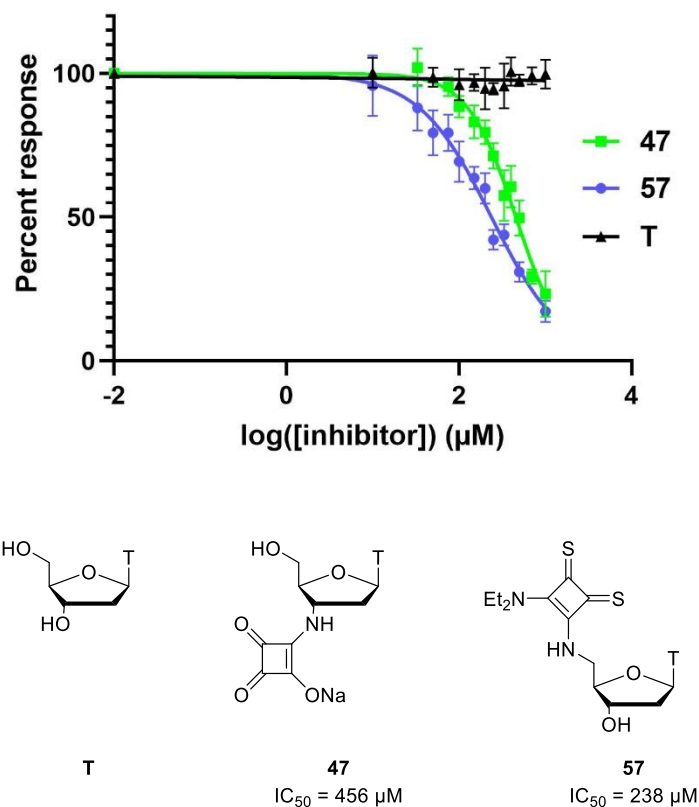


Figure 3.4. IC₅₀ determination of modified nucleosides **47** and **57** in a real-time fluorescence assay. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides at a range of concentrations for 6.5 min before the 20mer oligonucleotide substrate (125 nM) was added. Fluorescent readings were taken every 30 seconds for 35 min. For each concentration of inhibitor, fluorescence intensity was plotted against time, and the slope of the linear range of this graph was taken as the rate of enzyme activity. The rate of enzyme activity observed for the zero-inhibitor control was set as 100, and the other rates were normalised to this to obtain percent response values. Percent response was plotted against log([inhibitor]), and a nonlinear regression was performed to calculate the IC₅₀. Thymidine (**T**) is included as a control. Error bars were generated from 6 independent repeats.

Unmodified thymidine was tested in this assay as a control, showing no inhibition of SNM1A (Figure 3.4). Thymidine derivative **57** bearing a 5'-thiosquaramide again proved to be the most potent inhibitor, with an IC₅₀ of 238 μM. The IC₅₀ of thymidine derivative **47** with a squaric acid moiety at the 3'-position was found to be 456 μM.

Thymidine derivative **46** bearing a 3'-*N*-hydroxysquaramide and thymidine derivative **61** bearing a 3'-thiosquaramide were also tested in this assay (Figure 3.5) for comparison with squaric acid **47** and 5'-thiosquaramide **57**, to confirm the trends in inhibitory potency seen in the gel electrophoresis assays.

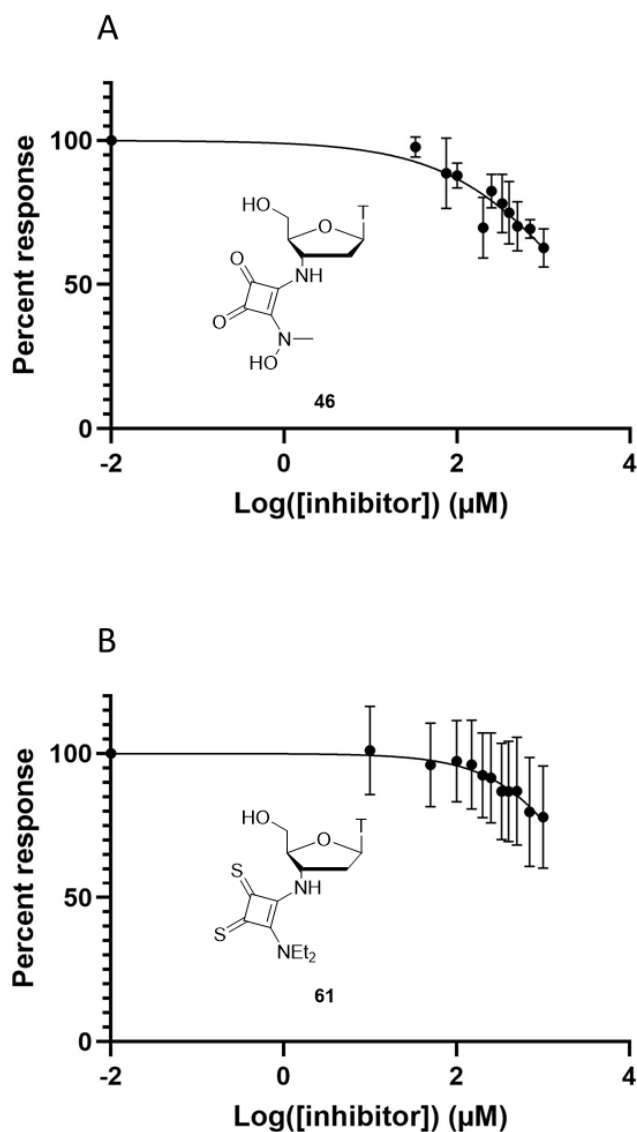


Figure 3.5. IC₅₀ measurement of modified nucleosides in a real-time fluorescence assay. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides at a range of concentrations for 6.5 min before the 20mer oligonucleotide substrate (125 nM) was added. Fluorescent readings were taken every 30 seconds for 35 min. For each concentration of inhibitor, fluorescence intensity was plotted

against time, and the slope of the linear range of this graph was taken as the rate of enzyme activity. The rate of enzyme activity observed for the zero-inhibitor control was set as 100, and the other rates were normalised to this to obtain percent response values. Percent response was plotted against $\log([\text{inhibitor}])$, and a nonlinear regression was performed. Error bars were generated from 6 independent repeats. A) IC_{50} measurement of *N*-hydroxysquaramide **46**. B) IC_{50} measurement of 3'-thiosquaramide **61**.

Some inhibition of SNM1A by *N*-hydroxysquaramide **46**, and to a lesser extent by thiosquaramide **61**, was observed (Figure 3.5), however the IC_{50} values for these compounds could not be accurately determined as they appeared to be above the concentration range tested. These results confirm the trend observed in the gel electrophoresis assays; 5'-thiosquaramide **57** is a more potent inhibitor than 3'-thiosquaramide **61**, and thymidine derivative **47** with a squaric acid moiety at the 3'-position is a more potent inhibitor than thymidine derivative **46** bearing an *N*-hydroxysquaramide at the 3'-position. 5'-Thiosquaramide **57** is the most effective of all the inhibitors tested.

3.3.3. UV-vis titrations

To investigate whether SNM1A inhibition occurs through coordination of active site zinc ions by the squaramide moieties, and to rationalise the relative inhibitory potency of compounds **46**, **47** and **57**, solutions of these compounds in MeCN were used in UV-vis titrations with $Zn(ClO_4)_2$.

The titration data were analysed by global non-linear regression using the ReactLab Equilibria software to elucidate binding modes. The experimental binding constants obtained are summarised in Table 3.1.

<u>Ligand-metal binding constants</u>			
<u>Ligand</u>	K _{1:1}	K _{1:2}	K _{2:1}
46	3.0x10 ⁴ M ⁻¹	6.6x10 ¹ M ⁻¹	-
47	4.0x10 ⁴ M ⁻¹	-	3.2x10 ⁶ M ⁻¹
57	6.8x10 ⁵ M ⁻¹	1.7x10 ² M ⁻¹	1.2x10 ⁶ M ⁻¹

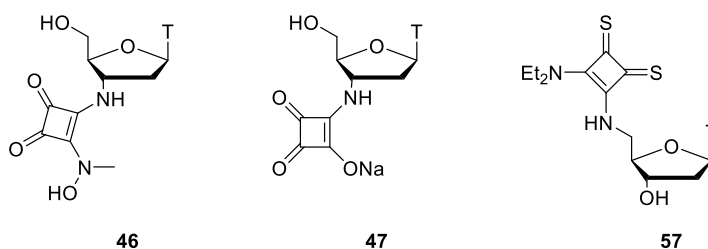


Table 3.1. Equilibrium constants for binding of squaramides to Zn²⁺ in MeCN calculated from UV-vis titration data.

For thymidine derivative **46** bearing a 3'-*N*-hydroxysquaramide, the best fit to the data was obtained using a two-species model in which **46** binds to a metal ion in both a 1:1 and 1:2 ligand-metal ratio. The equilibrium constant calculated for formation of the 1:1 species was 3.0x10⁴ M⁻¹, and the equilibrium constant for formation of the 1:2 species was significantly lower at 6.6x10¹ M⁻¹. Thymidine derivative **47** bearing a squaric acid at the 3'-position exhibited stronger binding to zinc, with an equilibrium constant of 4.0x10⁴ M⁻¹ for 1:1 ligand-metal binding. The data indicated squaric acid **47** could participate 2:1 ligand-metal binding in MeCN solution with an equilibrium constant of 3.2x10⁶ M⁻¹, although a 2:1 ligand-metal binding mode is unlikely to be accommodated within the SNM1A active site.

The strongest 1:1 zinc binding was observed when $\text{Zn}(\text{ClO}_4)_2$ was titrated against thymidine derivative **57** bearing a thiosquaramide at the 5'-position. The data from this titration were best fit by a three-species model in which ligand **57** binds to metal ions in a 1:1, 2:1, and 1:2 ratio. The equilibrium constant calculated for 1:1 ligand-metal binding was $6.8 \times 10^5 \text{ M}^{-1}$. For 2:1 ligand-metal binding the calculated equilibrium constant was $1.2 \times 10^6 \text{ M}^{-1}$, and 1:2 ligand-metal binding was less significant with an equilibrium constant of $1.7 \times 10^2 \text{ M}^{-1}$. Although these titrations should be considered as qualitative, given that MeCN solution is not a close representation of the SNM1A active site environment, the order of 1:1 binding strength to zinc ions observed, thiosquaramide **57** > squaric acid **47** > *N*-hydroxysquaramide **46**, is the same as the trend in the potency of SNM1A inhibition observed in both the gel electrophoresis and real-time fluorescence assays. This supports the hypothesis that the compounds inhibit SNM1A through binding to the active site zinc ion(s).

3.3.4. Membrane permeability assay

To evaluate the potential of the SNM1A inhibitors for future use in *in vivo* applications, a parallel artificial membrane permeability assay (PAMPA)³⁴⁹ was carried out to examine the membrane permeability of the biologically active compounds **46**, **47**, **52**, **57**, **61**, and **63** (Figure 3.6). Two drug molecules, carbamazepine and furosemide, were also tested in the PAMPA assay as controls. Carbamazepine was found to have a $\log P_e$ value of -5.3 ± 0.07 , indicating good permeability, and furosemide was found to have a $\log P_e$ value of -7.1 ± 0.007 , indicating impermeability. These results were in agreement with previously reported values.³⁴⁹ Thymidine derivative **46** bearing 3'-*N*-hydroxysquaramide, and thymidine derivative **47** bearing a squaric acid at the 3'-position were found to be membrane permeable, with $\log P_e$ values of -4.6 ± 0.4 and -4.1 ± 0.6 , respectively. Unlike its

thymidine analogue **46**, uridine derivative **52** bearing a 3'-*N*-hydroxysquaramide was found to be impermeable, with a $\log P_e$ of -7.0 ± 0.1 . Of the thiosquaramides, 5'-thiosquaramide **57** was found to have a low degree of membrane permeability, with a $\log P_e$ value of -6.5 ± 0.6 . 3'-Thiosquaramide **61** and bridging thiosquaramide **63** showed no membrane permeability, with $\log P_e$ values of -7.01 ± 0.04 and -7.084 ± 0.005 , respectively. The three most potent inhibitors, **46**, **47**, and **57**, all show some degree of membrane permeability, demonstrating their potential for future biological applications.

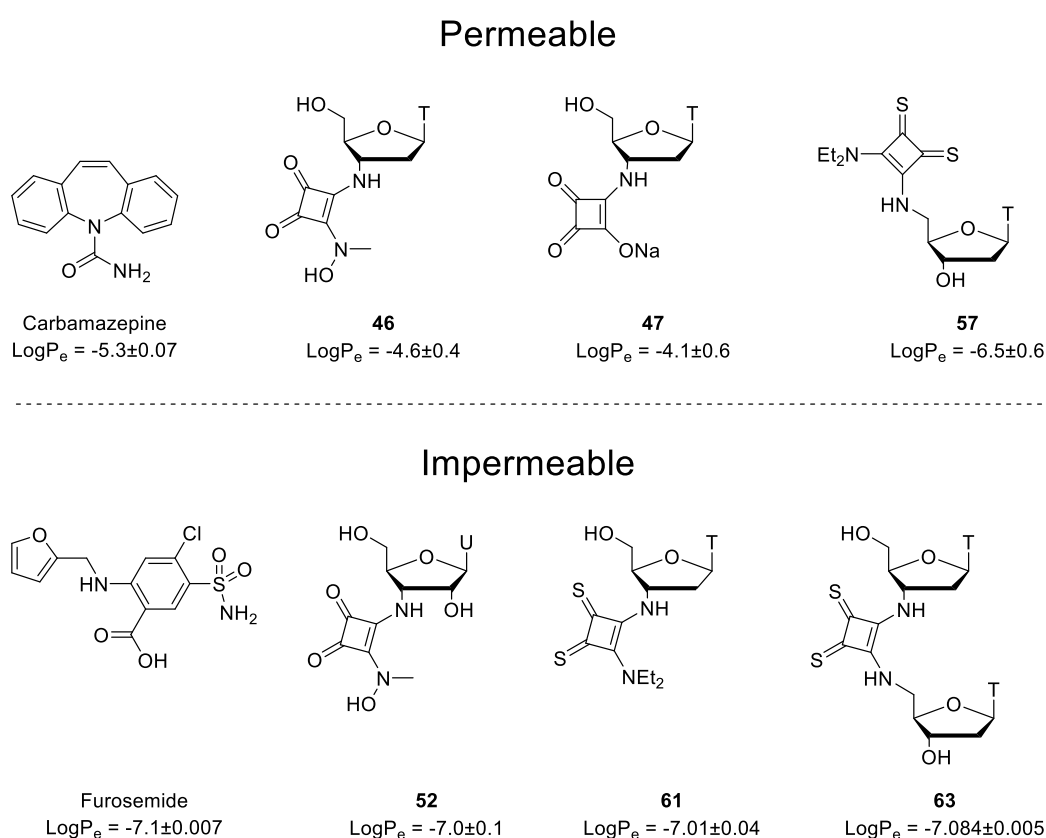


Figure 3.6. $\log P_e$ values for active SNM1A inhibitors obtained in a PAMPA assay. Carbamazepine and Furosemide were included as controls.

3.4. Conclusion

In summary, a range of nucleoside derivatives bearing squaramide and thiosquaramide modifications have been synthesised. These compounds were tested as inhibitors of

SNM1A in gel electrophoresis and real-time fluorescence assays. Their physical properties were also examined in a PAMPA assay to determine membrane permeability, and in UV-vis titrations to investigate their interaction with zinc ions. Although compounds bearing a squaramide group at the 5'-position proved ineffective, nucleoside derivatives **46**, **47**, and **52** bearing a squaramide moiety at the 3'-position demonstrated inhibition of SNM1A. Gel electrophoresis assays showed that thymidine derivative **46** bearing an *N*-hydroxysquaramide at the 3'-position is a more effective inhibitor of SNM1A than its uridine analogue **52**. This is consistent with the enzyme's preference for DNA substrates over RNA. Compound **47** bearing a squaric acid moiety at the 3'-position was a more effective inhibitor than either of the *N*-hydroxysquaramides **46** and **52**. UV-vis titrations showed that this could be attributed to a stronger interaction of squaric acid **47** with zinc ions compared to *N*-hydroxysquaramide **46**. Interestingly, thymidine derivative **57** bearing a thiosquaramide at the 5'-position proved to be a more effective inhibitor of SNM1A than compound **61** in which the thiosquaramide is placed at the 3'-position, contrary to the trend observed for the oxo squaramides. This may be due to displacement of the nucleobase and the deoxyribose ring from their ideal binding position in the SNM1A active site, caused by placement of the more sterically bulky thiosquaramide moiety at the 3'-position, resulting in steric clashes with amino acid residues. Although the testing of the oxo squaramides suggests that the ideal placement of a less bulky ZBG appears to be at the 3'-position, 5'-thiosquaramide **57** is nonetheless the most potent of any of the inhibitors tested, with an IC₅₀ value of 238 μM. UV-vis titrations with Zn(ClO₄)₂ indicate that this may be due to a much stronger interaction of the thiosquaramide with zinc ions compared to that observed for oxo squaramides. Finally, an assay was carried out to determine the membrane permeability of the SNM1A inhibitors to evaluate their potential for future *in*

in vivo applications. Squaramides **46** and **47** showed a high degree of membrane permeability, while the most potent inhibitor thiosquaramide **57** was permeable to a lesser degree. Taken together, these results demonstrate the potential of squaramide and thiosquaramide moieties for targeting zinc-dependent enzymes in biological applications, and provide insights useful for the further development of substrate-mimic inhibitors of SNM1A. The main results of this work are summarised in Table 3.2.

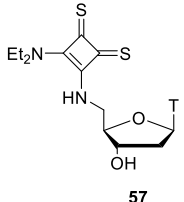
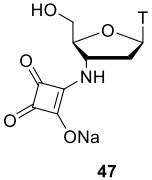
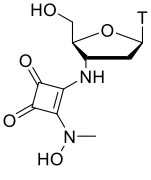
	 57	 47	 46
UV-vis	$\text{Zn} + \text{L} \xrightleftharpoons{K = 6.8 \times 10^5 \text{ M}^{-1}} \text{ZnL}$	$\text{Zn} + \text{L} \xrightleftharpoons{K = 4.0 \times 10^4 \text{ M}^{-1}} \text{ZnL}$	$\text{Zn} + \text{L} \xrightleftharpoons{K = 3.0 \times 10^4 \text{ M}^{-1}} \text{ZnL}$
Gel-based assays	Strongest inhibition	Intermediate inhibition	Intermediate inhibition
IC ₅₀	238 μM	456 μM	n.d.
PAMPA	LogP _e = -6.5±0.6	LogP _e = -4.1±0.6	LogP _e = -4.6±0.4

Table 3.2. Summary of data from UV-vis titrations, gel-electrophoresis-based assays, IC₅₀ determination, and PAMPA assays for the three most promising SNM1A inhibitors **46**, **47**, and **57**.

4. 5'-Phosphorylation increases the efficacy of nucleoside inhibitors of SNM1A

4.1. Introduction

Previous research on the development of nucleosides bearing ZBGs for targeting SNM1A, produced promising initial results, with several molecules showing efficacy as SNM1A inhibitors, as described in Chapters 2 and 3. The work presented in this chapter aimed to improve the potency of these compounds by exploiting non-covalent interactions with amino acid residues in a phosphate-binding pocket adjacent to the enzyme's active site,⁷⁹ in addition to binding to the catalytic zinc ion(s).

The phosphate-binding pocket in SNM1A recognises a terminal phosphate at the 5'-end of a DNA strand. This terminal phosphate moiety in the substrate is required for SNM1A activity.⁵² The closely related enzyme SNM1B also contains a phosphate-binding pocket, which has been characterised in greater detail, and amino acid residues in the phosphate-binding pocket are conserved between SNM1A and SNM1B.⁷⁹ A phosphate moiety at the 5'-end of a substrate molecule is also required for SNM1B activity, and mutation of residues in the phosphate-binding pocket substantially reduces SNM1B nuclease activity.⁸¹ A crystal structure (PDB: 7A1F) of SNM1B in complex with two 2'-deoxyadenosine 5'-monophosphate (AMP) molecules has been reported (Figure 4.1).⁸¹ The two AMP molecules are bound in the active site, in a position consistent with that predicted for the two 5'-terminal residues of a DNA substrate undergoing hydrolysis. Extensive hydrogen bonding interactions occur between the phosphate moiety of one co-crystallised AMP molecule (AMP-1) and amino acid residues in the phosphate-binding pocket of SNM1B. This crystal structure also shows π -cation interactions between the adenine nucleobase of AMP-1 and the guanidinium groups of two arginine residues in SNM1B. The ribose ring of AMP-1 does not appear to

form any direct interactions with the protein however. This suggests that modified nucleotides that retain a 5'-phosphate and unmodified nucleobase, but contain modifications appended to the ribose ring, may function well as inhibitors of SNM1B and the closely related enzyme SNM1A. As well as increasing binding-affinity to the enzyme, incorporation of a 5'-phosphate moiety in nucleoside-based inhibitors also affords the potential for future optimisation of lead compounds through masking of the polar phosphate moiety with biolabile protecting groups that confer membrane-permeability, a strategy that has already been developed for other nucleotide prodrugs.^{350–354}

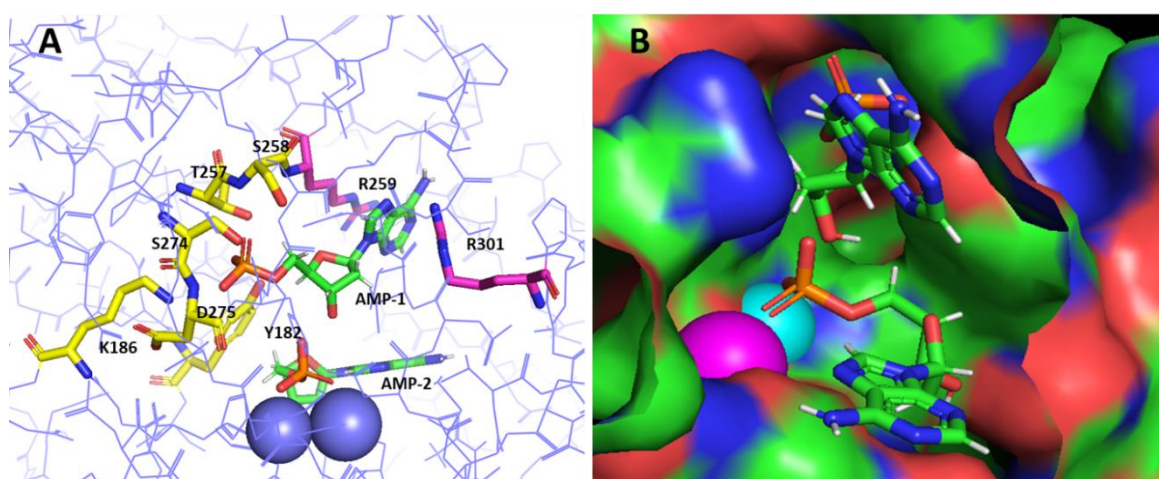


Figure 4.1. A crystal structure (PDB: 7A1F) of SNM1B in complex with two 2'-deoxyadenosine 5'-monophosphate (AMP) molecules,⁸¹ visualised using PyMOL. **A**) A view of the non-covalent interactions of the AMP molecules, highlighted in green, with SNM1B. Amino acid residues coloured yellow engage in hydrogen bonding interactions with the 5'-phosphate of AMP-1, forming the phosphate-binding pocket. Arginine residues highlighted in purple engage in π -cation interactions with the adenine nucleobase of AMP-1, while the two metal ions, coloured blue, interact with the 5'-phosphate moiety of AMP-2. **B**) A molecular surface representation of SNM1B in complex with two AMP molecules showing the active site and phosphate-binding pocket.

A series of nucleoside derivatives bearing phosphate moieties at the 5'-position, as well as ZBGs at the 3'-position, were therefore prepared. These compounds were designed to mimic the positioning of the AMP-1 molecule in the SNM1B active site, thus targeting the phosphate-binding pocket through their 5'-phosphate moieties, as well as targeting the active site metal ions through their 3'-ZBGs. A schematic view of the proposed binding mode of the compounds to the SNM1A active site is shown in Figure 4.2.

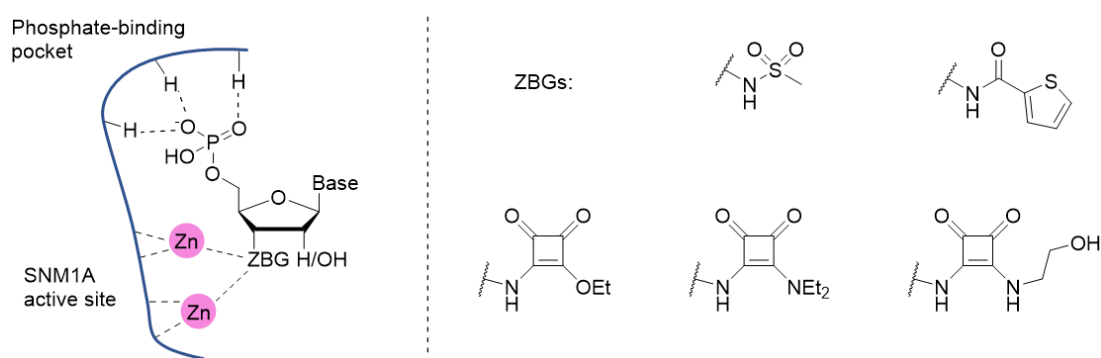


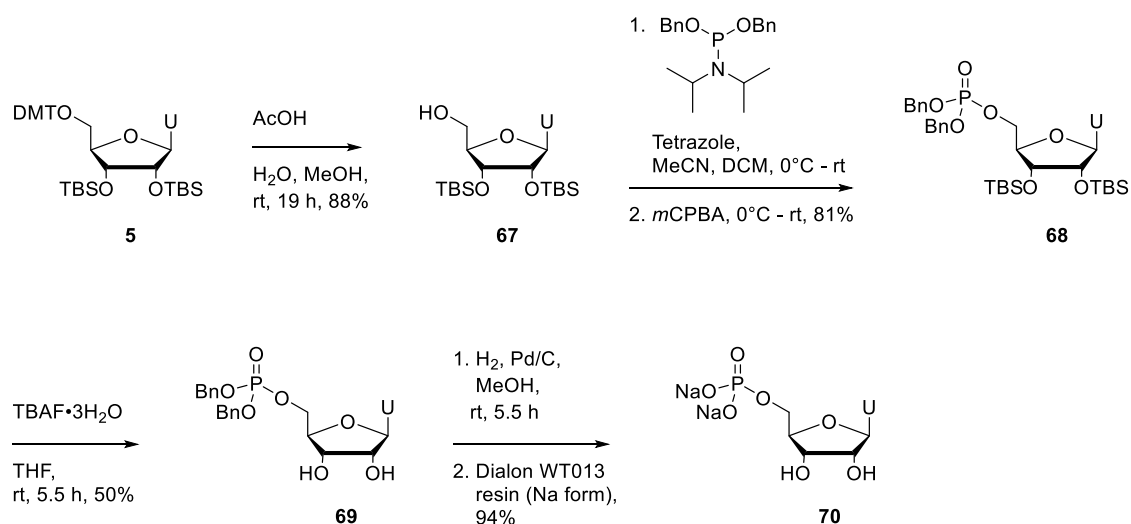
Figure 4.2. Schematic view of substrate-mimic inhibitors binding to the SNM1A active site and phosphate-binding pocket.

Phosphorylated nucleoside derivatives bearing ZBGs, such as squaramide and sulfonamide moieties, that had previously proven effective in analogous unphosphorylated compounds, were prepared. In order to compare the effect of a 5'-phosphate with the effect of a 3'-ZBG for binding to SNM1A, and thus investigate whether targeting the active site metal ions or targeting the phosphate-binding pocket is more effective, compounds bearing a 5'-phosphate and no ZBG, and a compound bearing a 2-thiophene carboxamide ZBG previously found to be ineffective (Figure 2.4, p. 64) were also prepared. Evaluation of these compounds in gel-electrophoresis and real-time fluorescence assays showed that incorporation of a 5'-phosphate dramatically increased the potency of SNM1A inhibition relative to analogous compounds lacking a phosphate moiety. Inhibitors based on both

thymidine and uridine scaffolds were prepared, and while thymidine derivatives show lower IC₅₀ values than analogous uridine derivatives, the presence of a 5'-phosphate and the choice of ZBG were found to be more important factors.

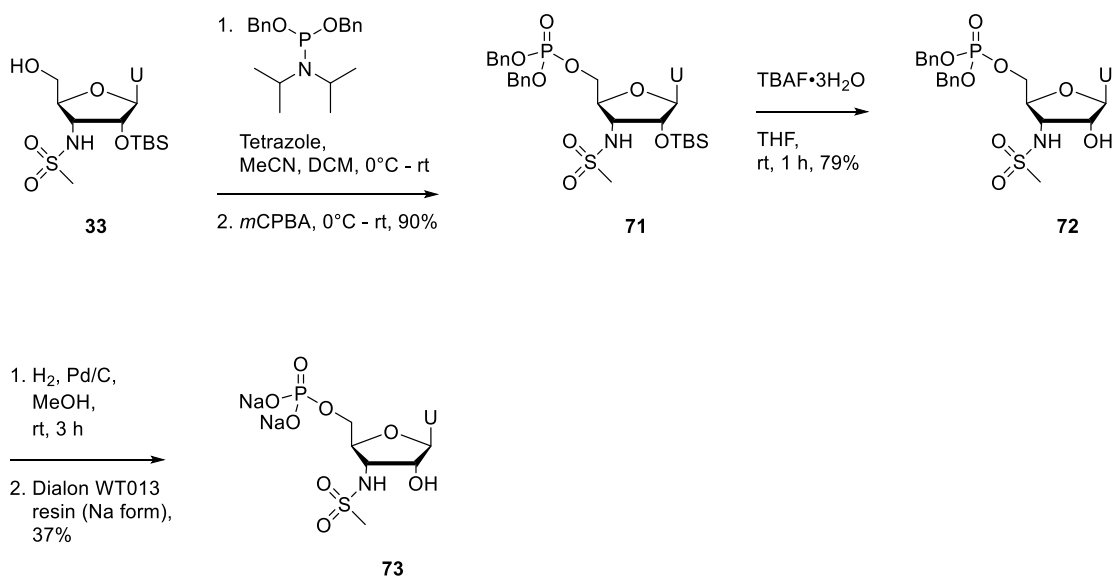
4.2. Synthesis of phosphorylated uridine derivatives

An established method for the generation of *O*-alkyl protected phosphates is the use of dialkyl *N,N*-diisopropylphosphoramidites, which can be reacted with alcohols to form phosphorous(III) species and then oxidised with peroxide reagents to form phosphotriesters.³⁵⁵ The *O*-alkyl groups of the phosphotriester facilitate further synthetic manipulations of the molecule before deprotection generates a polar phosphate moiety. To confirm the efficacy of this method for the synthesis of phosphorylated nucleoside derivatives, uridine 5'-monophosphate disodium salt **70** was first prepared (Scheme 4.1) before turning to target molecules bearing ZBGs. Compound **70** was also desired for use as a control in biochemical assays to differentiate the effect of a ZBG at the 3'-position from the effect of a 5'-phosphate in substrate mimic SNM1A inhibitors. Protected uridine derivative **5**, prepared as described in Chapter 2 (p. 49), was treated with acetic acid to remove the 5'-DMT group, generating 2',3'-bis-TBS protected uridine derivative **67** in 88% yield. Intermediate **67** was phosphorylated *via* treatment with dibenzyl *N,N*-diisopropylphosphoramidite followed by *m*CPBA to provide uridine derivative **68** in 81% yield. Removal of the TBS groups of **68** provided uridine derivative **69** in 50% yield. Hydrogenation of **69** followed by elution through Dialon WT013 ion-exchange resin (Na form) provided uridine 5'-monophosphate disodium salt **70** in 94% yield.



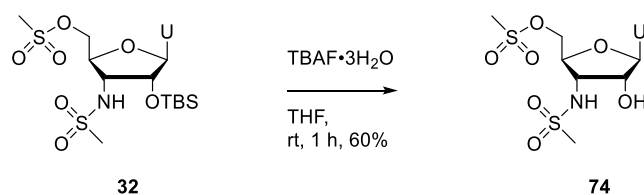
Scheme 4.1. Synthesis of uridine 5'-monophosphate disodium salt **70**.

Following the successful synthesis of uridine 5'-monophosphate disodium salt **70**, an analogous synthetic sequence was used to prepare uridine derivatives bearing a 5'-phosphate and a 3'-ZBG. A uridine derivative **34** bearing a sulfonamide at the 3'-position was previously found to inhibit SNM1A with moderate potency (Figure 2.4, p. 64), indicating a sulfonamide ZBG is effective for targeting this enzyme. To improve the inhibitory potency by exploiting interactions with SNM1A's phosphate-binding pocket as well as the catalytic zinc ion(s), compound **73** bearing a 5'-phosphate as well as a 3'-sulfonamide was prepared (Scheme 4.2). Sulfonamide-bearing uridine derivative **33** was treated with dibenzyl *N,N*-diisopropylphosphoramidite followed by *m*CPBA to provide the 5'-phosphorylated compound **71**. Reaction with tetrabutylammonium fluoride afforded intermediate **72** in 79% yield. Removal of the benzyl protecting groups of **72** through hydrogenation, followed by elution through Dialon WT013 ion-exchange resin (Na form) afforded the target sulfonamide-bearing nucleotide **73** in 37% yield.



Scheme 4.2. Synthesis of phosphorylated uridine derivative **73** bearing a sulfonamide ZBG.

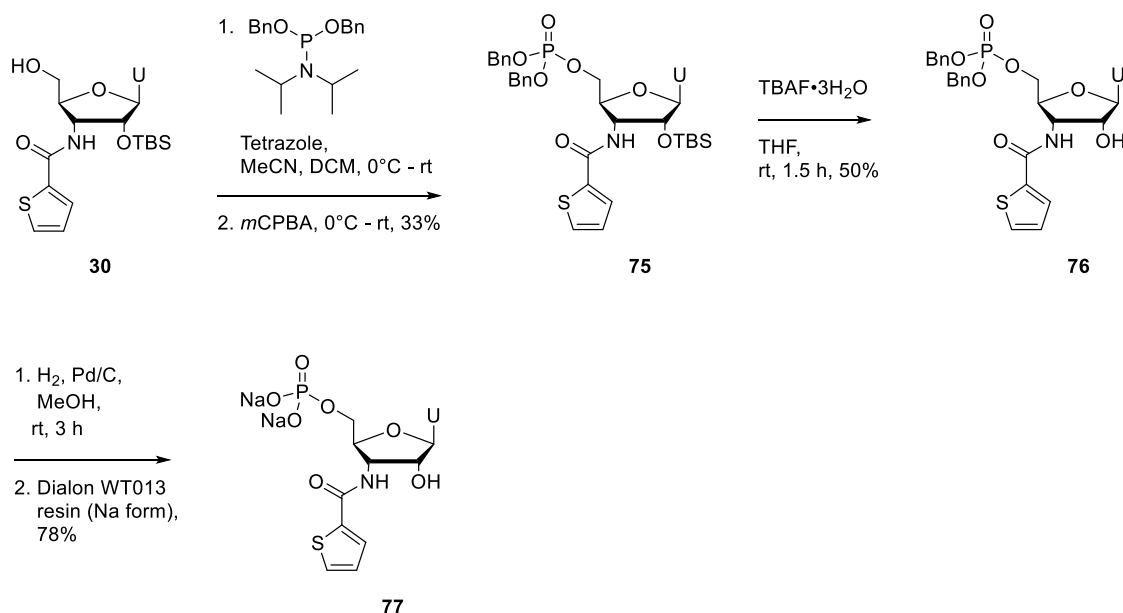
Uridine derivative **32** bearing a mesylate at the 5'-position and a sulfonamide at the 3'-position was obtained in the synthesis of sulfonamide **34**, as described in Chapter 2 (p. 57). In order to test whether the 5'-mesyl group could function as a phosphate mimic and thereby improve inhibition of SNM1A, compound **32** was deprotected through treatment with tetrabutylammonium fluoride to provide uridine derivative **74** in 60% yield (Scheme 4.3).



Scheme 4.3. Synthesis of uridine derivative **74** bearing a 5'-mesylate and a 3'-sulfonamide.

A uridine derivative **31** bearing a 2-thiophene carboxamide moiety appended at the 3'-position was previously tested as an inhibitor of SNM1A and found to be ineffective (Figure 2.4, p. 64). To investigate the relative importance of a 5'-phosphate and a 3'-ZBG in substrate-mimic SNM1A inhibitors, a uridine derivative **77** bearing a 5'-phosphate and

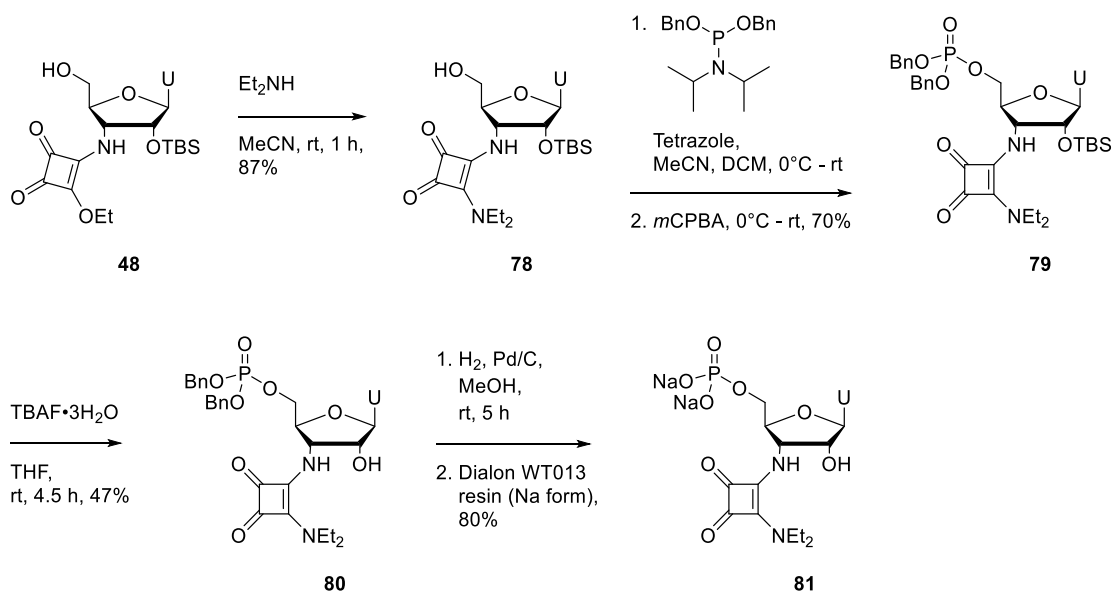
3'-thiophene was prepared for comparison in biochemical assays with sulfonamide-bearing nucleotide **73** (Scheme 4.4). Thiophene-bearing uridine derivative **30** was first phosphorylated to provide intermediate **75** in 33% yield. Silyl deprotection of **75** provided intermediate **76** in 50% yield. Further deprotection through catalytic hydrogenation, followed by elution through Dialon WT013 ion-exchange resin (Na form) provided the target thiophene-bearing nucleotide **77**.



Scheme 4.4. Synthesis of phosphorylated uridine derivative **77** bearing a 2-thiophene carboxamide ZBG.

As nucleoside derivatives bearing squaramide ZBGs were previously found to show promise as SNM1A inhibitors, as discussed in Chapter 3, a uridine derivative **81** bearing a 5'-phosphate as well as a 3'-squaramide was synthesised (Scheme 4.5). Squaryl monoamide **48**, prepared as described in Chapter 3 (p. 69), was reacted with diethylamine to afford squaryl diamide **78** in 87% yield. Phosphorylation of squaramide **78** was carried out under conditions similar to those used for phosphorylation of uridine derivatives **30** and **34** to provide the benzyl-protected phosphate **79** in 70% yield. Removal of the 2'-TBS

group of **79** through treatment with tetrabutylammonium fluoride afforded intermediate **80**, which was further deprotected *via* catalytic hydrogenation to obtain the target phosphorylated uridine derivative **81** bearing a squaramide ZBG.



Scheme 4.5. Synthesis of phosphorylated uridine derivative **81** bearing a squaramide ZBG.

4.3. Biochemical testing of phosphorylated uridine derivatives

The phosphorylated uridine derivatives **73**, **77** and **81** were screened as SNM1A inhibitors using a gel-electrophoresis-based assay (Figure 4.3). Uridine 5'-monophosphate disodium salt **70** was included as a control to distinguish the effect of the 5'-phosphate from the effect of the ZBGs. Uridine derivative **74** bearing a mesylate at the 5'-position in addition to a 3'-sulfonamide was also tested. Thiophene **77**, sulfonamide **73**, and squaramide **81** all appeared to fully inhibit SNM1A activity under the conditions of this assay. Interestingly, full inhibition of SNM1A was also observed in the presence of uridine 5'-monophosphate disodium salt **70** which lacks a ZBG. Uridine derivative **74** bearing a mesylate at the 5'-position as well as a 3'-sulfonamide did not appear to inhibit SNM1A. Indeed, the addition of a 5'-mesylate in **74** appears to have a deleterious effect on inhibitory potency

when compared to sulfonamide **34** with a free hydroxy group at the 5'-position (Figure 2.4, p. 64), showing that a mesylate does not function effectively as a phosphate mimic for targeting SNM1A.

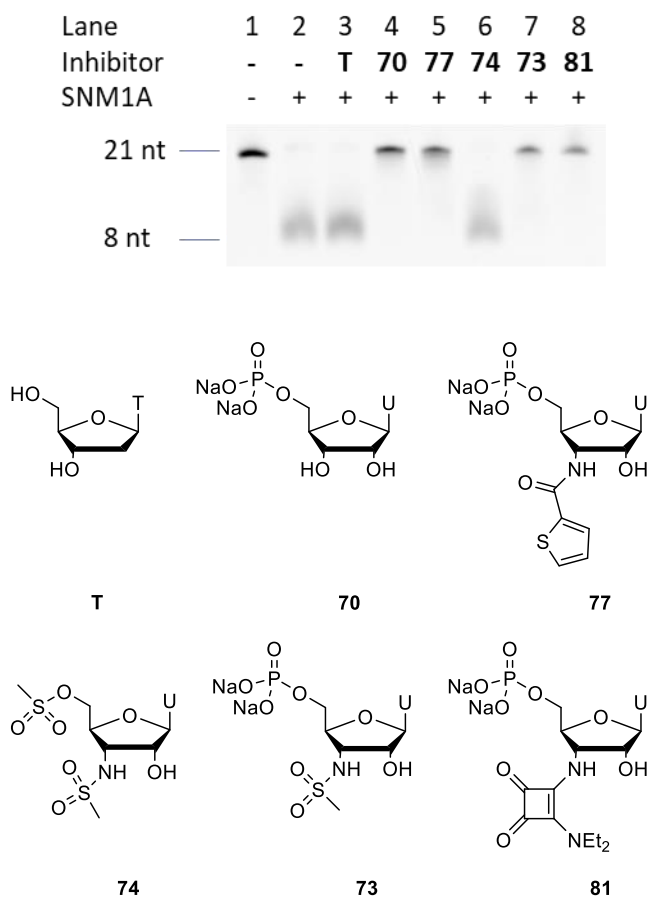


Figure 4.3. Evaluation of phosphorylated uridine derivatives **70**, **73**, **77** and **81**, and uridine derivative **74** bearing a 5'-mesylate, as SNM1A inhibitors through visualisation of the extent of digestion of a 21mer fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the candidate inhibitors (1 mM) for 5 minutes before the oligonucleotide substrate (80 nM) was added, and then incubated for a further 1 hour. Thymidine (T) was included in this assay as a control. nt = nucleotides.

Uridine derivatives **70**, **73**, **77** and **81** were tested individually in a gel-electrophoresis-based assay at a range of concentrations from 1 mM to 0.3 μ M (Figure 4.4). All four of the

phosphorylated uridine derivatives showed inhibition of SNM1A at a concentration of 33 μM , and in some cases a small degree of inhibition as low as 10 μM . Although a definitive distinction cannot be drawn between the different phosphorylated uridine derivatives based on these results, it is clear that they inhibit SNM1A with considerably higher potency than the analogous uridine derivatives that lack a 5'-phosphate. For example, uridine derivative **73** bearing a 5'-phosphate and 3'-sulfonamide shows partial inhibition of SNM1A at a concentration as low as 33 μM (Figure 4.4C), while the analogous nucleoside derivative **34** bearing a 3'-sulfonamide shows appreciable inhibition only at a concentration of 1 mM (Figure 2.5B, p. 65). Similarly, uridine derivative **77** bearing a 2-thiophene carboxamide group at the 3'-position and a 5'-phosphate shows inhibition of SNM1A at a concentration of 33 μM (Figure 4.4B), while uridine derivative **31** bearing only a 2-thiophene carboxamide group does not inhibit SNM1A even at 1 mM concentration (Figure 2.4, p. 64).

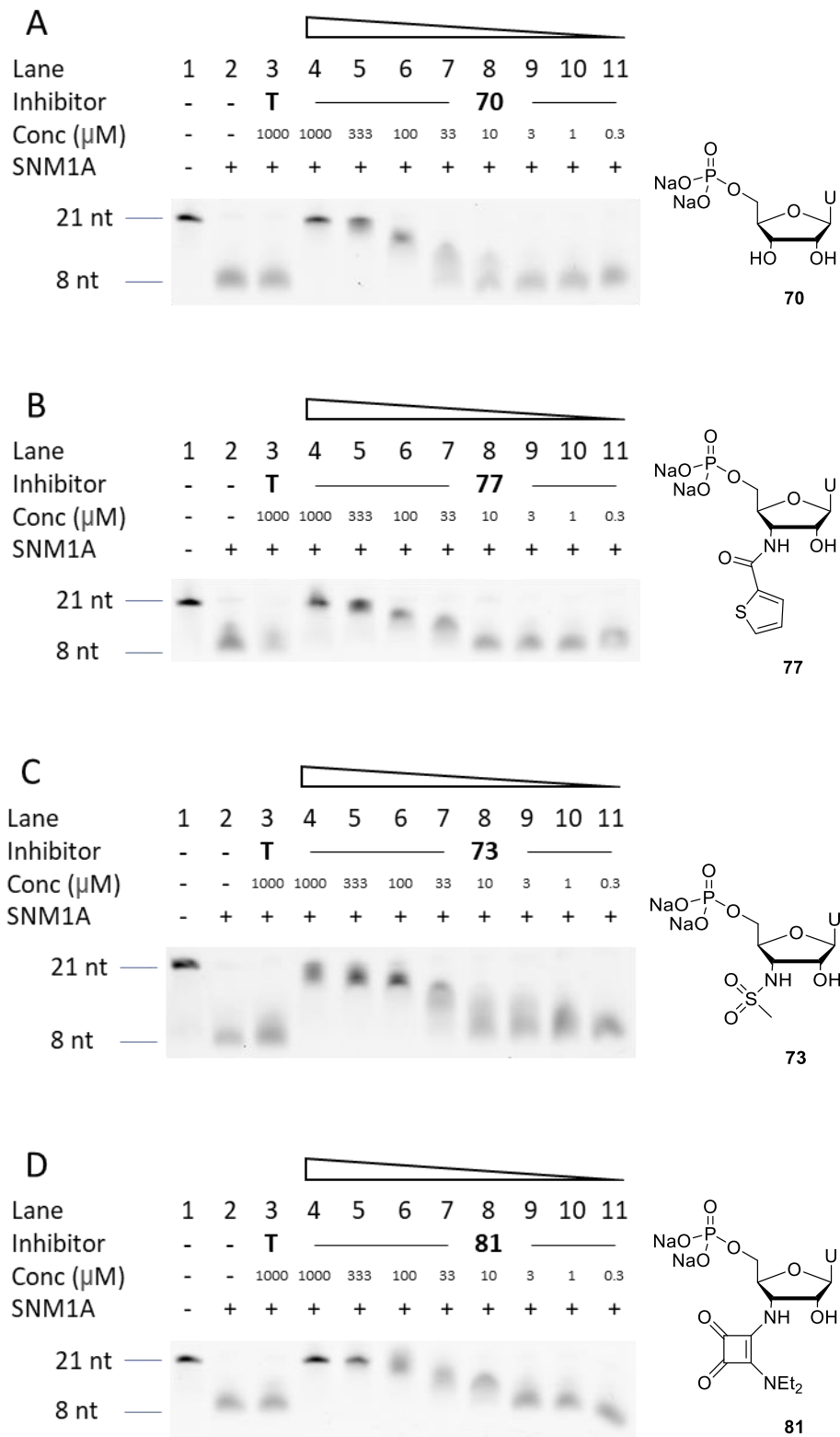


Figure 4.4. Evaluation of SNM1A inhibitors **70**, **73**, **77** and **81** at a range of concentrations from 1 mM to 0.3 μM through visualisation of the extent of digestion of a fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the modified nucleotides

for 5 min before the 21mer oligonucleotide substrate (80 nM) was added, and the mixture was incubated for a further 60 min. nt = nucleotides.

To quantitatively determine the relative efficacy of the different ZBGs incorporated into the phosphorylated uridine derivatives, a real-time fluorescence assay^{123,126} was carried out as before (Chapter 3, pp. 81-82) to measure IC₅₀ values for inhibition of SNM1A by compounds **70**, **73**, **77**, and **81** (Figure 4.5).

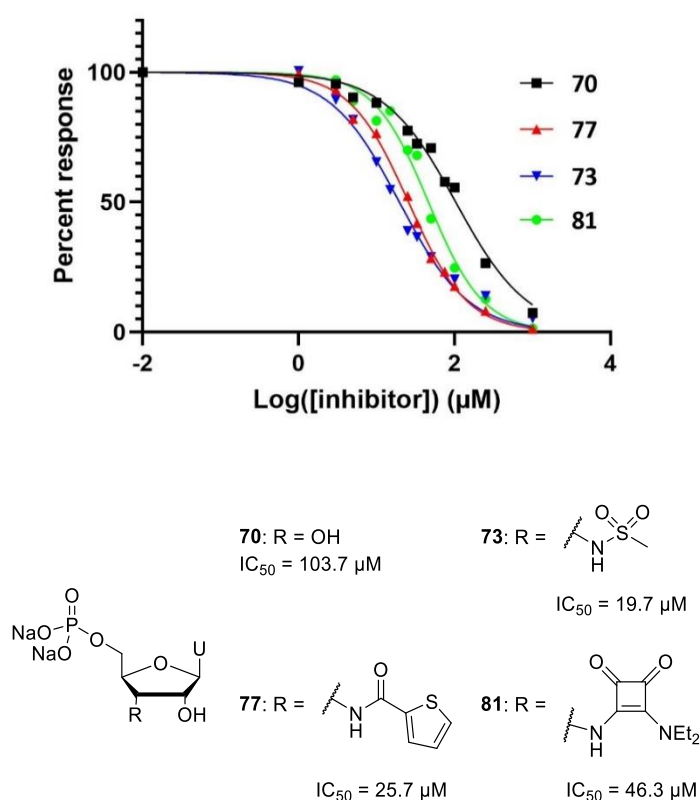


Figure 4.5. IC₅₀ determination of uridine derivatives **70**, **73**, **77**, and **81** in a real-time fluorescence assay. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides at a range of concentrations for 6.5 min before the 20mer oligonucleotide substrate (125 nM) was added. Fluorescent readings were taken every 30 seconds for 35 min. For each concentration of inhibitor, fluorescence intensity was plotted against time, and the slope of the linear range of this graph was taken as the rate of enzyme activity. The rate of enzyme activity observed for the zero-inhibitor control was set as 100, and the other rates were normalised to this to obtain percent response

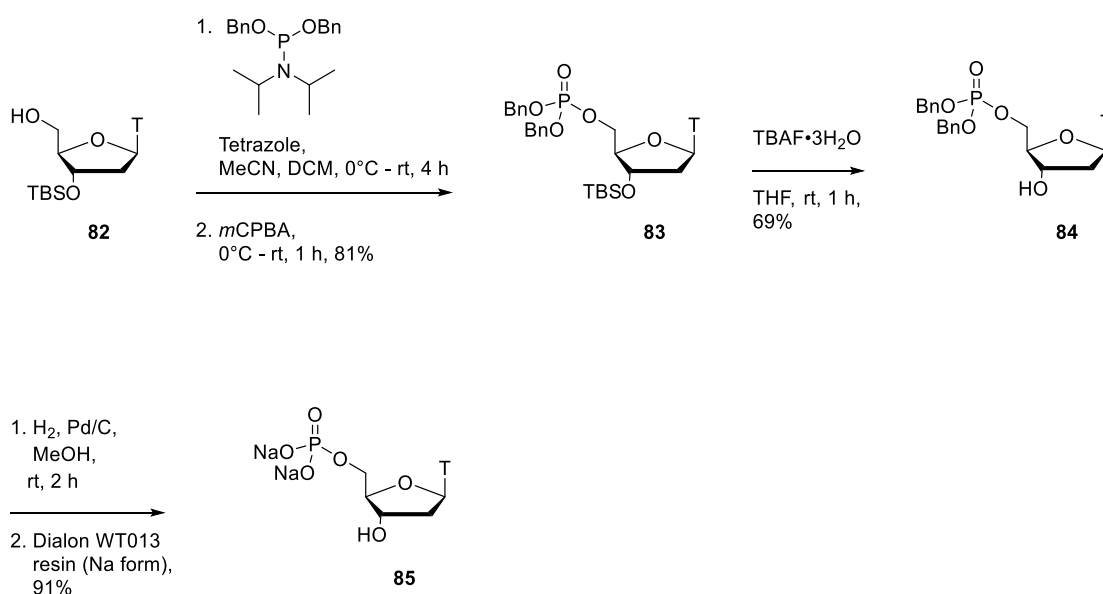
values. Percent response was plotted against $\log([\text{inhibitor}])$, and a nonlinear regression was performed to calculate the IC_{50} . Error bars were generated from 6 independent repeats.

Uridine 5'-monophosphate disodium salt **70**, tested as a control compound, was found to have an IC_{50} of 103.7 μM . This result illustrates the strong effect of the 5'-phosphate on inhibitory potency, as compound **70**, despite lacking a ZBG, has a lower IC_{50} than substrate-mimic SNM1A inhibitors bearing (thio)squaramide (Chapter 3) or hydroxamic acid¹²⁶ ZBGs. The presence of a ZBG in addition to a 5'-phosphate also has a significant effect on the inhibitory potency however, as uridine derivatives **73**, **77** and **81** bearing ZBGs at the 3'-position proved to be stronger inhibitors than the control **70**. Sulfonamide **73** proved to be the most effective inhibitor of SNM1A, with an IC_{50} of 19.7 μM . Thiophene-bearing uridine derivative **77** is also a potent inhibitor, with an IC_{50} of 25.7 μM . Squaramide-bearing uridine 5'-monophosphate derivative **81** has an IC_{50} of 46.3 μM , showing this is a much more potent inhibitor than squaramide-bearing nucleoside derivatives lacking a 5'-phosphate described in Chapter 3.

4.4. Synthesis of phosphorylated thymidine derivatives

Of a series of squaramide-bearing nucleoside derivatives previously tested as SNM1A inhibitors, two thymidine derivatives **46** and **47** proved to be more effective than the analogous uridine derivatives **52** and **53**, as discussed in Chapter 3 (p. 74). However phosphorylated squaramide-bearing uridine derivative **81**, proved to be a much more potent inhibitor, with an IC_{50} of 46.3 μM , compared to an IC_{50} of 456 μM for the most potent unphosphorylated thymidine derivative **47**. In an effort to develop more potent inhibitors, and to further investigate the relative impact of factors such as the choice of a thymidine or uridine scaffold, 5'-phosphorylation, and the use of differently substituted squaramides as ZBGs, a series of phosphorylated thymidine derivatives was prepared.

Before turning to more complex target molecules bearing squaramide ZBGs, thymidine 5'-monophosphate disodium salt **85** was first prepared for use as a control in biochemical assays (Scheme 4.6). Protected thymidine derivative **82**, provided by Dr. William Doherty, was phosphorylated *via* treatment with dibenzyl *N,N*-diisopropylphosphoramidite followed by *m*CPBA to provide the benzyl-protected phosphate **83** in 81% yield. Intermediate **83** was treated with tetrabutylammonium fluoride to generate thymidine derivative **84** in 69% yield. Deprotection of **84** through catalytic hydrogenation, followed by elution through Dialon WT013 resin (Na form), provided thymidine 5'-monophosphate disodium salt **85** in 91% yield.

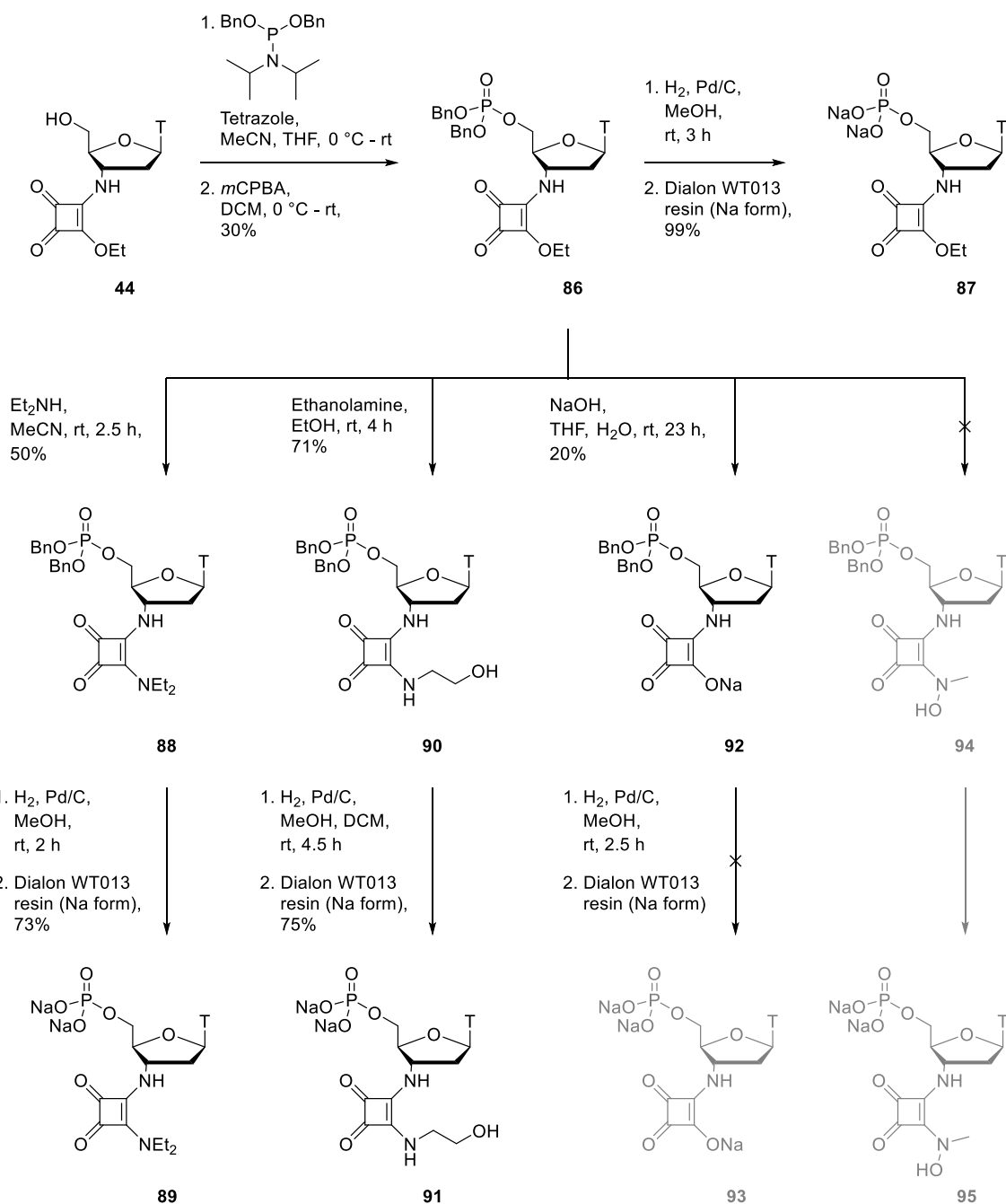


Scheme 4.6. Synthesis of thymidine 5'-monophosphate disodium salt **85**.

Phosphorylated thymidine derivatives bearing squaramide ZBGs at the 3'-position were next prepared (Scheme 4.7). Thymidine derivative **44** bearing a squaryl monoamide was generated as described in Chapter 3 (p. 68). Phosphorylation of squaryl monoamide **44** produced thymidine derivative **86**, a key intermediate for the preparation of the target compounds **87**, **89** and **91**. Intermediate **86** was hydrogenated to provide thymidine

5'-monophosphate derivative **87** bearing a squaryl monoamide at the 3'-position in 99% yield. Intermediate **86** was also reacted with diethylamine to provide squaryl diamide **88** in 50% yield. Hydrogenation of **88** generated thymidine 5'-monophosphate derivative **89**, an analogue of squaramide-bearing uridine derivative **81**, in 73% yield. Separately, intermediate **86** was reacted with ethanolamine to provide squaryl diamide **90** in 71% yield. Hydrogenation of **90** furnished squaramide-bearing nucleotide derivative **91**, a phosphorylated analogue of squaramide-bearing nucleoside derivative **45** described in Chapter 3.

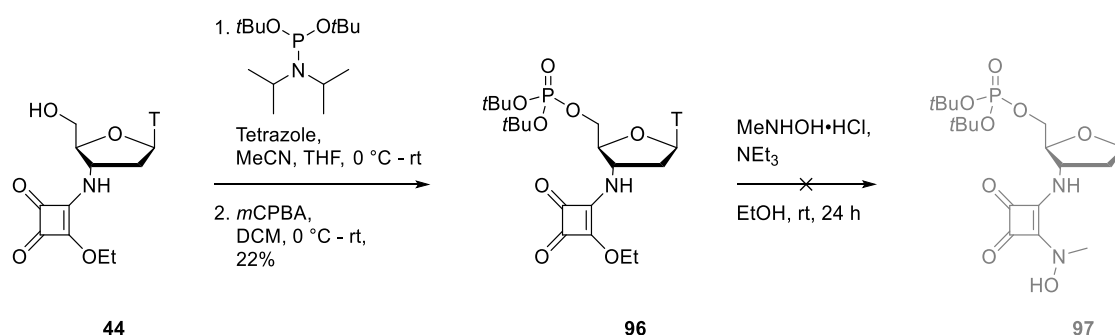
Thymidine derivatives bearing a squaric acid monoamide or *N*-hydroxysquaramide ZBG were previously found to inhibit SNM1A, while thymidine derivatives bearing other squaramide ZBGs did not show inhibitory potency, as described in Chapter 3 (p. 74). Phosphorylated thymidine derivatives **93** and **95** bearing a squaric acid monoamide or *N*-hydroxysquaramide moiety, respectively, were therefore promising target compounds for the development of SNM1A inhibitors of greater efficacy. Intermediate **86** bearing a squaryl monoamide was hydrolysed under basic conditions to provide thymidine derivative **92** bearing a squaric acid monoamide moiety as well as a protected phosphate at the 5'-position in a modest yield of 20% (Scheme 4.7). Attempts to deprotect intermediate **92** through catalytic hydrogenation were unsuccessful however, and appeared to result in decomposition of the compound. Attempts to prepare thymidine derivative **94** bearing an *N*-hydroxysquaramide through treatment of squaryl monoamide **86** with *N*-methylhydroxylamine hydrochloride under a variety of conditions resulted in no reaction (Scheme 4.7).



Scheme 4.7. Synthesis of phosphorylated thymidine derivatives **87**, **89**, and **91** bearing squaramide ZBGs.

Efforts to react intermediate **86** with *N*-methylhydroxylamine hydrochloride to generate *N*-hydroxysquaramide **94** were hindered by differences in solubility of the two reagents in a variety of solvent systems. Furthermore, it was anticipated that upon formation of the desired product **94**, removal of the benzyl protecting groups of the phosphate moiety

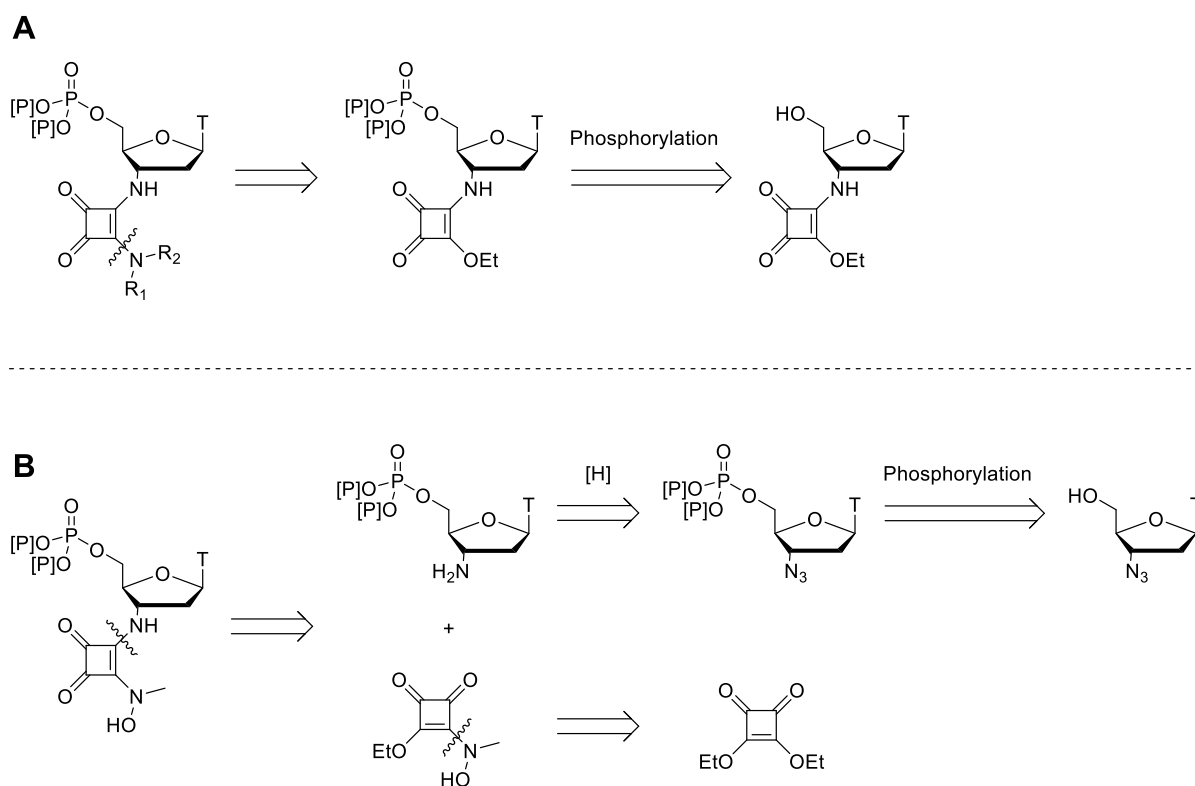
through catalytic hydrogenation could also lead to cleavage of the N-O bond in the *N*-hydroxysquaramide ZBG, precluding formation of the target compound **95**. Therefore, to modulate solubility and avoid the need for catalytic hydrogenation, an alternative synthesis utilising *tert*-butyl protecting groups in place of benzyl protecting groups was attempted (Scheme 4.8). Squaryl monoamide **44** was reacted with di-*tert*-butyl *N,N*-diisopropylphosphoramidite, and oxidised using *m*CPBA to form thymidine derivative **96** bearing a *tert*-butyl protected phosphate and a 3'-squaryl monoamide. However, treatment of the *tert*-butyl protected intermediate **96** with *N*-methylhydroxylamine hydrochloride resulted in no reaction.



Scheme 4.8. Attempted synthesis of *N*-hydroxysquaramide **97** bearing a *tert*-butyl protected phosphate.

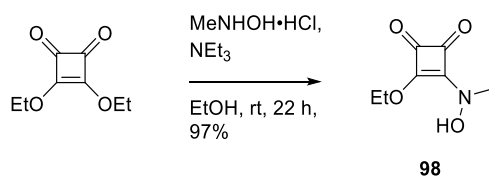
Due to the difficulty in reacting either of the squaryl monoamides **86** and **96** with *N*-methylhydroxylamine hydrochloride, an alternative synthetic approach was undertaken. The strategy of first preparing a thymidine derivative bearing a protected phosphate at the 5'-position and a squaryl monoamide at the 3'-position, summarised in Scheme 4.9A, allowed for a number of target compounds (**87**, **89** and **91**) to be prepared from this one squaryl monoamide intermediate, minimising the number of synthetic steps required. However, as dialkyl squarates are more electrophilic than squaryl monoamides, for the

challenging *N*-hydroxysquaramide target **95**, a strategy of first reacting the *N*-methylhydroxylamine hydrochloride nucleophile with a dialkyl squarate, followed by reaction of the resulting *N*-hydroxysquaryl monoamide product with a 3'-aminonucleotide derivative was employed (Scheme 4.9B).



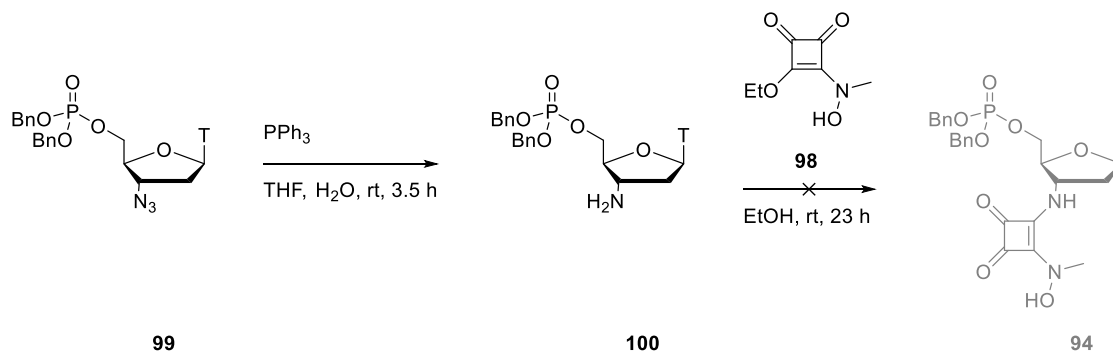
Scheme 4.9. Two alternative synthetic strategies for the preparation of a phosphorylated thymidine derivative bearing an *N*-hydroxysquaramide ZBG.

Diethyl squarate was successfully reacted with *N*-methylhydroxylamine hydrochloride to form *N*-hydroxysquaryl monoamide **98** in 97% yield (Scheme 4.10).



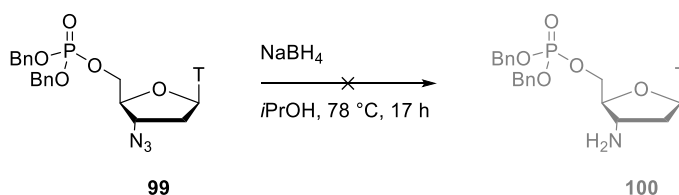
Scheme 4.10. Synthesis of *N*-hydroxysquaryl monoamide **98**.

3'-Azidothymidine derivative **99** bearing a benzyl-protected phosphate group was provided by Manav T Manoj. Staudinger reduction of azide **99** provided amine **100** (Scheme 4.11), however attempts to purify **100** were not successful in fully removing the triphenylphosphine oxide by-product of the reduction reaction. Reaction of impure amine **100** with *N*-hydroxysquaryl monoamide **98** did not yield the desired product **94**.



Scheme 4.11. Attempted synthesis of *N*-hydroxysquaramide **94**.

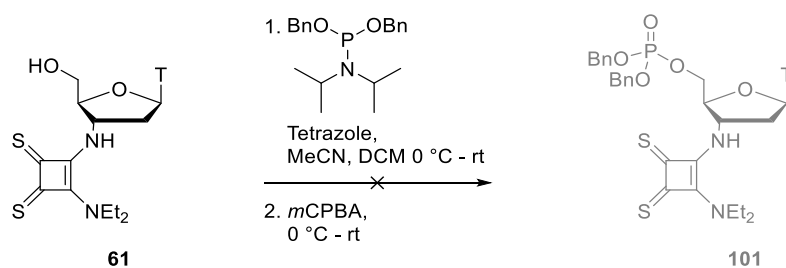
To avoid formation of triphenylphosphine oxide, an alternative method of reducing azide **99** using sodium borohydride was attempted (Scheme 4.12). This however did not yield the desired product **100**, instead appearing to cleave one of the benzyl protecting groups of the phosphate moiety in addition to reducing the azide, producing a product that was difficult to purify due to its high polarity.



Scheme 4.12. Attempted reduction of 3'-azidothymidine derivative **99** using sodium borohydride.

Thymidine derivative **61** bearing a thiosquaramide ZBG at the 3'-postion was previously found to inhibit SNM1A, as described in Chapter 3 (p. 74). The synthesis of a

phosphorylated thymidine derivative bearing a 3'-thiosquaramide was therefore attempted. Compound **61** was treated with dibenzyl *N,N*-diisopropylphosphoramidite followed by *m*CPBA (Scheme 4.13), however the desired product **101** was not obtained. This may have been due to reactivity of the thiocarbonyl moieties of **61** to *m*CPBA, as other thiocarbonyls have previously been reported to undergo oxidation by peroxide reagents to form unstable products.³⁵⁶



Scheme 4.13. Attempted synthesis of a phosphorylated thymidine derivative bearing a thiosquaramide ZBG.

4.5. Biochemical testing of thymidine derivatives

The squaramide-bearing thymidine 5'-monophosphate derivatives **87**, **89**, and **91** were screened for inhibitory activity against SNM1A using a gel-electrophoresis-based assay (Figure 4.6). Thymidine 5'-monophosphate disodium salt **85** was included as a control to differentiate the effect of a 5'-phosphate from the effects of the squaramide modifications. All of the compounds tested fully inhibited SNM1A at a concentration of 1 mM.

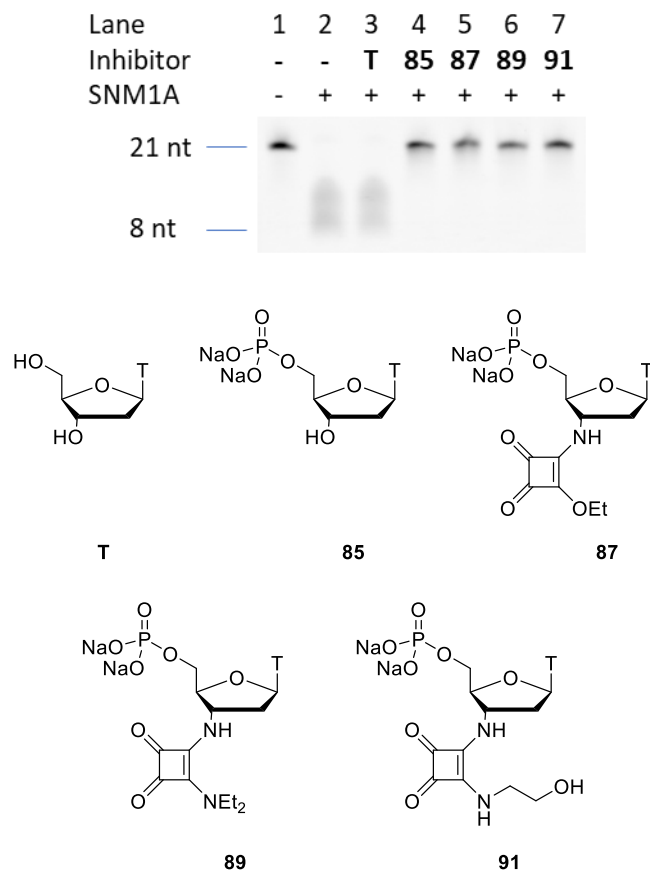


Figure 4.6. Evaluation of thymidine derivatives **85**, **87**, **89**, and **91** as SNM1A inhibitors through visualisation of the extent of digestion of a 21mer fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the modified nucleotides (1 mM) for 5 minutes before the oligonucleotide substrate (80 nM) was added, and then incubated for a further 1 hour. Thymidine (T) was included in this assay as a control. nt = nucleotides.

The thymidine derivatives **85**, **87**, **89** and **91** were tested individually in gel-electrophoresis-based assays at a range of concentrations from 1 mM to 0.3 μ M (Figure 4.7). All of the compounds showed moderate inhibition of SNM1A at 33 μ M, and in some cases partial inhibition as low as 10 μ M. This shows that the phosphorylated thymidine derivatives **87** and **91** are much more potent inhibitors than their unphosphorylated analogues **45** and **45** described in Chapter 3. The thymidine derivatives **85**, **87**, **89** and **91** all inhibit SNM1A with similar efficacy, and the results of this assay do not allow a clear distinction to be drawn

between them, or between these compounds and the phosphorylated uridine derivatives

70, 73, 77 and 81.

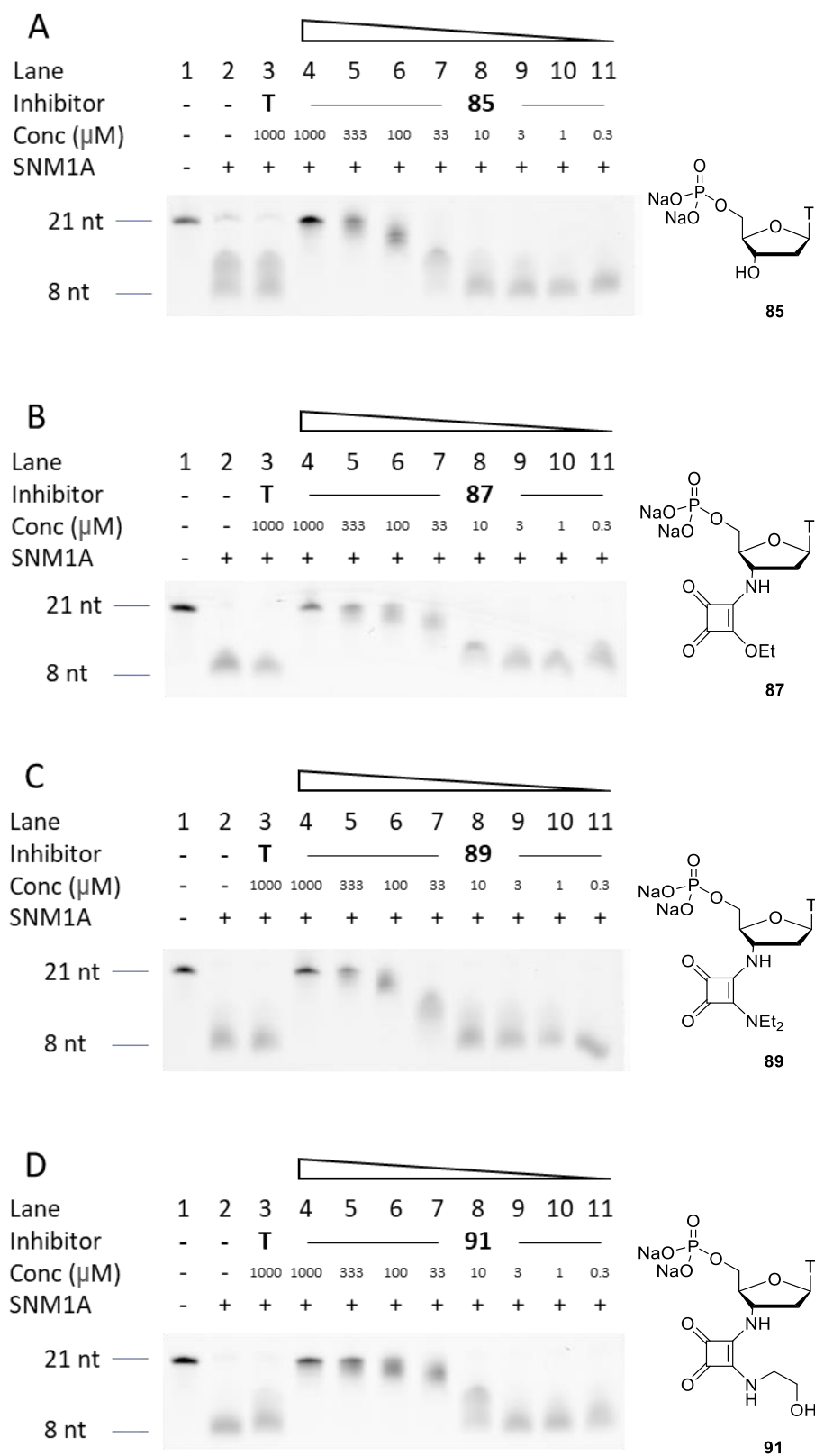


Figure 4.7. Evaluation of SNM1A inhibitors **85**, **87**, **89** and **91** at a range of concentrations from 1 mM to 0.3 μ M through visualisation of the extent of digestion of a fluorescent oligonucleotide substrate by denaturing PAGE. SNM1A (2.5 nM) was pre-incubated with the modified nucleotides for 5 min before the 21mer oligonucleotide substrate (80 nM) was added, and the mixture was incubated for a further 60 min. nt = nucleotides.

To quantitatively determine the efficacy of the phosphorylated thymidine derivatives **85**, **87**, **89**, and **91**, IC_{50} values for inhibition of SNM1A were determined (Figure 4.8) using a real-time fluorescence assay^{123,126} as before (Chapter 3, pp. 81-82). Thymidine 5'-monophosphate **85** was found to have an IC_{50} of 54 μ M, significantly lower than the value of 103.7 μ M determined for uridine 5'-monophosphate **70**. An IC_{50} of 46.7 μ M was determined for squaryl monoamide **87**, showing this ZBG does not significantly improve inhibition compared to control compound **85**. Squaramide **89** however was found to have a lower IC_{50} of 31.7 μ M. The greater inhibitory effect of squaryl diamide **89** relative to squaryl monoamide **87** can be attributed to delocalisation of nitrogen lone pairs leading to greater electron density on the squaramide carbonyl groups in **89**, improving zinc binding. Squaramide-bearing thymidine derivative **89** is also a more potent SNM1A inhibitor than the corresponding uridine derivative **81**, which has an IC_{50} of 46.3 μ M. Squaramide **91** substituted with an ethanolamine moiety showed the strongest inhibition of any of the thymidine derivatives, with an IC_{50} of 26.2 μ M. This may be caused by additional coordination of the side-chain hydroxy group of **91** to the catalytic metal ion(s) in the active site of SNM1A.³⁵⁷

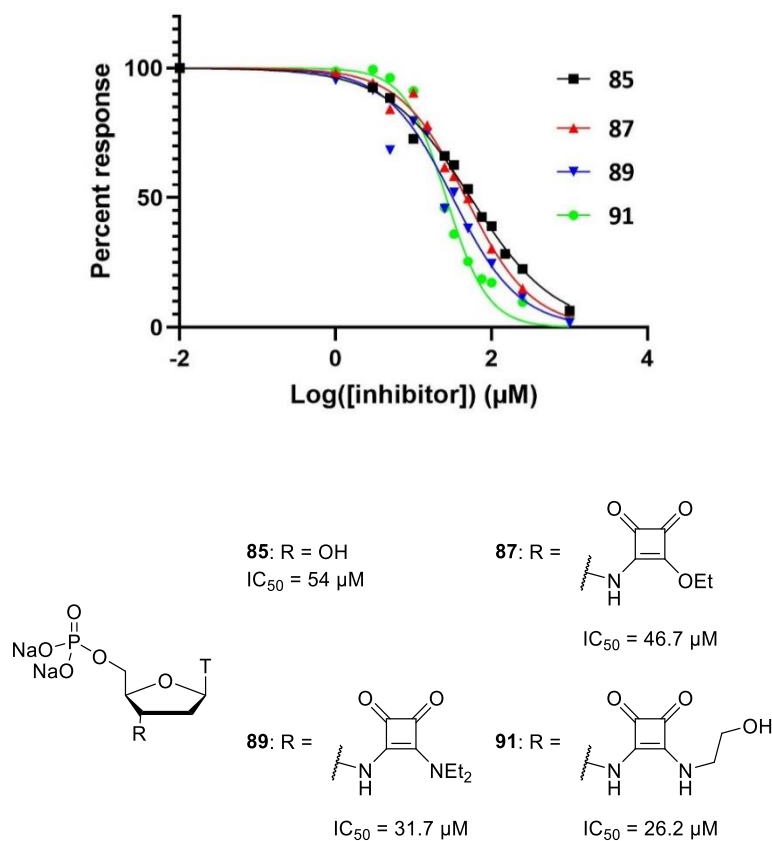


Figure 4.8. IC₅₀ determination of thymidine derivatives **85**, **87**, **89**, and **91** in a real-time fluorescence assay. SNM1A (2.5 nM) was pre-incubated with the modified nucleosides at a range of concentrations for 6.5 min before the 20mer oligonucleotide substrate (125 nM) was added. Fluorescent readings were taken every 30 seconds for 35 min. For each concentration of inhibitor, fluorescence intensity was plotted against time, and the slope of the linear range of this graph was taken as the rate of enzyme activity. The rate of enzyme activity observed for the zero-inhibitor control was set as 100, and the other rates were normalised to this to obtain percent response values. Percent response was plotted against log([inhibitor]), and a nonlinear regression was performed to calculate the IC₅₀. Error bars were generated from 6 independent repeats.

4.6. Conclusion

In summary, several series of nucleoside derivatives bearing ZBGs have been synthesised and tested as inhibitors of the DNA repair enzyme SNM1A. Inhibitors bearing squaramide and sulfonamide ZBGs, and an inhibitor bearing a 2-thiophene carboxamide, which is not

well established as a ZBG, have proven efficacious. These results also show that targeting the enzyme's phosphate-binding pocket through the inclusion of a 5'-phosphate in nucleoside derivatives greatly improves inhibitory potency. For example, nucleotide derivatives **77** and **73**, bearing a 2-thiophene carboxamide and sulfonamide ZBG, respectively, inhibit SNM1A at a concentration as low as 33 μM (Figure 4.4), while the corresponding nucleoside derivative **34** bearing a sulfonamide shows inhibition only at 1 mM concentration (Figure 2.5B, p. 65), and thiophene-bearing nucleoside **31** does not inhibit SNM1A in gel electrophoresis assays (Figure 2.4, p. 64). Similarly, thymidine derivatives **87** and **91** bearing a 5'-phosphate and a squaramide ZBG at the 3'-position are effective inhibitors of SNM1A, while their unphosphorylated analogues **44** and **45** described in Chapter 3 were not. These results also indicate that deoxyribonucleotide derivatives inhibit SNM1A more effectively than ribonucleotide derivatives, which is consistent with the enzyme's preference for DNA substrates over RNA. For example, squaramide-bearing uridine derivative **81** has an IC_{50} of 46.3 μM , while the analogous thymidine derivative **89** has a lower IC_{50} value of 31.7 μM . Ribonucleotide inhibitors are also effective however, and the choice of ZBG appears to be a more important factor, as sulfonamide-bearing uridine derivative **73** proved to be the most potent of any of the compounds tested, with an IC_{50} of 19.7 μM . The use of a ribonucleoside scaffold in place of a deoxyribonucleoside also offers the potential for further functionalisation at the 2'-position for future optimisation of lead compounds. Further development of inhibitors bearing a 5'-phosphate for *in vivo* use may be facilitated by investigations of prodrug strategies for cellular delivery of nucleoside 5'-monophosphates already reported in the literature.^{350–354} Overall, the work presented in this chapter demonstrates a marked

improvement in the potency of substrate-mimic SNM1A inhibitors, and offers insights useful for further development of compounds of this type.

5. Conclusion and Future Work

This work details the development of substrate-mimic inhibitors for targeting the DNA repair enzyme SNM1A. The inhibitors are based on a nucleoside scaffold, appended with a ZBG at the 3'- or 5'-position. A dinucleoside bearing a ZBG as an internucleoside linkage was also investigated. The active inhibitors are shown in Figure 5.1.

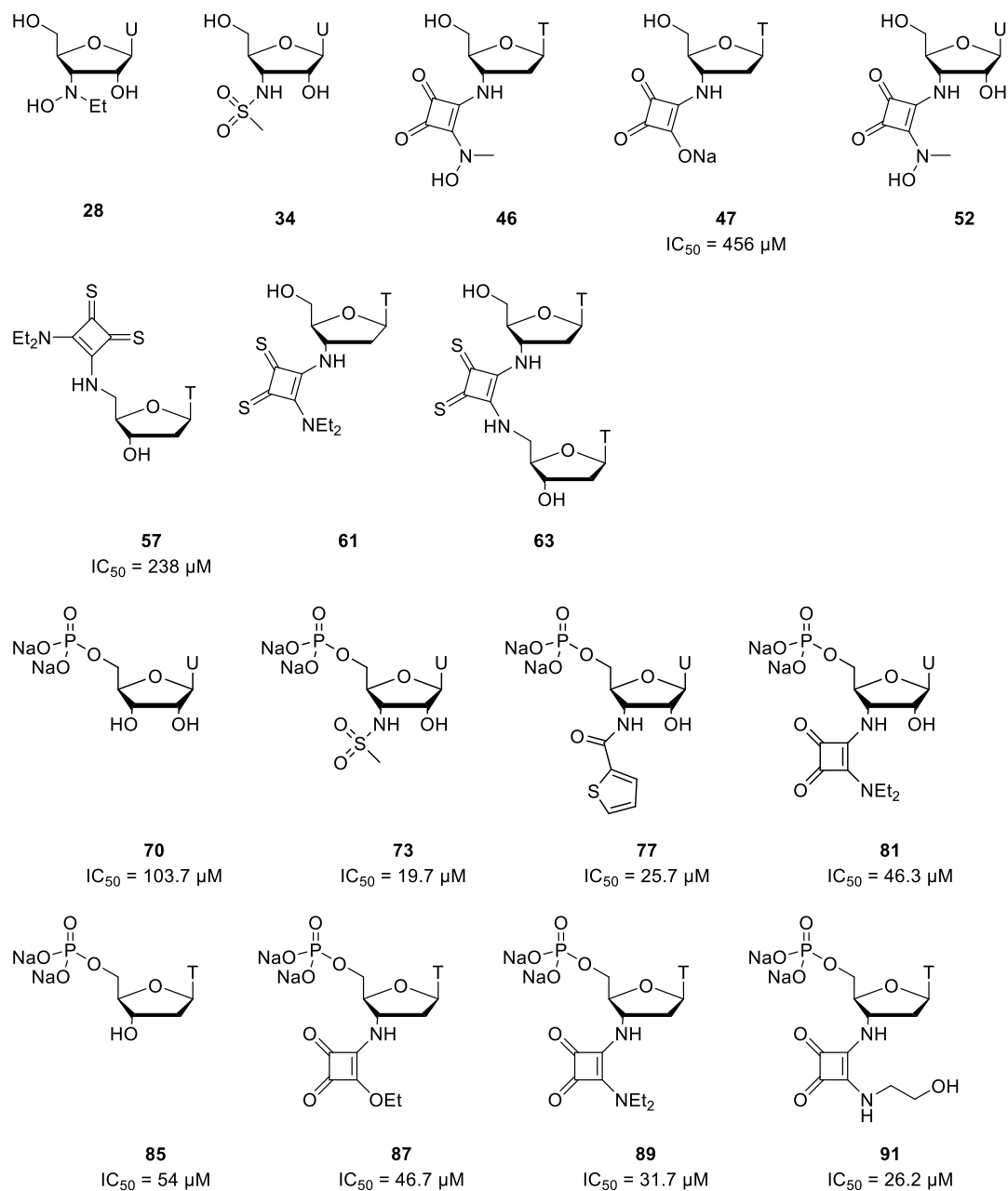


Figure 5.1. Inhibitors of SNM1A developed in this work.

The selection of an appropriate ZBG is a key factor in the development of molecules for targeting SNM1A. This work therefore aimed to test a series of novel ZBGs, as well as established ZBGs previously untried against this enzyme. Uridine derivatives bearing oxime, hydroxylamine, 2-thiophene carboxamide, and sulfonamide ZBGs were prepared and tested in gel-electrophoresis-based assays, as described in Chapter 2. Hydroxylamine **28** and sulfonamide **34** proved efficacious as SNM1A inhibitors, although neither compound showed appreciable inhibition of the enzyme at concentrations lower than 333 μM .

Squaramides bearing various substituents were also tested as novel ZBGs for targeting SNM1A, as detailed in Chapter 3. Nucleoside derivatives **46**, **47**, and **52** bearing squaramides at the 3'-position proved effective in gel electrophoresis assays, while analogous compounds bearing squaramides at the 5'-position showed no inhibitory potency. Thymidine derivative **46** bearing an *N*-hydroxysquaramide showed greater efficacy than the analogous uridine derivative **52**, consistent with the enzyme's preference for DNA substrates over RNA. Interestingly, thymidine derivative **57** bearing a 5'-thiosquaramide inhibited SNM1A with greater potency than compound **61** bearing a 3'-thiosquaramide, the reverse of the trend observed for oxo squaramides. This was likely because a 3'-thiosquaramide was less readily accommodated in the SNM1A active site due to the large steric bulk of this ZBG. Although malonate-linked dinucleosides previously reported were found to inhibit SNM1A more effectively than mononucleosides bearing a 3'-malonate ZBG,¹²⁷ the reverse trend was again observed in the case of thiosquaramide-linked dinucleoside **63**, which inhibited SNM1A less effectively than either of the thiosquaramide-bearing mononucleosides **57** and **61**. The rigidity of the thiosquaramide linkage likely restricts dinucleoside **63** to an unfavourable conformation for SNM1A-

binding. These results show that steric size and conformational flexibility are important considerations in the selection of ZBGs for SNM1A inhibitors. UV-vis titrations however showed that zinc-binding strength is also well correlated with inhibitory potency, as squaric acid **47** was found to bind to zinc ions with greater affinity than *N*-hydroxysquaramide **46**, a weaker inhibitor of SNM1A. Thiosquaramide **57**, the most potent inhibitor tested in this chapter, exhibited the strongest interaction with zinc ions. Three of the active inhibitors; squaramides **46** and **47**, and to a lesser extent thiosquaramide **57**, were found to be membrane permeable. The effect of these compounds on cancer cells in conjunction with a crosslinking agent may be tested as a future direction of this work.

SNM1A contains a phosphate-binding pocket adjacent to the active site, and the presence of a 5'-phosphate in a substrate molecule is essential for SNM1A activity. The results presented in Chapter 4 show that nucleoside-based inhibitors can exploit interactions with both the catalytic zinc ion(s) and the phosphate-binding pocket, through inclusion of both a 5'-phosphate moiety and a 3'-ZBG. Phosphorylation of the 5'-position of the inhibitors led to a dramatic improvement in potency when compared to analogous compounds described in Chapters 2 and 3 which contain identical ZBGs but lack phosphate groups. For example, sulfonamide-bearing nucleotide derivative **73** inhibited SNM1A at a concentration as low as 33 μ M in a gel-electrophoresis assay, while sulfonamide-bearing nucleoside **34** showed appreciable inhibition only at 1 mM concentration. Furthermore, 2-thiophene carboxamide **77**, and squaramides **87** and **91** inhibit SNM1A with IC₅₀ values in the 25-50 μ M range, while analogous compounds lacking a 5'-phosphate do not inhibit SNM1A at 1 mM concentration. A prodrug strategy utilising cell permeable, biocleavable phosphate protecting groups may allow the SNM1A inhibitors developed in this chapter to be tested *in vivo* in the future.

Overall, as well as producing a number of SNM1A inhibitors of improved potency, this work has generated a number of insights that will inform the further development of molecules for targeting this enzyme. A number of effective ZBGs were identified, including hydroxylamine, sulfonamide, squaramide, thiosquaramide, and 2-thiophene carboxamide moieties. The ideal positioning of a ZBG in a nucleoside derivative was explored, and 5'-phosphorylation was found to be an important factor. Incorporation of the most potent inhibitors into oligonucleotide probes to study SNM1A is a promising future direction of this work, as previous research in the McGouran lab showed that incorporation of thymidine derivatives bearing 5'-squaramides into oligonucleotides greatly enhanced their affinity for SNM1A. To allow for a 5'-phosphate to be incorporated, a ZBG may be better placed at the 3'-position of the 5'-terminal residue of an oligonucleotide. The use of a ribonucleotide scaffold will enable attachment of the 5'-terminal residue to the oligonucleotide *via* a 2'-5' linkage, as this work showed that uridine derivatives have only slightly lower affinity for SNM1A than analogous thymidine derivatives. The presence of 2'-5' linkages may also facilitate *in vivo* use of oligonucleotide probes by conferring resistance to digestion by some nucleases,³⁵⁸⁻³⁶⁰ and reducing their immunogenicity.³⁶¹ 2'-5' linked DNA strands can act as a template for DNA polymerases,³⁶² and 2'-5' linkages can be tolerated by ribozymes,³⁶³ showing that these backbone modifications do not necessarily distort the structure of nucleic acids so greatly as to disrupt their function or recognition by enzymes. Although there have been two reports of synthetic siRNAs containing 2'-5' linkages which were successfully used to silence gene expression in *in vitro* studies,^{361,364} the full potential for use of this backbone modification in biological applications has yet to be explored.

6. Experimental

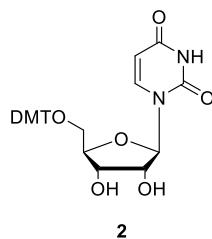
6.1. General experimental

^1H , ^{13}C and ^{31}P NMR spectra were recorded using an Agilent 400-MR, Bruker 400 MHz or Bruker 600 MHz system. Spectra were referenced relative to residual DMSO, CHCl_3 , water, MeOH, or acetone as appropriate. Infrared (IR) spectra were measured using a Perkin Elmer spectrophotometer. ESI mass spectra were acquired using a Bruker microTOF-Q III spectrometer interfaced to a Dionex UltiMate 3000 LC in positive and negative modes as required. The instrument was calibrated using a tune mix solution (Agilent Technologies ESI-I Low concentration tuning mix) which was also used as an internal lock mass. Operating conditions were as follows: end-plate offset 500 V, capillary 4500 V, nebulizer 2.0 Bar, dry gas 8.0 L/min, and dry temperature 180 °C. MicroTof control 3.2 and HyStar 3.2 software were used to carry out the analysis. APCI experiments were carried out on the same machines. Agilent tuning mix APCI-TOF was used to calibrate the system. Operating conditions were as follows: Capillary voltage 4000 V, corona 4000 nA, nebulizer gas 2.0 Bar, dry gas 3.0 L/min, dry gas temperature 100-200 °C, vapour temperature 100-400°C. MicroTof control 3.2 and HyStar 3.2 software were used to carry out the analysis. MALDI-TOF spectra were acquired using a Waters MALDI Q-TOF Premier. The instrument was operated in positive or negative mode as required. The laser operated at 337 nm. Samples were run using DCTB, CHCA, DHB and HPA matrices. 1 μL of matrix/sample was loaded onto a MALDI plate and allowed to air dry. The instrument was calibrated using PEG. The internal lock mass used was [Glu] Fibrinopeptide B. MassLynx 4.1 software was used to carry out the analysis. Melting points were measured using a Griffin melting point apparatus. Flash chromatography was carried out using silica gel, particle size 40-60 μm . TLC analysis was performed on precoated 60F₂₅₄ slides and visualised by UV irradiation or staining with KMnO_4 , ninhydrin, anisaldehyde, or iodine. Unless stated otherwise chemicals were purchased from commercial sources and used as received. Anhydrous MeCN, THF and CH_2Cl_2 were prepared using a PureSolv MD solvent purification system.

6.2. Synthetic methods

6.2.1. Synthetic methods for Chapter 2

5'-O-(4,4'-Dimethoxytrityl)uridine³³⁵ (**2**)



Uridine (**1**) (5.04 g, 20.6 mmol) was dissolved in pyridine (67 mL). DMTCl (10.66 g, 31.5 mmol) was added in portions over 80 minutes. The reaction mixture was stirred under argon at room temperature for 19 hours. After this time TLC analysis (CH₂Cl₂-MeOH, 94:6) showed formation of the product ($R_f = 0.3$) and consumption of the uridine starting material ($R_f = 0.0$). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 94:6). The product **2** was obtained as a white foam (7.93 g, 70%).

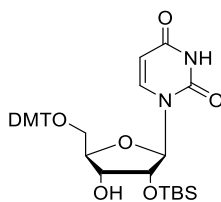
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3377 (N-H), 3056, 2928, 1675, 1607, 1508, 1460, 1247, 1175, 1105, 1031, 904, 825, 754, 727, 701, 581.

¹H NMR (600 MHz, DMSO-d₆): $\delta = 3.20$ - 3.27 (m, 2H, H-5'), 3.74 (s, 6H, 2 x OCH₃), 3.92-3.97 (m, 1H, H-4'), 4.04-4.12 (m, 2H, H-2', H-3'), 5.14 (d, $J_{3',\text{OH}-3'} = 5.3$ Hz, 1H, OH-3'), 5.31 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.49 (d, $J_{2',\text{OH}-2'} = 4.4$ Hz, 1H, OH-2'), 5.74 (d, $J_{1',2'} = 3.3$ Hz, 1H, H-1'), 6.89 (d, $J = 6.90$ Hz, 4H, H^{Ar}), 7.22-7.39 (m, 9H, H^{Ar}), 7.71 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.35 (s, 1H, NH) ppm.

¹³C NMR (151 MHz, DMSO-d₆): $\delta = 55.1$ (OCH₃), 63.0 (C-5'), 69.6 (C-3'), 73.4 (C-2'), 82.4 (C-4'), 85.9 (qC^{DMT}), 88.9 (C-1'), 101.5 (C-5), 113.3, 123.9, 126.8, 127.7, 127.9, 129.8, 135.2, 135.4, 136.1, 140.6 (C-6), 144.7, 149.6, 150.5 (C-2), 158.1, 163.0 (C-4) ppm.

HRMS (ESI⁺): m/z calc. 569.1894 [M+Na]⁺, found: 569.1881.

5'-O-(4,4'-Dimethoxytrityl)-2'-O-(*tert*-butyldimethylsilyl)uridine (**4**)



5'-O-(4,4'-Dimethoxytrityl)uridine (**2**) (7.93 g, 14.5 mmol) was dissolved in anhydrous DMF (55 mL). Imidazole (3.33 g, 48.9 mmol) was added, followed by TBDMSCl (2.62 g, 17.4 mmol). The reaction mixture was stirred under argon at room temperature for 85 minutes, after which time TLC analysis (petroleum ether-EtOAc 1:1) showed formation of 5'-O-(4,4'-dimethoxytrityl)-2'-O-(*tert*-butyldimethylsilyl)uridine (**4**) ($R_f = 0.4$), 5'-O-(4,4'-dimethoxytrityl)-3'-O-(*tert*-butyldimethylsilyl)uridine (**5**) ($R_f = 0.2$) and 5'-O-(4,4'-dimethoxytrityl)-2',3'-bis-O-(*tert*-butyldimethylsilyl)uridine (**29**) ($R_f = 0.5$). The reaction was quenched by addition of methanol (7.5 mL). The reaction mixture was diluted with EtOAc (100 mL) and washed with water (125 mL). The organic layer was dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (petroleum ether-EtOAc, 2:1). 5'-O-(4,4'-Dimethoxytrityl)-2'-O-(*tert*-butyldimethylsilyl)uridine (**4**) was obtained as a white foam (3.23 g, 34%). 5'-O-(4,4'-Dimethoxytrityl)-3'-O-(*tert*-butyldimethylsilyl)uridine (**3**) was obtained as a white foam (2.63 g, 28%). 5'-O-(4,4'-Dimethoxytrityl)-2',3'-bis-O-(*tert*-butyldimethylsilyl)uridine (**29**) was obtained as a white foam (0.65 g, 7%).

5'-O-(4,4'-Dimethoxytrityl)-3'-O-(*tert*-butyldimethylsilyl)uridine (**3**) (2.63 g, 3.99 mmol) was dissolved in pyridine (45.5 mL) and water (5.4 mL), and stirred at room temperature for 4 days. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (petroleum ether-EtOAc, 2:1). 5'-O-(4,4'-dimethoxytrityl)-2'-O-(*tert*-butyldimethylsilyl)uridine (**4**) was obtained as a white foam (2.14 g). The total yield of 5'-O-(4,4'-dimethoxytrityl)-2'-O-(*tert*-butyldimethylsilyl)uridine (**4**) was 5.37 g (56%).

5'-O-(4,4'-Dimethoxytrityl)-2'-O-(tert-butyldimethylsilyl)uridine³³⁶ (4)

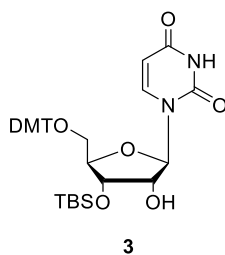
$\nu_{\max}/\text{cm}^{-1}$ (neat) 2932, 2857, 2172, 1678, 1611, 1508, 1463, 1248, 1174, 1067, 1034, 829, 776, 699, 587.

^1H NMR (400 MHz, DMSO- d_6): δ = 0.06 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.07 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.86 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.25 (dd, $J_{5'a,5'b} = 10.7$ Hz, $J_{4',5'a} = 2.5$ Hz, 1H, H-5'a), 3.32 (m, 1H, H-5'b), 3.74 (s, 6H, 2 x OCH_3), 4.00 (m, 1H, H-4'), 4.07 (app t, $J = 4.5$ Hz, 1H, H-3'), 4.21 (app t, $J = 4.5$ Hz, 1H, H-2'), 5.12 (br s, 1H, OH-3'), 5.30 (dd, $J_{5,6} = 8.1$ Hz, $J_{5,\text{NH}} = 2.0$ Hz, 1H, H-5), 5.76 (d, $J_{1',2'} = 4.5$ Hz, 1H, H-1'), 6.91 (d, $J = 8.9$ Hz, 4H, H^{Ar}), 7.22-7.41 (m, 9H, H^{Ar}), 7.75 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.37 (d, $J_{5,\text{NH}} = 2.0$ Hz, 1H, NH) ppm.

^{13}C NMR (100 MHz, DMSO- d_6): δ = -5.1 ($\text{CH}_3^{\text{TBDMS}}$), -4.7 ($\text{CH}_3^{\text{TBDMS}}$), 17.9 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.6 ($t\text{Bu}^{\text{TBDMS}}$), 55.1 (OCH_3), 62.7 (C-5'), 69.4 (C-3'), 75.6 (C-2'), 82.7 (C-4'), 86.0 (qC^{DMT}), 88.6 (H-1'), 101.5 (C-5), 112.8 (Ar), 113.3 (Ar), 126.8 (Ar), 127.7 (Ar), 127.9 (Ar), 128.9 (Ar), 129.8 (Ar), 135.0 (Ar), 135.3 (Ar), 140.1 (C-6), 144.7 (Ar), 150.4 (C-2), 158.2 (Ar), 163.0 (C-4)

HRMS (ESI⁺): m/z calc. 683.2759 [M+H]⁺, found: 683.2746.

5'-O-(4,4'-Dimethoxytrityl)-3'-O-(tert-butyldimethylsilyl)uridine³³⁶ (3)



$\nu_{\max}/\text{cm}^{-1}$ (neat) 3321, 2983, 1697, 1661, 1503, 1381, 1239, 1144, 1113, 1029, 949, 856, 819, 802, 696.

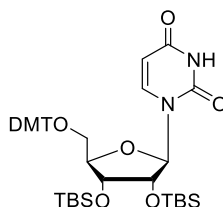
^1H NMR (600 MHz, DMSO- d_6): δ = -0.03 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.02 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.78 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.15-3.19 (m, 1H, H-5'a), 3.33 (m, 1H, H-5'b), 3.74 (s, 6H, 2 x OCH_3), 3.92 (m, 1H, H-4'), 4.10 (m, 1H, H-2'), 4.15 (app t, $J = 5.3$ Hz, 1H, H-3'), 5.33 (d, $J_{2',\text{OH}-2'} = 6.0$ Hz, 1H, OH-2'), 5.37 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.73 (d, $J_{1',2'} = 4.3$ Hz, 1H, H-1'), 6.87-6.93 (m, 4H, H^{Ar}), 7.22-7.40 (m, 9H, H^{Ar}), 7.76 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.36 (s, 1H, NH) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = -5.2 ($\text{CH}_3^{\text{TBDMS}}$), -4.5 ($\text{CH}_3^{\text{TBDMS}}$), 17.9 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.7 ($t\text{Bu}^{\text{TBDMS}}$), 55.1 (OCH_3), 62.6 (C-5'), 71.3 (C-3'), 72.8 (C-2'), 82.7 (C-4'), 86.0 (qC^{DMT}), 89.1

(C-1'), 101.6 (C-5), 113.2 (Ar), 126.9 (Ar), 127.7 (Ar), 127.9 (Ar), 129.8 (Ar), 135.1 (Ar), 135.2 (Ar), 140.7 (C-6), 144.5 (Ar), 150.6 (C-2), 158.2 (Ar), 163.1 (C-4) ppm.

HRMS (ESI⁺): *m/z* calc. 683.2759 [M+Na]⁺, found: 683.2739.

5'-O-(4,4'-Dimethoxytrityl)-2',3'-bis-O-(*tert*-butyldimethylsilyl)uridine³⁶⁵ (5)



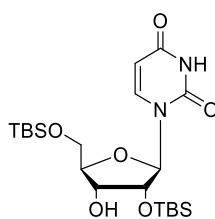
$\nu_{\max}/\text{cm}^{-1}$ (neat) 2929, 2854, 1681, 1608, 1509, 1462, 1250, 1174, 1128, 1068, 1035, 1002, 865, 830, 777, 701.

¹H NMR (600 MHz, DMSO-d₆): δ = -0.06 (s, 3H, CH₃^{TBDMS}), 0.01 (s, 3H, CH₃^{TBDMS}), 0.03 (s, 3H, CH₃^{TBDMS}), 0.04 (s, 3H, CH₃^{TBDMS}), 0.75 (s, 9H, tBu^{TBDMS}), 0.84 (s, 9H, tBu^{TBDMS}), 3.22 (dd, J_{5'a,5'b} = 11.0 Hz, J_{4',5'a} = 3.5 Hz, 1H, H-5'a), 3.44 (dd, J_{5'a,5'b} = 11.0 Hz, J_{4',5'b} = 3.0 Hz, 1H, H-5'b), 3.74 (s, 6H, 2 x OCH₃), 3.97 (m, 1H, H-4'), 4.07 (app t, J = 4.4 Hz, 1H, H-3'), 4.27 (app t, J = 4.4 Hz, 1H, H-2'), 5.36 (d, J_{5,6} = 8.1 Hz, 1H, H-5), 5.73 (d, J_{1',2'} = 4.4 Hz, 1H, H-1'), 6.89-6.91 (m, 4H, H^{Ar}), 7.20-7.39 (m, 9H, H^{Ar}), 7.82 (d, J_{5,6} = 8.1 Hz, 1H, H-6), 11.39 (s, 1H, NH) ppm.

¹³C NMR (151 MHz, DMSO-d₆): δ = -5.2 (CH₃^{TBDMS}), -4.91 (CH₃^{TBDMS}), -4.86 (CH₃^{TBDMS}), -4.4 (CH₃^{TBDMS}), 17.59 (qC, tBu^{TBDMS}), 17.61 (qC, tBu^{TBDMS}), 25.6 (2 x tBu^{TBDMS}), 55.1 (OCH₃), 62.2 (C-5'), 71.1 (C-3'), 74.5 (C-2'), 82.8 (C-4'), 86.3 (qC^{DMT}), 88.3 (C-1'), 101.6 (C-5), 113.23 (Ar), 113.25 (Ar), 127.0 (Ar), 127.7 (Ar), 127.9 (Ar), 129.8 (Ar), 129.9 (Ar), 134.8 (Ar), 135.0 (Ar), 140.1 (C-6), 144.4 (Ar), 150.5 (C-2), 158.27 (Ar), 158.28 (Ar), 163.0 (C-4) ppm.

HRMS (ESI⁺): *m/z* calc. 797.3624 [M+Na]⁺, found: 797.3648.

2',5'-Bis-*O*-(*tert*-butyldimethylsilyl)uridine³³⁷ (**6**)



6

Uridine (**1**) (5.01 g, 20.5 mmol) was dissolved in pyridine (42 mL). TBDMSCl (9.73 g, 64.6 mmol) was added. The reaction mixture was stirred at room temperature under argon for 17 hours. TLC analysis (CH₂Cl₂-MeOH, 95:5) after this time showed formation of the desired product (*R*_f = 0.5), but incomplete consumption of starting material (*R*_f = 0.0). Another portion of TBDMSCl (1.53 g, 10.2 mmol) was added and the reaction mixture was stirred under argon for a further 4.5 hours. After this time another portion of TBDMSCl (1.51 g, 10.0 mmol) was added and the reaction mixture was stirred for a further 2.5 hours. Another portion of TBDMSCl (1.51 g, 10.0 mmol) was added and the reaction was stirred for a further 17.5 hours. The reaction was then quenched by addition of MeOH (5 mL), and the solvent was removed under reduced pressure. The residue was purified by flash chromatography (CH₂Cl₂-MeOH, 98:2), and the product **6** was obtained as a white foam (8.60 g, 89%).

Note: compound **6** exhibits rotamers in NMR spectroscopy.

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 2962, 2929, 2859, 1685, 1460, 1254, 1123, 1063, 1002, 831, 777.

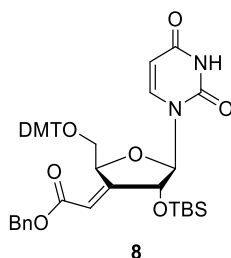
¹H NMR (600 MHz, DMSO-*d*₆): δ = 0.00-0.11 (m, 12H, 4 x CH₃^{TBDMS}), 0.81-0.93 (m, 18H, 2 x *t*Bu^{TBDMS}), 3.64-3.98 (m, 3.6H, 0.6 x H-3', H-4', 2 x H-5'), 4.09 (m, 1.4H, H-2', 0.4 x H-3'), 5.09 (d, *J*_{OH-3',3'} = 5.4 Hz, 0.6H, OH-3'), 5.36 (d, *J*_{OH-3',3'} = 5.4 Hz, 0.4H, OH-3'), 5.60 (d, *J*_{5,6} = 8.1 Hz, 1H, H-5), 5.76 (d, *J*_{1',2'} = 4.9 Hz, 0.4H, H-1'), 5.81 (d, *J*_{1',2'} = 4.9 Hz, 0.6H, H-1'), 7.73 (d, *J*_{5,6} = 8.1, 0.4H, H-6), 7.80 (d, *J*_{5,6} = 8.1, 0.6H, H-6), 11.37 (s, 1H, NH) ppm.

¹³C NMR (151 MHz, DMSO-*d*₆): δ = -5.60 (CH₃^{TBDMS}), -5.56 (CH₃^{TBDMS}), -5.2 (CH₃^{TBDMS}), -5.1 (CH₃^{TBDMS}), -4.8 (CH₃^{TBDMS}), -4.5 (CH₃^{TBDMS}), 17.87 (qC, *t*Bu^{TBDMS}), 17.99 (qC, *t*Bu^{TBDMS}), 18.01 (qC, *t*Bu^{TBDMS}), 18.03 (qC, *t*Bu^{TBDMS}), 25.6 (*t*Bu^{TBDMS}), 25.8 (*t*Bu^{TBDMS}), 62.4 (C-5'), 62.6 (C-5'), 69.6 (C-3'), 71.6 (C-3'), 72.7 (C-2'), 75.9 (C-2'), 84.5 (C-4'), 87.6 (C-1'), 87.7 (C-1'), 101.66 (C-

5), 101.71 (C-5), 139.7 (C-6), 140.3 (C-6), 150.5 (C-2), 150.7 (C-2), 162.9 (C-4), 163.0 (C-4) ppm.

HRMS (ESI⁺): *m/z* calc. 473.2498 [M+H]⁺, found: 473.2499.

5'-O-(4,4'-Dimethoxytrityl)-3'-(benzylcarboxy)ylidene-2'-O-(*tert*-butyldimethylsilyl)-3'-deoxyuridine (8**)**



5'-O-(4,4'-Dimethoxytrityl)-2'-O-(*tert*-butyldimethylsilyl)uridine (**4**) (1.02 g, 1.54 mmol) was dissolved in anhydrous CH₂Cl₂ (5 mL). Dess-Martin periodinane (1.03 g, 2.42 mmol) was dissolved in anhydrous CH₂Cl₂ (17.5 mL), and this solution was added to the reaction mixture at 0 °C. The reaction mixture was stirred under argon at 0 °C for 25 minutes, and was then allowed to warm to room temperature and stirred under argon for a further 3 hours. TLC analysis (petroleum ether-EtOAc 1:1) at this time showed complete consumption of starting material (*R_f* = 0.5) and formation of 5'-O-(4,4'-dimethoxytrityl)-2'-O-(*tert*-butyldimethylsilyl)-3'-oxouridine intermediate (**7**) (*R_f* = 0.6). The reaction mixture was diluted with EtOAc (50 mL), and an ice cold aqueous solution of 10% (w/v) sodium thiosulfate (30 mL) was added, followed by an ice cold saturated NaHCO₃ solution (30 mL). The organic layer was separated and washed with saturated NaHCO₃ solution (50 mL), followed by water (50 mL). The organic layer was dried over MgSO₄ and then filtered. The solvent was removed under reduced pressure.

The residue was redissolved in anhydrous CH₂Cl₂ (20 mL). Benzyl(triphenylphosphoranylidene)acetate (0.70 g, 1.70 mmol) was added, and the reaction mixture was stirred under argon at room temperature. After 4 hours additional benzyl(triphenylphosphoranylidene)acetate (0.23 g, 0.57 mmol) was added. NMR analysis of a 0.2 mL aliquot of the reaction mixture 90 minutes later showed complete consumption of the starting material. The solvent was removed under reduced pressure, and the residue purified by flash chromatography to give the product **8** as a white foam (1.08 g, 99%).

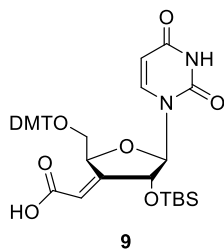
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3369, 3058, 2864, 1697, 1649, 1607, 1508, 1461, 1247, 1174, 1064, 1029, 829, 735, 695.

$^1\text{H NMR}$ (400 MHz, DMSO- d_6): δ = -0.05 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.09 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.87 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.11 (dd, $J_{5'a,5'b} = 10.1$ Hz, $J_{4',5'a} = 1.8$ Hz, 1H, H-5'a), 3.64 (dd, $J_{5'a,5'b} = 10.1$ Hz, $J_{4',5'b} = 3.5$ Hz, 1H, H-5'b), 3.73 (s, 3H, OCH_3), 3.74 (s, 3H, OCH_3) 5.04-5.06 (m, 2H, H-2', $\text{H}^{\text{CH}_2\text{Ph-a}}$), 5.10 (d, $J_{\text{CH}_2\text{Ph-a}, \text{CH}_2\text{Ph-b}} = 12.3$ Hz, 1H, $\text{H}^{\text{CH}_2\text{Ph-b}}$), 5.33 (dd, $J_{4',5'b} = 3.5$ Hz, $J_{4',5'a} = 1.8$ Hz, 1H, H-4'), 5.57 (dd, $J_{5,6} = 8.1$ Hz, $J_{5,\text{NH}} = 2.1$ Hz, 1H, H-5), 5.82 (d, $J_{1',2'} = 8.0$ Hz, 1H, H-1'), 5.87 (app t, $J = 2.2$ Hz, 1H, CH^{Vinyl}), 6.83-6.90 (m, 5H, Ar), 7.17-7.35 (m, 13H, Ar), 7.73 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.55 (d, $J_{5,\text{NH}} = 2.1$ Hz, 1H, NH) ppm.

$^{13}\text{C NMR}$ (100 MHz, DMSO- d_6): δ = -5.4 ($\text{CH}_3^{\text{TBDMS}}$), -5.1 ($\text{CH}_3^{\text{TBDMS}}$), 17.6 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.3 ($t\text{Bu}^{\text{TBDMS}}$), 55.0 (OCH_3), 64.4 (C-5'), 65.9 (CH_2^{Bn}), 75.5 (C-2'), 77.9 (C-4'), 84.7 (C-1'), 86.1 (qC $^{\text{DMT}}$), 102.3 (C-5), 112.8 (Ar), 113.25 (CH^{Vinyl}), 113.26 (Ar), 127.4 (Ar), 127.9 (Ar), 128.2 (Ar), 128.3 (Ar), 128.5 (Ar), 128.9 (Ar), 129.6 (Ar), 134.7 (Ar), 135.2 (Ar), 135.6 (Ar), 139.2 (C-6), 144.6 (Ar), 150.6 (C-2), 157.8 (Ar), 158.2 (Ar), 159.5 (Ar), 162.6 (C-4), 163.9 (CO^{Ester}) ppm.

HRMS (APCI $^-$): m/z calc. 789.3213 [M-H] $^-$, found: 789.3204.

5'-O-(4,4'-Dimethoxytrityl)-3'-ylideneacetic acid-2'-O-(*tert*-butyldimethylsilyl)-3'-deoxyuridine (9)



5'-O-(4,4'-Dimethoxytrityl)-3'-(benzylcarboxy)ylidene-2'-O-(*tert*-butyldimethylsilyl)-3'-deoxyuridine (**8**) (1.08 g, 1.37 mmol) was suspended in a mixture of *t*BuOH (30 mL) and water (30 mL). The reaction mixture was degassed by bubbling with argon, and 10% Pd/C (0.30 g) was added. The reaction vessel was purged with H_2 , and the reaction mixture was stirred under an atmosphere of H_2 for 5 hours. After this time TLC analysis (petroleum ether-EtOAc 1:1) showed consumption of starting material ($R_f = 0.6$) and formation of product ($R_f = 0.1$) The reaction mixture was filtered through celite, washing through with

MeOH. The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH₂Cl₂-MeOH 96:4). The product **9** was obtained as a white foam (0.81 g, 85%).

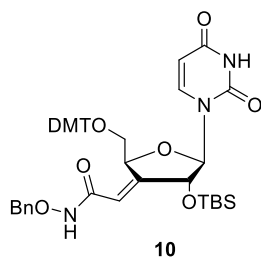
$\nu_{\max}/\text{cm}^{-1}$ (neat) 2934, 2858, 2819, 1696, 1508, 1452, 1387, 1248, 1174, 1032, 994, 907, 835, 630, 540.

¹H NMR (400 MHz, DMSO-d₆): δ = -0.03 (s, 3H, CH₃^{TBDMS}), 0.11 (s, 3H, CH₃^{TBDMS}), 0.89 (s, 9H, tBu^{TBDMS}), 3.13 (dd, $J_{5'a,5'b}$ = 9.8 Hz, $J_{5'a,4'}$ = 1.7 Hz, 1H, H-5'a), 3.72 (dd, $J_{5'a,5'b}$ = 9.8 Hz, $J_{5'b,4'}$ = 3.2 Hz, 1H, H-5'b), 3.74 (s, 6H, OCH₃), 5.03 (app dt, J = 7.9 Hz, J = 2.1 Hz, 1H, H-2'), 5.37 (m, 1H, H-4'), 5.53 (dd, $J_{5,6}$ = 8.2 Hz, $J_{5,\text{NH}}$ = 1.8 Hz, 1H, H-5), 5.81 (d, $J_{1',2'}$ = 7.9 Hz, 1H, H-1'), 5.84 (app t, J = 2.1 Hz, 1H, CH^{vinyl}), 6.87-6.91 (m, 4H, Ar), 7.20-7.35 (m, 9H, Ar), 7.73 (d, $J_{5,6}$ = 8.2 Hz, 1H, H-6), 11.54 (d, $J_{5,\text{NH}}$ = 1.8 Hz, 1H, NH) ppm.

¹³C NMR (100 MHz, DMSO-d₆): δ = -5.4 (CH₃^{TBDMS}), -5.1 (CH₃^{TBDMS}), 17.6 (qC, tBu^{TBDMS}), 25.3 (tBu^{TBDMS}), 55.1 (OCH₃), 64.5 (C-5'), 75.5 (C-2'), 77.8 (C-4'), 84.7 (C-1'), 86.0 (qC^{DMT}), 102.7 (C-5), 113.3, 126.9, 127.4 (Ar), 127.9 (Ar), 128.4 (Ar), 129.6, 129.2, 129.3 (Ar), 131.9, 132.5, 134.8, 135.2, 139.3 (C-6), 144.8, 150.6 (C-2), 158.1 (C-3'), 158.2 (Ar), 162.6 (C-4), 165.8 (CO₂H) ppm.

HRMS (ESI⁺): m/z calc. 975.3403 [M+K]⁺, found: 975.3438.

5'-O-(4,4'-Dimethoxytrityl)-3'-(O-benzyl-N'-acetylhydroxylamine)ylidene-2'-O-(tert-butylidimethylsilyl)-3'-deoxyuridine (10)



5'-O-(4,4'-dimethoxytrityl)-3'-ylideneacetic acid-2'-O-(tert-butylidimethylsilyl)-3'-deoxyuridine (**9**) (102 mg, 0.14 mmol) was dissolved in anhydrous DMF (1.7 mL). EDCI.HCl (34 mg, 0.18 mmol) was added, followed by HOAt (27 mg, 0.20 mmol), *o*-benzylhydroxylamine hydrochloride (21 mg, 0.17 mmol), and DIPEA (32 μ L, 0.18 mmol). The reaction mixture was stirred under argon at room temperature for 17 hours. TLC

analysis (petroleum ether-EtOAc 1:1) after this time showed consumption of starting material ($R_f = 0.1$) and formation of product ($R_f = 0.3$). The reaction mixture was diluted with CH_2Cl_2 (10 mL) and water (10 mL). The organic layer was separated and washed with brine (10 mL), then dried over MgSO_4 , and filtered. The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (petroleum ether:EtOAc 1:1) to obtain the product **10** as a white foam (31 mg, 26%).

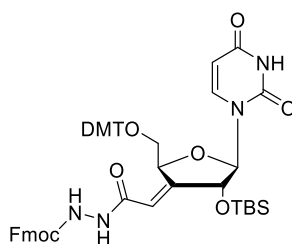
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3257, 2930, 1678, 1461, 1251, 1112, 1069, 834, 778, 559.

^1H NMR (400 MHz, DMSO-d_6): $\delta = -0.02$ (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.14 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.88 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.11 (dd, $J_{5'a,5'b} = 9.8$ Hz, $J_{4',5'a} = 1.6$ Hz, 1H, H-5'a), 3.70 (s, 3H, OCH_3), 3.72 (s, 3H, OCH_3), 3.77 (dd, $J_{5'a,5'b} = 9.8$ Hz, $J_{4',5'b} = 3.4$ Hz, 1H, H-5'b), 4.80 (s, 2H, CH_2^{Bn}), 5.01 (m, 1H, H-2'), 5.47 (d, $J_{5,6} = 8.1$ Hz, H-5), 5.50 (m, 1H, H-4'), 5.79 (d, $J_{1',2'} = 7.9$ Hz, 1H, H-1'), 5.84 (app t, $J = 2.1$ Hz, 1H, CH^{Vinyl}), 6.87-6.90 (m, 4H, Ar), 7.21-7.40 (m, 14H, Ar), 7.75 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.52 (s, 1H, NH^{U}), 11.66 (s, 1H, $\text{NH}^{\text{Hydroxamate}}$) ppm.

^{13}C NMR (100 MHz, DMSO-d_6): $\delta = -5.5$ ($\text{CH}_3^{\text{TBDMS}}$), -4.9 ($\text{CH}_3^{\text{TBDMS}}$), 17.6 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.4 ($t\text{Bu}^{\text{TBDMS}}$), 55.00 (OCH_3), 55.03 (OCH_3), 64.8 (C-5'), 75.6 (C-2'), 77.1 (CH_2^{Bn}), 77.8 (C-4'), 84.9 (C-1'), 86.0 (qC $^{\text{DMT}}$), 102.6 (C-5), 113.2 (CH^{Vinyl}), 113.3 (Ar), 126.9 (Ar), 127.4 (Ar), 127.9 (Ar), 128.4 (Ar), 128.8 (Ar), 129.65 (Ar), 129.69 (Ar), 134.8 (Ar), 135.3 (Ar), 135.8 (Ar), 139.4 (C-6), 144.8 (Ar), 150.6 (C-2), 154.1 (Ar), 158.1 (Ar), 158.2 (Ar), 161.8 ($\text{CO}^{\text{Hydroxamate}}$), 162.6 (C-4) ppm.

HRMS (ESI $^+$): m/z calc. 828.3287 $[\text{M}+\text{Na}]^+$, found: 828.3280.

5'-O-(4,4'-Dimethoxytrityl)-3'-(N-Fmoc-N'-acetylhydrazine)ylidene-2'-O-(tert-butylidimethylsilyl)-3'-deoxyuridine (11)



5'-O-(4,4'-Dimethoxytrityl)-3'-ylideneacetic acid-2'-O-(tert-butylidimethylsilyl)-3'-deoxyuridine (**9**) (80 mg, 0.11 mmol) was dissolved in anhydrous DMF (1.35 mL) and stirred

at 0 °C under argon. HOAt (21 mg, 0.15 mmol) was added, followed by EDCI.HCl (27 mg, 0.14 mmol). After stirring for ten minutes, Fmoc-hydrazine (32 mg, 0.13 mmol) was added. The reaction mixture was allowed to warm to room temperature and stirred under argon for a further 20 hours. TLC analysis after this time (petroleum ether-EtOAc 1:1) showed complete consumption of starting material ($R_f = 0.1$) and formation of product ($R_f = 0.4$). The reaction mixture was diluted with CH_2Cl_2 (10 mL), and water (10 mL). The aqueous layer was separated and extracted with CH_2Cl_2 (5 mL). The combined organic layers were washed with brine (10 mL), dried over MgSO_4 , and filtered. The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (petroleum ether-EtOAc 1:1) to obtain the product **11** as a white foam (69 mg, 65%).

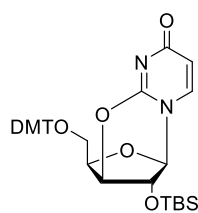
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3232, 3041, 2930, 2855, 1685, 1508, 1453, 1248, 1175, 1032, 835, 740.

^1H NMR (400 MHz, DMSO-d_6): $\delta = 0.01$ (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.21 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.91 (s, 9H, $\text{tBu}^{\text{TBDMS}}$), 3.03 (d, $J_{5'a,5'b} = 8.7$ Hz, 1H, H-5'a), 3.71 (s, 3H, OCH_3), 3.73 (s, 3H, OCH_3), 3.87 (d, $J_{5'a,5'b} = 8.7$ Hz, 1H, H-5'b), 4.28 (t, $J_{\text{CHFmoc},\text{CH}_2\text{Fmoc}} = 6.1$ Hz, 1H, CH^{Fmoc}), 4.42 (d, $J_{\text{CHFmoc},\text{CH}_2\text{Fmoc}} = 6.1$ Hz, 2H, $\text{CH}_2^{\text{Fmoc}}$), 5.04 (d, $J_{1',2'} = 7.0$ Hz, 1H, H-2'), 5.36 (d, $J_{5,6} = 7.7$ Hz, 1H, H-5), 5.43 (s, 1H, H-4'), 5.79 (d, $J_{1',2'} = 7.0$ Hz, 1H, H-1'), 6.09 (s, 1H, CH^{Vinyl}), 6.84-6.91 (m, 4H, Ar), 7.16-7.43 (m, 13H, Ar), 7.68 (d, $J_{5,6} = 7.7$ Hz, H-6), 7.74 (d, $J = 7.2$ Hz, 2H, Ar), 7.90 (d, $J = 7.2$ Hz, 2H, Ar), 9.44 (s, 1H, NH^{Fmoc}), 10.38 (s, 1H, $\text{NH}^{\text{Hydrazide}}$), 11.51 (s, 1H, NH^{U}) ppm.

^{13}C NMR (100 MHz, DMSO-d_6): $\delta = -5.0$ ($\text{CH}_3^{\text{TBDMS}}$), -4.4 ($\text{CH}_3^{\text{TBDMS}}$), 18.1 (qC, $\text{tBu}^{\text{TBDMS}}$), 25.9 ($\text{tBu}^{\text{TBDMS}}$), 47.0 (CH^{Fmoc}), 55.4 (OCH_3), 55.5 (OCH_3), 65.2 (C-5'), 66.6 ($\text{CH}_2^{\text{Fmoc}}$), 76.3 (C-2'), 78.4 (C-4'), 85.3 (C-1'), 86.5 (qC $^{\text{DMT}}$), 103.0 (C-5), 113.8 (Ar), 113.9 (Ar), 114.4 (CH^{Vinyl}), 120.6 (Ar), 125.7 (Ar), 127.3 (Ar), 127.6 (Ar), 127.9 (Ar), 128.2 (Ar), 128.4 (Ar), 129.8 (Ar), 130.3 (Ar), 139.8 (C-6), 141.2 (Ar), 144.1 (Ar), 145.3 (Ar), 151.1 (C-2), 155.4 (C-3'), 156.6 (CO^{Fmoc}), 158.5 (Ar), 158.7 (Ar), 163.0 (C-4), 163.9 ($\text{CO}^{\text{Hydrazide}}$) ppm.

HRMS (ESI $^+$): m/z calc. 975.3403 $[\text{M}+\text{K}]^+$, found: 975.3438.

5'-O-(4,4'-Dimethoxytrityl)-2'-O-(tert-butylidimethylsilyl)-2,3'-anhydrouridine³⁶⁶ (12**)**



5'-O-(4,4'-Dimethoxytrityl)-2'-O-(tert-butylidimethylsilyl)uridine (**4**) (15 mg, 0.23 mmol) was dissolved in anhydrous DMF (0.7 mL). Triphenylphosphine (12 mg, 0.46 mmol) was added, followed by DIAD (90 μ L, 0.46 mmol). The reaction mixture was stirred under argon at room temperature for 23 hours, and then at 100 °C for 2 hours. TLC analysis (toluene-acetone 1:1) after this time showed consumption of starting material ($R_f = 0.6$) and formation of product ($R_f = 0.2$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (toluene-acetone 1:1) to provide the product **12** as a white foam (62 mg, 42%).

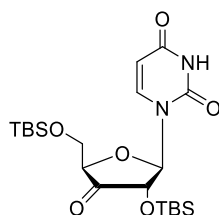
$\nu_{\max}/\text{cm}^{-1}$ (neat) 2930, 2860, 1695, 1508, 1464, 1249, 1071, 1033, 835, 778.

^1H NMR (400 MHz, DMSO- d_6): $\delta = -0.02$ (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.07 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.81 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 2.85 (dd, $J_{5'a,5'b} = 10.6$ Hz, $J_{4',5'a} = 5.8$ Hz, 1H, H-5'a), 3.05 (dd, $J_{5'a,5'b} = 10.6$ Hz, $J_{4',5'b} = 4.5$ Hz, 1H, H-5'b), 3.72 (s, 3H, OCH₃), 3.73 (s, 3H, OCH₃), 4.12 (m, 1H, H-4'), 4.49 (dd, $J_{3',4'} = 4.2$ Hz, $J_{2',3'} = 2.0$ Hz, 1H, H-3'), 5.25 (dd, $J_{1',2'} = 5.9$ Hz, $J_{2',3'} = 2.0$ Hz, 1H, H-2'), 5.88 (d, $J_{5,6} = 7.5$ Hz, 1H, H-5), 6.34 (d, $J_{1',2'} = 5.9$ Hz, 1H, H-1'), 6.82-6.86 (m, 4H, Ar), 7.13-7.30 (m, 9H, Ar), 7.93 (d, $J_{5,6} = 7.5$ Hz, 1H, H-6) ppm.

^{13}C NMR (100 MHz, DMSO- d_6): $\delta = -5.2$ ($\text{CH}_3^{\text{TBDMS}}$), -5.0 ($\text{CH}_3^{\text{TBDMS}}$), 17.6 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.5 ($t\text{Bu}^{\text{TBDMS}}$), 55.0 (OCH₃), 61.9 (C-5'), 76.4 (C-3'), 85.55 (qC^{DMT}), 85.64 (C-4'), 88.4 (C-2'), 89.1 (C-1'), 108.9 (C-5), 113.24 (Ar), 113.25 (Ar), 126.7 (Ar), 127.4 (Ar), 127.9 (Ar), 129.4 (Ar), 129.5 (Ar), 135.1 (Ar), 135.2 (Ar), 136.7 (C-6), 144.5 (Ar), 158.1 (Ar), 159.2 (C-2), 170.9 (C-4) ppm.

HRMS (MALDI-TOF): m/z calc. 665.2659 [M+Na]⁺, found: 665.2663.

2',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-3'-oxouridine³⁶⁷ (14**)**



14

2',5'-Bis-*O*-(*tert*-butyldimethylsilyl)uridine (**6**) (2.16 g, 4.56 mmol) was dissolved in anhydrous CH₂Cl₂ (14 mL). A solution of Dess-Martin periodinane (3.10 g, 7.31 mmol) in anhydrous CH₂Cl₂ (50 mL) was added to the reaction mixture at 0 °C. The reaction mixture was stirred under argon at 0 °C for 15 minutes, and then at room temperature for a further 20 hours. TLC analysis (petroleum ether-EtOAc, 1:1) after this time showed complete consumption of starting material (*R*_f = 0.3) and formation of product (*R*_f = 0.5). The reaction mixture was diluted with EtOAc (130 mL) and washed with an ice cold aqueous solution of 10% (w/v) sodium thiosulfate (78 mL), followed by an ice cold saturated aqueous NaHCO₃ solution (130 mL). The organic layer was dried over MgSO₄ and filtered. The solvent was removed under reduced pressure to provide the product **14** as a white foam (2.14 g, 99%).

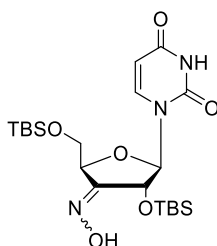
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3109, 2954, 2930, 2858, 1766, 1686 1463, 1410, 1381, 1263, 1083, 897, 834, 777, 667.

¹H NMR (600 MHz, DMSO-*d*₆): δ = -0.030 (s, 3H, CH₃^{TBDMS}), 0.031 (s, 3H, CH₃^{TBDMS}), 0.033 (s, 3H, CH₃^{TBDMS}), 0.06 (s, 3H, CH₃^{TBDMS}), 0.82 (s, 9H, *t*Bu^{TBDMS}), 0.85 (s, 9H, *t*Bu^{TBDMS}), 3.83 (d, *J*_{4',5'} = 2.8 Hz, 2H, H-5'), 4.26 (d, *J*_{1',2'} = 8.0 Hz, 1H, H-2'), 4.42 (t, *J*_{4',5'} = 2.8 Hz, 1H, H-4'), 5.84 (dd, *J*_{5,6} = 8.1 Hz, *J*_{NH,5} = 1.5 Hz, 1H, H-5), 6.21 (d, *J*_{1',2'} = 8.0 Hz, 1H, H-1'), 7.77 (d, *J*_{5,6} = 8.1 Hz, 1H, H-6), 11.56 (br s, 1H, NH) ppm.

¹³C NMR (151 MHz, DMSO-*d*₆): δ = -5.8 (CH₃^{TBDMS}), -5.6 (CH₃^{TBDMS}), -5.5 (CH₃^{TBDMS}), -4.9 (CH₃^{TBDMS}), 17.85 (qC, *t*Bu^{TBDMS}), 17.88 (qC, *t*Bu^{TBDMS}), 25.2 (*t*Bu^{TBDMS}), 25.6 (*t*Bu^{TBDMS}), 62.6 (C-5'), 76.4 (C-2'), 81.7 (C-4'), 84.0 (C-1'), 103.2 (C-5), 139.1 (C-6), 150.5 (C-2), 162.6 (C-4), 208.4 (C-3') ppm.

HRMS (APCI⁺): *m/z* calc. 471.2341 [M+H]⁺, found: 471.2342.

2',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine³⁴⁰ (**17/18**)



17, 18

2',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-3'-oxouridine (**14**) (8.15 g, 17.3 mmol) was dissolved in anhydrous pyridine (160 mL). Hydroxylamine hydrochloride (6.22 g, 89.6 mmol) was added. The reaction mixture was stirred at room temperature under argon for 17 hours. After this time TLC analysis (petroleum ether-EtOAc, 1:1, visualised under UV irradiation) showed consumption of starting material ($R_f = 0.7$, turns brown upon heating) and formation of product ($R_f = 0.7$, does not change colour upon heating). The solvent was removed under reduced pressure and the residue was co-evaporated with toluene (150 mL), and then redissolved in EtOAc (300 mL) and water (300 mL). The organic layer was separated, dried over $MgSO_4$, and filtered. The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (petroleum ether-EtOAc, 3:1) to obtain the product as a mixture of isomers **17** and **18** as a white foam (7.26 g, 86%). A side product, 3',5'-bis-*O*-(*tert*-butyldimethylsilyl)-2'-hydroxyimino-3'-deoxyuridine (**15/16**) was also obtained in a trace amount as a mixture of E/Z isomers as a white foam. The dominant isomer was isolated for characterisation.

2',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine³⁴⁰ (**17/18**)

ν_{max}/cm^{-1} (neat) 3234, 2930, 2858, 1658, 1461, 1382, 1276, 1252, 1110, 1071, 952, 895, 834, 778, 718, 672.

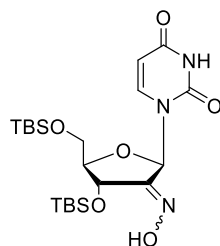
1H NMR (600 MHz, DMSO- d_6): Isomer A: $\delta = -0.04-0.08$ (m, 12H, CH_3^{TBDMS}), 0.81-0.84 (m, 18H, tBu^{TBDMS}), 3.77-3.82 (m, 1H, H-5'a), 3.91 (dd, $J_{5'a,5'b} = 11.6$ Hz, $J_{4',5'a} = 2.1$ Hz, H-5'b), 4.68-4.72 (m, 1H, H-2'), 4.92-4.93 (m, 1H, H-4'), 5.63 (d, $J_{5,6} = 8.0$ Hz, H-5), 5.77 (d, $J_{1',2'} = 4.1$ Hz, 1H, H-1'), 7.74 (d, $J_{5,6} = 8.0$ Hz, 1H, H-6), 11.34 (s, 1H, NH), 11.48 (br s, 1H, N-OH) ppm. Isomer B: $\delta = -0.04-0.08$ (m, 12H, CH_3^{TBDMS}), 0.81-0.84 (m, 18H, tBu^{TBDMS}), 3.84 (dd, $J_{5'a,5'b} = 11.3$ Hz, $J_{4',5'a} = 1.6$ Hz, 1H, H-5'a), 4.02 (dd, $J_{5'a,5'b} = 11.3$ Hz, $J_{4',5'b} = 2.2$ Hz, 1H, H-5'b), 4.59

(dd, $J_{1',2'} = 7.4$ Hz, $J_{2',4'} = 1.6$ Hz, 1H, H-2'), 4.94-4.95 (m, 1H, H-4'), 5.77 (d, $J_{5,6} = 8.0$ Hz, 1H, H-5), 5.91 (d, $J_{1',2'} = 7.4$ Hz, 1H, H-1'), 7.78 (d, $J_{5,6} = 8.0$ Hz, 1H, H-6), 11.47 (s, 1H, NH), 11.48 (br s, 1H, N-OH) ppm .

^{13}C NMR (100 MHz, DMSO- d_6): $\delta = -5.6$ ($\text{CH}_3^{\text{TBDMS}}$), -5.42 ($\text{CH}_3^{\text{TBDMS}}$), -5.41 ($\text{CH}_3^{\text{TBDMS}}$), -5.3 ($\text{CH}_3^{\text{TBDMS}}$), -5.2 ($\text{CH}_3^{\text{TBDMS}}$), -4.9 ($\text{CH}_3^{\text{TBDMS}}$), -4.7 ($\text{CH}_3^{\text{TBDMS}}$), 18.0 (qC, $t\text{Bu}^{\text{TBDMS}}$), 18.1 (qC, $t\text{Bu}^{\text{TBDMS}}$), 18.3 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.4 ($t\text{Bu}^{\text{TBDMS}}$), 25.6 ($t\text{Bu}^{\text{TBDMS}}$), 25.8 ($t\text{Bu}^{\text{TBDMS}}$), 25.9 ($t\text{Bu}^{\text{TBDMS}}$), 62.1 (C-5' isomer B), 64.5 (C-5' isomer A), 70.6 (C-4' isomer A), 74.1 (C-2' isomer B), 77.3 (C-4' isomer B), 79.1 (C-2' isomer A), 86.0 (C-1' isomer B), 91.2 (C-1' isomer A), 102.4 (C-5 isomer A), 103.1 (C-5 isomer B), 139.4 (C-6 isomer B), 141.6 (C-6 isomer A), 150.6 (C-2 isomer A), 150.7 (C-2 isomer B), 154.4 (C-3' isomer A), 155.5 (C-3' isomer B), 162.9 (C-4 isomer B), 163.2 (C-2 isomer A) ppm.

HRMS (APCI⁺): m/z calc. 486.2450 $[\text{M}+\text{H}]^+$, found: 486.2446.

3',5'-bis-O-(*tert*-butyldimethylsilyl)-2'-hydroxyimino-3'-deoxyuridine (15/16)



15, 16

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3228, 2953, 2929, 2857, 1686, 1461, 1388, 1253, 1091, 982, 954, 901, 832, 777, 671, 630.

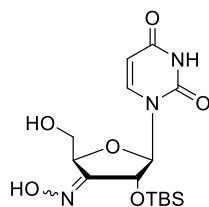
^1H NMR (600 MHz, DMSO- d_6): 0.06 (s, 6H, 2 x $\text{CH}_3^{\text{TBDMS}}$), 0.1 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.14 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.86 (s, 18H, $t\text{Bu}^{\text{TBDMS}}$), 3.80 (d, $J_{4',5'} = 3.4$ Hz, 2H, H-5'a, H-5'b), 4.01 (dd, $J_{3',4'} = 6.6$ Hz, $J_{4',5'} = 3.4$ Hz, 1H, H-4'), 4.89-4.91 (m, 1H, H-3'), 5.62 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 6.54 (s, 1H, H-1'), 7.54 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.47 (br s, 1H, NH^{U}), 11.80 (s, 1H, N-OH) ppm .

^{13}C NMR (151 MHz, DMSO- d_6): $\delta = -5.61$ ($\text{CH}_3^{\text{TBDMS}}$), -5.60 ($\text{CH}_3^{\text{TBDMS}}$), -5.2 ($\text{CH}_3^{\text{TBDMS}}$), -4.9 ($\text{CH}_3^{\text{TBDMS}}$), 17.9 (qC, $t\text{Bu}^{\text{TBDMS}}$), 18.0 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.6 ($t\text{Bu}^{\text{TBDMS}}$), 25.7 ($t\text{Bu}^{\text{TBDMS}}$), 63.2 (C-

5'), 65.8 (C-3'), 81.4 (C-1'), 85.9 (C-4'), 102.2 (C-5), 141.7 (C-6), 150.5 (C-2), 154.9 (C-2'), 162.8 (C-4) ppm.

HRMS (APCI⁺): *m/z* calc. 486.2450 [M+H]⁺, found: 486.2455.

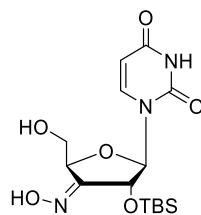
2'-*O*-(*tert*-Butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine³⁴⁰ (19**, **20**)**



19, 20

2',5'-Bis-*O*-(*tert*-butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine (**17/18**) (6.19 g, 12.7 mmol) was dissolved in a mixture of THF (93 mL) and water (68 mL). Trifluoroacetic acid (12 mL) was added, and the reaction mixture was stirred at 0 °C for 4.5 hours. TLC analysis after this time (CH₂Cl₂-MeOH, 90:10) showed consumption of starting material (*R_f* = 0.6) and formation of product isomers **19** (*R_f* = 0.46) and **20** (*R_f* = 0.38). The reaction was then quenched by addition of saturated aqueous NaHCO₃ solution (300 mL). The reaction mixture was extracted with EtOAc (300 mL), and the organic layer was then washed with saturated aqueous NaHCO₃ solution (3 x 250 mL), followed by water (200 mL) and brine (200 mL). The organic layer was dried over MgSO₄ and filtered. The solvent was removed under reduced pressure. The residue was purified by flash chromatography (CH₂Cl₂-MeOH, 96:4) to obtain the product as a mixture of isomers **19/20** as a white foam (3.41 g, 72%). A small amount of each isomer was isolated for separate characterisation.

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine (19**)**



19

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3296, 2929, 2867, 1677, 1460, 1392, 1249, 1149, 1086, 958, 836, 779, 673.

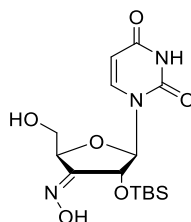
^1H NMR (600 MHz, DMSO- d_6): δ = -0.03 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.04 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.81 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.67 (ddd, $J_{5'a,5'b}$ = 11.7 Hz, $J_{5'a,\text{OH}-5'}$ = 5.4 Hz, $J_{4',5'a}$ = 2.2 Hz, 1H, H-5'a), 3.83 (ddd, $J_{5'a,5'b}$ = 11.7 Hz, $J_{5'b,\text{OH}-5'}$ = 5.4 Hz, $J_{4',5'b}$ = 3.0 Hz, 1H, H-5'b), 4.70 (dd, $J_{1',2'}$ = 7.5 Hz, $J_{2',4'}$ = 1.7 Hz, 1H, H-2'), 4.86 (m, 1H, H-4'), 5.18 (app t, J = 5.4 Hz, 1H, OH-5'), 5.80 (dd, $J_{5,6}$ = 8.1 Hz, $J_{5,\text{NH}}$ = 1.8 Hz, 1H, H-5), 5.85 (d, $J_{1',2'}$ = 7.5 Hz, 1H, H-1'), 7.94 (d, $J_{5,6}$ = 8.1 Hz, H-6), 11.35 (s, 1H, N-OH), 11.47 (d, $J_{5,\text{NH}}$ = 1.8 Hz, NH^{U}) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = -5.5 ($\text{CH}_3^{\text{TBDMS}}$), -4.8 ($\text{CH}_3^{\text{TBDMS}}$), 17.8 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.4 ($t\text{Bu}^{\text{TBDMS}}$), 59.8 (C-5'), 73.3 (C-2'), 77.6 (C-4'), 85.9 (C-1'), 103.0 (C-5), 140.2 (C-6), 150.7 (C-2), 155.9 (C-3'), 162.8 (C-4) ppm.

HRMS (APCI): m/z calc. 370.1439 $[\text{M}+\text{H}]^+$, found: 370.1448.

Note: ROESY NMR showed correlations between the N-OH proton and the H-4' and H-5'b protons, allowing the stereochemistry of the oxime moiety of **19** to be assigned.

2'-O-(*tert*-Butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine (**20**)



20

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3237, 2930, 2858, 1681, 1462, 1382, 1275, 1252, 1109, 950, 894, 835, 813, 779.

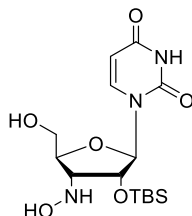
^1H NMR (400 MHz, DMSO- d_6): δ = 0.04 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.07 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.81 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.57-3.63 (m, 1H, H-5'a), 3.73 (ddd, $J_{5'a,5'b}$ = 12.1 Hz, $J_{5'b,\text{OH}-5'}$ = 5.4 Hz, $J_{5'b,4'}$ = 2.4 Hz, 1H, H-5'b), 4.64-4.65 (m, 1H, H-4'), 4.95-4.98 (m, 2H, H-2', OH-5'), 5.67 (d, $J_{5,6}$ = 8.0 Hz, 1H, H-5), 5.77 (d, $J_{1',2'}$ = 4.3 Hz, 1H, H-1'), 7.82 (d, $J_{5,6}$ = 8.0 Hz, H-6), 11.30 (s, 1H, N-OH), 11.44 (br s, 1H, NH) ppm.

^{13}C NMR (100 MHz, DMSO- d_6): δ = -5.3 ($\text{CH}_3^{\text{TBDMS}}$), -5.1 ($\text{CH}_3^{\text{TBDMS}}$), 17.8 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.5 ($t\text{Bu}^{\text{TBDMS}}$), 62.0 (C-5'), 70.3 (C-2'), 79.2 (C-4'), 90.6 (C-1'), 102.2 (C-5), 141.6 (C-6), 150.5 (C-2), 154.5 (C-3'), 163.0 (C-4) ppm.

HRMS (APCI): m/z calc. 370.1439 $[M+H]^+$, found: 370.1437.

Note: ROESY NMR showed a correlation between the N-OH proton and the H-2' proton, allowing the stereochemistry of the oxime moiety of **20** to be assigned.

2'-O-(*tert*-Butyldimethylsilyl)-3'-hydroxyamino-3'-deoxyuridine³⁴⁰ (22**)**



22

2'-O-(*tert*-Butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine (**19/20**) (461 mg, 1.24 mmol) was dissolved in AcOH (16 mL). NaBH₄ (196 mg, 5.19 mmol) was added. The reaction mixture was stirred at room temperature for 2.5 hours. TLC analysis (CH₂Cl₂-MeOH, 93:7) after this time showed consumption of starting material ($R_f = 0.4$) and formation of product ($R_f = 0.3$). The reaction mixture was then diluted with EtOAc (50 mL) and washed with saturated aqueous NaHCO₃ solution (3 x 50 mL), water (50 mL), and brine (50 mL). The organic layer was dried over MgSO₄ and filtered. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 94:6) to provide the product **22** as a white foam (391 mg, 84%). A side product, 2'-O-(*tert*-butyldimethylsilyl)-3'-(*N*-hydroxy-*N*-ethyl)amino-3'-deoxyuridine (**21**), was obtained in a trace amount as a white foam.

2'-O-(*tert*-Butyldimethylsilyl)-3'-hydroxyamino-3'-deoxyuridine (**22**)

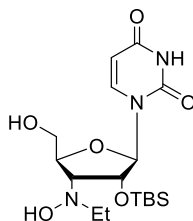
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3363, 2934, 2859, 1676, 1461, 1386, 1255, 1104, 1052, 842, 779.

¹H NMR (600 MHz, DMSO-*d*₆): $\delta = 0.03$ (s, 3H, CH₃^{TBDMS}), 0.06 (s, 3H, CH₃^{TBDMS}), 0.85 (s, 9H, *t*Bu^{TBDMS}), 3.38 (app t, $J = 5.1$ Hz, 1H, H-4'), 3.53-3.55 (m, 1H, H-5'a), 3.69-3.72 (m, 1H, H-5'b), 3.98 (m, 1H, H-3'), 4.32 (app t, $J = 4.9$ Hz, 1H, H-2'), 5.23 (app t, $J = 4.6$ Hz, 1H, NH-3'), 5.41 (br s, 1H, N-OH), 5.67 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.82 (d, $J_{1',2'} = 4.9$ Hz, 1H, H-1'), 7.58 (s, 1H, OH-5'), 8.02 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.33 (s, 1H, NH^U) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = -5.4 ($\text{CH}_3^{\text{TBDMS}}$), -5.2 ($\text{CH}_3^{\text{TBDMS}}$), 17.7 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.6 ($t\text{Bu}^{\text{TBDMS}}$), 61.5 (C-5'), 62.7 (C-4'), 74.5 (C-2'), 81.9 (C-3'), 88.4 (C-1'), 101.8 (C-5), 140.3 (C-6), 150.6 (C-2), 163.1 (C-4) ppm.

HRMS (APCI $^-$): m/z calc. 372.1596 [M-H] $^-$, found: 372.1596.

2'-O-(*tert*-Butyldimethylsilyl)-3'-(*N*-hydroxy-*N*-ethyl)amino-3'-deoxyuridine (21)



21

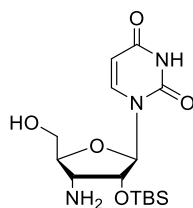
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3442, 3055, 2932, 2858, 1674, 1464, 1387, 1255, 1146, 1112, 1082, 836, 778, 760.

^1H NMR (600 MHz, DMSO- d_6): δ = 0.01 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.05 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.83 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 1.02 (t, $J_{\text{CH}_2, \text{CH}_3}$ = 6.9 Hz, 3H, CH_3^{Et}), 2.62 (dq, $J_{\text{CH}_2\text{Et}_a, \text{CH}_2\text{Et}_b}$ = 13.7 Hz, $J_{\text{CH}_2, \text{CH}_3}$ = 6.9 Hz, 1H, $\text{CH}_2^{\text{Et-a}}$), 2.82 (dq, $J_{\text{CH}_2\text{Et}_a, \text{CH}_2\text{Et}_b}$ = 13.7 Hz, $J_{\text{CH}_2, \text{CH}_3}$ = 6.9 Hz, 1H, $\text{CH}_2^{\text{Et-b}}$), 3.15 (dd, $J_{2', 3'}$ = 6.2 Hz, $J_{3', 4'}$ = 3.9 Hz, 1H, H-3'), 3.57 (d, $J_{5'a, 5'b}$ = 10.9 Hz, 1H, H-5'a), 3.68 (d, $J_{5'a, 5'b}$ = 10.9 Hz, 1H, H-5'b), 4.25 (app t, J = 6.2 Hz, 1H, H-2'), 4.38-4.41 (m, 1H, H-4'), 5.11 (br s, 1H, OH-5'), 5.67 (d, $J_{5, 6}$ = 8.1 Hz, 1H, H-5), 5.93 (d, $J_{1', 2'}$ = 6.2 Hz, 1H, H-1'), 7.72 (s, 1H, N-OH), 7.97 (d, $J_{5, 6}$ = 8.1 Hz, 1H, H-6), 11.29 (s, 1H, NH^{U}) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = -5.1 ($\text{CH}_3^{\text{TBDMS}}$), -5.0 ($\text{CH}_3^{\text{TBDMS}}$), 12.9 (CH_3^{Et}), 17.9 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.6 ($t\text{Bu}^{\text{TBDMS}}$), 51.6 (CH_2^{Et}), 62.7 (C-5'), 66.0 (C-3'), 75.7 (C-2'), 80.1 (C-4'), 87.8 (C-1'), 101.8 (C-5), 140.3 (C-6), 150.8 (C-2), 163.1 (C-4) ppm.

HRMS (APCI $^+$): m/z calc. 402.2055 [M+H] $^+$, found: 402.2050.

2'-O-(*tert*-Butyldimethylsilyl)-3'-amino-3'-deoxyuridine³⁴⁰ (23**)**



23

2'-O-(*tert*-Butyldimethylsilyl)-3'-hydroxyamino-3'-deoxyuridine (**22**) (360 mg, 0.74 mmol) was dissolved in a mixture of AcOH (8.9 mL) and water (1 mL). The solution was degassed by bubbling with argon. 10% Pd/C (75 mg) was added. The reaction vessel was purged with H₂, and the reaction mixture was stirred under an atmosphere of H₂ for 6 hours. TLC analysis (CH₂Cl₂-MeOH, 90:10) after this time showed complete consumption of starting material (R_f = 0.4) and formation of product (R_f = 0.3). The reaction mixture was filtered through celite, washing through with MeOH. The solvent was removed under reduced pressure. The residue was co-evaporated with toluene (50 mL) and then purified by flash chromatography (CH₂Cl₂-MeOH-NEt₃, 91:8:1) to afford the product **23** as a white foam (250 mg, 73%).

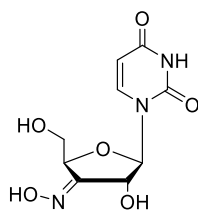
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3371, 2930, 2863, 1685, 1460, 1378, 1254, 1105, 1060, 993, 838, 779, 699.

¹H NMR (600 MHz, DMSO-*d*₆): δ = 0.09 (s, 3H, CH₃^{TBDMS}), 0.11 (s, 3H, CH₃^{TBDMS}), 0.88 (s, 9H, *t*Bu^{TBDMS}), 3.20 (dd, $J_{3',4'} = 7.6$ Hz, $J_{2',3'} = 4.9$ Hz, 1H, H-3'), 3.60-3.67 (m, 2H, H-5'a, H-4'), 3.73-3.76 (m, 1H, H-5'b), 4.06 (dd, $J_{2',3'} = 4.9$ Hz, $J_{1',2'} = 2.1$ Hz, 1H, H-2'), 5.12 (br s, 1H, OH-5'), 5.59 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.66 (d, $J_{1',2'} = 2.1$ Hz, 1H, H-1'), 8.07 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6) ppm.

¹³C NMR (151 MHz, DMSO-*d*₆): δ = -5.1 (CH₃^{TBDMS}), -4.8 (CH₃^{TBDMS}), 17.8 (qC, *t*Bu^{TBDMS}), 25.7 (*t*Bu^{TBDMS}), 51.8 (C-3'), 59.6 (C-5'), 77.1 (C-2'), 85.1 (C-4'), 89.3 (C-1'), 101.1 (C-5), 140.4 (C-6), 150.5 (C-2), 163.2 (C-4) ppm.

HRMS (APCI⁺): m/z calc. 356.1647 [M+H]⁺, found: 356.1638.

3'-Hydroxyimino-3'-deoxyuridine (**26**)



26

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine (**19**) (93 mg, 250 μ mol) was dissolved in THF (2.7 mL). Tetrabutylammonium fluoride trihydrate (120 mg, 379 μ mol) was added and the reaction mixture was stirred at room temperature for 2 hours. TLC analysis (EtOAc-MeOH, 88:12) after this time showed complete consumption of starting material ($R_f = 0.9$) and formation of product ($R_f = 0.6$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (EtOAc-MeOH, 97:3) to provide the product **26** as a white foam (60 mg, 92%).

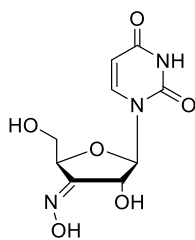
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3243, 2927, 1663, 1462, 1385, 1274, 1249, 1112, 1078, 1051, 1022, 957, 897, 810, 760, 717, 684, 640, 616.

^1H NMR (400 MHz, DMSO- d_6): $\delta = 3.65$ (ddd, $J_{4',5'a} = 2.2$ Hz, $J_{5'a,\text{OH}-5'} = 5.3$ Hz, $J_{5'a,5'b} = 11.7$ Hz, 1H, H-5'a), 3.83 (ddd, $J_{4',5'b} = 3.6$ Hz, $J_{5'b,\text{OH}-5'} = 5.3$ Hz, $J_{5'a,5'b} = 11.7$ Hz, 1H, H-5'b), 4.54-5.59 (m, 1H, H-2'), 4.80 (dd, $J_{4',5'a} = 2.2$ Hz, $J_{4',5'b} = 3.6$ Hz, 1H, H-4'), 5.13 (app t, $J = 5.3$ Hz, 1H, OH-5'), 5.74 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.80 (d, $J_{1',2'} = 7.7$ Hz, 1H, H-1'), 5.96 (d, $J_{2',\text{OH}-2'} = 6.6$ Hz, 1H, OH-2'), 7.91 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.21 (s, 1H, N-OH), 11.41 (br s, 1H, NH^{U}) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): $\delta = 59.9$ (C-5'), 71.9 (C-2'), 77.5 (C-4'), 86.3 (C-1'), 102.6 (C-5), 140.6 (C-6), 150.8 (C-2), 156.6 (C-3'), 162.9 (C-4) ppm.

HRMS (APCI $^+$): m/z calc. 256.0575 [$\text{M}-\text{H}$] $^-$, found: 256.0579.

3'-Hydroxyimino-3'-deoxyuridine (**27**)



27

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-hydroxyimino-3'-deoxyuridine (**20**) (73 mg, 197 μ mol) was dissolved in THF (2.1 mL). Tetrabutylammonium fluoride trihydrate (96 mg, 303 μ mol) was added and the reaction mixture was stirred at room temperature for 3 hours. TLC analysis (EtOAc-MeOH, 98:2) after this time showed complete consumption of starting material ($R_f = 0.8$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (EtOAc-MeOH, 98:2) to provide the product **27** as a white foam (46 mg, 91%).

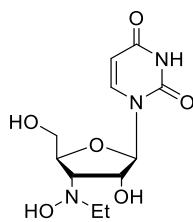
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3227, 2927, 2884, 1652, 1463, 1386, 1271, 1045, 964, 811, 762.

^1H NMR (400 MHz, acetone- d_6): $\delta = 3.80$ (ddd, $J_{4',5'a} = 4.4$ Hz, $J_{5'a,\text{OH}-5'} = 5.9$ Hz, $J_{5'a,5'b} = 12.1$ Hz, 1H, H-5'a), 3.88 (ddd, $J_{4',5'b} = 2.7$ Hz, $J_{5'b,\text{OH}-5'} = 5.9$ Hz, $J_{5'a,5'b} = 12.1$ Hz, 1H, H-5'b), 4.13 (app t, $J = 5.9$ Hz, 1H, OH-5'), 4.72 (ddd, $J_{2',4'} = 1.7$ Hz, $J_{4',5'b} = 2.7$ Hz, $J_{4',5'a} = 4.4$ Hz, 1H, H-4'), 5.01 (d, $J_{2',\text{OH}-2'} = 4.3$ Hz, 1H, OH-2'), 5.12 (app td, $J = 1.7$ Hz, $J = 4.3$ Hz, 1H, H-2'), 5.66 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.97 (d, $J_{1',2'} = 4.3$ Hz, 1H, H-1'), 7.79 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 10.17 (br s, 1H, NH^{H}), 10.58 (s, 1H, N-OH) ppm.

^{13}C NMR (151 MHz, acetone- d_6): $\delta = 63.5$ (C-5'), 71.5 (C-2'), 80.3 (C-4'), 91.9 (C-1'), 103.1 (C-5), 142.3 (C-6), 151.5 (C-2), 158.2 (C-3'), 163.6 (C-4) ppm.

HRMS (APCI $^+$): m/z calc. 258.0721 $[\text{M}-\text{H}]^-$, found: 258.0726.

3'-*N*-(Ethyl)hydroxyamino-3'-deoxyuridine (**28**)



28

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-*N*-(ethyl)hydroxyamino-3'-deoxyuridine (**21**) (107 mg, 267 μ mol) was dissolved in THF (2.5 mL). Tetrabutylammonium fluoride trihydrate (125 mg, 395 μ mol) was added and the reaction mixture was stirred at room temperature for 1.5 hours. TLC analysis (CH₂Cl₂-MeOH, 96:4) after this time showed complete consumption of starting material (R_f = 0.3) and formation of product (R_f = 0.0). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 92:8) to obtain the product **28** as a white amorphous solid (41 mg, 54%).

Note: compound **28** exhibits rotamers in NMR spectroscopy.

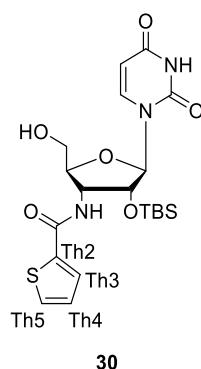
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3203, 3051, 2522, 2160, 2023, 1672, 1596, 1468, 1425, 1386, 1346, 1270, 1252, 994, 864, 823, 740, 695.

¹H NMR (600 MHz, DMSO-*d*₆): δ = 1.00-1.06 (m, 3H, CH₃^{Et}), 2.68 (dq, $J_{\text{CH}_2\text{Et}_a, \text{CH}_2\text{Et}_b}$ = 13.7 Hz, $J_{\text{CH}_2, \text{CH}_3}$ = 6.9 Hz, 1H, CH₂^{Et-a}), 2.79 (dq, $J_{\text{CH}_2\text{Et}_a, \text{CH}_2\text{Et}_b}$ = 13.7 Hz, $J_{\text{CH}_2, \text{CH}_3}$ = 6.9 Hz, 1H, CH₂^{Et-b}), 3.20 (dd, $J_{2',3'}$ = 7.0 Hz, $J_{3',4'}$ = 4.2 Hz, 0.8H, H-3'), 3.44 (dd, $J_{2',3'}$ = 6.9 Hz, $J_{3',4'}$ = 3.8 Hz, 0.2H, H-3'), 3.49-3.60 (m, 1H, H-5'a), 3.66 (ddd, $J_{5'a,5'b}$ = 11.5 Hz, $J_{5'b, \text{OH}-5'}$ = 5.0 Hz, $J_{4',5'b}$ = 3.0 Hz, 1H, H-5'b), 4.00-4.15 (m, 1H, H-2'), 4.29-4.40 (m, 1H, H-4'), 4.70 (d, $J_{2', \text{OH}-2'}$ = 7.1 Hz, 0.2H, OH-2'), 4.83 (d, $J_{2', \text{OH}-2'}$ = 7.1 Hz, 0.8H, OH-2'), 5.11 (app t, J = 5.0 Hz, 0.8H, OH-5'), 5.15 (app t, J = 5.0 Hz, 0.2H, OH-5'), 5.65 (d, $J_{5,6}$ = 8.0 Hz, 1H, H-5), 5.87 (d, $J_{1',2'}$ = 6.0 Hz, 1H, H-1'), 7.86 (s, 0.2H, N-OH), 7.92 (d, $J_{5,6}$ = 8.0 Hz, 1H, H-6), 7.95 (s, 0.8H, N-OH), 11.30 (s, 1H, NH^U) ppm.

¹³C NMR (151 MHz, DMSO-*d*₆): δ = 12.7 (CH₃^{Et}), 51.6 (CH₂^{Et}), 61.8 (C-3'), 62.3 (C-5'), 62.6 (C-5'), 65.6 (C-3'), 73.8 (C-2'), 79.1 (C-4'), 79.3 (C-4'), 88.3 (C-1'), 88.5 (C-1'), 101.75 (C-5), 101.79 (C-5), 140.5 (C-6), 150.8 (C-2), 163.1 (C-4) ppm.

HRMS (ESI⁺): m/z calc. 310.1010 [M+Na]⁺, found: 310.1007.

***N*-(2'-*O*-(*tert*-Butyldimethylsilyl)-3'-deoxyuridine-3'-yl)thiophene-2-carboxamide (**30**)**



2'-*O*-(*tert*-Butyldimethylsilyl)-3'-amino-3'-deoxyuridine (**23**) (50 mg, 140 μ mol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL). Anhydrous pyridine (13 μ L, 160 μ mol) was added. A solution of 2-thiophenecarbonyl chloride (18 μ L, 17 μ mol) in anhydrous CH_2Cl_2 (0.9 mL) was added dropwise to the reaction mixture over 20 minutes at 0 °C. The reaction mixture was then allowed to warm to room temperature and stirred under argon for 20 hours. TLC analysis (CH_2Cl_2 -MeOH, 90:10) after this time showed consumption of starting material ($R_f = 0.4$) and formation of product ($R_f = 0.6$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (petroleum ether-EtOAc, 1:1 then 1:2). The product **30** was obtained as a white foam (61 mg, 93%).

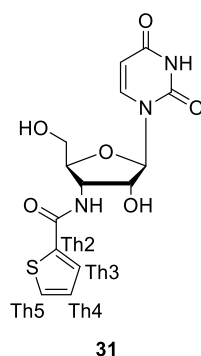
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3441, 3056, 2928, 2856, 1692, 1534, 1462, 1256, 1103, 837, 781, 715.

^1H NMR (600 MHz, DMSO-d_6): $\delta = -0.10$ (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.00 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.75 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.60 (ddd, $J_{5'a,5'b} = 12.2$ Hz, $J_{5'a,\text{OH}-5'} = 4.9$ Hz, $J_{4',5'a} = 3.4$ Hz, 1H, H-5'a), 3.73 (ddd, $J_{5'a,5'b} = 12.2$ Hz, $J_{5'b,\text{OH}-5'} = 4.9$ Hz, $J_{4',5'b} = 2.4$ Hz, 1H, H-5'b), 4.18 (m, 1H, H-4'), 4.31 (dd, $J_{2',3'} = 6.1$ Hz, $J_{1',2'} = 4.1$ Hz, 1H, H-2'), 4.43 (m, 1H, H-3'), 5.26 (app t, $J = 4.9$ Hz, 1H, OH-5'), 5.69 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.87 (d, $J_{1',2'} = 4.1$ Hz, 1H, H-1'), 7.16 (dd, $J_{\text{Th4,Th5}} = 4.8$ Hz, $J_{\text{Th3,Th4}} = 3.6$ Hz, 1H, H-Th4), 7.76 (d, $J_{\text{Th4,Th5}} = 4.8$ Hz, 1H, H-Th5), 7.85 (d, $J_{\text{Th3,Th4}} = 3.6$ Hz, 1H, H-Th3), 8.05 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 8.47 (d, $J_{3',\text{NH}-3'} = 8.0$ Hz, 1H, NH-3'), 11.39 (s, 1H, NH^{U}) ppm.

^{13}C NMR (151 MHz, DMSO-d_6): $\delta = -5.4$ ($\text{CH}_3^{\text{TBDMS}}$), -5.3 ($\text{CH}_3^{\text{TBDMS}}$), 17.6 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.4 ($t\text{Bu}^{\text{TBDMS}}$), 50.3 (C-3'), 60.3 (C-5'), 74.5 (C-2'), 82.1 (C-4'), 88.8 (C-1'), 101.8 (C-5), 127.8 (C-Th4), 128.4 (C-Th3), 131.0 (C-Th5), 139.6 (C-Th2), 140.3 (C-6), 150.6 (C-2), 161.2 (CO^{Amide}) 163.0 (C-4) ppm.

HRMS (ESI⁺): m/z calc. 466.1474 [M-H]⁻, found: 466.14680.

***N*-(3'-Deoxyuridine-3'-yl)thiophene-2-carboxamide (31)**



N-(2'-*O*-(*tert*-Butyldimethylsilyl)-3'-deoxyuridine-3'-yl)thiophene-2-carboxamide (**30**) (44 mg, 94 μmol) was dissolved in THF (1 mL). Tetrabutylammonium fluoride trihydrate (45 mg, 140 μmol) was added and the reaction mixture was stirred at room temperature for 1.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 93:7) after this time showed complete consumption of starting material ($R_f = 0.5$) and formation of product ($R_f = 0.4$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 95:5) to provide the product **31** as a white amorphous solid (30 mg, 90%).

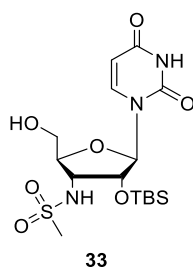
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3240, 3063, 2971, 2930, 1793, 1672, 1573, 1518, 1454, 1441, 1381, 1256, 1082, 994, 878, 813, 736, 696.

^1H NMR (400 MHz, acetone- d_6): $\delta = 3.81$ (ddd, $J_{5'a,5'b} = 12.7$ Hz, $J_{5'a,\text{OH}-5'} = 5.8$ Hz, $J_{4',5'a} = 2.4$ Hz, 1H, H-5'a), 3.96 (ddd, $J_{5'a,5'b} = 12.7$ Hz, $J_{5'b,\text{OH}-5'} = 5.8$ Hz, $J_{4',5'a} = 2.4$ Hz, 1H, H-5'b), 4.20 (app dt, $J = 2.4$ Hz, $J = 8.3$ Hz, 1H, H-4'), 4.39 (app t, $J = 5.8$ Hz, 1H, OH-5'), 4.44-4.49 (m, 1H, H-2'), 4.64 (app td, $J = 8.3$ Hz, $J = 5.2$ Hz, 1H, H-3'), 5.45 (br s, 1H, OH-2'), 5.62 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.87 (d, $J_{1',2'} = 2.0$ Hz, 1H, H-1'), 7.14 (dd, $J_{\text{Th4,Th5}} = 5.0$ Hz, $J_{\text{Th3,Th4}} = 3.8$ Hz, 1H, H-Th4), 7.61 (d, $J_{3',\text{NH}-3'} = 8.3$ Hz, 1H, NH-3'), 7.71 (dd, $J_{\text{Th4,Th5}} = 5.0$ Hz, $J_{\text{Th3,Th5}} = 1.0$ Hz, 1H, H-Th5), 7.83 (dd, $J_{\text{Th3,Th4}} = 3.8$ Hz, $J_{\text{Th3,Th5}} = 1.0$ Hz, 1H, H-Th3), 8.15 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 10.05 (br s, 1H, NH^{U}) ppm.

^{13}C NMR (100 MHz, acetone- d_6): $\delta = 51.1$ (C-3'), 61.2 (C-5'), 75.3 (C-2'), 83.9 (C-4'), 91.8 (C-1'), 102.1 (C-5), 128.6 (C-Th4), 129.4 (C-Th3), 131.7 (C-Th5), 140.1 (C-Th2), 141.1 (C-6), 151.4 (C-2), 162.9 (CO^{Amide}), 163.7 (C-4) ppm.

HRMS (APCI $^+$): m/z calc. 354.0754 $[\text{M}+\text{H}]^+$, found: 354.0745.

2'-O-(*tert*-Butyldimethylsilyl)-3'-N-(methanesulfonyl)amino-3'-deoxyuridine (**33**)



2'-O-(*tert*-Butyldimethylsilyl)-3'-amino-3'-deoxyuridine (**23**) (102 mg, 290 μ mol) was dissolved in anhydrous CH_2Cl_2 (2.5 mL). A solution of mesyl chloride (22 μ L, 290 μ mol) in anhydrous CH_2Cl_2 (1 mL) was added dropwise to the reaction mixture over 10 minutes at 0 $^\circ\text{C}$. The reaction mixture was allowed to warm to room temperature and stirred under argon for 2.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 92:8) after this time showed incomplete consumption of starting material ($R_f = 0.3$) and formation of product ($R_f = 0.5$). Pyridine (20 μ L, 250 μ mol) was added and the reaction mixture was stirred for a further 1.5 hours. Additional mesyl chloride (11 μ L, 140 μ mol) was added at 0 $^\circ\text{C}$. After another hour, another aliquot of mesyl chloride (5 μ L, 70 μ mol) was added at 0 $^\circ\text{C}$. The reaction mixture was stirred at room temperature for 17 hours. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 98:2 – 95:5) to provide the product **33** as a white foam (54 mg, 43%). A side product, 2'-O-(*tert*-butyldimethylsilyl)-3'-N-(methanesulfonyl)amino-5'-O-methanesulfonyl-3'-deoxyuridine **32** was also obtained as a white amorphous solid (11 mg, 8%).

2'-O-(*tert*-Butyldimethylsilyl)-3'-N-(methanesulfonyl)amino-3'-deoxyuridine (**33**)

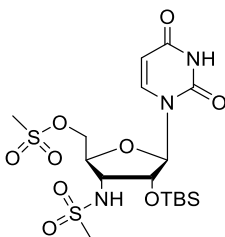
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3240, 2934, 2863, 1685, 1461, 1370, 1330, 1253, 1114, 978, 830, 780.

^1H NMR (600 MHz, DMSO-d_6): δ = 0.096 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.100 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.87 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 2.95 (s, 3H, $\text{CH}_3^{\text{Mesyl}}$), 3.60 (m, 1H, H-5'a), 3.78-3.82 (m, 2H, H-5'b, H-3'), 3.97-3.99 (m, 1H, H-4'), 4.20 (dd, $J_{2',3'} = 5.2$ Hz, $J_{1',2'} = 2.7$ Hz, 1H, H-2'), 5.34 (app t, $J = 4.5$ Hz, 1H, OH-5'), 5.64 (dd, $J_{5,6} = 8.1$ Hz, $J_{5,N-U} = 1.3$ Hz, 1H, H-5), 5.69 (d, $J_{1',2'} = 2.7$ Hz, 1H, H-1'), 7.21 (d, $J_{\text{NH-3',3'}} = 8.2$ Hz, 1H, NH-3'), 8.01 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.36 (br s, 1H, NH^{U}) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = -5.0 ($\text{CH}_3^{\text{TBDMS}}$), -4.8 ($\text{CH}_3^{\text{TBDMS}}$), 17.9 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.7 ($t\text{Bu}^{\text{TBDMS}}$), 40.9 ($\text{CH}_3^{\text{Mesyl}}$), 53.0 (C-3'), 59.0 (C-5'), 75.2 (C-2'), 82.0 (C-4'), 89.2 (C-1'), 101.4 (C-5), 140.1 (C-6), 150.4 (C-2), 163.1 (C-4) ppm.

HRMS (ESI $^+$): m/z calc. 458.1388 $[\text{M}+\text{Na}]^+$, found: 458.1392.

2'-O-(*tert*-Butyldimethylsilyl)-3'-N-(methanesulfonyl)amino-5'-O-methanesulfonyl-3'-deoxyuridine (32)



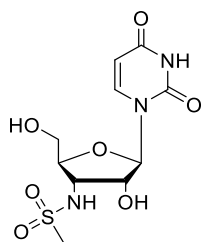
32

^1H NMR (400 MHz, DMSO- d_6): δ = 0.08 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.10 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.87 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 2.99 (s, 3H, $\text{CH}_3^{\text{Mesyl}}$), 3.25 (s, 3H, $\text{CH}_3^{\text{Mesyl}}$), 3.83 (app td, J = 8.3 Hz, J = 6.0 Hz, 1H, H-3'), 4.17 (ddd, $J_{4',5'b}$ = 2.7 Hz, $J_{4',5'a}$ = 5.6 Hz, $J_{3',4'}$ = 8.3 Hz, 1H, H-4'), 4.35 (dd, $J_{2',3'}$ = 6.0 Hz, $J_{1',2'}$ = 3.2 Hz, 1H, H-2'), 4.45 (dd, $J_{4',5'a}$ = 5.6 Hz, $J_{5'a,5'b}$ = 11.5 Hz, 1H, H-5'a), 4.52 (dd, $J_{4',5'b}$ = 2.7 Hz, $J_{5'a,5'b}$ = 11.5 Hz, 1H, H-5'b), 5.68 (dd, $J_{5,\text{NH-U}}$ = 2.0 Hz, $J_{5,6}$ = 8.1 Hz, 1H, H-5), 5.72 (d, $J_{2',3'}$ = 3.2 Hz, 1H, H-1'), 7.36 (d, $J_{3',\text{NH-3'}}$ = 8.3 Hz, 1H, NH-3'), 7.67 (d, $J_{5,6}$ = 8.1 Hz, 1H, H-6), 11.43 (d, $J_{5,\text{NH-U}}$ = 2.0 Hz, 1H, NH^{U}) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = -5.1 ($\text{CH}_3^{\text{TBDMS}}$), -4.9 ($\text{CH}_3^{\text{TBDMS}}$), 17.8 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.7 ($t\text{Bu}^{\text{TBDMS}}$), 36.9 ($\text{CH}_3^{\text{Mesyl}}$), 41.2 ($\text{CH}_3^{\text{Mesyl}}$), 53.6 (C-3'), 68.3 (C-5'), 73.9 (C-2'), 78.7 (C-4'), 90.1 (C-1'), 102.0 (C-5), 140.3 (C-6), 150.4 (C-2), 163.0 (C-4) ppm.

HRMS (ESI $^+$): m/z calc. 536.1163 $[\text{M}+\text{Na}]^+$, found: 536.1153.

3'-*N*-(Methanesulfonyl)amino-3'-deoxyuridine (**34**)



34

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-*N*-(methanesulfonyl)amino-3'-deoxyuridine (**33**) (47 mg, 109 μ mol) was dissolved in THF (1 mL). Tetrabutylammonium fluoride trihydrate (51 mg, 160 μ mol) was added and the reaction mixture was stirred at room temperature for 2 hours. TLC analysis (CH₂Cl₂-MeOH, 93:7) after this time showed complete consumption of starting material (R_f = 0.4) and formation of product (R_f = 0.2). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 92:8) to afford the product **34** as a white crystalline solid (34 mg, 97%); mp 148-162 °C decomp.

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3200, 2931, 1655, 1609, 1546, 1445, 1419, 1328, 1274, 1202, 1177, 1148, 1101, 1062, 981, 950, 829, 807, 759, 709.

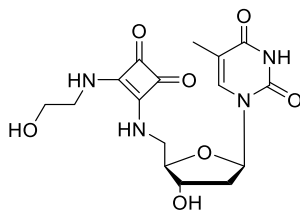
¹H NMR (600 MHz, DMSO-*d*₆): δ = 2.97 (s, 3H, CH₃^{MesyI}), 3.56 (dd, $J_{5'a,5'b}$ = 12.3 Hz, $J_{4',5'a}$ = 2.9 Hz, 1H, H-5'a), 3.77 (dd, $J_{5'a,5'b}$ = 12.3 Hz, $J_{4',5'b}$ = 1.5 Hz, 1H, H-5'b), 3.81 (app td, J = 8.7 Hz, J = 5.5 Hz, 1H, H-3'), 3.86-3.91 (m, 1H, H-4'), 4.14 (dd, $J_{2',3'}$ = 5.5 Hz, $J_{1',2'}$ = 2.1 Hz, 1H, H-2'), 5.62 (dd, $J_{5,6}$ = 8.1 Hz, $J_{5,\text{NH-U}}$ = 1.8 Hz, 1H, H-5), 5.70 (d, $J_{1',2'}$ = 2.1 Hz, 1H, H-1'), 6.05 (br s, 1H, OH-2'), 7.17 (d, $J_{3',\text{NH-3}'}$ = 8.7 Hz, 1H, NH-3'), 7.99 (d, $J_{5,6}$ = 8.1 Hz, 1H, H-6), 11.32 (br s, 1H, NH^U) ppm.

¹³C NMR (151 MHz, DMSO-*d*₆): δ = 41.1 (CH₃^{MesyI}), 53.0 (C-3'), 59.2 (C-5'), 72.7 (C-2'), 82.3 (C-4'), 90.1 (C-1'), 101.2 (C-5), 140.5 (C-6), 150.4 (C-2), 163.2 (C-4) ppm.

HRMS (APCI⁻): m/z calc. 320.0558 [M-H]⁻, found: 320.0560.

6.2.2. Synthetic methods for Chapter 3

5'-*N*-(2-(2-Hydroxyethyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-5'-deoxythymidine (38)



38

5'-*N*-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-5'-deoxythymidine (**37**) (36 mg, 99 μ mol) was suspended in EtOH (1 mL). Ethanolamine (9 μ L, 149 μ mol) was added and the reaction mixture was stirred at room temperature for 2 hours. TLC analysis (EtOAc-MeOH, 85:15) after this time showed very little consumption of starting material ($R_f = 0.6$) and minimal formation of product ($R_f = 0.3$). MeCN (1 mL) was added and the reaction mixture was stirred at room temperature for a further 18 hours. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (EtOAc-MeOH, 88:12 then 85:15) to provide the product **38** as a white crystalline solid (35 mg, 93%); mp 197-202 $^{\circ}$ C (decomp).

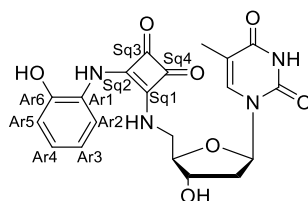
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3229, 2926, 1802, 1659, 1586, 1535, 1474, 1425, 1349, 1268, 1050, 959.

^1H NMR (600 MHz, DMSO- d_6): $\delta = 1.78$ (s, 3H, CH_3^{T}), 2.05-2.10 (m, 1H, H-2'a), 2.14-2.20 (m, 1H, H-2'b), 3.50 (br s, 2H, CH_2), 3.55 (br s, 2H, CH_2), 3.69 (br s, 1H, H-5'a), 3.79-3.83 (m, 1H, H-4'), 3.89 (br s, 1H, H-5'b), 4.19 (br s, 1H, H-3'), 4.92 (br s, 1H, OH), 5.42 (br s, 1H, OH), 6.19 (app t, $J = 6.9$ Hz, H-1'), 7.44 (s, 1H, H-6), 7.83 (br s, 2H, 2 x NH^{Sq}), 10.97 (br s, 1H, NH^{T}) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): $\delta = 12.0$ (CH_3^{T}), 38.4 (C-2'), 45.3 (C-5'), 45.8 (CH_2), 60.8 (CH_2), 70.6 (C-3'), 83.6 (C-1'), 85.2 (C-4'), 109.9 (C-5), 135.9 (C-6), 150.5 (C-2), 163.7 (C-4), 167.8 (C-Sq), 168.1 (C-Sq), 182.4 (C-Sq), 182.6 (C-Sq) ppm.

HRMS (APCI $^-$): m/z calc. 379.1259 [M-H] $^-$, 379.1257.

5'-N-(2-(2-Hydroxyphenyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-5'-deoxythymidine (39)



39

5'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-5'-deoxythymidine (**37**) (32 mg, 88 μmol) was suspended in water (1 mL) and EtOH (1 mL). 2-Aminophenol (51 mg, 467 μmol) was added and the reaction mixture was stirred at room temperature for 24 hours. After this time MeCN (1 mL) was added and the reaction mixture was stirred at room temperature for a further 3 hours. After this time DMF (0.8 mL) was added to fully dissolve the reaction mixture. The reaction mixture was stirred at room temperature for a further 21 hours. TLC analysis (CH_2Cl_2 -MeOH, 90:10) after this time showed consumption of starting material ($R_f = 0.4$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure. The residue was co-evaporated with toluene and then purified by flash chromatography (CH_2Cl_2 -MeOH, 95:5-90:10) to obtain the product **39** as a brown crystalline solid (31 mg, 81%); mp 201-207 $^\circ\text{C}$.

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3242, 3066, 2975, 1795, 1661, 1586, 1531, 1455, 1366, 1269, 1210, 1086, 1022, 840, 747.

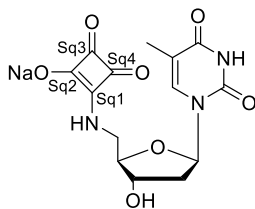
^1H NMR (600 MHz, DMSO-d_6): $\delta = 1.76$ (s, 3H, CH_3^{T}), 2.11 (ddd, $J_{2'a,2'b} = 13.6$ Hz, $J_{1',2'a} = 6.6$ Hz, $J = 4.0$ Hz, 1H, H-2'a), 2.22 (app dt, $J_{2'a,2'b} = 13.6$ Hz, $J_{1',2'b} = 6.6$ Hz, 1H, H-2'b), 3.71-3.79 (m, 1H, H-5'a), 3.84 (app dt, $J = 7.6$ Hz, $J = 3.7$ Hz, 1H, H-4'), 4.03 (ddd, $J_{5'a,5'b} = 10.3$ Hz, $J_{5'b,\text{NH-5}'} = 6.1$ Hz, $J_{4',5'b} = 3.7$ Hz, 1H, H-5'b), 4.20-4.24 (m, 1H, H-3'), 5.46 (br s, 1H, OH-3'), 6.23 (app t, $J = 6.6$ Hz, 1H, H-1'), 6.76 (app dt, $J = 4.1$ Hz, $J = 8.1$ Hz, $\text{H}^{\text{Ar}4}$), 6.86 (app d, $J = 4.1$ Hz, 2H, $\text{H}^{\text{Ar}3}$, $\text{H}^{\text{Ar}5}$), 7.43 (s, 1H, H-6), 7.76 (d, $J_{\text{Ar}2,\text{Ar}4} = 8.1$ Hz, 1H, $\text{H}^{\text{Ar}2}$), 8.45 (t, $J_{5',\text{NH-5}'} = 6.1$ Hz, 1H, NH-5'), 9.31 (s, 1H, NH^{Sq}), 10.18 (br s, 1H, OH-Ar), 11.32 (br s, 1H, NH^{T}) ppm.

^{13}C NMR (151 MHz, DMSO-d_6): $\delta = 12.0$ (CH_3^{T}), 38.4 (C-2'), 45.6 (C-5'), 70.6 (C-3'), 83.5 (C-1'), 85.0 (C-4'), 110.0 (C-5), 115.0 (C-Ar5), 119.3 (C-Ar4), 120.0 (C-Ar2), 123.6 (C-Ar3), 127.0

(C-Ar1), 136.0 (C-6), 146.7 (C-Ar6), 150.5 (C-2), 163.65 (C-4), 163.72 (C-Sq2), 169.2 (C-Sq1), 180.3 (C-Sq3), 184.1 (C-Sq4) ppm.

HRMS (APCI): m/z calc. 427.1259 [M-H]⁻, 427.1258.

5'-N-(2-Hydroxy-3,4-dioxocyclobut-1-en-1-yl)amino-5'-deoxythymidine³¹⁶ (40)



40

5'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-5'-deoxythymidine (**37**) (40 mg, 109 μ mol) was suspended in water (1.4 mL). NaOH (10 mg, 260 μ mol) was added and the reaction mixture was stirred at room temperature for 4 hours. TLC analysis (EtOAc-MeOH, 85:15) after this time showed complete consumption of starting material ($R_f = 0.6$) and formation of product ($R_f = 0.1$). The reaction mixture was eluted through Dialon WT01S ion-exchange resin (H form) and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (water-*i*PrOH-EtOAc, 1:5:4) to provide the product **40** as a white crystalline solid (13 mg, 33%); mp 218-222 °C (decomp).

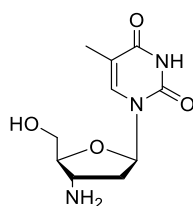
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3299, 2908, 1792, 1637, 1618, 1515, 1414, 1367, 1273, 1237, 1133, 1052, 1035, 979, 931, 795, 751.

¹H NMR (400 MHz, D₂O): $\delta = 1.84$ (s, 3H, CH₃^T), 2.38-2.43 (m, 2H, H-2'a, H-2'b), 3.80 (dd, $J_{5'a,5'b} = 14.6$ Hz, $J_{4',5'a} = 4.7$ Hz, 1H, H-5'a), 3.98 (dd, $J_{5'a,5'b} = 14.6$ Hz, $J_{4',5'b} = 4.1$ Hz, 1H, H-5'b), 4.09 (dd, $J_{4',5'a} = 4.7$ Hz, $J_{4',5'b} = 4.1$ Hz, 1H, H-4'), 4.48-4.56 (m, 1H, H-3'), 6.25 (app t, $J = 6.6$ Hz, 1H, H-1'), 7.35 (s, 1H, H-6) ppm.

¹³C NMR (100 MHz, D₂O): $\delta = 11.5$ (CH₃^T), 37.8 (C-2'), 44.3 (C-5'), 70.4 (C-3'), 84.7 (C-1'), 84.8 (C-4'), 111.6 (C-5), 136.9 (C-6), 151.7 (C-2), 166.3 (C-4), 181.8 (C-Sq1), 188.1 (C-Sq3), 195.1 (C-Sq2, C-Sq4) ppm.

HRMS (ESI⁺): m/z calc. 382.0622 [M+Na]⁺, found: 382.0625.

3'-Amino-3'-deoxythymidine¹²⁷ (**43**)



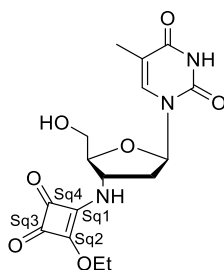
43

AZT (**42**) (1.08 g, 4.49 mmol) was dissolved in MeOH (20 mL). The reaction vessel was purged with argon, and 10% Pd/C (0.134 g) was added. The reaction vessel was purged with H₂. The reaction mixture was stirred under an atmosphere of H₂ for 4 hours. TLC analysis (EtOAc) after this time showed complete consumption of starting material ($R_f = 0.6$) and formation of product ($R_f = 0.0$). The reaction mixture was filtered through celite, washing through with MeOH and water. The solvent was removed under reduced pressure to provide the product as a white foam (0.960 g, 98%).

¹H NMR (400 MHz, DMSO-d₆): $\delta = 1.76$ (s, 3H, CH₃^T), 1.94-2.01 (m, 1H, H-2'a), 2.08 (ddd, $J_{1',2'b} = 5.2$ Hz, $J_{2'b,3'} = 7.1$ Hz, $J_{2'a',2'b} = 12.6$ Hz, 1H, H-2'b), 3.37-3.43 (m, 1H, H-3'), 3.49-3.54 (m, 1H, H-4'), 3.57 (dd, $J_{4',5'a} = 3.8$ Hz, $J_{5'a,5'b} = 11.9$ Hz, 1H, H-5'a), 3.65 (dd, $J_{4',5'b} = 3.0$ Hz, $J_{5'a,5'b} = 11.9$ Hz, 1H, 5'b), 4.10 (br s, 1H, OH-5'), 4.98 (br s, 2H, NH₂-3'), 6.08 (dd, $J_{1',2'b} = 5.2$ Hz, $J_{1,2'a'} = 6.6$ Hz 1H, H-1'), 7.75 (s, 1H, H-6), 8.31 (s, 1H, NH^T) ppm.

HRMS (APCI⁺): m/z calc. 242.1063 [M+H]⁺, found: 242.1135.

3'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (**44**)



44

3'-Amino-3'-deoxythymidine (**43**) (502 mg, 2.08 mmol) was suspended in a mixture of EtOH (12 mL) and CH₂Cl₂ (6 mL). Diethyl squarate (0.46 mL, 3.11 mmol) was added, and the reaction mixture was stirred at room temperature. After 10 minutes the suspension

dissolved. TLC analysis (CH₂Cl₂-EtOH, 92:8) after 45 minutes showed complete consumption of starting material ($R_f = 0.0$) and formation of product ($R_f = 0.8$). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 96:4). The product **44** was obtained as a white powder (602 mg, 79%); mp 172-176 °C.

Note: compound **44** exhibits rotamers in NMR spectroscopy.

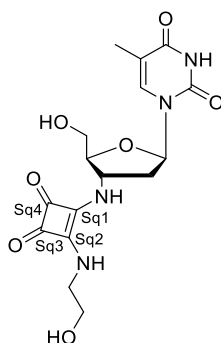
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3197, 3040, 1805, 1686, 1596, 1421, 1269, 993, 736, 695, 598.

¹H NMR (400 MHz, DMSO-d₆): $\delta = 1.37$ (br s, 3H, CH₃^{Et}), 1.78 (s, 3H, CH₃^T), 2.22 - 2.37 (m, 2H, H-2'), 3.56 - 3.67 (m, 2H, H-5'), 3.90 (s, 0.5H, H-4'), 3.93 (s, 0.5H, H-4'), 4.21 (br s, 0.5H, H-3'), 4.57 - 4.75 (m, 2.5H, CH₂^{Et}, H-3'), 5.10-5.17 (m, 1H, OH-5'), 6.22 (app t, $J = 6.3$ Hz, 1H, H-1'), 7.74 (s, 1H, H-6), 8.96 (d, $J_{3',\text{NH}} = 6.3$ Hz, 0.5H, NH-3'), 9.15 (d, $J_{3',\text{NH}} = 6.3$ Hz, 0.5H, NH-3'), 11.30 (s, 1H, NH^T) ppm.

¹³C NMR (100 MHz, DMSO-d₆): $\delta = 12.3$ (CH₃^T), 15.6 (CH₃^{Et}), 15.7 (CH₃^{Et}), 37.6 (C-2'), 37.9 (C-2'), 54.1 (C-3'), 54.4 (C-3'), 61.3 (C-5'), 61.4 (C-5'), 69.1 (CH₂^{Et}), 83.48 (C-1'), 83.53 (C-1'), 84.5 (C-4'), 84.9 (C-4'), 109.5 (C-5), 136.3 (C-6), 150.4 (C-2), 163.8 (C-4), 171.8 (C-Sq2), 172.2 (C-Sq2), 177.1 (C-Sq1), 177.5 (C-Sq1), 182.3 (C-Sq3), 182.8 (C-Sq3), 189.0 (C-Sq4) ppm.

HRMS (APCI⁺): m/z calc. 366.1296 [M+H]⁺, found: 366.1296.

3'-N-(2-(2-Hydroxyethyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (45)



45

3'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (**44**) (35 mg, 95 μmol) was dissolved in EtOH (0.8 mL) and CH₂Cl₂ (1 mL). Ethanolamine (8 μL , 132 μmol)

was added. The reaction mixture was stirred at room temperature for 2.5 hours, after which time a white solid precipitate had formed. Water (1 mL) was added to dissolve the precipitate and the reaction mixture was stirred for a further 18 hours. TLC analysis (CH₂Cl₂-MeOH, 92:8) after this time showed consumption of starting material ($R_f = 0.5$) and formation of product ($R_f = 0.0$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (EtOAc-MeOH, 75:25) to provide the product **45** as a white crystalline solid (34 mg, 94%); mp 176-182 °C.

Note: compound **45** exhibits rotamers in NMR spectroscopy.

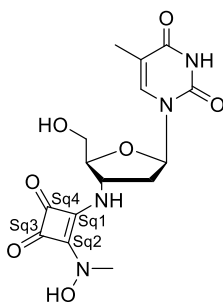
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3169, 2925, 1801, 1670, 1572, 1536, 1468, 1432, 1348, 1272, 1212, 1090, 1053, 965, 865, 763, 734.

¹H NMR (400 MHz, DMSO-d₆): $\delta = 1.21$ -1.26 (m, 2H, CH₂-O), 1.79 (s, 3H, CH₃^T), 2.26-2.31 (m, 1H, H-2'a), 2.35-2.41 (m, 1H, H-2'b), 3.50-3.54 (m, 1H x H-5'a, 0.7H, CH₂-N), 3.54-3.60 (m, 1.3H, CH₂-N), 3.61-3.70 (m, 1H, H-5'b), 3.87-3.91 (1H, H-4'), 4.56 (br s, 1H, H-3'), 4.94 (br s, 1H, OH), 5.18 (br s, 1H, OH), 6.20 (app t, $J = 6.3$ Hz, 1H, H-1'), 7.69 (br s, 1H, NH), 7.77 (s, 1H, H-6), 8.08 (br s, 1H, NH), 11.31 (s, 1H, NH^T) ppm.

¹³C NMR (100 MHz, DMSO-d₆): $\delta = 12.3$ (CH₃^T), 29.0 (CH₂-O), 38.2 (C-2'), 45.9 (CH₂-N), 53.6 (C-3'), 60.7 (C-5'), 83.3 (C-1'), 85.5 (C-4'), 109.5 (C-5), 136.1 (C-6), 150.5 (C-2), 163.7 (C-4), 166.8 (C-Sq), 168.5 (C-Sq), 182.2 (C-Sq), 182.7 (C-Sq) ppm.

HRMS (ESI⁺): m/z calc. 403.1224 [M+Na]⁺, found: 403.1220.

3'-N-(2-(N-Methyl)hydroxylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (46)



46

3'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (**44**) (31 mg, 86 μmol) was dissolved in EtOH (1 mL) and CH_2Cl_2 (1 mL). *N*-Methylhydroxylamine hydrochloride (12 mg, 140 μmol) was added, followed by triethylamine (35 μL , 251 μmol). The reaction mixture was stirred at room temperature for 24 hours. TLC analysis (EtOAc-MeOH, 75:25) after this time showed complete consumption of starting material ($R_f = 0.8$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (toluene-acetone, 1:9) to provide the product **46** as a white foam (24 mg, 76%).

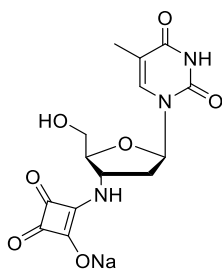
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3231, 2917, 1689, 1656, 1546, 1453, 1402, 1376, 1270, 1094, 1069, 884, 813, 786, 716, 676.

^1H NMR (400 MHz, D_2O): $\delta = 1.91$ (d, $J_{6,\text{CH}_3} = 1.1$ Hz, 3H, CH_3^{T}), 2.59-2.66 (m, 2H, H-2'a, H-2'b), 3.51 (s, 3H, $\text{CH}_3\text{-N}$), 3.82 (dd, $J_{5'a,5'b} = 12.8$ Hz, $J_{4',5'a} = 4.6$ Hz, 1H, H-5'a), 3.93 (dd, $J_{5'a,5'b} = 12.8$ Hz, $J_{4',5'b} = 2.3$ Hz, 1H, H-5'b), 4.07-4.14 (m, 1H, H-4'), 4.81-4.89 (m, 1H, H-3'), 6.28 (app t, $J = 5.9$ Hz, 1H, H-1'), 7.71 (app d, $J = 1.1$ Hz, 1H, H-6) ppm.

^{13}C NMR (100 MHz, D_2O): $\delta = 12.5$ (CH_3^{T}), 38.4 (C-2'), 42.0 ($\text{CH}_3\text{-N}$), 54.3 (C-3'), 61.2 (C-5'), 85.0 (C-1'), 85.6 (C-4'), 112.4 (C-5), 138.8 (C-6), 152.6 (C-2), 165.5 (C-Sq), 166.7 (C-Sq), 167.6 (C-4), 178.4 (C-Sq), 179.2 (C-Sq) ppm.

HRMS (ESI⁺): m/z calc. 389.1068 [$\text{M}+\text{Na}$]⁺, found: 389.1067.

3'-N-(2-Hydroxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine³¹⁶ (47)



47

3'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (**44**) (48 mg, 132 μmol) was dissolved in EtOH (1.3 mL) and water (0.3 mL). NaOH (18 mg, 447 μmol) was added, and the reaction mixture was stirred at room temperature for 4 hours. TLC analysis (EtOAc-MeOH, 90:10) after this time showed formation of product ($R_f = 0.0$), but a large amount of starting material ($R_f = 0.5$) remaining. Additional NaOH (17 mg, 425 μmol) was added, and the reaction mixture was stirred at room temperature for a further 1.5 hours. TLC analysis after this time showed complete consumption of starting material and formation of product. The reaction mixture was eluted through Dialon WT01S ion-exchange resin (H form) and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (water-*i*PrOH-EtOAc, 4:56:40) to obtain the product **47** as a white crystalline solid (10 mg, 22%); mp 112-118 °C.

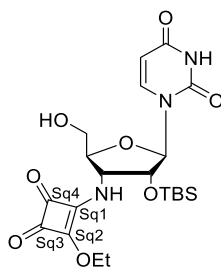
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3368, 2508, 1651, 1618, 1529, 1499, 1469, 1445, 1239, 1096, 1070, 897, 766.

^1H NMR (400 MHz, D_2O): $\delta = 1.86$ (d, $J_{6,\text{CH}_3} = 1.0$ Hz, 3H, CH_3^{T}), 2.52-2.57 (m, 2H, H-2'a, H-2'b), 3.76 (dd, $J_{5'a,5'b} = 13.0$ Hz, $J_{4',5'a} = 4.5$ Hz, 1H, H-5'a), 3.86 (dd, $J_{5'a,5'b} = 13.0$ Hz, $J_{4',5'b} = 2.6$ Hz, 1H, H-5'b), 4.01 (ddd, $J_{3',4'} = 7.9$ Hz, $J_{4',5'a} = 4.5$ Hz, $J_{4',5'b} = 2.6$ Hz, 1H, H-4'), 4.73-4.79 (m, 1H, H-3'), 6.23 (app t, $J = 5.9$ Hz, 1H, H-1'), 7.69 (app d, $J = 1.0$ Hz, 1H, H-6) ppm.

^{13}C NMR (100 MHz, D_2O): $\delta = 11.5$ (CH_3^{T}), 37.4 (C-2'), 52.3 (C-3'), 60.0 (C-5'), 84.3 (C-1'), 84.4 (C-4'), 111.3 (C-5), 137.7 (C-6), 151.6 (C-2), 166.5 (C-4), 180.6 (C-Sq), 188.3 (C-Sq), 195.2 (C-Sq) ppm.

HRMS (APCI): m/z calc. 336.0837 [$\text{M}-\text{Na}$] $^-$, found: 336.0837.

2'-O-(*tert*-Butyldimethylsilyl)-3'-N-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (48)



2'-O-(*tert*-Butyldimethylsilyl)-3'-amino-3'-deoxyuridine (**23**) (56 mg, 0.16 mmol) was dissolved in EtOH (1 mL). Diethyl squarate (35 μ L, 0.24 mmol) was added, and the reaction mixture was stirred at room temperature under argon for 17.5 hours. TLC analysis (CH₂Cl₂-MeOH, 90:10) after this time showed formation of product (R_f = 0.4) and consumption of starting material (R_f = 0.3). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 97:3) to obtain the product **48** as a white foam (73 mg, 98%).

Note: compound **48** exhibits rotamers in NMR spectroscopy.

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3224, 2931, 2858, 1807, 1681, 1595, 1464, 1428, 1380, 1333, 1264, 1216, 1090, 1061, 835, 778.

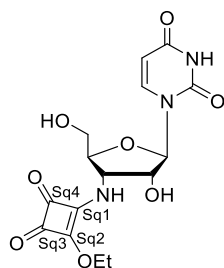
¹H NMR (600 MHz, DMSO-*d*₆): δ = -0.082 (s, 1.5H, CH₃^{TBDMS}), -0.078 (s, 1.5H, CH₃^{TBDMS}), -0.038 (s, 1.5H, CH₃^{TBDMS}), -0.035 (s, 1.5H, CH₃^{TBDMS}), -0.77 (s, 4.5H, *t*Bu^{TBDMS}), -0.78 (s, 4.5H, *t*Bu^{TBDMS}), 1.33 (t, $J_{\text{CH}_2, \text{CH}_3}$ = 7.0 Hz, 1.5H, CH₃^{Et}), 1.37 (t, $J_{\text{CH}_2, \text{CH}_3}$ = 7.0 Hz, 1.5H, CH₃^{Et}), 3.57-3.61 (m, 1H, H-5'a), 3.65-3.70 (m, 1H, H-5'b), 4.09-4.15 (m, 1.5H, H-3', 0.5 x H-4'), 4.21-4.34 (m, 1H, H-2'), 4.57-4.73 (m, 2.5H, CH₂^{Et}, 0.5 x H-4'), 5.29 (dd, $J_{5'a, \text{OH}-5'}$ = 11.7 Hz, $J_{5'a, \text{OH}-5'}$ = 5.1 Hz, 1H, OH-5'), 5.71-5.74 (2 x d, 1H, H-5), 5.93 (d, $J_{1', 2'}$ = 5.1 Hz, 0.5H, H-1') 5.97 (d, $J_{1', 2'}$ = 5.9 Hz, 0.5H, H-1'), 7.93 (app t, J = 9.0 Hz, 1H, H-6), 9.03 (d, $J_{3', \text{NH}^{\text{Sq}}}$ = 9.3 Hz, 0.5H, NH^{Sq}), 9.24 (d, $J_{3', \text{NH}^{\text{Sq}}}$ = 7.9 Hz, 0.5H, NH^{Sq}), 11.43 (s, 1H, NH^U) ppm.

¹³C NMR (151 MHz, DMSO-*d*₆): δ = -5.5 (CH₃^{TBDMS}), -5.4 (CH₃^{TBDMS}), -5.3 (CH₃^{TBDMS}), -5.2 (CH₃^{TBDMS}), 15.5 (CH₃^{Et}), 15.7 (CH₃^{Et}), 17.45 (qC, *t*Bu^{TBDMS}), 17.49 (qC, *t*Bu^{TBDMS}), 25.3 (*t*Bu^{TBDMS}), 55.3 (C-4'), 55.4 (C-4'), 60.2 (C-5'), 60.8 (C-5'), 69.0 (CH₂^{Et}), 69.1 (CH₂^{Et}), 74.6 (C-

2'), 75.0 (C-2'), 82.1 (C-3'), 82.5 (C-3'), 86.8, (C-1'), 87.3 (C-1'), 102.1 (C-5), 102.4 (C-5), 140.1 (C-6), 140.2 (C-6), 150.6 (C-2), 150.7 (C-2), 162.88 (C-4), 162.91 (C-4), 172.5 (C-Sq2), 172.8 (C-Sq2), 176.8 (C-Sq1), 177.1 (C-Sq1), 182.3 (C-Sq3), 182.5 (C-Sq3), 188.7 (C-Sq4), 189.2 (C-Sq4) ppm.

HRMS (APCI⁻): m/z calc. 480.1807 [M+H]⁻, found: 480.180.

3'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**49**)



49

2'-O-(*tert*-Butyldimethylsilyl)-3'-N-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**48**) (64 mg, 133 μ mol) was dissolved in THF (1 mL). Tetrabutylammonium fluoride trihydrate (60 mg, 190 μ mol) was added and the reaction mixture was stirred at room temperature for 2 hours. TLC analysis (CH₂Cl₂-MeOH, 90:10) after this time showed consumption of starting material (R_f = 0.7) and formation of both rotamers of the product (R_f = 0.33, 0.41). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 93:7) to obtain the product **49** as a white powder (40 mg, 83%); mp 225-235 °C (decomp).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3214, 2055, 2926, 1806, 1673, 1590, 1425, 1382, 1345, 1258, 1097, 1052, 987, 864, 812, 766.

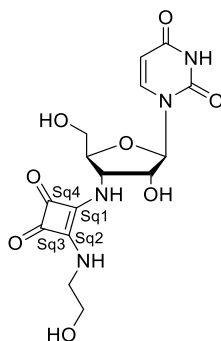
¹H NMR (600 MHz, acetone-d₆): δ = 1.40 (t, $J_{\text{CH}_2, \text{CH}_3}$ = 7.0 Hz, 3H, CH₃^{Et}), 3.83-3.89 (m, 1H, H-5'a), 3.93-4.01 (m, 1H, H-5'b), 4.22-4.37 (m, 1H, H-4'), 4.43-4.53 (m, 2H, H-2', OH-5'), 4.71 (q, $J_{\text{CH}_2, \text{CH}_3}$ = 7.0 Hz, 2H, CH₂^{Et}), 4.80 (br s, 1H, H-3'), 5.46 (br s, 1H, OH-2'), 5.63 (d, $J_{5,6}$ = 8.2 Hz, 1H, H-5), 5.91 (d, $J_{1',2'}$ = 1.9 Hz, 1H, H-1'), 7.60 (br s, 0.5H, NH^{Sq}), 7.68 (br s, 0.5H, NH^{Sq}), 8.09 (br s, 1H, H-6), 10.07 (br s, 1H, NH^U) ppm.

¹³C NMR (151 MHz, acetone-d₆): δ = 16.0 (CH₃^{Et}), 55.2 (C-3'), 55.9 (C-3'), 61.1 (C-5'), 70.6 (CH₂^{Et}), 75.8 (C-2'), 83.5 (C-4'), 84.0 (C-4'), 91.2 (C-1'), 102.3 (C-5), 141.2 (C-6), 151.5 (C-2),

163.7 (C-4), 173.5 (C-Sq1), 174.4 (C-Sq1), 178.6 (C-Sq2), 179.4 (C-Sq2), 183.7 (C-Sq4), 184.5 (C-Sq4), 189.5 (C-Sq3), 190.6 (C-Sq3) ppm.

HRMS (APCI⁺): *m/z* calc. 368.1088 [M+H]⁺, found: 368.1091.

3'-*N*-(2-(2-Hydroxyethyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (50)



50

3'-*N*-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**49**) (41 mg, 0.113 mmol) was dissolved in EtOH (2 mL). Ethanolamine (9 μ L, 0.147 mmol) was added. The reaction mixture was stirred at room temperature under argon. After 5 minutes a white precipitate formed. After 2 hours TLC analysis (EtOAc-MeOH, 87:13) showed complete consumption of both rotamers of the starting material (R_f = 0.66, 0.73) and formation of product (R_f = 0.3). The solvent was removed under reduced pressure. The residue was purified by flash chromatography (EtOAc-MeOH, 88:12) to afford the product **50** as a white foam (41 mg, 94%).

Note: compound **50** exhibits rotamers in NMR spectroscopy.

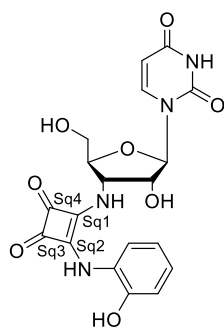
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3472, 3331, 3169, 2963, 2795, 1804, 1697, 1675, 1648, 1562, 1493, 1466, 1415, 1379, 1357, 1308, 1270, 1206, 1170, 1111, 1046, 945, 868, 830, 710, 626.

¹H NMR (600 MHz, DMSO-*d*₆): δ = 3.30 (m, 0.4H, CH₂-N), 3.50-3.66 (m, 1.6H x CH₂-N, 2H x CH₂-O, 1H x H-5'a), 3.72 (dd, $J_{5'a,5'b}$ = 12.1 Hz, $J_{5'b,OH-5'}$ = 5.1 Hz, 1H, H-5'b), 3.93 (br s, 1H, H-4'), 4.21 (br s, 0.2H, H-2'), 4.26 (br s, 0.8H, H-2'), 4.38 (br s, 0.2H, H-3'), 4.49 (br s, 0.8H, H-3'), 4.94 (m, 1H, OH), 5.24 (app t, J = 5.1 Hz, OH-5'), 5.67 (d, $J_{5,6}$ = 8.1 Hz, 1H, H-5), 5.73 (d, $J_{1',2'}$ = 3.6 Hz, 0.2H, H-1'), 5.76 (m, 0.8H, H-1'), 6.03 (br s, 0.2H, OH-2'), 6.37 (br s, 0.8H, OH-2'), 7.75 (d, $J_{3',NH-3'}$ = 7.1 Hz, 1H, NH-3'), 7.84 (br s, 0.2H, NH^{Sq}), 7.91 (br s, 0.8H, NH^{Sq}), 7.99 (d, $J_{5,6}$ = 8.1 Hz, 1H, H-6), 11.35 (s, 1H, NH^U) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = 45.9 (CH₂-N), 53.9 (C-3'), 55.2 (C-3'), 60.0 (C-5'), 60.7 (CH₂-O), 71.0 (C-2'), 73.6 (C-2'), 82.3 (C-4'), 83.8 (C-4'), 89.1 (C-1'), 89.7 (C-1'), 101.6 (C-5), 140.4 (C-6), 150.5 (C-2), 152.9 (C-Sq2), 163.2 (C-4), 166.9 (C-Sq), 168.7 (C-Sq), 182.5 (C-Sq) ppm.

HRMS (APCI⁻): m/z calc. 381.1052 [M-H]⁻, found: 381.1050.

3'-N-(2-(2-Hydroxyphenyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (51)



51

3'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**51**) (35 mg, 95 μmol) was dissolved in EtOH (2 mL). 2-Aminophenol (15 mg, 137 μmol) was added. The reaction mixture was stirred at room temperature for 48 hours. More 2-aminophenol (15 mg, 137 μmol) was added, and the reaction mixture was stirred at room temperature for a further 20 hours. 2-Aminophenol (11 mg, 101 μmol) was again added and the reaction mixture was stirred for a further 7 hours. TLC analysis (CH₂Cl₂-MeOH, 90:10) after this time showed consumption of both rotamers of the starting material (R_f = 0.33, 0.41) and formation of product (R_f = 0.26). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 92:8) to provide the product **51** as a brown crystalline solid (35 mg, 85%); mp 199-206 °C.

Note: compound **51** exhibits rotamers in NMR spectroscopy.

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3240, 3063, 2971, 2932, 1795, 1672, 1573, 1518, 1454, 1440, 1381, 1256, 1082, 995, 875, 814, 737, 696.

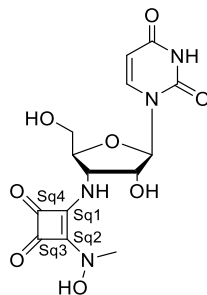
^1H NMR (600 MHz, DMSO- d_6): δ = 3.54-3.59 (m, 0.2H, H-5'), 3.60-3.64 (m, 0.2H, H-5'), 3.64-3.69 (m, 0.8H, H-5'), 3.71-3.76 (m, 0.8H, H-5'), 3.79-3.84 (m, 0.2H, H-4'), 3.98-4.03 (m,

0.8H, H-4'), 4.23-4.27 (m, 0.2H, H-2'), 4.29-4.34 (m, 0.8H, H-2'), 4.50-4.56 (m, 0.2H, H-3'), 4.59-4.69 (m, 0.8H, H-3'), 4.96 (app t, $J = 5.2$ Hz, 0.2H, OH-5'), 5.24 (app t, $J = 5.2$ Hz, 0.8H, OH-5'), 5.70 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.77 (d, $J_{1',2'} = 3.8$ Hz, 0.2H, H-1'), 5.81 (d, $J_{1',2'} = 3.8$ Hz, 0.8H, H-1'), 6.00 (d, $J_{2',\text{OH-}2'} = 4.9$ Hz, 0.2H, OH-2'), 6.32 (d, $J_{2',\text{OH-}2'} = 4.9$ Hz, 0.8H, OH-2'), 6.74-6.81 (m, 1H, H^{Ar}), 6.84-6.93 (m, 2H, H^{Ar}), 7.71-7.83 (m, 1H, H^{Ar}), 7.99 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 8.59 (d, $J_{3',\text{NH-}3'} = 8.6$ Hz, 1H, NH-3'), 9.50 (br s, 0.2H, NH^{Sq}), 9.53 (br s, 0.8H, NH^{Sq}), 10.14 (br s, 1H, OH-Ar), 10.31 (br s, 0.2H, NH^U), 11.38 (br s, 0.8H, NH^U) ppm.

¹³C NMR (151 MHz, DMSO-d₆): $\delta = 54.7$ (C-3'), 55.8 (C-3'), 60.2 (C-5'), 61.1 (C-5'), 70.9 (C-2'), 73.6 (C-2'), 81.9 (C-4'), 83.5 (C-4'), 88.9 (C-1'), 89.5 (C-1'), 101.7 (C-5), 115.0 (C-Ar), 119.3 (C-Ar), 120.05 (C-Ar), 120.12 (C-Ar), 123.6 (C-Ar), 123.7 (C-Ar), 126.9 (C-Ar), 127.0 (C-Ar), 140.4 (C-6), 146.8 (C-Ar), 150.6 (C-2), 163.2 (C-4), 164.0 (C-Sq2), 164.2 (C-Sq2), 168.9 (C-Sq1), 169.1 (C-Sq1), 180.5 (C-Sq3), 180.6 (C-Sq3), 184.0 (C-Sq4) ppm.

HRMS (ESI⁻): m/z calc. 429.1052 [M-H]⁻, found: 429.1059.

3'-*N*-(2-(*N*-Methyl)hydroxylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**52**)



52

3'-*N*-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**49**) (41 mg, 112 μmol) was dissolved in EtOH (2 mL). *N*-Methylhydroxylamine hydrochloride (14 mg, 169 μmol) and triethylamine (69 μL , 494 μmol) were added. The reaction mixture was stirred at room temperature for 2 hours. TLC analysis (CH₂Cl₂-MeOH, 90:10) after this time showed complete consumption of starting material ($R_f = 0.3$) and formation of product ($R_f = 0.1$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 90:10, then 87:13) to provide the product **52** as a white crystalline solid (21 mg, 51%); mp 200-215 °C (decomp).

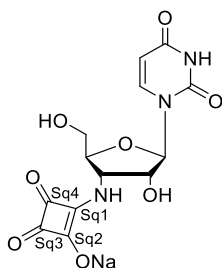
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3448, 3238, 3045, 2926, 2797, 1801, 1683, 1667, 1557, 1451, 1411, 1380, 1263, 1252, 1226, 1155, 1096, 1062, 1025, 892, 844, 781, 675, 628.

^1H NMR (600 MHz, DMSO- d_6): δ = 3.37 (s, 3H, CH₃), 3.57-3.63 (m, 1H, H-5'a), 3.70 (d, $J_{5'a,5'b}$ = 11.6 Hz, 1H, H-5'b), 4.12-4.14 (m, 1H, H-4'), 4.15-4.18 (m, 1H, H-2'), 4.56-4.61 (m, 1H, H-3'), 5.16 (br s, 1H, OH-5'), 5.65 (d, $J_{5,6}$ = 8.0 Hz, 1H, H-5), 5.83 (d, J = 3.6 Hz, 1H, H-1'), 6.06 (br s, 1H, OH-2'), 7.32 (br s, 1H, NH-3'), 7.93 (d, $J_{5,6}$ = 8.0 Hz, 1H, H-6), 10.76 (br s, 1H, N-OH), 11.35 (br s, 1H, NH^U) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = 40.9 (CH₃), 54.3 (C-3'), 60.3 (C-5'), 73.7 (C-2'), 82.4 (C-4'), 88.9 (C-1'), 101.7 (C-5), 140.6 (C-6), 150.6 (C-2), 163.2 (C-4), 165.9 (C-Sq2), 166.8 (C-Sq1), 178.8 (C-Sq), 179.9 (C-Sq) ppm.

HRMS (ESI⁺): m/z calc. 391.0860 [M+Na]⁺, found: 391.0862.

3'-N-(2-Hydroxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**53**)



53

3'-N-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**49**) (36 mg, 98 μmol) was dissolved in EtOH (0.96 mL) and water (0.24 mL). NaOH (11 mg, 285 μmol) was added and the reaction mixture was stirred at room temperature for 3 hours. TLC analysis (EtOAc-MeOH, 85:15) after this time showed consumption of starting material (R_f = 0.6) and formation of product (R_f = 0.0). The reaction mixture was eluted through Dialon WT01S ion-exchange resin (H form) and then eluted through Dialon WT01S ion exchange resin (Na form). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (water-*i*PrOH-EtOAc, 3:57:40) to obtain the product **53** as a white crystalline solid (16 mg, 40%); mp 188-194 °C (decomp).

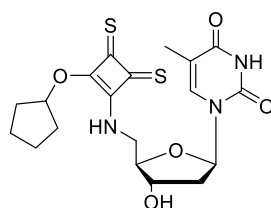
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3260, 2982, 1631, 1509, 1439, 1353, 1294, 1223, 1132, 1094, 1029, 988, 819, 792, 741, 701, 653.

^1H NMR (600 MHz, D_2O): δ = 3.85 (dd, $J_{5'a,5'b}$ = 13.2 Hz, $J_{4',5'a}$ = 4.1 Hz, 1H, H-5'a), 3.98 (dd, $J_{5'a,5'b}$ = 13.2 Hz, $J_{4',5'b}$ = 2.1 Hz, 1H, H-5'b), 4.18-4.22 (m, 1H, H-4'), 4.45 (dd, $J_{2',3'}$ = 5.7 Hz, $J_{1',2'}$ = 1.8 Hz, 1H, H-2'), 4.63 (dd, $J_{3',4'}$ = 9.1 Hz, $J_{2',3'}$ = 5.7 Hz, 1H, H-3'), 5.83 (d, $J_{5,6}$ = 7.7 Hz, 1H, H-5), 5.94 (d, $J_{1',2'}$ = 1.8 Hz, 1H, H-1'), 7.82 (d, $J_{5,6}$ = 7.7 Hz, 1H, H-6) ppm.

^{13}C NMR (151 MHz, D_2O): δ = 53.6 (C-3'), 60.0 (C-5'), 74.5 (C-2'), 82.1 (C-4'), 91.4 (C-1'), 102.4 (C-5), 140.4 (C-6), 159.3 (C-2), 177.1 (C-4), 180.9 (C-Sq1), 188.4 (C-Sq3), 195.7 (C-Sq2, C-Sq4) ppm.

HRMS (ESI⁺): m/z calc. 384.0414 [M+Na]⁺, found: 384.0417.

5'-N-(2-Cyclopentyloxy-3,4-dithionecyclobut-1-en-1-yl)amino-5'-deoxythymidine (56)



56

5'-Amino-5'-deoxythymidine (**54**) (100 mg, 440 μmol) was dissolved in anhydrous DMF (1.2 mL). Dicyclopentyl dithiosquarate (**55**) (125 mg, 440 μmol) was added and the reaction mixture was stirred under argon at room temperature for 4 hours. TLC analysis (water-*i*PrOH-EtOAc, 1:2:2) after this time showed complete consumption of starting material (R_f = 0.0) and formation of product (R_f = 0.4). The solvent was removed under reduced pressure, and the residue was co-evaporated with toluene (50 mL) to provide the product **56** as a yellow amorphous solid (190 mg, 99%). The product was used in the next step without further purification.

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3236, 2958, 2518, 2160, 2024, 1659, 1577, 1465, 1404, 1263, 1230, 1046, 959, 883, 764, 692.

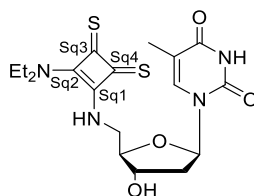
^1H NMR (400 MHz, DMSO-d_6) δ = 1.46 – 1.72 (m, 9H, 4 x CH_2^{Cp} , CH^{Cp}), 1.78 (d, $J_{\text{CH}_3,6}$ = 1.0 Hz, 3H, CH_3^{T}), 2.10 (ddd, $J_{2'a,2'b}$ = 13.6 Hz, $J_{1',2'a}$ = 6.6 Hz, $J_{2'a,3'}$ = 3.6 Hz, 1H, H-2'a), 2.23 (dt, J = 6.6 Hz, J = 3.6 Hz, 1H, H-2'b), 3.86-3.93 (m, 1H, H-4'), 4.12 (ddd, $J_{5'a,5'b}$ = 13.3 Hz, $J_{\text{NH},5'a}$ = 8.4 Hz, $J_{4',5'a}$ = 5.4 Hz, 1H, H-5'a), 4.21 (app dt, J = 6.6 Hz, J = 3.6 Hz, 1H, H-3'), 4.62 (ddd, $J_{5'a,5'b}$ = 13.3 Hz, $J_{\text{NH},5'b}$ = 7.2 Hz, $J_{4',5'b}$ = 4.1 Hz, 1H, H-5'b), 6.22 (app t, J = 6.6 Hz, 1H, H-1'),

7.46 (app d, $J = 1.0$ Hz, 1H, H-6), 8.68 (dd, $J_{\text{NH},5'a} = 8.4$ Hz, $J_{\text{NH},5'b} = 7.2$ Hz, 1H, NH-5'), 11.32 (s, 1H, NH^T) ppm.

¹³C NMR (100 MHz, DMSO-d₆) $\delta = 12.1$ (CH₃^T), 22.9 (CH₂^{Cp}), 23.1 (CH₂^{Cp}), 23.25 (CH₂^{Cp}), 23.30 (CH₂^{Cp}), 30.1 (CH^{Cp}), 38.2 (C-2'), 45.0 (C-5'), 70.8 (C-3'), 83.8 (C-1'), 84.5 (C-4'), 110.1 (C-5), 136.1 (C-6), 150.5 (C-2), 163.6 (C-4), 170.5 (C-Sq), 204.2 (C-Sq) ppm.

HRMS (ESI⁻): m/z calc. 436.1006 [M-H]⁻, found: 436.1017.

5'-N-(2-Diethylamino-3,4-dithionecyclobut-1-en-1-yl)amino-5'-deoxythymidine (**57**)



57

5'-N-(2-Cyclopentyloxy-3,4-dithionecyclobut-1-en-1-yl)amino-5'-deoxythymidine (**56**) (43 mg, 98 μmol) was dissolved in anhydrous DMF (1 mL). Diethylamine (99 μL , 957 μmol) was added and the reaction mixture was stirred under argon at room temperature for 23.5 hours. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (EtOAc-*i*PrOH, 75:25). Remaining impurities were removed through further purification by flash chromatography (CH₂Cl₂-MeOH, 95:5 then 90:10, then EtOAc-*i*PrOH, 75:25) to provide the product **57** as a yellow amorphous solid (11 mg, 27%).

Note: compound **57** exhibits rotamers in NMR spectroscopy.

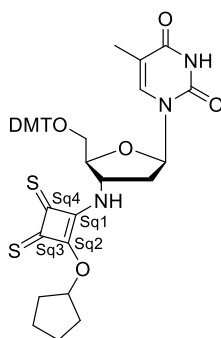
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3229, 3053, 2969, 2926, 1686, 1561, 1461, 1432, 1379, 1320, 1263, 1234, 1199, 1178, 1127, 1075, 1022, 951, 804, 764, 613.

¹H NMR (400 MHz, DMSO-d₆): $\delta = 1.19$ -1.27 (m, 6H, 2 x CH₃^{Et}), 1.78 (s, 3H, CH₃^T), 2.05-2.14 (m, 2H, H-2'a, H-2'b), 3.51-3.60 (m, 2H, CH₂^{Et}), 3.87-3.91 (m, 0.2H, H-4'), 4.00 (td, $J_{4',5'} = 7.0$ Hz, $J_{3',4'} = 3.0$ Hz, 0.8H, H-4'), 4.11-4.32 (m, 4H, H-3', H-5'a, CH₂^{Et}), 4.53-4.62 (m, 0.2H, H-5'b), 4.67-4.77 (m, 0.8H, H-5'b), 5.38 (d, $J_{3',\text{OH}-3'} = 4.3$ Hz, 0.8H, OH-3'), 5.44 (d, $J_{3',\text{OH}-3'} = 4.3$ Hz, 0.2H, OH-3'), 6.16 (app t, $J = 7.0$ Hz, 0.8H, H-1'), 6.21 (app t, $J = 7.0$ Hz, 0.2H, H-1'), 7.44 (s, 0.8H, H-6), 7.47 (s, 0.2H, H-6), 8.79 (br s, 0.8H, NH-5'), 8.90 (br s, 0.2H, NH-5'), 11.32 (br s, 1H, NH^T) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): δ = 12.1 (CH_3^{T}), 15.0 (CH_3^{Et}), 15.3 (CH_3^{Et}), 38.4 (C-2'), 43.7 (CH_2^{Et}), 45.0 (C-5'), 45.2 (CH_2^{Et}), 70.7 (C-3'), 84.0 (C-1'), 84.5 (C-4'), 109.8 (C-5), 135.8 (C-6), 150.4 (C-2), 163.7 (C-4), 169.1 (C-Sq1), 169.9 (C-Sq2), 201.7 (C-Sq4), 204.9 (C-Sq3) ppm.

HRMS (APCI $^-$): m/z calc. 423.1166 [M-H] $^-$, found: 423.1162.

5'-O-(4,4'-Dimethoxytrityl)-3'-(2-cyclopentoxy-3,4-dithionecyclobuten-1-yl)-amino-3'-deoxythymidine (59)



59

Dicyclopentyl dithiosquarate **55** (286 mg, 1.01 mmol) was dissolved in anhydrous CH_2Cl_2 (3 mL). The reaction mixture was cooled to 0 °C and stirred under argon. 5'-O-(4,4'-Dimethoxytrityl)-3'-amino-3'-deoxythymidine (**58**) (500 mg, 0.92 mmol) was added and the reaction mixture was stirred for 15 minutes at 0 °C and then for 4 hours at room temperature. After this time, TLC analysis (CH_2Cl_2 -MeOH, 90:10) showed complete consumption of the starting material (R_f = 0.3) and formation of product (R_f = 0.7). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 97:3) to provide the product **59** as a yellow foam (544 mg, 80%).

Notes: in the ^1H NMR spectrum the signal for H-2' is obscured by the residual protonated DMSO- d_6 peak and the signal for H-5'b is obscured by the water peak. Compound **59** exhibits four rotamers in NMR spectroscopy

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3160, 2953, 1676, 1607, 1508, 1444, 1362, 1286, 1248, 1175, 1112, 1031, 960.

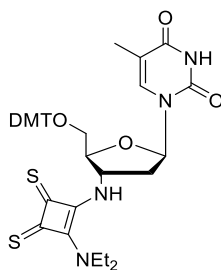
^1H NMR (600 MHz, DMSO- d_6) δ = 1.48-1.69 (m, 6.5H, CH_2^{Cp} , CH_3^{T}), 1.71-2.08 (m, 4.5H, CH_2^{Cp} , CH_3^{T}), 2.34-2.40 (m, 0.15H, H-2'a), 2.44 (app dt, J = 13.6 Hz, J = 6.3 Hz, 0.35H, H-2'a),

2.47-2.55 (m, 1.5 H, H-2'a, H-2'b), 3.25 (dd, $J_{5'a,5'b} = 10.7$ Hz, $J_{4',5'a} = 3.5$ Hz, 0.5H, H-5'a), 3.27-3.28 (m, 0.5H, H-5'a), 3.34-3.38 (m, 0.5H, H-5'b), 3.72 (s, 3H, OCH₃), 3.73 (s, 3H, OCH₃), 4.01-4.05 (m, 0.15H, H-4'), 4.05-4.11 (m, 0.85H, H-4'), 4.41-4.50 (m, 0.7H, H-3'), 5.77-5.85 (m, 0.3H, H-3'), 6.22 (app t, $J = 6.3$ Hz, 0.5H H-1'), 6.24 (app t, $J = 6.3$ Hz, 0.1H, H-1'), 6.27-6.38 (m, 1.3H, CH^{CP}, H-1'), 6.36-6.40 (m, 0.1H, CH^{CP}), 6.82-6.89 (m, 4H, H^{Ar}), 7.06-7.10 (m, 0.5H, H^{Ar}), 7.17-7.25 (m, 4.5H, H^{Ar}), 7.28 (app t, $J = 7.6$ Hz, 2H, H^{Ar}), 7.35 (app t, $J = 8.3$ Hz, 2H, H^{Ar}), 7.49 (s, 0.5H, H-6), 7.57 (s, 0.35H, H-6), 7.74 (s, 0.05H, H-6), 7.79 (s, 0.1H, H-6), 9.98 (d, $J_{3',NH-3'} = 8.5$ Hz, 0.4H, NH-3'), 10.06 (d, $J_{3',NH-3'} = 8.5$ Hz, 0.5H, NH-3'), 10.08 (d, $J_{3',NH-3'} = 8.5$ Hz, 0.1H, NH-3'), 11.32 (s, 0.05H, NH^T), 11.33 (s, 0.1H, NH^T), 11.35 (s, 0.05H, NH^T), 11.38 (s, 0.8H, NH^T) ppm.

¹³C NMR (150 MHz, DMSO-d₆): $\delta = 11.9$ (CH₃^T), 11.9 (CH₃^T), 12.2 (CH₃^T), 23.1 (CH₂^{CP}), 23.2 (CH₂^{CP}), 23.30 (CH₂^{CP}), 23.32 (CH₂^{CP}), 33.4 (CH₂^{CP}), 33.7 (CH₂^{CP}), 33.8 (CH₂^{CP}), 33.9 (CH₂^{CP}), 37.3 (C-2'), 37.7 (C-2'), 53.4 (C-3'), 54.4 (C-3'), 55.0 (OCH₃), 55.1 (OCH₃), 55.3 (C-3'), 62.6 (C-5'), 63.1 (C-5'), 81.6 (C-4'), 82.0 (C-4'), 83.4 (C-1'), 83.8 (C-4'), 84.0 (C-1'), 86.0 (qC^{DMT}), 86.1 (qC^{DMT}), 87.41 (CH^{CP}), 87.43 (CH^{CP}), 109.6 (C-5), 109.8 (C-5), 112.8 (Ar), 113.1 (Ar), 113.17 (Ar), 113.21 (Ar), 126.4 (Ar), 126.8 (Ar), 127.4 (Ar), 127.5 (Ar), 127.6 (Ar), 127.7 (Ar), 127.8 (Ar), 127.9 (Ar), 128.9 (Ar), 129.6 (Ar), 129.7 (Ar), 135.14 (Ar), 135.18 (Ar), 135.25 (Ar), 135.8 (C-6), 136.0 (C-6), 140.2 (Ar), 144.5 (Ar), 144.6 (Ar), 148.3 (Ar), 150.2 (C-2), 150.4 (C-2), 150.5 (C-2), 157.8 (Ar), 158.1 (Ar), 158.2 (Ar), 163.6 (C-4), 163.7 (C-4), 172.7 (C-Sq2), 174.8 (C-Sq2), 182.1 (C-Sq1), 182.4 (C-Sq1), 182.6 (C-Sq1), 206.01 (C-Sq3), 206.04 (C-Sq3), 216.9 (C-Sq4), 217.5 (C-Sq4) ppm.

HRMS (ESI⁺): m/z calc. 762.2278 [M+Na]⁺, found: 762.2257.

5'-O-(4,4'-Dimethoxytrityl)-3'-N-(2-diethylamino-3,4-dithionecyclobut-1-en-1-yl)amino-3'-deoxythymidine (60)



60

5'-O-(4,4'-Dimethoxytrityl)-3'-N-(2-cyclopentyloxy-3,4-dithionecyclobut-1-en-1-yl)amino-3'-deoxythymidine (**59**) (41 mg, 56 μmol) was dissolved in CH_2Cl_2 (2 mL). Diethylamine (87 μL , 841 μmol) was added and the reaction mixture was stirred under argon at room temperature for 2.5 hours. TLC analysis (toluene-acetone, 65:35) after this time showed complete consumption of starting material ($R_f = 0.5$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (toluene-acetone, 65:35) to provide the product **60** as an orange amorphous solid (26 mg, 64%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3200, 2943, 1800, 1665, 1568, 1509, 1441, 1331, 1249, 1203, 1176, 1076, 1026, 826, 728, 698.

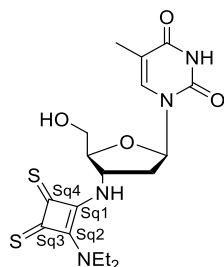
^1H NMR (600 MHz, acetone- d_6): $\delta = 1.22$ - 1.34 (m, 6H, 2 x CH_3^{Et}), 1.65 (d, $J_{6,\text{CH}_3} = 0.7$ Hz, 3H, CH_3^{T}), 2.62- 2.68 (m, 2H, H-2'a, H-2'b), 3.45 (dd, $J_{5'a,5'b} = 10.6$ Hz, $J_{4',5'a} = 3.3$ Hz, 1H, H-5'a), 3.49 (dd, $J_{5'a,5'b} = 10.6$ Hz, $J_{4',5'b} = 3.3$ Hz, 1H, H-5'b), 3.56 (br s, 1H, CH_2^{Et}), 3.65 (br s, 1H, CH_2^{Et}), 3.772 (s, 3H, OCH_3), 3.774 (s, 3H, OCH_3), 4.21 (app dt, $J = 6.7$ Hz, $J = 3.3$ Hz, 1H, H-4'), 4.30 (br s, 1H, CH_2^{Et}), 4.45 (br s, 1H, CH_2^{Et}), 6.32 (app t, $J = 6.4$ Hz, 1H, H-1'), 6.69 (m, 1H, H-3'), 6.82- 6.89 (m, 4H, H^{Ar}), 7.19- 7.25 (m, 2H, H^{Ar}), 7.27- 7.32 (m, 2H, H^{Ar}), 7.34- 7.39 (m, 3H, H^{Ar}), 7.47- 7.54 (m, 2H, H^{Ar}), 7.68 (app d, $J = 0.7$ Hz, 1H, H-6), 7.75 (d, $J_{3',\text{NH-3}'} = 9.1$ Hz, 1H, NH-3'), 9.97 (br s, 1H, NH^{T}) ppm.

^{13}C NMR (151 MHz, acetone- d_6): $\delta = 12.5$ (CH_3^{T}), 15.5 (CH_3^{Et}), 15.8 (CH_3^{Et}), 40.3 (C-2'), 44.9 (CH_2^{Et}), 46.2 (CH_2^{Et}), 54.6 (C-3'), 55.5 (OCH_3), 63.3 (C-5'), 82.9 (C-4'), 83.9 (C-1'), 87.4 (qC^{DMT}), 111.5 (C-5), 113.93 (Ar), 113.94 (Ar), 127.6 (Ar), 128.7 (Ar), 129.1 (Ar), 131.11 (Ar),

131.13 (Ar), 136.5 (C-6), 145.9 (Ar), 151.4 (C-2), 159.66 (Ar), 159.68 (Ar), 164.2 (C-4), 170.3 (C-Sq), 171.1 (C-Sq), 204.4 (C-Sq), 209.2 (C-Sq) ppm.

HRMS (APCI⁻): *m/z* calc. 725.2473 [M-H]⁻, found: 725.2484.

3'-*N*-(2-Diethylamino-3,4-dithionecyclobut-1-en-1-yl)amino-3'-deoxythymidine (**61**)



61

5'-*O*-(4,4'-Dimethoxytrityl)-3'-*N*-(2-diethylamino-3,4-dithionecyclobut-1-en-1-yl)amino-3'-deoxythymidine (**60**) (56 mg, 77 μ mol) was suspended in MeOH (1.5 mL). AcOH (1.5 mL) was added and the reaction mixture was stirred at room temperature for 3.5 hours. TLC analysis (toluene-acetone, 1:1) after this time showed complete consumption of starting material ($R_f = 0.5$) and formation of product ($R_f = 0.1$). The solvent was removed under reduced pressure, co-evaporating with toluene. The residue was purified by flash chromatography (toluene-acetone, 1:1 then 2:3) to provide the product as a yellow amorphous solid (25 mg, 76%).

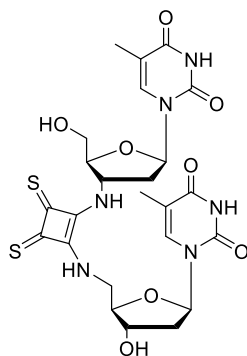
$\nu_{\max}/\text{cm}^{-1}$ (neat) 3191, 3050, 2972, 2927, 1795, 1664, 1537, 1441, 1331, 1267, 1246, 1229, 1120, 1178, 1086, 1024, 962, 803, 753.

$^1\text{H NMR}$ (400 MHz, acetone- d_6): $\delta = 1.29$ (br s, 3H, CH_3^{Et}), 1.37 (br s, 3H, CH_3^{Et}), 1.85 (d, $J_{6,\text{CH}_3} = 1.1$ Hz, 3H, CH_3^{T}), 2.57 (m, 1H, H-2'a), 2.65 (app dt, $J = 14.0$ Hz, $J = 7.0$ Hz, 1H, H-2'b), 3.69 (br s, 2H, CH_2^{Et}), 3.88 (ddd, $J_{5'a,5'b} = 12.1$ Hz, $J_{5'a,\text{OH}-5'} = 5.7$ Hz, $J_{4',5'a} = 2.6$ Hz, 1H, H-5'a), 3.94 (ddd, $J_{5'a,5'b} = 12.1$ Hz, $J_{5'b,\text{OH}-5'} = 5.7$ Hz, $J_{4',5'b} = 3.3$ Hz, 1H, H-5'b), 4.15 (m, 1H, H-4'), 4.27 (app t, $J = 5.7$ Hz, 1H, OH-5'), 4.38 (br s, 2H, CH_2^{Et}), 6.27 (dd, $J_{1',2'b} = 7.0$ Hz, $J_{1',2'a} = 5.2$ Hz, 1H, H-1'), 6.38 (m, 1H, H-3'), 7.84 (d, $J_{3',\text{NH}-3'} = 9.0$ Hz, 1H, NH-3'), 7.87 (app d, $J = 1.1$ Hz, 1H, H-6), 9.96 (br s, 1H, NH^{T}) ppm.

^{13}C NMR (151 MHz, acetone- d_6): δ = 12.7 (CH_3^{T}), 15.6 (CH_3^{Et}), 15.8 (CH_3^{Et}), 40.1 (C-2'), 45.0 (CH_2^{Et}), 46.5 (CH_2^{Et}), 54.2 (C-3'), 61.5 (C-5'), 84.3 (C-1'), 85.6 (C-4'), 111.0 (C-5), 137.0 (C-6), 151.5 (C-2), 164.3 (C-4), 170.6 (C-Sq2), 171.0 (C-Sq1), 204.1 (C-Sq4), 208.7 (C-Sq3) ppm.

HRMS (APCI $^-$): m/z calc. 423.1166 [M-H] $^-$, found: 423.1174.

***N*-(5'-Deoxythymidine-5'-yl)-*N'*-(3'-deoxythymidine-3'-yl)-1,2-diamino-3,4-dithionecyclobut-1-ene (63)**



63

N-(5'-Deoxythymidine-5'-yl)-*N'*-(5'-*O*-(4,4'-dimethoxytrityl)-3'-deoxythymidine-3'-yl)-1,2-diamino-3,4-dithionecyclobut-1-ene (**62**) (31 mg, 52 μmol) was suspended in MeOH (1 mL). AcOH (1 mL) was added and the reaction mixture was stirred at room temperature for 1.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 85:15) after this time showed consumption of starting material (R_f = 0.7) and formation of product (R_f = 0.1). The reaction mixture was diluted with water (5 mL) and EtOAc (5 mL). The organic layer was separated and extracted with water (4 x 5 mL). The aqueous layers were combined, and the solvent was removed under reduced pressure. The residue was purified by flash chromatography (toluene-acetone-MeOH, 4:5:1) to provide the product **63** as a yellow amorphous solid (20 mg, 98%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3223, 2969, 1682, 1560, 1461, 1432, 1320, 1263, 1234, 1199, 1075, 1023, 950, 804.

^1H NMR (600 MHz, DMSO- d_6): δ = 1.788 (s, 3H, CH_3^{T}), 1.794 (s, 3H, CH_3^{T}), 2.07-2.12 (m, 1H, $^3\text{T}\text{H-2}'\text{a}$), 2.20-2.27 (m, 1H, $^3\text{T}\text{H-2}'\text{b}$), 2.36-2.41 (m, 1H, $^5\text{T}\text{H-2}'\text{a}$), 2.43-2.48 (m, 1H, $^5\text{T}\text{H-2}'\text{b}$), 3.69 (br s, 2H, $^5\text{T}\text{H-5}'\text{a}$, $^5\text{T}\text{H-5}'\text{b}$), 3.90 (app dt, J = 7.7 Hz, J = 3.7 Hz, 1H, $^3\text{T}\text{H-4}'$), 3.99-4.03 (m, 1H, $^5\text{T}\text{H-4}'$), 4.20 (dd, $J_{5'\text{a},5'\text{b}}$ = 13.5 Hz, $J_{4',5'\text{a}}$ = 7.7 Hz, 1H, $^3\text{T}\text{H-5}'\text{a}$), 4.25-4.28 (m, 1H, $^3\text{T}\text{H-3}'$), 4.53 (dd, $J_{5'\text{a},5'\text{b}}$ = 13.5 Hz, $J_{4',5'\text{b}}$ = 3.7 Hz, 1H, $^3\text{T}\text{H-5}'\text{b}$), 5.18 (br s, 1H, OH), 5.45 (br s, 1H,

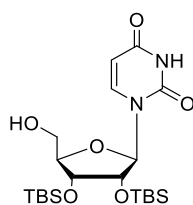
OH), 5.47-5.54 (m, 1H, $^5\text{T H-3}'$), 6.20-6.27 (m, 2H, $^3\text{T H-1}'$, $^5\text{T H-1}'$), 7.51 (s, 1H, $^3\text{T H-6}$), 7.75 (s, 1H, $^5\text{T H-6}$), 8.87 (br s, 1H, NH^{Sq}), 9.41 (br s, 1H, NH^{Sq}), 11.32 (br s, 2H, $^3\text{T NH}^{\text{T}}$, $^5\text{T NH}^{\text{T}}$) ppm.

^{13}C NMR (151 MHz, DMSO-d_6): δ = 12.1 ($^3\text{T CH}_3^{\text{T}}$), 12.3 ($^5\text{T CH}_3^{\text{T}}$), 38.3 ($^3\text{T C-2}'$), 38.5 ($^5\text{T C-2}'$), 44.9 ($^3\text{T C-5}'$), 53.7 ($^5\text{T C-3}'$), 60.7 ($^5\text{T C-5}'$), 70.7 ($^3\text{T C-3}'$), 83.3 ($^5\text{T C-1}'$), 83.8 ($^3\text{T C-1}'$), 84.5 ($^3\text{T C-4}'$), 85.0 ($^5\text{T C-4}'$), 109.6 ($^5\text{T C-5}$), 110.1 ($^3\text{T C-5}$), 136.0 ($^5\text{T C-6}$), 136.1 ($^3\text{T C-6}$), 150.48 ($^3\text{T C-2}$), 150.49 ($^5\text{T C-2}$), 163.68 ($^3\text{T C-4}$), 163.71 ($^5\text{T C-4}$), 170.1 (C-Sq), 170.8 (C-Sq), 204.0 (C-Sq), 204.5 (C-Sq) ppm.

HRMS (APCI): m/z calc. 591.1337 [M-H] $^-$, found: 591.1335.

6.2.3. Synthetic methods for Chapter 4

2',3'-Bis-*O*-(*tert*-butyldimethylsilyl)uridine³⁶⁸ (**67**)



67

5'-*O*-(4,4'-Dimethoxytrityl)-2',3'-bis-*O*-(*tert*-butyldimethylsilyl)uridine (**5**) (406 mg, 0.52 mmol) was dissolved in a mixture of water (0.5 mL) and AcOH (2.1 mL). The reaction was stirred at room temperature for 18.5 hours. MeOH (2 mL) was added, and the reaction mixture was stirred for a further 30 minutes. TLC analysis (CH_2Cl_2 -MeOH, 98:2) showed complete consumption of starting material (R_f = 0.4) and formation of product (R_f = 0.2). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 98:2). The product **67** was obtained as a white foam (218 mg, 88%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3459, 3059, 2929, 2858, 1711, 1691, 1463, 1392, 1258, 1153, 1097, 1073, 956, 868, 832, 772, 681.

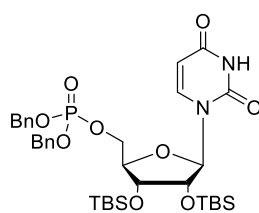
^1H NMR (600 MHz, DMSO-d_6): δ = -0.03 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.02 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.08 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.09 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.83 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 0.89 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 3.56 (ddd, $J_{5'a,5'b}$ = 12.0 Hz, $J_{5',\text{OH-5}'}$ = 4.6 Hz, $J_{4',5'a}$ = 2.9 Hz, 1H, H-5'a), 3.63-3.68 (m, 1H, H-5'b), 3.88 (m, 1H, H-4'), 4.13 (dd, $J_{2',3'}$ = 4.4 Hz, $J_{3',4'}$ = 3.0 Hz, 1H, H-3'), 4.25 (dd, $J_{1',2'}$ = 5.9 Hz, $J_{2',3'}$ = 4.4

Hz, 1H, H-2'), 5.23 (t, $J_{5',OH-5'} = 4.6$ Hz, 1H, OH-5'), 5.69 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.81 (d, $J_{1,2'} = 5.9$ Hz, 1H, H-1'), 7.93 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 11.35 (s, 1H, NH) ppm.

^{13}C NMR (151 MHz, DMSO- d_6): $\delta = -5.0$ ($\text{CH}_3^{\text{TBDMS}}$), -4.9 ($\text{CH}_3^{\text{TBDMS}}$), -4.8 ($\text{CH}_3^{\text{TBDMS}}$), -4.6 ($\text{CH}_3^{\text{TBDMS}}$), 17.6 (qC, $t\text{Bu}^{\text{TBDMS}}$), 17.8 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.6 ($t\text{Bu}^{\text{TBDMS}}$), 25.7 ($t\text{Bu}^{\text{TBDMS}}$), 60.4 (C-5'), 71.9 (C-3'), 74.5 (C-2'), 85.5 (C-4'), 86.8 (C-1'), 102.0 (C-5), 140.4 (C-6), 150.8 (C-2), 163.0 (C-4) ppm.

HRMS (APCI $^+$): m/z calc. 473.2498 [M+H] $^+$, found: 473.2503.

***O,O*-Dibenzyl-*O*-(2',3'-bis-*O*-(*tert*-butyldimethylsilyl)-5'-deoxyuridine-5'-yl)phosphate
(68)**



68

2',3'-Bis-*O*-(*tert*-butyldimethylsilyl)uridine (**67**) (90 mg, 190 μmol) was suspended in a solution of tetrazole (3-4 wt%) in MeCN (1.67 mL). Anhydrous CH_2Cl_2 (1 mL) was added. Dibenzyl *N,N*-diisopropylphosphoramidite (96 μL , 290 μmol) was added at 0 $^\circ\text{C}$. After 5 minutes, the reaction mixture was allowed to warm to room temperature and stirred under argon. After 4 hours TLC analysis (CH_2Cl_2 -MeOH, 96:4) showed formation of intermediate ($R_f = 0.9$) but incomplete consumption of starting material ($R_f = 0.5$). Another portion of *N,N*-diisopropylphosphoramidite (26 μL , 77 μmol) was added. The reaction mixture was stirred under argon for a further 1 hour, and *m*-chloroperoxybenzoic acid (99 mg, 77% purity) was then added. The reaction mixture was stirred for a further 1 hour, after which time a further portion of *m*-chloroperoxybenzoic acid (129 mg, 77% purity) was added. After another hour TLC analysis (CH_2Cl_2 -MeOH, 96:4) showed consumption of intermediate ($R_f = 0.9$) and formation of product ($R_f = 0.5$). The solvent was removed under reduced pressure. The residue was redissolved in CH_2Cl_2 (5 mL) and washed with saturated NaHCO_3 solution (5 mL) followed by brine (5 mL). The organic layer was dried over MgSO_4 and filtered. The solvent was removed under reduced pressure and the residue was purified by

flash chromatography (CH₂Cl₂, then CH₂Cl₂-MeOH, 98:2). The product **68** was obtained as a yellow oil (113 mg, 81%).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3445, 3036, 2952, 2930, 2857, 1691, 1458, 1380, 1252, 1166, 1129, 1072, 998, 866, 836, 777, 738, 697.

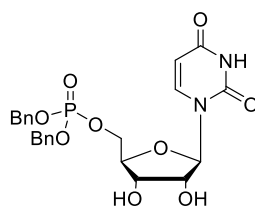
¹H NMR (600 MHz, DMSO-d₆): δ = -0.05 (s, 3H, CH₃^{TBDMS}), 0.00 (s, 3H, CH₃^{TBDMS}), 0.049 (s, 3H, CH₃^{TBDMS}), 0.053 (s, 3H, CH₃^{TBDMS}), 0.81 (s, 9H, tBu^{TBDMS}), 0.86 (s, 9H, tBu^{TBDMS}), 4.01-4.04 (m, 1H, H-4'), 4.08-4.11 (m, 1H, H-3'), 4.13-4.18 (m, 1H, H-5'a), 4.26 (dd, $J_{5'a,5'b}$ = 11.0 Hz, $J_{4',5'b}$ = 6.1 Hz, 1H, H-5'b), 4.28-4.31 (m, 1H, H-2'), 5.05 (d, J_{P,CH_2} = 3.5 Hz, 2H, CH₂^{Bn}) 5.06 (d, J_{P,CH_2} = 3.5 Hz, 2H, CH₂^{Bn}), 5.53 (dd, $J_{5,6}$ = 8.1 Hz, $J_{5,NH-U}$ = 1.8 Hz, 1H, H-5), 5.79 (d, $J_{1',2'}$ = 6.0 Hz, 1H, H-1'), 7.35-7.39 (m, 9H, H^{Ar}), 7.65 (d, $J_{5,6}$ = 8.1 Hz, 1H, H-6), 7.88-7.90 (m, 1H, H^{Ar}) ppm.

¹³C NMR (151 MHz, DMSO-d₆): δ = -5.09 (CH₃^{TBDMS}), -5.06 (CH₃^{TBDMS}), -4.8 (CH₃^{TBDMS}), -4.7 (CH₃^{TBDMS}), 17.6 (qC, tBu^{TBDMS}), 17.7 (qC, tBu^{TBDMS}), 25.6 (tBu^{TBDMS}), 25.7 (tBu^{TBDMS}), 66.25 (d, $J_{P,5'}$ = 5.1 Hz, C-5'), 68.76 (d, J_{P,CH_2} = 5.4 Hz, CH₂^{Bn}), 68.78 (d, J_{P,CH_2} = 5.4 Hz, CH₂^{Bn}), 71.3 (C-3'), 73.5 (C-2'), 82.6 (d, $J_{P,4'}$ = 7.6 Hz, C-4'), 87.5 (C-1'), 102.1 (C-5), 127.80 (Ar), 127.82 (Ar), 127.9 (Ar), 128.46 (Ar), 128.48 (Ar), 128.49 (Ar), 128.50 (Ar), 128.8 (Ar), 130.6 (Ar), 132.6 (Ar), 135.8 (Ar), 135.9 (Ar), 140.3 (C-6), 150.7 (C-2), 162.8 (C-4) ppm.

³¹P NMR (162 MHz, DMSO-d₆): δ = -0.87 ppm.

HRMS (APCI⁺): m/z calc. 733.3010 [M+H]⁺, found: 733.3116.

***O,O*-Dibenzyl-*O*-(5'-deoxyuridine-5'-yl)phosphate (**69**)**



69

O,O-Dibenzyl-*O*-(2',3'-bis-*O*-(*tert*-butyldimethylsilyl)-5'-deoxyuridine-5'-yl)phosphate (**68**) (97 mg, 130 μ mol) was dissolved in THF (1.3 mL). Tetrabutylammonium fluoride trihydrate (108 mg, 340 μ mol) was added and the reaction mixture was stirred at room temperature for 5.5 hours. TLC analysis (CH₂Cl₂-MeOH, 93:7) after this time showed complete

consumption of starting material ($R_f = 0.8$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 94:6) to give the product **69** as a white amorphous solid (33 mg, 50%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3188, 3064, 2924, 2851, 1677, 1498, 1456, 1385, 1249, 1108, 992, 874, 808, 735, 695.

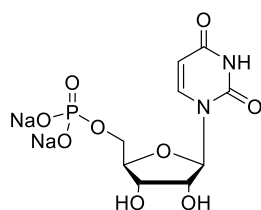
^1H NMR (400 MHz, acetone- d_6): $\delta = 4.14$ (m, 1H, H-4'), 4.20-4.36 (m, 4H, H-2', H-3', H-5'a, H-5'b), 4.51 (br s, 1H, OH-3'), 4.75 (br s, 1H, OH-2'), 5.11 (d, $J_{\text{P,CH}_2} = 8.2$ Hz, 4H, 2 x CH_2^{Bn}), 5.50 (d, $J_{5,6} = 8.0$ Hz, 1H, H-5), 5.91 (d, $J_{1',2'} = 4.0$ Hz, 1H, H-1'), 7.30-7.46 (m, 10H, H^{Ar}), 7.65 (d, $J_{5,6} = 8.0$ Hz, 1H, H-6), 10.04 (br s, 1H, NH) ppm.

^{13}C NMR (100 MHz, acetone- d_6): $\delta = 67.6$ (d, $J_{\text{P},5'} = 5.6$ Hz, C-5'), 69.9 (d, $J_{\text{P,CH}_2} = 2.2$ Hz, CH_2^{Bn}), 70.0 (d, $J_{\text{P,CH}_2} = 2.2$ Hz, 2 x CH_2^{Bn}), 70.9 (C-3'), 74.7 (C-2'), 83.2 (d, $J_{\text{P},4'} = 7.7$ Hz, C-4'), 90.4 (C-1'), 102.9 (C-5), 128.8 (d, $J = 1.7$ Hz, Ar), 129.3 (d, $J = 1.4$ Hz, Ar), 129.4 (d, $J = 0.7$ Hz, Ar), 137.2 (d, $J = 0.8$ Hz, Ar), 137.3 (d, $J = 0.9$ Hz, Ar), 141.1 (C-6), 151.5 (C-2), 163.4 (C-4) ppm.

^{31}P NMR (162 MHz, acetone- d_6): $\delta = -0.74$ ppm.

HRMS (APCI $^+$): m/z calc. 505.1370 $[\text{M}+\text{H}]^+$, found: 505.1368.

Uridine 5'-monophosphate disodium salt (**70**)



70

O,O-Dibenzyl-*O*-(5'-deoxyuridine-5'-yl)phosphate (**69**) (54 mg, 108 μmol) was dissolved in MeOH (1.2 mL). The reaction vessel was purged with argon, and 10% Pd/C (11 mg) was added. The reaction vessel was purged with H_2 . The reaction mixture was stirred under an atmosphere of H_2 for 5.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 93:7) after this time showed complete consumption of starting material ($R_f = 0.3$) and formation of a new product ($R_f = 0.0$). The reaction mixture was filtered through celite, washing through with MeOH,

and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure to provide the product **70** as a white amorphous solid (37 mg, 94%).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3060, 2927, 1676, 1639, 1597, 1548, 1452, 1411, 1068, 975, 844, 820, 790, 765, 706, 680.

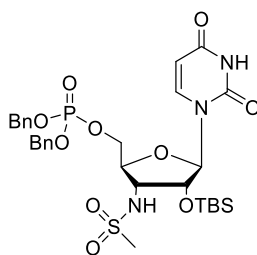
^1H NMR (400 MHz, acetone- d_6): δ = 3.92 (ddd, $J_{5'a,5'b}$ = 11.8 Hz, $J_{4',5'a}$ = 5.0 Hz, $J_{p,5'a}$ = 3.3 Hz, 1H, H-5'a), 3.97 (app dt, J = 11.8 Hz, J = 3.3 Hz, 1H, H-5'b), 4.18-4.23 (m, 1H, H-4'), 4.31 (app t, J = 5.0 Hz, 1H, H-3'), 4.37 (app t, J = 5.0 Hz, 1H, H-2'), 5.91 (d, $J_{5,6}$ = 7.9 Hz, 1H, H-5), 5.98 (d, $J_{1',2'}$ = 5.0 Hz, 1H, H-1'), 8.00 (d, $J_{5,6}$ = 7.9 Hz, 1H, H-6) ppm.

^{13}C NMR (100 MHz, acetone- d_6): δ = 63.8 (d, $J_{p,5'}$ = 4.2 Hz, C-5'), 70.8 (C-3'), 74.5 (C-2'), 84.6 (d, $J_{p,4'}$ = 8.8 Hz, C-4'), 89.0 (C-1'), 103.5 (C-5), 142.2 (C-6), 155.9 (C-2), 164.5 (C-4) ppm.

^{31}P NMR (162 MHz, acetone- d_6): δ = 3.76 ppm.

HRMS (APCI⁺): m/z calc. 390.9890 [M+Na]⁺, found: 390.9882.

***O,O*-Dibenzyl-*O*-(2'-*O*-(*tert*-butyldimethylsilyl)-3'-*N*-(methanesulfonyl)amino-3'-deoxyuridine-5'-yl)phosphate (**71**)**



71

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-*N*-(methanesulfonyl)amino-3'-deoxyuridine (**34**) (101 mg, 232 μmol) was dissolved in CH_2Cl_2 (2 mL). A solution of tetrazole (3-4 wt%) in MeCN (2 mL) was added. The reaction mixture was cooled to 0 °C and dibenzyl *N,N*-diisopropylphosphoramidite (200 μL , 593 μmol) was added. After ten minutes, the reaction mixture was allowed to warm to room temperature. The reaction mixture was stirred for a further 2.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 94:6) after this time showed consumption of starting material (R_f = 0.3) and formation of intermediate (R_f = 0.7). The reaction mixture was cooled to 0 °C and *m*-chloroperoxybenzoic acid (152 mg, 70-75%

purity) was added. The reaction mixture was stirred for a further 30 minutes, after which time TLC analysis (CH₂Cl₂-MeOH, 96:4) showed consumption of intermediate (R_f = 0.5) and formation of product (R_f = 0.3). The reaction was quenched by addition of 10% (w/v) sodium thiosulfate solution (20 mL). The reaction mixture was diluted with CH₂Cl₂ (15 mL). The organic layer was separated and washed with saturated NaHCO₃ solution (15 mL). The organic layer was dried over MgSO₄ and filtered. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 96:4) to provide the product **71** as a colourless oil (145 mg, 90%).

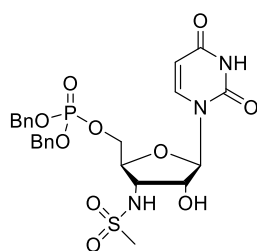
¹H NMR (400 MHz, acetone-d₆): δ = 0.21 (s, 3H, CH₃^{TBDMS}), 0.22 (s, 3H, CH₃^{TBDMS}), 0.95 (s, 9H, tBu^{TBDMS}), 3.01 (s, 3H, CH₃^{Mesyl}), 4.10-4.17 (m, 1H, H-3'), 4.29 (ddd, J_{3',4'} = 8.5 Hz, J_{4',5'a} = 4.0 Hz, J_{4',5'b} = 2.2 Hz, 1H, H-4'), 4.35 (ddd, J_{5'a,5'b} = 11.7 Hz, J_{p,5'a} = 5.7 Hz, J_{4',5'a} = 4.0 Hz, 1H, H-5'a), 4.49 (ddd, J_{5'a,5'b} = 11.7 Hz, J_{p,5'b} = 5.7 Hz, J_{4',5'b} = 2.2 Hz, 1H, H-5'b), 4.58 (dd, J_{2',3'} = 4.9 Hz, J_{1',2'} = 1.8 Hz, 1H, H-2'), 5.10-5.16 (m, 4H, 2 x CH₂^{Bn}), 5.48 (d, J_{5,6} = 8.1 Hz, 1H, H-5), 5.78 (d, J_{1',2'} = 1.8 Hz, 1H, H-1'), 5.97 (d, J_{NH-3',3'} = 8.9 Hz, 1H, NH-3'), 7.32-7.46 (m, 10H, H^{Ar}), 7.79 (d, J_{5,6} = 8.1 Hz, 1H, H-6), 10.08 (br s, 1H, NH^U) ppm.

¹³C NMR (100 MHz, acetone-d₆): δ = -4.7 (CH₃^{TBDMS}), -4.4 (CH₃^{TBDMS}), 18.7 (qC, tBu^{TBDMS}), 26.3 (tBu^{TBDMS}), 41.7 (CH₃^{Mesyl}), 54.5 (C-3'), 66.1 (d, J_{p,5'} = 5.2 Hz, C-5'), 70.0 (d, J_{p,CH2} = 5.5 Hz, 2 x CH₂^{Bn}), 76.0 (C-2'), 81.1 (d, J_{p,4'} = 8.2 Hz, H-4'), 92.3 (C-1'), 102.6 (C-5), 128.9 (Ar), 129.3 (d, J_{p,Ar} = 2.6 Hz, Ar), 129.4 (Ar), 137.2 (d, J_{p,Ar} = 7.0 Hz, Ar), 140.8 (C-6), 151.3 (C-2), 163.4 (C-4) ppm.

³¹P NMR (162 MHz, acetone-d₆): δ = -0.84 ppm.

HRMS (APCI⁺): *m/z* calc. 696.2171 [M+H]⁺, found: 696.2178.

***O,O*-Dibenzyl-*O*-(3'-*N*-(methanesulfonyl)amino-3'-deoxyuridine-5'-yl)phosphate (**72**)**



72

O,O-Dibenzyl-*O*-(2'-*O*-(*tert*-butyldimethylsilyl)-3'-*N*-(methanesulfonyl)amino-3'-deoxyuridine-5'-yl)phosphate (**71**) (122 mg, 175 μmol) was dissolved in THF (1.6 mL). Tetrabutylammonium fluoride trihydrate (82 mg, 253 μmol) was added. The reaction mixture was stirred at room temperature for 1 hour. TLC analysis (CH_2Cl_2 -MeOH, 94:6) after this time showed complete consumption of starting material ($R_f = 0.5$) and formation of product ($R_f = 0.2$). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 94:6) to provide the product **72** as a white amorphous solid (81 mg, 79%).

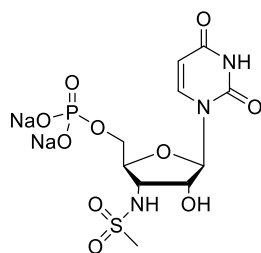
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3234, 2921, 2544, 2161, 2035, 1983, 1681, 1456, 1390, 1325, 1109, 989, 737, 696.

^1H NMR (400 MHz, acetone- d_6): $\delta = 3.03$ (s, 3H, $\text{CH}_3^{\text{Mesyl}}$), 4.11-4.17 (m, 1H, H-3'), 4.18-4.23 (m, 1H, H-4'), 4.36 (ddd, $J_{5'a,5'b} = 11.7$ Hz, $J_{p,5'a} = 5.7$ Hz, $J_{4',5'a} = 4.3$ Hz, 1H, H-5'a), 4.48 (ddd, $J_{5'a,5'b} = 11.7$ Hz, $J_{p,5'b} = 5.7$ Hz, $J_{4',5'b} = 1.9$ Hz, 1H, H-5'b), 4.56 (dd, $J_{2',3'} = 5.7$ Hz, $J_{1',2'} = 1.6$ Hz, 1H, H-2'), 5.10-5.14 (m, 4H, 2 x CH_2^{Bn}), 5.52 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.71 (br s, 1H, OH-2'), 5.83 (d, $J_{1',2'} = 1.6$ Hz, 1H, H-1'), 6.33 (br s, 1H, NH-3'), 7.31-7.45 (m, 10H, H^{Ar}), 7.75 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 10.22 (br s, 1H, NH^{U}) ppm.

^{13}C NMR (100 MHz, acetone- d_6): $\delta = 41.3$ ($\text{CH}_3^{\text{Mesyl}}$), 54.5 (C-3'), 66.5 (d, $J_{p,5'} = 5.5$ Hz, C-5'), 70.0 (d, $J_{p,\text{CH}_2} = 5.9$ Hz, 2 x CH_2^{Bn}), 73.8 (C-2'), 81.5 (d, $J_{p,4'} = 8.0$ Hz, C-4'), 93.0 (C-1'), 102.5 (C-5), 128.9 (d, $J_{p,\text{Ar}} = 2.2$ Hz, Ar), 129.3 (d, $J_{p,\text{Ar}} = 3.4$ Hz, Ar), 129.4 (Ar), 137.1 (d, $J_{p,\text{Ar}} = 4.8$ Hz, Ar), 141.3 (C-6), 151.3 (C-2), 163.8 (C-4) ppm.

^{31}P NMR (162 MHz, acetone- d_6): $\delta = -0.99$ ppm.

***O*-(3'-*N*-(Methanesulfonyl)amino-3'-deoxyuridine-5'-yl)phosphate disodium salt (**73**)**



73

O,O-Dibenzyl-*O*-(3'-*N*-(methanesulfonyl)amino-3'-deoxyuridine-5'-yl)phosphate (**72**) (69 mg, 119 μ mol) was dissolved in MeOH (5 mL). The reaction vessel was purged with argon, and 10% Pd/C (18 mg) was added. The reaction vessel was purged with H₂. The reaction mixture was stirred under an atmosphere of H₂ for 3 hours. TLC analysis (CH₂Cl₂-MeOH, 88:12) after this time showed complete consumption of starting material ($R_f = 0.6$) and formation of product ($R_f = 0.0$). The reaction mixture was filtered through celite, washing through with MeOH, and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure to provide the product **73** as a white amorphous solid (20 mg, 37%).

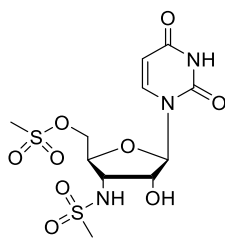
¹H NMR (600 MHz, D₂O): $\delta = 3.19$ (s, 3H, CH₃^{Mesyl}), 4.04-4.08 (m, 1H, H-5'a), 4.09 (dd, $J_{3',4'} = 8.5$ Hz, $J_{2',3'} = 5.5$ Hz, 1H, H-3'), 4.26-4.30 (m, 2H, H-4', H-5'b), 4.46 (dd, $J_{1',2'} = 2.3$ Hz, $J_{2',3'} = 5.5$ Hz, 1H, H-2'), 5.87 (d, $J_{1',2'} = 2.3$ Hz, 1H, H-1'), 5.96 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 8.11 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6) ppm.

¹³C NMR (100 MHz, D₂O): $\delta = 40.5$ (CH₃^{Mesyl}), 53.1 (C-3'), 62.3 (d, $J_{p,5'} = 4.2$ Hz, C-5'), 73.3 (C-2'), 81.2 (d, $J_{p,4'} = 9.2$ Hz, C-4'), 90.7 (C-1'), 101.9 (C-5), 141.9 (C-6), 151.5 (C-2), 166.4 (C-4) ppm.

³¹P NMR (162 MHz, D₂O): $\delta = 1.42$ ppm.

HRMS (ESI⁺): m/z calc. 467.9825 [M+Na]⁺, found: 467.9828.

3'-*N*-(Methanesulfonyl)amino-5'-*O*-methanesulfonyl-3'-deoxyuridine (**74**)



74

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-*N*-(methanesulfonyl)amino-5'-*O*-methanesulfonyl-3'-deoxyuridine (**32**) (26 mg, 50.6 μ mol) was dissolved in THF (1 mL). Tetrabutylammonium fluoride trihydrate (25 mg, 79.2 μ mol) was added and the reaction mixture was stirred at room temperature for 1 hour. TLC analysis (CH_2Cl_2 -MeOH, 94:6) after this time showed consumption of starting material ($R_f = 0.4$) and formation of product ($R_f = 0.1$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 92:8) to provide the product **74** as a white amorphous solid (12 mg, 60%).

Note: compound **74** exhibits rotamers in NMR spectroscopy.

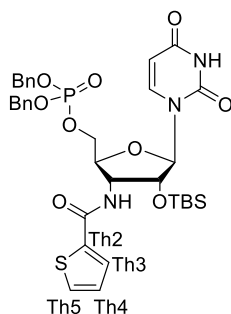
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3259, 3026, 2928, 1668, 1456, 1410, 1322, 1270, 1214, 1172, 1096, 970, 891, 812.

^1H NMR (400 MHz, acetone- d_6): $\delta = 3.05$ (s, 2.25H, $\text{CH}_3^{\text{Mesyl}}$), 3.05 (s, 0.75H, $\text{CH}_3^{\text{Mesyl}}$), 3.16 (s, 0.75H, $\text{CH}_3^{\text{Mesyl}}$), 3.18 (s, 2.25H, $\text{CH}_3^{\text{Mesyl}}$), 3.99-4.04 (m, 0.25H, H-3'), 4.06-4.10 (m, 0.25H, H-4'), 4.12-4.17 (m, 0.75H, H-3'), 4.26 (ddd, $J_{3',4'} = 8.7$ Hz, $J_{4',5'a} = 4.7$ Hz, $J_{4',5'b} = 2.0$ Hz, 0.75H, H-4'), 4.42 (dd, $J_{5'a,5'b} = 11.5$ Hz, $J_{4',5'a} = 4.7$ Hz, 0.25H, H-5'a), 4.51 (dd, $J_{5'a,5'b} = 11.5$ Hz, $J_{4',5'a} = 4.7$ Hz, 0.75H, H-5'a), 4.55 (br s, 0.25H, H-2'), 4.58 (dd, $J_{5'a,5'b} = 11.5$ Hz, $J_{4',5'b} = 2.0$ Hz, 0.25H, H-5'b), 4.60 (br s, 0.75H, H-2'), 4.66 (dd, $J_{5'a,5'b} = 11.5$ Hz, $J_{4',5'b} = 2.0$ Hz, 0.75H, H-5'b), 5.36 (br s, 0.25H, OH-2'), 5.61 (d, $J_{5,6} = 8.1$ Hz, 0.75H, H-5), 5.64 (br s, 0.75H, OH-2'), 5.66 (d, $J_{5,6} = 8.1$ Hz, 0.25H, H-5), 5.78 (d, $J_{1',2'} = 1.2$ Hz, 0.25H, H-1'), 5.82 (d, $J_{1',2'} = 1.9$ Hz, 0.75H, H-1'), 6.24 (d, $J_{\text{NH-3}',3'} = 7.9$ Hz, 0.25H, NH-3'), 6.30 (d, $J_{\text{NH-3}',3'} = 7.9$ Hz, 0.75H, NH-3'), 7.73 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 9.09 (br s, 0.25H, NH^{U}), 10.07 (br s, 0.75H, NH^{U}) ppm.

^{13}C NMR (100 MHz, acetone- d_6): $\delta = 37.4$ ($\text{CH}_3^{\text{Mesyl}}$), 41.2 ($\text{CH}_3^{\text{Mesyl}}$), 54.7 (C-3'), 68.9 (C-5'), 73.7 (C-2'), 80.7 (C-4'), 93.4 (C-1'), 102.6 (C-5), 141.4 (C-6), 151.3 (C-2), 163.5 (C-4) ppm.

HRMS (APCI): m/z calc. 398.0333 $[M-H]^-$, found: 398.0335.

***O,O*-Dibenzyl-*O*-(2'-*O*-(*tert*-butyldimethylsilyl)-3'-(thiophene-2-carboxamido)-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (75)**



75

N-(2'-*O*-(*tert*-Butyldimethylsilyl)-3'-deoxyuridine-3'-yl)thiophene-2-carboxamide (**30**) (36 mg, 77 μ mol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL). A solution of tetrazole (3-4 wt%) in MeCN (0.67 mL) was added. The reaction mixture was cooled to 0 $^\circ\text{C}$ and dibenzyl *N,N*-diisopropylphosphoramidite (90 μ L, 267 μ mol) was added. After 10 minutes the reaction mixture was allowed to warm to room temperature and stirred under argon for 2.5 hours. TLC analysis (petroleum ether-EtOAc, 1:2) after this time showed complete consumption of starting material ($R_f = 0.3$) and formation of intermediate ($R_f = 0.8$). The reaction mixture was cooled to 0 $^\circ\text{C}$ and *m*-chloroperoxybenzoic acid (46 mg, 70-75% purity) was added. TLC analysis (petroleum ether-EtOAc, 1:2) after 15 minutes showed consumption of intermediate ($R_f = 0.8$) and formation of product ($R_f = 0.6$). The solvent was removed under reduced pressure. The residue was redissolved in CH_2Cl_2 (5 mL) and washed with saturated NaHCO_3 solution (5 mL) and brine (5 mL). The organic layer was dried over MgSO_4 and filtered. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (petroleum ether-EtOAc, 2:3) to give the product **75** as a white foam (19 mg, 34%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3066, 2954, 2928, 2856, 1680, 1650, 1544, 1456, 1421, 1377, 1261, 1120, 1090, 997, 983, 833, 779, 735, 695.

^1H NMR (400 MHz, acetone- d_6): $\delta = 0.03$ (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.13 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.86 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 4.31-4.38 (m, 1H, H-5'a), 4.39-4.48 (m, 2H, H-4', H-5'b), 4.58 (dd, $J_{2',3'} = 6.2$ Hz, $J_{1',2'} = 3.1$ Hz, 1H, H-2'), 4.72 (app td, $J = 6.2$ Hz, $J = 8.2$ Hz, 1H, H-3'), 5.09-5.12 (m, 4H, 2

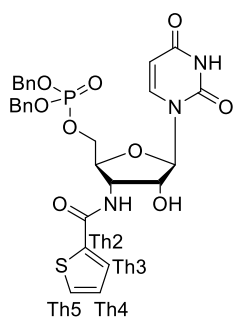
x CH₂^{Bn}), 5.55 (dd, J_{5,6} = 8.1 Hz, J_{5,NH-U} = 1.9 Hz, 1H, H-5), 5.92 (d, J_{1',2'} = 3.1 Hz, 1H, H-1'), 7.15 (dd, J_{Th4,Th5} = 5.0 Hz, J_{Th3,Th4} = 3.8 Hz, 1H, H-Th4), 7.28-7.44 (m, 10H, H^{Ar}), 7.68 (d, J_{3',NH-3'} = 8.2 Hz, 1H, NH-3'), 7.72 (dd, J_{Th4,Th5} = 5.0 Hz, J_{Th3,Th5} = 1.0 Hz, 1H, H-Th5), 7.76 (dd, J_{Th3,Th4} = 3.8 Hz, J_{Th3,Th5} = 1.0 Hz, 1H, H-Th3), 7.81 (d, J_{5,6} = 8.1 Hz, 1H, H-6), 10.15 (br s, 1H, NH^U) ppm.

¹³C NMR (100 MHz, acetone-d₆): δ = -4.9 (CH₃^{TBDMS}), -4.8 (CH₃^{TBDMS}), 18.5 (qC, tBu^{TBDMS}), 26.1 (tBu^{TBDMS}), 51.7 (C-3'), 67.2 (d, J_{P,5'} = 5.4 Hz, C-5'), 69.88 (d, J_{P,CH2} = 2.5 Hz, CH₂^{Bn}), 69.94 (d, J_{P,CH2} = 2.5 Hz, CH₂^{Bn}), 75.6 (C-2'), 81.2 (d, J = 7.9 Hz, C-4'), 91.6 (C-1'), 103.0 (C-5), 128.5 (C-Th4), 128.8 (Ar), 129.18 (C-Th3), 129.21 (Ar), 129.25 (Ar), 129.34 (Ar), 129.35 (Ar), 129.37 (Ar), 131.7 (C-Th5), 137.2 (d, J_{31P,qC} = 7.0 Hz, qC-Bn), 140.2 (C-Th2), 140.7 (C-6), 151.4 (C-2), 162.3 (CO^{Amide}), 163.4 (C-4) ppm.

³¹P NMR (162 MHz, acetone-d₆): δ = -1.03 ppm.

HRMS (APCI⁺): *m/z* calc. 728.2221 [M+H]⁺, found: 728.2229.

***O,O*-Dibenzyl-*O*-(3'-(thiophene-2-carboxamido)-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (**76**)**



76

O,O-Dibenzyl-*O*-(2'-*O*-(*tert*-butyldimethylsilyl)-3'-(thiophene-2-carboxamido)-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (**75**) (32 mg, 44 μmol) was dissolved in THF (1 mL). Tetrabutylammonium fluoride trihydrate (21 mg, 65 μmol) was added and the reaction mixture was stirred at room temperature for 1.5 hours. TLC analysis (CH₂Cl₂-MeOH, 95:5) after this time showed consumption of starting material (*R*_f = 0.3) and formation of product (*R*_f = 0.1). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 95:5). The product **76** was obtained as a white amorphous solid (16 mg, 58%).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3332, 2970, 1666, 1467, 1408, 1379, 1309, 1160, 1128, 1107, 950, 816.

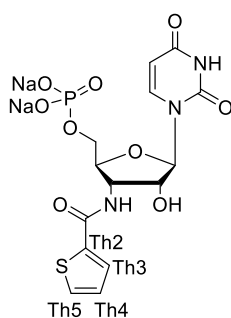
$^1\text{H NMR}$ (600 MHz, CDCl_3): δ = 4.21-4.26 (m, 1H, H-5'a), 4.27-4.30 (m, 1H, H-4'), 4.32 (d, $J_{2',3'}$ = 5.1 Hz, 1H, H-2'), 4.37 (dd, $J_{5'a,5'b}$ = 10.6 Hz, $J_{P,5'b}$ = 5.1 Hz, 1H, H-5'b), 4.48-4.52 (m, 1H, H-3'), 4.95-5.08 (m, 4H, 2 x CH_2^{Bn}), 5.53 (d, $J_{5,6}$ = 8.1 Hz, H-5), 5.76 (s, 1H, H-1'), 6.97-6.99 (m, 1H, H-Th4), 7.19 (d, $J_{3',\text{NH-3}'}$ = 7.7 Hz, 1H, NH-3'), 7.20 (br s, 1H, OH-2'), 7.22-7.29 (m, 10H, H^{Ar}), 7.41 (d, $J_{\text{Th4,Th5}}$ = 4.8 Hz, 1H, H-Th5), 7.52 (d, $J_{\text{Th3,Th4}}$ = 3.3 Hz, 1H, H-Th3), 7.77 (d, $J_{5,6}$ = 8.1 Hz, 1H, H-6), 10.31 (br s, 1H, NH^{U}) ppm.

$^{13}\text{C NMR}$ (151 MHz, CDCl_3): δ = 49.8 (C-3'), 66.2 (d, $J_{P,5'}$ = 4.8 Hz, C-5'), 69.81 (d, J_{P,CH_2} = 3.1 Hz, CH_2^{Bn}), 69.84 (d, J_{P,CH_2} = 3.1 Hz, CH_2^{Bn}), 74.8 (C-2'), 81.6 (d, $J_{P,4'}$ = 7.7 Hz, C-4'), 92.2 (C-1'), 102.4 (C-5), 127.9 (Ar), 128.2 (d, $J_{P,\text{Ar}}$ = 3.3 Hz, Ar), 128.77 (d, $J_{P,\text{Ar}}$ = 4.0 Hz, Ar), 128.80 (C-Th4), 128.9 (Ar), 129.0 (C-Th3), 131.0 (C-Th5), 135.65 (d, $J_{P,\text{Ar}}$ = 6.9 Hz, Ar), 135.69 (d, $J_{P,\text{Ar}}$ = 6.8 Hz, Ar), 138.2 (C-Th2), 140.1 (C-6), 151.0 (C-2), 162.3 (CO^{Amide}), 164.2 (C-4) ppm.

$^{31}\text{P NMR}$ (162 MHz, CDCl_3): δ = -1.15 ppm.

HRMS (ESI⁺): m/z calc. 636.1176 $[\text{M}+\text{Na}]^+$, found: 636.1175.

***O*-(3'-(Thiophene-2-carboxamido)-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate disodium salt (77)**



77

O,O-Dibenzyl-*O*-(3'-(thiophene-2-carboxamido)-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (76) (13 mg, 21 μmol) was dissolved in MeOH (1.6 mL). The reaction vessel was purged with argon, and 10% Pd/C (8 mg) was added. The reaction vessel was purged with H_2 . The reaction mixture was stirred under an atmosphere of H_2 for 4.5 hours. TLC analysis (CH_2Cl_2 -EtOH, 90:10) after this time showed complete consumption of starting material (R_f = 0.5) and formation of product (R_f = 0.0). The reaction mixture was filtered through celite,

washing through with MeOH, and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure to obtain the product **77** as a white amorphous solid (8 mg, 78%).

Note: compound **77** exhibits rotamers in NMR spectroscopy.

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3247, 1659, 1535, 1504, 1420, 1357, 1270, 1088, 975, 920, 861, 813, 718.

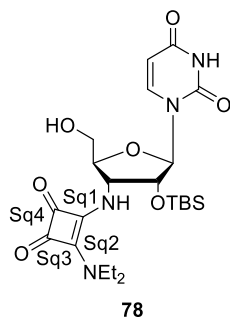
^1H NMR (400 MHz, D_2O): δ = 3.93-4.03 (m, 1H, H-5'a), 4.16-4.29 (m, 1H, H-5'b), 4.41-4.65 (m, 3H, H-2', H-3', H-4'), 5.83 (d, $J_{5,6}$ = 8.1 Hz, 0.3H, H-5), 5.88 (d, $J_{1',2'}$ = 1.6 Hz, 0.3H, H-1'), 5.98 (d, $J_{1',2'}$ = 3.0 Hz, 0.7H, H-1'), 5.99 (d, $J_{5,6}$ = 8.1 Hz, 0.7H, H-5), 7.19-7.23 (m, 1H, H-Th4), 7.73-7.77 (m, 1.3H, 1 x H-Th5, 0.3 x H-Th3), 7.81 (dd, $J_{\text{Th3,Th4}}$ = 3.7 Hz, $J_{\text{Th3,Th5}}$ = 0.7 Hz, 0.7H, H-Th3), 7.95 (d, $J_{5,6}$ = 8.1 Hz, 0.3H, H-6), 8.17 (d, $J_{5,6}$ = 8.1 Hz, 0.7H, H-6) ppm.

^{13}C NMR (100 MHz, D_2O): δ = 48.8 (C-3'), 51.3 (C-3'), 63.1 (d, $J_{\text{P},5'}$ = 4.3 Hz, C-5') 63.3 (d, $J_{\text{P},5'}$ = 4.9 Hz, C-5'), 73.4 (C-2'), 73.5 (C-2'), 79.9 (d, $J_{\text{P},4'}$ = 9.0 Hz, C-4'), 81.2 (d, $J_{\text{P},4'}$ = 8.6 Hz, C-4'), 90.6 (C-1'), 91.3 (C-1'), 101.9 (C-5), 102.2 (C-5), 128.16 (C-Th4), 128.21 (C-Th4), 130.3 (C-Th3), 130.4 (C-Th3), 131.9 (C-Th5), 132.0 (C-Th5), 136.5 (C-Th2), 136.6 (C-Th2), 141.6 (C-6), 142.1 (C-6), 151.4 (C-2), 151.6 (C-2), 164.7 (CO^{Amide}), 164.9 (CO^{Amide}), 166.3 (C-4), 166.4 (C-4) ppm.

^{31}P NMR (162 MHz, D_2O): δ = 0.07, 2.79 ppm.

HRMS (ESI⁺): m/z calc. 478.0057 [$\text{M}+\text{H}$]⁺, found: 478.0043.

2'-O-(*tert*-Butyldimethylsilyl)-3'-N-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**78**)



2'-O-(*tert*-Butyldimethylsilyl)-3'-N-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**48**) (54 mg, 0.11 mmol) was dissolved in MeCN (1.5 mL). Et_2NH (0.17 mL, 1.64 mmol) was added, and the reaction mixture was stirred at room temperature under

argon for 1 hour. TLC analysis (CH₂Cl₂-MeOH, 94:6) after this time showed complete consumption of starting material (R_f = 0.3) and formation of product (R_f = 0.4). The solvent was removed under reduced pressure to obtain the product **78** as a white foam (50 mg, 87%).

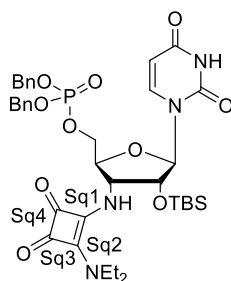
$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3276, 3057, 2934, 2863, 1671, 1571, 1517, 1465, 1445, 1386, 1263, 1098, 837, 782.

¹H NMR (600 MHz, DMSO-d₆): δ = -0.08 (s, 3H, CH₃^{TBDMS}), -0.07 (s, 3H, CH₃^{TBDMS}), 0.78 (s, 9H, tBu^{TBDMS}), 1.15 (t, J_{CH₃,CH₂} = 6.9 Hz, 6H, 2 x CH₃^{Et}), 3.56 (br s, 4H, 2 x CH₂^{Et}), 3.60-3.69 (m, 2H, H-5'a, H-5'b), 4.18 (m, 1H, H-4'), 4.27 (app t, J = 6.1 Hz, 1H, H-2'), 4.90 (app dt, J = 9.0 Hz, J = 6.1 Hz, 1H, H-3'), 5.25 (app t, J = 4.8 Hz, 1H, OH-5'), 5.72 (d, J_{5,6} = 8.1 Hz, 1H, H-5), 5.96 (d, J_{1',2'} = 6.1 Hz, 1H, H-1'), 7.56 (d, J_{3',NH-3'} = 9.0 Hz, 1H, NH-3'), 7.96 (d, J_{5,6} = 8.1 Hz, 1H, H-6), 11.37 (s, 1H, NH^U) ppm.

¹³C NMR (151 MHz, DMSO-d₆): δ = -5.4 (CH₃^{TBDMS}), -5.3 (CH₃^{TBDMS}), 14.9 (2 x CH₃^{Et}), 17.5 (qC, tBu^{TBDMS}), 25.4 (tBu^{TBDMS}), 43.4 (2 x CH₂^{Et}), 54.7 (C-3'), 60.8 (C-5'), 75.0 (C-2'), 82.4 (C-4'), 87.2 (C-1'), 102.3 (C-5), 140.4 (C-6), 150.7 (C-2), 162.9 (C-4), 166.7 (C-Sq1, C-Sq2), 181.8 (C-Sq4), 183.0 (C-Sq3) ppm.

HRMS (ESI⁺): *m/z* calc. 828.3287 [M+Na]⁺, found: 828.3280.

***O,O*-Dibenzyl-*O*-(2'-*O*-(*tert*-butyldimethylsilyl)-3'-*N*-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (**79**)**



79

2'-*O*-(*tert*-Butyldimethylsilyl)-3'-*N*-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine (**78**) (44 mg, 86 μmol) was dissolved in anhydrous CH₂Cl₂ (1 mL). A solution of tetrazole (3-4 wt%) in MeCN (0.74 mL) was added. Dibenzyl *N,N*-diisopropylphosphoramidite (57 μL , 169 μmol) was added at 0 °C. After 5 minutes the

reaction mixture was allowed to warm to room temperature and stirred under argon for 3 hours. Another portion of dibenzyl *N,N*-diisopropylphosphoramidite (20 μL , 59 μmol) was added and the reaction mixture was stirred under argon for a further 1 hour. Another portion of dibenzyl *N,N*-diisopropylphosphoramidite (45 μL , 133 μmol) was then added and the reaction mixture was stirred under argon for a further 1 hour. TLC analysis (CH_2Cl_2 -MeOH, 95:5) after this time showed complete consumption of starting material ($R_f = 0.3$) and formation of intermediate ($R_f = 0.4$). The reaction mixture was cooled to 0 $^\circ\text{C}$ and *m*-chloroperoxybenzoic acid (60 mg, 77% purity) was added. After 20 minutes the reaction mixture was allowed to warm to room temperature and stirred for a further 30 minutes. After this time TLC analysis (CH_2Cl_2 -MeOH, 95:5) showed consumption of intermediate ($R_f = 0.4$) and formation of product ($R_f = 0.2$). The solvent was removed under reduced pressure. The residue was redissolved in CH_2Cl_2 (5 mL) and washed with saturated NaHCO_3 solution (5 mL) followed by brine (5 mL). The organic layer was dried over MgSO_4 and filtered. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 98:2 – 96:4) to afford the product **79** as a white amorphous solid (46 mg, 70%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3242, 2932, 2858, 1793, 1710, 1689, 1672, 1572, 1512, 1484, 1455, 1439, 1381, 1316, 1260, 1232, 1128, 1097, 1061, 1010, 982, 912, 836, 777, 738, 695.

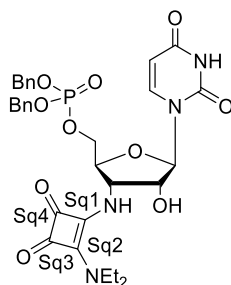
^1H NMR (400 MHz, acetone- d_6): $\delta = 0.06$ (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.08 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.88 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 1.23 (t, $J_{\text{CH}_2,\text{CH}_3} = 7.2$ Hz, 6H, 2 x CH_3^{Et}), 3.55-3.79 (m, 4H, 2 x CH_2^{Et}), 4.37-4.43 (m, 3H, H-4', H-5'a, H-5'b), 4.51 (dd, $J_{2',3'} = 6.7$ Hz, $J_{1',2'} = 4.5$ Hz, 1H, H-2'), 5.11 (d, $J_{\text{P},\text{CH}_2} = 7.9$ Hz, 4H, 2 x CH_2^{Bn}), 5.18 (app dt, $J = 6.7$ Hz, $J = 9.1$ Hz, 1H, H-3'), 5.56 (dd, $J_{5,6} = 8.1$ Hz, $J_{5,\text{NH-U}} = 1.7$ Hz, 1H, H-5), 5.96 (d, $J_{1',2'} = 4.5$ Hz, 1H, H-1'), 6.71 (d, $J_{3',\text{NH-3}'} = 9.1$ Hz, 1H, NH^{Sq}), 7.25-7.49 (m, 10H, H^{Ar}), 7.73 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 10.10 (br s, 1H, NH^{U}) ppm.

^{13}C NMR (100 MHz, acetone- d_6): $\delta = -4.82$ ($\text{CH}_3^{\text{TBDMS}}$), -4.77 ($\text{CH}_3^{\text{TBDMS}}$), 15.4 (2 x CH_3^{Et}), 18.6 (qC, $t\text{Bu}^{\text{TBDMS}}$), 26.1 ($t\text{Bu}^{\text{TBDMS}}$), 44.5 (2 x CH_2^{Et}), 55.7 (C-3'), 67.3 (d, $J_{\text{P},5'} = 5.3$ Hz, C-5'), 69.90 (d, $J = 5.1$ Hz, CH_2^{Bn}), 69.94 (d, $J = 5.1$ Hz, CH_2^{Bn}), 76.0 (C-2'), 81.3 (d, $J_{\text{P},4'} = 7.7$ Hz, C-4'), 90.0 (C-1'), 103.5 (C-5), 128.8 (Ar), 129.2 (d, $J = 3.2$ Hz, Ar), 129.4 (d, $J = 3.0$, Ar), 137.3 (d, $J = 7.1$ Hz, Ar), 140.6 (C-6), 151.5 (C-2), 163.1 (C-4), 167.2 (C-Sq1), 168.8 (C-Sq2), 183.4 (C-Sq4), 184.8 (C-Sq3) ppm.

^{31}P NMR (162 MHz, acetone- d_6): $\delta = -0.57$ ppm.

HRMS (APCI $^+$): m/z calc. 769.3028 $[\text{M}+\text{H}]^+$, found: 769.3013.

***O,O*-Dibenzyl-*O*-(3'-*N*-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (**80**)**



80

O,O-Dibenzyl-*O*-(2'-*O*-(*tert*-butyldimethylsilyl)-3'-*N*-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (**79**) (35 mg, 45.3 μmol) was dissolved in THF (1 mL). Tetrabutylammonium fluoride trihydrate (19 mg, 60.2 μmol) was added. The reaction mixture was stirred at room temperature for 4.5 hours. TLC analysis after this time (CH_2Cl_2 -MeOH, 95:5) showed consumption of starting material ($R_f = 0.5$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 95:5) to give the product **80** as a colourless oil (14 mg, 47%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3239, 3063, 2963, 2925, 2853, 1794, 1672, 1572, 1517, 1456, 1439, 1381, 1256, 1082, 994, 877, 812, 736, 696.

^1H NMR (400 MHz, acetone- d_6): $\delta = 1.23$ (t, $J_{\text{CH}_3,\text{CH}_2} = 7.1$ Hz, 6H, 2 x CH_3^{Et}), 3.64 (br s, 4H, CH_2^{Et}), 4.33-4.44 (m, 3H, H-4', H-5'a, H-5'b), 4.44-4.51 (m, 1H, H-2'), 5.06-5.18 (m, 5H, H-3', 2 x CH_2^{Bn}), 5.51 (d, $J_{5,6} = 8.1$ Hz, 1H, H-5), 5.58 (br s, 1H, OH-2'), 5.93 (d, $J_{1',2'} = 3.2$ Hz, 1H, H-1'), 6.71 (app t, $J = 9.8$ Hz, 1H, NH^{Sq}), 7.30-7.44 (m, 10H, H^{Ar}), 7.70 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6), 10.15 (br s, 1H, NH^{U}) ppm.

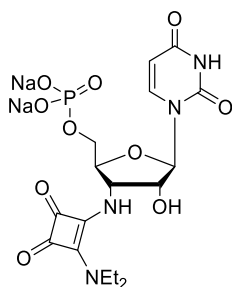
^{13}C NMR (100 MHz, acetone- d_6): $\delta = 15.5$ (2 x CH_3^{Et}), 44.8 (2 x CH_2^{Et}), 55.5 (C-3'), 67.2 (d, $J_{\text{P},5'} = 4.7$ Hz, C-5'), 69.9 (d, $J_{\text{P},\text{CH}_2} = 5.5$ Hz, 2 x CH_2^{Bn}), 74.9 (C-2'), 81.1 (d, $J_{\text{P},4'} = 7.9$ Hz, C-4'), 91.4 (C-1'), 103.0 (C-5), 128.8 (Ar), 129.2 (d, $J = 2.0$ Hz, Ar), 129.4 (Ar), 137.2 (d, $J = 2.0$ Hz,

Ar), 137.3 (d, $J = 2.0$ Hz, Ar), 141.0 (C-6), 151.4 (C-2), 163.4 (C-4), 167.4 (C-Sq1), 168.9 (C-Sq2), 183.3 (C-Sq4), 184.8 (C-Sq3) ppm.

^{31}P NMR (162 MHz, acetone- d_6): $\delta = -1.14$ ppm.

HRMS (APCI $^+$): m/z calc. 655.2164 $[\text{M}+\text{H}]^+$, found: 655.2152.

***O*-(3'-*N*-(2-Diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate disodium salt (**81**)**



81

O,O-Dibenzyl-*O*-(3'-*N*-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-uridine-5'-yl)phosphate (**80**) (22 mg, 34 μmol) was dissolved in MeOH (1 mL). This solution was added to a reaction vessel containing 10% Pd/C (5 mg) under argon. The reaction vessel was purged with H_2 and the reaction mixture was stirred under an atmosphere of H_2 for 5 hours. TLC analysis (CH_2Cl_2 -MeOH, 95:5) after this time showed consumption of starting material ($R_f = 0.3$) and formation of product ($R_f = 0.0$). The reaction mixture was filtered through celite, washing through with MeOH, and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure to obtain the product **81** as a white amorphous solid (14 mg, 80%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3276, 2924, 1664, 1575, 1436, 1273, 1055, 1033, 974, 765.

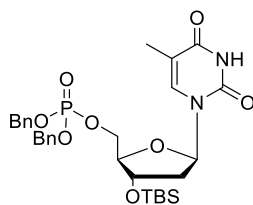
^1H NMR (600 MHz, D_2O): $\delta = 1.28$ (t, $J_{\text{CH}_2,\text{CH}_3} = 6.9$ Hz, 6H, 2 x CH_3^{Et}), 3.67 (br s, 4H, 2 x CH_2^{Et}), 4.05 (app dt, $J = 11.8$ Hz, $J = 4.6$ Hz, 1H, H-5'a), 4.20 (ddd, $J_{5'a,5'b} = 11.8$ Hz, $J_{\text{P},5'b} = 4.6$ Hz, $J_{4',5'b} = 2.3$ Hz, 1H, H-5'b), 4.49-4.57 (m, 2H, H-2', H-4'), 4.90 (app t, $J = 6.7$ Hz, 1H, H-3'), 5.97-6.00 (m, 2H, H-5, H-1'), 8.10 (d, $J_{5,6} = 8.1$ Hz, 1H, H-6) ppm.

^{13}C NMR (151 MHz, D_2O): δ = 14.3 (CH_3^{Et}), 45.0 (CH_2^{Et}), 55.2 (C-3'), 63.5 (d, $J_{\text{P},5'} = 4.2$ Hz, C-5'), 74.3 (C-2'), 81.1 (d, $J_{\text{P},4'} = 8.6$ Hz, C-4'), 90.1 (C-1'), 102.4 (C-5), 142.0 (C-6), 151.6 (C-2), 166.3 (C-4), 167.2 (C-Sq), 167.7 (C-Sq), 180.5 (C-Sq), 183.4 (C-Sq) ppm.

^{31}P NMR (162 MHz, D_2O): δ = 2.04 ppm.

HRMS (ESI $^+$): m/z calc. 519.0863 $[\text{M}+\text{H}]^+$, found: 519.0867.

***O,O*-Dibenzyl-*O*-(3'-*O*-(*tert*-butyldimethylsilyl)-5'-deoxythymidine-5'-yl)phosphate³⁶⁹
(83)**



83

3'-*O*-(*tert*-Butyldimethylsilyl)thymidine (**82**), (92 mg, 257 μmol) was dissolved in CH_2Cl_2 (1.3 mL). A solution of tetrazole (3-4 wt%) in MeCN (2.2 mL) was added. The reaction mixture was cooled to 0 $^\circ\text{C}$ and dibenzyl *N,N*-diisopropylphosphoramidite (130 μL , 385 μmol) was added. After ten minutes the reaction mixture was allowed to warm to room temperature. The reaction mixture was stirred under argon for a further 2.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 94:6) after this time showed formation of intermediate ($R_f = 0.7$) but incomplete consumption of starting material ($R_f = 0.5$). Additional dibenzyl *N,N*-diisopropylphosphoramidite (65 μL , 193 μmol) was added, and the reaction mixture was stirred for a further 1.5 hours. The reaction mixture was diluted with CH_2Cl_2 (2.5 mL) and cooled to 0 $^\circ\text{C}$. *m*-Chloroperoxybenzoic acid (100 mg, 70-75% purity) was added and the reaction mixture was stirred for a further 1 hour. TLC analysis (CH_2Cl_2 -MeOH, 94:6) after this time showed complete consumption of intermediate ($R_f = 0.7$) and formation of product ($R_f = 0.6$). The reaction mixture was diluted with CH_2Cl_2 (10 mL) and quenched by addition of 10% (w/v) sodium thiosulfate (25 mL). The organic layer was separated and washed with saturated NaHCO_3 solution (25 mL), water (25 mL), and brine (25 mL). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 97:3). The product **83** was obtained as a colourless oil (124 mg, 78%).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3175, 2952, 2929, 1683, 1575, 1456, 1251, 1214, 1103, 997, 885, 831, 778, 734, 696, 673.

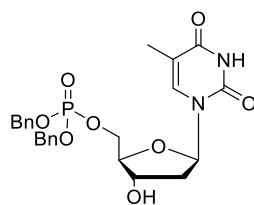
$^1\text{H NMR}$ (400 MHz, CDCl_3): δ = -0.06 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), -0.05 (s, 3H, $\text{CH}_3^{\text{TBDMS}}$), 0.77 (s, 9H, $t\text{Bu}^{\text{TBDMS}}$), 1.74 (s, 3H, CH_3^{T}), 1.79 (dd, $J_{2'a,2'b} = 13.6$ Hz, $J_{1',2'a} = 6.8$ Hz, 1H, H-2'a), 2.06 (ddd, $J_{2'a,2'b} = 13.6$ Hz, $J_{1',2'b} = 5.9$ Hz, $J_{2'b,3'} = 2.9$ Hz, 1H, H-2'b), 3.83-3.87 (m, 1H, H-4'), 3.96-4.09 (m, 2H, H-5'a, H-5'b), 4.18-4.24 (m, 1H, H-3'), 4.90-5.02 (m, 4H, 2 x CH_2^{Bn}), 6.16-6.22 (m, 1H, H-1'), 7.19-7.33 (m, 11H, 10 x H^{Ar} , H-6), 8.21 (br s, 1H, NH^{T}) ppm.

$^{13}\text{C NMR}$ (100 MHz, CDCl_3): δ = -4.8 ($\text{CH}_3^{\text{TBDMS}}$), -4.6 ($\text{CH}_3^{\text{TBDMS}}$), 12.5 (CH_3^{T}), 18.0 (qC, $t\text{Bu}^{\text{TBDMS}}$), 25.8 ($t\text{Bu}^{\text{TBDMS}}$), 40.9 (C-2'), 66.7 (d, $J_{\text{P},5'} = 5.5$ Hz, C-5'), 69.8 (d, $J_{\text{P},\text{CH}_2} = 5.4$ Hz, CH_2^{Bn}), 69.9 (d, $J_{\text{P},\text{CH}_2} = 5.6$ Hz, CH_2^{Bn}), 72.0 (C-3'), 85.0 (C-1'), 85.5 (d, $J_{\text{P},4'} = 8.4$ Hz, C-4'), 111.4 (C-5), 128.17 (Ar), 128.19 (Ar), 128.9 (Ar), 129.0 (Ar), 135.5 (C-6), 150.3 (C-2), 163.5 (C-4) ppm.

$^{31}\text{P NMR}$ (162 MHz, DMSO-d_6): δ = -0.46 ppm.

HRMS (APCI⁺): m/z calc. 617.2443 $[\text{M}+\text{H}]^+$, found: 617.2444.

***O,O*-Dibenzyl-*O*-(5'-deoxythymidine-5'-yl)phosphate³⁶⁹ (**84**)**



84

O,O-Dibenzyl-*O*-(3'-*O*-(*tert*-butyldimethylsilyl)-5'-deoxythymidine-5'-yl)phosphate (**83**) (110 mg, 178 μmol) was dissolved in THF (2 mL). Tetrabutylammonium fluoride trihydrate (82 mg, 253 μmol) was added. The reaction mixture was stirred at room temperature for 1 hour. TLC analysis (CH_2Cl_2 -MeOH, 95:5) after this time showed complete consumption of starting material ($R_f = 0.6$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 96:4) to provide the product **84** as a colourless oil (62 mg, 69%).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3378, 3036, 2953, 2160, 1682, 1456, 1381, 1268, 1082, 993, 963, 881, 735, 696.

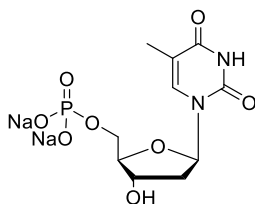
^1H NMR (400 MHz, CDCl_3): δ = 1.82 (s, 3H, CH_3^{T}), 2.00 (app dt, $J = 13.7$ Hz, $J = 6.5$ Hz, 1H, H-2'a), 2.29 (ddd, $J_{2'a,2'b} = 13.7$ Hz, $J_{1',2'a} = 6.5$ Hz, $J_{2'a,3'} = 4.1$ Hz, 1H, H-2'b), 3.95-4.01 (m, 1H, H-4'), 4.11-4.21 (m, 2H, H-5'a, H-5'b), 4.30-4.35 (m, 1H, H-3'), 4.99-5.12 (m, 4H, 2 x CH_2^{Bn}), 6.28 (app t, $J = 6.5$ Hz, 1H, H-1'), 7.31 (s, 1H, H-6), 7.32-7.38 (m, 10H, 10 x H^{Ar}), 8.68 (br s, 1H, NH^{T}) ppm.

^{13}C NMR (100 MHz, CDCl_3): δ = 12.5 (CH_3^{T}), 40.2 (C-2'), 66.7 (d, $J_{\text{P},5'} = 5.8$ Hz, C-5'), 69.98 (d, $J_{\text{P},\text{CH}_2} = 5.0$ Hz, CH_2^{Bn}), 70.03 (d, $J_{\text{P},\text{CH}_2} = 5.0$ Hz, CH_2^{Bn}), 71.0 (C-3'), 84.7 (d, $J_{\text{P},4'} = 7.3$ Hz, C-4'), 84.8 (C-1'), 111.4 (C-5), 128.3 (Ar), 128.9 (Ar), 129.08 (Ar), 129.11 (Ar), 135.5 (C-6), 150.4 (C-2), 163.7 (C-4) ppm.

^{31}P NMR (162 MHz, CDCl_3): δ = -0.24 ppm.

HRMS (ESI $^+$): m/z calc. 503.1578 $[\text{M}+\text{H}]^+$, found: 503.1581.

Thymidine 5'-monophosphate disodium salt (**85**)



85

O,O-Dibenzyl-*O*-(5'-deoxythymidine-5'-yl)phosphate (**84**) (68 mg, 136 μmol) was dissolved in MeOH (6 mL). The reaction vessel was purged with argon. 10% Pd/C (24 mg) was added. The reaction vessel was purged with H_2 and stirred under an atmosphere of H_2 for 2 hours. TLC analysis (CH_2Cl_2 -MeOH, 95:5) after this time showed complete consumption of starting material ($R_f = 0.3$) and formation of product ($R_f = 0.0$). The reaction mixture was filtered through celite, washing through with MeOH and water, and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure to provide the product **85** as a white amorphous solid (45 mg, 91%).

$\nu_{\text{max}}/\text{cm}^{-1}$ (neat) 3201, 2779, 1672, 1646, 1473, 1430, 1254, 1131, 1049, 973, 750.

^1H NMR (400 MHz, CDCl_3): δ = 1.94 (d, $J_{\text{CH}_3,6} = 1.0$ Hz, 3H, CH_3^{T}), 2.35 (ddd, $J_{2'a,2'b} = 14.1$ Hz, $J_{1',2'a} = 6.7$ Hz, $J_{2'a,3'} = 3.4$ Hz, 1H, H-2'a), 2.37-2.45 (m, 1H, H-2'b), 3.98-4.01 (m, 2H, H-5'a,

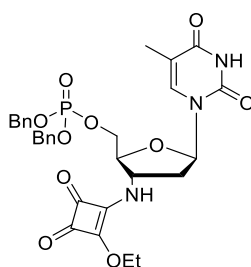
H-5'b), 4.14-4.18 (m, 1H, H-4'), 6.35 (app t, $J = 6.7$ Hz, 1H, H-1'), 7.81 (app d, $J = 1.0$ Hz, 1H, H-6) ppm.

^{13}C NMR (100 MHz, CDCl_3): $\delta = 11.6$ (CH_3^{T}), 38.5 (C-2'), 64.0 (d, $J_{\text{P},5'} = 4.4$ Hz, C-5'), 71.2 (C-3'), 84.9 (C-1'), 85.9 (d, $J_{\text{P},4'} = 8.7$ Hz, C-4'), 111.7 (C-5), 137.6 (C-6), 151.8 (C-2), 166.6 (C-4) ppm.

^{31}P NMR (162 MHz, CDCl_3): $\delta = 2.53$ ppm.

HRMS (ESI⁺): m/z calc. 367.0278 $[\text{M}+\text{H}]^+$, found: 367.0282.

***O,O*-Dibenzyl-*O*-(3'-*N*-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl)phosphate (86)**



86

3'-*N*-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (**44**) (327 mg, 0.895 mmol) was suspended in anhydrous THF (10 mL) under argon at 0 °C. Dibenzyl *N,N*-diisopropylphosphoramidite (0.54 mL, 1.61 mmol) was added to the solution. Tetrazole (3-4 wt%) in MeCN (6.5 mL) was then added. This caused the suspension to dissolve and after 2 minutes a white precipitate formed. The reaction mixture was allowed to warm to room temperature and stirred under argon. TLC analysis (CH_2Cl_2 -EtOH, 92:8) after 3 hours showed the formation of intermediate ($R_f = 0.8$) but incomplete consumption of starting material ($R_f = 0.5$). Another portion of dibenzyl *N,N*-diisopropylphosphoramidite (0.18 mL, 0.536 mmol) was added. The reaction mixture was stirred for a further 40 min after which TLC analysis showed full consumption of the starting material and formation of intermediate. The reaction mixture was cooled to -40 °C and a solution of *m*-chloroperoxybenzoic acid (360 mg, 77% purity) in CH_2Cl_2 (9 mL) was added. The reaction mixture was allowed to reach room temperature gradually. TLC analysis (CH_2Cl_2 -EtOH, 92:8) after 1 hour showed full consumption of intermediate ($R_f = 0.8$) and formation of product ($R_f = 0.7$). The reaction mixture was quenched by addition of 10% (w/v) sodium

thiosulfate solution (50 mL). The aqueous and organic layers were separated. The aqueous layer was extracted with CH₂Cl₂ (3 x 50 mL). The combined organic layers were washed with saturated NaHCO₃ (2 x 100 mL). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-EtOH, 96:4 then 94:6). The product **86** was obtained as a white amorphous solid (474 mg, 84%).

Note: compound **86** exhibits rotamers in NMR spectroscopy.

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3196, 3047, 1807, 1664, 1585, 1420, 1343, 1272, 1095, 786.

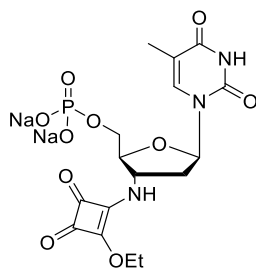
¹H NMR (400 MHz, DMSO-d₆): δ = 1.28-1.44 (m, 3H, CH₃^{Et}), 1.70 (s, 3H, CH₃^T), 2.25-2.42 (m, 2H, H-2'), 4.08 (br s, 1H, H-4'), 4.20 (br s, 2H, H-5'a, H-5'b), 4.28 (br s, 0.5H, H-3'), 4.56-4.73 (m, 2.5H, H-3', CH₂^{Et}), 5.03 (m, 4H, CH₂^{Bn}), 6.24 (app t, J = 6.8 Hz, 1H, H-1'), 7.28-7.40 (m, 10H, H^{Ar}), 7.50 (s, 0.5H, H-6), 7.52 (s, 0.5H, H-6), 8.96 (br s, 0.5H, NH-3'), 9.17 (br s, 0.5H, NH-3'), 11.34 (s, 1H, NH^T) ppm.

¹³C NMR (100MHz, DMSO-d₆): δ = 12.0 (CH₃^T), 15.5 (CH₃^{Et}), 15.6 (CH₃^{Et}), 36.6 (C-2'), 37.0 (C-2'), 53.7 (C-3'), 54.2 (C-3'), 66.7 (d, J_{P,5'} = 5.3 Hz, C-5'), 68.7 (d, J_{P,CH₂Bn} = 5.1 Hz, 2 x CH₂^{Bn}), 69.1 (CH₂^{Et}), 81.8 (d, J_{P,4'} = 6.5 Hz, C-4'), 82.0 (d, J_{P,4'} = 6.5 Hz, C-4'), 83.8 (C-1'), 84.0 (C-1'), 109.9 (C-5), 127.81 (Ar), 127.83 (Ar), 128.5 (Ar), 135.8 (Ar), 135.9 (Ar), 136.1 (C-6), 150.3 (C-2), 163.7 (C-4), 171.7 (C-Sq2), 172.0 (C-Sq2), 176.9 (C-Sq1), 177.5 (C-Sq1), 182.3 (C-Sq3), 182.8 (C-Sq3), 188.89 (C-Sq4), 188.93 (C-Sq4) ppm.

³¹P NMR (162 MHz, DMSO-d₆): δ = -1.11, -1.03 ppm.

HRMS (ESI⁺): *m/z* calc. 648.1718 [M+Na]⁺, found: 648.1722.

3'-(*N*-(2-Ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl) phosphate disodium salt (87**)**



87

O,O-Dibenzyl-*O*-(3'-*N*-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl)phosphate (**86**) (53 mg, 84.4 μ mol) was dissolved in MeOH (3 mL) and CH₂Cl₂ (2 mL). The reaction vessel was purged with argon. 10% Pd/C (13 mg) was added. The reaction vessel was purged with H₂ and stirred under an atmosphere of H₂ for 3 hours. TLC analysis (CH₂Cl₂-EtOH, 92:8) after this time showed complete consumption of starting material (R_f = 0.5) and formation of product (R_f = 0.0). The reaction mixture was filtered through celite, washing through with MeOH. The reaction mixture was concentrated under reduced pressure and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure to provide the product **87** as a white amorphous solid (41 mg, 99%).

Note: compound **87** exhibits rotamers in NMR spectroscopy.

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3214, 2927, 1807, 1692, 1590, 1436, 1347, 1052, 975, 768.

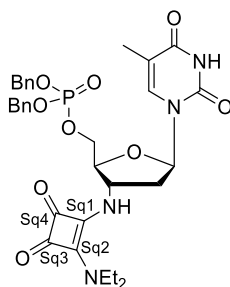
¹H NMR (600 MHz, D₂O): δ = 1.42-1.54 (m, 3H, CH₃^{Et}), 1.95 (s, 3H, CH₃^T), 2.54-2.60 (m, 1H, H-2'a), 2.60-2.67 (m, 1H, H-2'b), 4.02-4.07 (m, 1H, H-5'a), 4.12 (ddd, $J_{4',5'b}$ = 2.8 Hz, $J_{P,5'b}$ = 4.7 Hz, $J_{5'a,5'b}$ = 11.5 Hz, 1H, H-5'b), 4.29 (br s, 0.5H, H-4'), 4.33 (br s, 0.5H, H-4'), 4.58-4.62 (m, 0.5H, H-3'), 4.72-4.85 (m, 2.5H, CH₂^{Et}, H-3'), 6.28-6.38 (m, 1H, H-1'), 7.79 (s, 1H, H-6) ppm.

¹³C NMR (151 MHz, D₂O): δ = 11.6 (CH₃^T), 15.1 (CH₃^{Et}), 37.5 (C-2'), 37.8 (C-2'), 54.3 (C-3'), 54.6 (C-3'), 64.0 (d, $J_{P,5'}$ = 4.6 Hz, C-5'), 70.8 (CH₂^{Et}), 70.9 (CH₂^{Et}), 83.4 (d, $J_{P,4'}$ = 8.6 Hz, C-4'), 84.86 (C-1'), 84.94 (C-1'), 111.7 (C-5), 137.7 (C-6), 151.6 (C-2), 166.6 (C-4), 173.0 (C-Sq), 177.5 (C-Sq), 184.1 (C-Sq), 188.7 (C-Sq) ppm.

^{31}P NMR (162 MHz, D_2O): $\delta = 1.72, 1.88$ ppm.

HRMS (ESI $^-$): m/z calc. 444.0814 [$\text{M}+\text{H}-2\text{Na}$] $^-$, found: 444.0823.

***O,O*-Dibenzyl-*O*-(3'-*N*-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-thymidine-5'-yl)phosphate (**88**)**



88

O,O-Dibenzyl-*O*-(3'-*N*-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl)phosphate (**86**) (73 mg, 0.146 mmol) was dissolved in MeCN (3 mL). Diethylamine (0.18 mL, 1.74 mmol) was added, and the reaction mixture was stirred at room temperature for 2.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 95:5) after this time showed complete consumption of starting material ($R_f = 0.4$) and formation of product ($R_f = 0.3$). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH_2Cl_2 -MeOH, 96:4) to provide the product **88** as a white foam (39 mg, 50%).

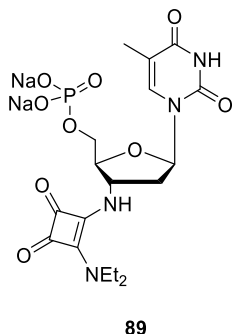
^1H NMR (600 MHz, CDCl_3): $\delta = 1.21$ (t, $J_{\text{CH}_2,\text{CH}_3} = 7.2$ Hz, 6H, 2 x CH_3^{Et}), 1.82 (d, $J_{\text{CH}_3,6} = 0.6$ Hz, 3H, CH_3^{T}), 2.25-2.34 (m, 1H, H-2'a), 2.46-2.57 (m, 1H, H-2'b), 3.58 (br s, 4H, 2 x CH_2^{Et}), 4.20-4.33 (m, 2H, H-4', H-5'a), 4.33-4.47 (m, 1H, H-5'b), 4.93-4.99 (m, 1H, H-3'), 5.00-5.09 (m, 4H, 2 x CH_2^{Bn}), 6.30 (app t, $J = 6.8$ Hz, 1H, H-1'), 6.98 (d, $J_{3',\text{NH}-3'} = 7.1$ Hz, 1H, NH-3'), 7.28-7.35 (m, 10H, H^{Ar}), 7.37-7.39 (m, 1H, H-6), 9.83 (s, 1H, NH^{T}) ppm.

^{13}C NMR (151 MHz, CDCl_3): $\delta = 12.5$ (CH_3^{T}), 15.2 (2 x CH_3^{Et}), 38.4 (C-2'), 44.5 (2 x CH_2^{Et}), 54.9 (C-3'), 67.3 (d, $J_{\text{P},5'} = 5.5$ Hz, C-5'), 69.9 (d, $J_{\text{P},\text{CH}_2} = 5.4$ Hz, CH_2^{Bn}), 70.0 (d, $J_{\text{P},\text{CH}_2} = 5.6$ Hz, CH_2^{Bn}), 83.5 (d, $J_{\text{P},4'} = 7.1$ Hz, C-4'), 84.8 (C-1'), 112.0 (C-5), 128.1 (Ar), 128.2 (Ar), 128.8 (Ar), 128.9 (Ar), 134.7 (C-6), 135.5 (d, $J_{\text{P},\text{Ar}} = 6.6$ Hz, Ar), 150.9 (C-2), 164.2 (C-4), 165.8 (C-Sq), 168.2 (C-Sq), 182.3 (C-Sq), 183.7 (C-Sq) ppm.

^{31}P NMR (162 MHz, CDCl_3): $\delta = -1.21$ ppm.

HRMS (ESI⁺): *m/z* calc. 675.2190 [M+Na]⁺, found: 675.2196.

***O*-(3'-*N*-(2-Diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-thymidine-5'-yl)phosphate disodium salt (**89**)**



O,O-Dibenzyl-*O*-(3'-*N*-(2-diethylamino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-thymidine-5'-yl)phosphate (**88**) (33 mg, 51 μmol) was dissolved in MeOH (4 mL). The reaction vessel was purged with argon. 10% Pd/C (8 mg) was added. The reaction vessel was purged with H₂ and stirred under an atmosphere of H₂ for 2 hours. TLC analysis (CH₂Cl₂-MeOH, 95:5) after this time showed complete consumption of starting material (*R_f* = 0.1) and formation of product (*R_f* = 0.0). The reaction mixture was filtered through celite, washing through with MeOH and water, and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure to provide the product **89** as a white amorphous solid (19 mg, 73%).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3220, 2930, 2120, 1791, 1661, 1575, 1520, 1486, 1439, 1382, 1366, 1274, 1077, 972, 915, 831, 781, 767.

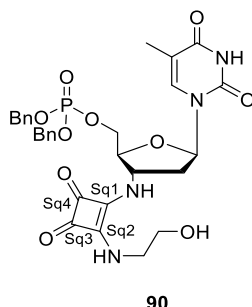
¹H NMR (600 MHz, D₂O): δ = 1.27 (t, *J*_{CH₂,CH₃} = 7.1 Hz, 6H, 2 x CH₃^{Et}), 1.95 (s, 3H, CH₃^T), 2.58-2.68 (m, 2H, H-2'a, H-2'b), 3.56 (br s, 2H, CH₂^{Et}), 3.74 (br s, 2H, CH₂^{Et}), 4.01-4.08 (m, 1H, H-5'a), 4.11-4.19 (m, 1H, H-5'b), 4.30-4.37 (m, 1H, H-4'), 4.92-4.99 (m, 1H, H-3'), 6.35 (app t, *J* = 6.3 Hz, 1H, H-1'), 7.78 (s, 1H, H-6) ppm.

¹³C NMR (151 MHz, D₂O): δ = 11.6 (CH₃^T), 14.3 (2 x CH₃^{Et}), 38.0 (C-2'), 45.0 (2 x CH₂^{Et}), 54.0 (C-3'), 63.9 (d, *J*_{P,5'} = 3.7 Hz, C-5'), 83.1 (d, *J*_{P,4'} = 8.3 Hz, C-4'), 84.7 (C-1'), 111.7 (C-5), 137.8 (C-6), 151.6 (C-2), 166.6 (C-4), 166.8 (C-Sq), 167.8 (C-Sq), 180.3 (C-Sq), 182.9 (C-Sq) ppm.

³¹P NMR (162 MHz, D₂O): δ = 2.05 ppm.

HRMS (ESI⁺): *m/z* calc. 517.1071 [M+H]⁺, found: 517.1067.

***O,O*-Dibenzyl-*O*-(3'-*N*-(2-(2-Hydroxyethyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl) phosphate (**90**)**



O,O-Dibenzyl-*O*-(3'-*N*-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl)phosphate (**86**) (70 mg, 112 μ mol) was dissolved in a mixture of EtOH (2 mL) and MeCN (1 mL). Ethanolamine (6 μ L, 99 μ mol) was added. The reaction mixture was stirred at room temperature. After 45 minutes a white precipitate formed. After 3.5 hours TLC analysis (CH₂Cl₂-EtOH, 92:8) showed the formation of product (*R_f* = 0.1) but incomplete consumption of starting material (*R_f* = 0.6). Another portion of ethanolamine (5 μ L, 83 μ mol) was added. The reaction mixture was stirred for a further 40 minutes after which time TLC analysis showed complete consumption of starting material. The solvent was removed under reduced pressure. The residue was purified by flash chromatography (CH₂Cl₂-EtOH, 90:10) to afford the product **90** as a white amorphous solid (51 mg, 71%).

Note: compound **90** exhibits rotamers in NMR spectroscopy.

ν_{\max} /cm⁻¹ (neat) 3488, 3166, 3065, 2961, 1700, 1647, 1581, 1464, 1352, 1273, 1013, 731, 693.

¹H NMR (400 MHz, DMSO-*d*₆): δ = 1.70 (s, 3H, CH₃^T), 2.24-2.34 (m, 1H, H-2'a), 2.35-2.45 (m, 1H, H-2'b), 3.49-3.55 (m, 2H, CH₂-O), 3.55-3.64 (m, 2H, CH₂-N), 4.08 (dd, *J*_{4',5'} = 4.2 Hz, *J*_{3',4'} = 8.6 Hz, 1H, H-4'), 4.21-4.27 (m, 2H, H-5'), 4.62 (br s, 0.5H, H-3'), 4.97 (br s, 1H, NH^{Sq}), 5.01-5.06 (m, 4.5H, H-3', 2 x CH₂^{Bn}), 6.24 (app t, *J* = 6.7 Hz, 1H, H-1'), 7.32-7.36 (m, 10H, H^{Ar}), 7.51 (s, 1.5H, H-6, NH-3'), 7.84 (br s, 0.5H, NH-3'), 11.35 (s, 1H, NH^T) ppm.

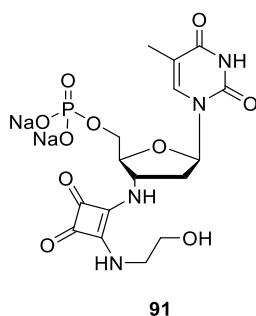
¹³C NMR (100 MHz, DMSO-*d*₆): δ = 12.1 (CH₃^T), 37.4 (C-2'), 46.0 (CH₂-N), 53.5 (C-3'), 60.8 (CH₂-O), 66.9 (d, *J*_{P,5'} = 6.2 Hz, C-5'), 68.8 (d, *J*_{P,CH₂Bn} = 5.4 Hz, 2 x CH₂^{Bn}), 82.7 (d, *J*_{P,4'} = 3.0

Hz, C-4'), 83.7 (C-1'), 110.0 (C-5), 127.9 (Ar), 128.49 (Ar), 128.54 (Ar), 135.9 (Ar), 136.0 (C-6), 150.5 (C-2), 163.8 (C-4), 166.7 (C-Sq2), 168.6 (C-Sq1), 182.2 (C-Sq3), 183.1 (C-Sq4) ppm.

³¹P NMR (162 MHz, DMSO-d₆): δ = -1.08 ppm.

HRMS (APCI+): m/z calc. 641.2007 [M+H]⁺, found: 641.2018.

***O*-(3'-*N*-(2-(2-Hydroxyethyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-3',5'-bis-(deoxy)-thymidine-5'-yl)phosphate disodium salt (**91**)**



O,O-Dibenzyl-*O*-(3'-*N*-(2-(2-Hydroxyethyl)amino-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl) phosphate (**90**) (40 mg, 62 μmol) was dissolved in a mixture of MeOH (1 mL) and CH₂Cl₂ (2 mL). The reaction vessel was purged with argon. 10% Pd/C (8 mg) was added. The reaction vessel was purged with H₂, and the reaction mixture was stirred at room temperature for 4.5 hours. TLC analysis (CH₂Cl₂-EtOH, 92:8) showed complete consumption of starting material (R_f = 0.1) and formation of product (R_f = 0). The reaction mixture was filtered through celite, washing through with MeOH and CH₂Cl₂. This solution was concentrated under reduced pressure and eluted through Dialon WT01S Resin (Na form). The solvent was removed under reduced pressure to provide the product **91** as a white amorphous solid (26 mg, 75%).

$\nu_{\max}/\text{cm}^{-1}$ (neat) 3196, 3047, 1807, 1654, 1585, 1420, 1343, 1272, 1095.

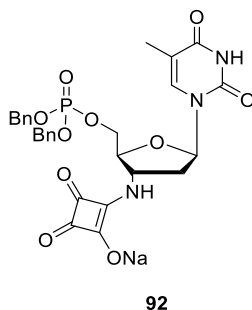
¹H NMR (400 MHz, D₂O): δ = 1.94 (s, 3H, CH₃^T), 2.52-2.70 (m, 2H, H-2'a, H-2'b), 3.76 (br s, 4H, CH₂-O, CH₂-N), 4.02-4.21 (m, 2H, H-5'a, H-5'b), 4.30 (br s, 1H, H-4'), 4.79 (br s, 1H, H-3'), 6.34 (app t, J = 6.2 Hz, 1H, H-1'), 7.78 (s, 1H, H-6) ppm.

¹³C NMR (100MHz, D₂O): δ = 12.2 (CH₃^T), 38.6 (C-2'), 46.9 (CH₂-N), 54.8 (C-3'), 61.6 (CH₂-O), 64.7 (d, J_{P,5'} = 3.3 Hz, C-5'), 84.1 (d, J_{P,C-4'} = 7.7 Hz, C-4'), 85.4 (C-1'), 112.3 (C-5), 138.3 (C-6), 152.3 (C-2), 167.2 (C-4), 168.2 (C-Sq), 182.3 (C-Sq), 183.1 (C-Sq) ppm.

^{31}P NMR (162 MHz, D_2O): $\delta = -1.30$ ppm.

HRMS (APCI⁺): m/z calc. 506.0634 $[\text{M}+\text{H}]^+$, found: 505.0707.

***O,O*-Dibenzyl-*O*-(3'-*N*-(2-hydroxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl)phosphate (92)**



O,O-Dibenzyl-*O*-(3'-*N*-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxyuridine-5'-yl)phosphate (86) (42 mg, 67 μmol) was dissolved in THF (1 mL) and water (0.5 mL). NaOH (9 mg, 225 μmol) was added. The reaction mixture was stirred at room temperature for 1.5 hours. TLC analysis (water-isopropanol-EtOAc, 10:50:40) after this time showed formation of product ($R_f = 0.7$), but incomplete consumption of starting material ($R_f = 0.9$). Another portion of NaOH (9 mg, 225 μmol) was added and the reaction mixture was stirred at room temperature for a further 3 hours. The reaction mixture was eluted through Dialon WT01S ion-exchange resin (H form) and then eluted through Dialon WT01S ion-exchange resin (Na form). The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (water-isopropanol-EtOAc, 6:54:40, then 8:52:40) to obtain the product as a white amorphous solid (8 mg, 19%).

^1H NMR (600 MHz, DMSO-d_6): $\delta = 1.70$ (d, $J_{6,\text{CH}_3} = 1.0$ Hz, 3H, CH_3^{T}), 2.27 (ddd, $J_{1',2'a} = 6.4$ Hz, $J_{2'a,3'} = 8.7$ Hz, $J_{2'a,2'b} = 13.4$ Hz, 1H, H-2'a), 2.35 (app dt, $J = 6.4$ Hz, $J = 13.4$ Hz, 1H, H-2'b), 4.03-4.08 (m, 1H, H-4'), 4.19-4.26 (m, 2H, H-5'a, H-5'b), 4.51-4.57 (m, 1H, H-3'), 5.02 (d, $J_{\text{P},\text{CH}_2} = 5.5$ Hz, 2H, CH_2^{Bn}), 5.03 (d, $J_{\text{P},\text{CH}_2} = 5.5$ Hz, 2H, CH_2^{Bn}), 6.24 (app t, $J = 6.4$ Hz, 1H, H-1'), 7.29-7.37 (m, 11H, NH-3', 10 x H^{Ar}), 7.54 (d, $J_{6,\text{CH}_3} = 1.0$ Hz, 1H, H-6), 11.29 (br s, 1H, NH^{T}) ppm.

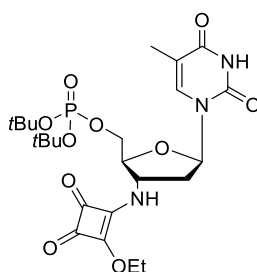
^{13}C NMR (151 MHz, DMSO-d_6): $\delta = 12.0$ (CH_3^{T}), 37.5 (C-2'), 52.5 (C-3'), 67.1 (d, $J_{\text{P},5'} = 5.3$ Hz, C-5'), 68.6 (d, $J_{\text{P},\text{CH}_2} = 4.9$ Hz, CH_2^{Bn}), 68.7 (d, $J_{\text{P},\text{CH}_2} = 3.7$ Hz, CH_2^{Bn}), 82.6 (d, $J_{\text{P},4'} = 7.4$ Hz, C-4'), 83.5 (C-1'), 109.8 (C-5), 127.1 (Ar), 127.9 (Ar), 128.1 (Ar), 128.4 (Ar), 128.5 (Ar), 135.9

(d, $J_{P,Ar} = 2.6$ Hz, Ar), 136.0 (d, $J_{P,Ar} = 3.2$ Hz, Ar), 136.2 (C-6), 150.4 (C-2), 163.7 (C-4), 180.9 (C-Sq), 188.6 (C-Sq) ppm.

^{31}P NMR (162 MHz, DMSO- d_6): $\delta = -1.17$ ppm.

HRMS (ESI $^-$): m/z calc. 596.1440 $[\text{M-Na}]^-$, found: 596.1436.

***O,O*-tert-Butyl-*O*-(3'-*N*-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine-5'-yl)phosphate (96)**



96

3'-*N*-(2-ethoxy-3,4-dioxocyclobut-1-en-1-yl)amino-3'-deoxythymidine (**44**) (160 mg, 438 μmol) was suspended in anhydrous THF (1 mL). A solution of tetrazole (3-4 wt%) in acetonitrile (3 mL) was added. The reaction mixture was cooled to 0 $^{\circ}\text{C}$ and di-*tert*-butyl *N,N*-diisopropylphosphoramidite (250 μL , 792 μmol) was added. After five minutes the reaction mixture was allowed to warm to room temperature. CH_2Cl_2 (1 mL) and THF (4 mL) were added to fully dissolve the reaction mixture. The reaction mixture was stirred under argon for a further 2.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 90:10) after this time showed formation of intermediate ($R_f = 0.5$) but incomplete consumption of starting material ($R_f = 0.4$). Additional di-*tert*-butyl *N,N*-diisopropylphosphoramidite (200 μL , 634 μmol) was added, and the reaction mixture was stirred for a further 20 hours. The reaction mixture was cooled to 0 $^{\circ}\text{C}$ and *m*-chloroperoxybenzoic acid (303 mg, 70-75% purity) was added. The reaction mixture was stirred for a further 1.5 hours. TLC analysis (CH_2Cl_2 -MeOH, 94:6) after this time showed complete consumption of intermediate and formation of product ($R_f = 0.2$). The reaction mixture was quenched by addition of 10% (w/v) sodium thiosulfate (50 mL). The reaction mixture was diluted with CH_2Cl_2 (20 mL). The aqueous layer was separated and extracted with CH_2Cl_2 (2 x 20 mL) and EtOAc (2 x 20 mL). The combined organic layers were concentrated under reduced pressure and the residue was purified by

flash chromatography (CH₂Cl₂-MeOH, 96:4 – 92:8) to provide the product **96** as a white foam (53 mg, 22%).

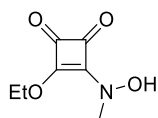
¹H NMR (600 MHz, acetone-d₆): δ = 1.26-1.34 (m, 3H, CH₃^{Et}), 1.48-1.51 (m, 18H, CH₃^{tBu}), 1.89 (s, 3H, CH₃^T), 2.51-2.59 (m, 2H, H-2'a, H-2'b), 4.21-4.36 (m, 3H, H-4', H-5'a, H-5'b), 4.60 (br s, 1H, H-3'), 4.75 (br s, 2H, CH₂^{Et}), 6.42 (app t, J = 6.9 Hz, 1H, H-1'), 7.68 (s, 1H, H-6) ppm.

¹³C NMR (151 MHz, acetone-d₆): δ = 12.6 (CH₃^T), 16.1 (qC, tBu), 30.1 (CH₃^{tBu}), 31.9 (CH₃^{Et}), 38.7 (C-2'), 55.9 (C-3'), 67.1 (C-5'), 70.1 (CH₂^{Et}), 83.0 (d, J_{P,4'} = 7.0 Hz, C-4'), 85.2 (C-1'), 111.4 (C-5), 134.9 (C-6), 151.3 (C-2), 164.2 (C-4), 166.4 (C-Sq), 173.2 (C-Sq) ppm.

³¹P NMR (162 MHz, acetone-d₆): δ = -9.10 ppm.

HRMS (APCI⁻): *m/z* calc. 556.2066 [M-H]⁻, found: 556.2068.

Ethoxy-2-(*N*-methyl)hydroxylamino-3,4-dioxocyclobut-1-en (98**)**



98

N-Methylhydroxylamine hydrochloride (58 mg, 693 μmol) was suspended in EtOH (2.5 mL). Triethylamine (290 μL, 2.1 mmol) was added. Diethyl squarate (100 μL, 676 μmol) was added and the reaction mixture was stirred at room temperature for 22 hours. TLC analysis (CH₂Cl₂-MeOH, 90:10) after this time showed consumption of starting material (R_f = 0.8) and formation of product (R_f = 0.5). The solvent was removed under reduced pressure and the residue was purified by flash chromatography (CH₂Cl₂-MeOH, 98:2) to provide the product **98** as a white powder (113 mg, 97%).

*v*_{max}/cm⁻¹ (neat) 3078, 2755, 1797, 1699, 1567, 1498, 1407, 1385, 1345, 1242, 1202, 1168, 1133, 1092, 901, 822.

¹H NMR (400 MHz, DMSO-d₆): δ = 1.36 (t, J_{CH₂,CH₃} = 7.1 Hz, 3H, CH₃^{Et}), 3.38 (s, 3H, N-CH₃), 4.65 (q, J_{CH₂,CH₃} = 7.1 Hz, 2H, CH₂^{Et}), 10.98 (s, 1H, OH) ppm.

HRMS (ESI⁻): *m/z* calc. 170.0459 [M-H]⁻, found: 170.0455.

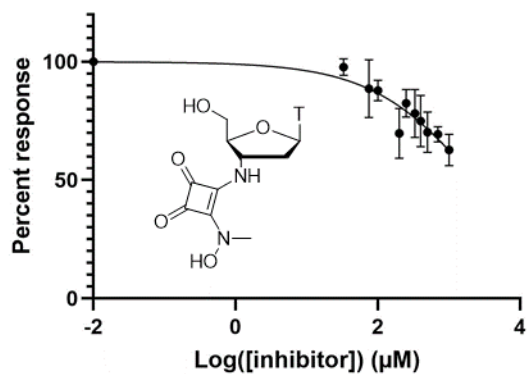
6.3. Biochemical evaluation

6.3.1. Gel electrophoresis assay

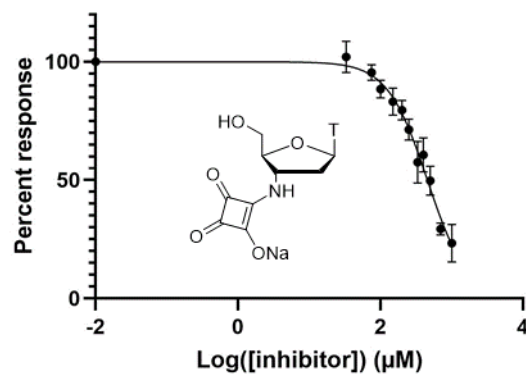
Truncated human SNM1A (698-1040) was stored as a 1.0 μM (0.04 mg/mL) solution in reaction buffer (20 mM HEPES-KOH pH 7.5, 50 mM KCl, 10 mM MgCl_2 , 0.05% Triton-X, 0.1 mg/mL BSA, 5% glycerol, 0.5 mM). Thymidine (control) and modified nucleosides (1 mM in reaction mixture or as specified) were treated with SNM1A (25 fmol) in reaction buffer (10 μL) containing 4% DMSO on ice, and then incubated at 37 °C for 5 minutes. A solution of the oligonucleotide substrate (1 μL , 0.8 pmol/ μL) was added and each reaction was incubated at 37 °C for a further 60 minutes. The reactions were stopped by addition of 2 μL of stop solution (95% formamide, 10 mM EDTA) followed by heating at 95 °C for 3 minutes. Oligonucleotides were separated on 15% acrylamide 6.5 M urea gels (2.9 g urea, 2.7 mL 40% acrylamide-bisacrylamide 25:1, 0.7 mL 10X TBE (0.9 M Tris, 0.9 M boric acid, 0.02 M EDTA pH 8.0), 1.3 mL H_2O) in 1X TBE at 150 V for between 75 and 90 minutes, alongside bromophenol blue and xylene cyanol as markers for 8 nt and 28 nt, respectively. The gels were imaged using a Typhoon FLA 9500 biomolecular imager.

6.3.2. Real-time fluorescence assay

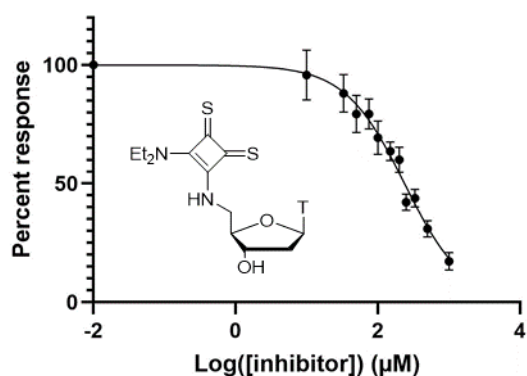
Real-time fluorescence assays were performed using a modified literature procedure^{123,126} utilising a 20-nucleotide ssDNA substrate of the following sequence: 5'-A[FamT]AATTTGA[BHQT]CATCTATTAT-3' (Eurogentec). This oligonucleotide contained a fluorescein-conjugated thymine (FamT) as the second residue from the 5'-end, and a black-hole quencher moiety conjugated to a thymine residue (BHQT) eight nucleotides away. The oligonucleotide substrate was phosphorylated at the 5'-end using T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's protocol, and made up to 1.25 μM for addition to reactions. Nuclease reactions were carried out in black 384-well microplates in a total volume of 25 μL in nuclease buffer (20 mM HEPES-KOH, pH 7.5, 50 mM KCl, 10 mM MgCl_2 , 0.5 mM DTT, 0.05% (v/v) Triton-X100, 5% (v/v) glycerol), with 125 nM oligonucleotide substrate and 2.5 nM SNM1A (698-1040). Reactions were carried out in the presence of twelve different concentrations (0–1000 μM) of each inhibitor. Six replicates of each reaction were performed. SNM1A was incubated with the inhibitor in the above nuclease buffer for 6.5 minutes at room temperature, before the reactions were started by the addition of the DNA substrate. The fluorescence spectra were measured at 37 °C using a SpectraMax i3x microplate reader (excitation at 495 nm, emission at 525 nm) with 7 readings taken every 30 seconds. The fluorescence intensity for each reaction was plotted against time, and the rate of increase was determined and normalised to the zero-inhibitor control. This was plotted against inhibitor concentration and the data were fitted using a "log[inhibitor] vs. normalised response" nonlinear regression algorithm on GraphPad Prism software (GraphPad Software, Inc., La Jolla, CA, USA) to calculate the IC_{50} values.



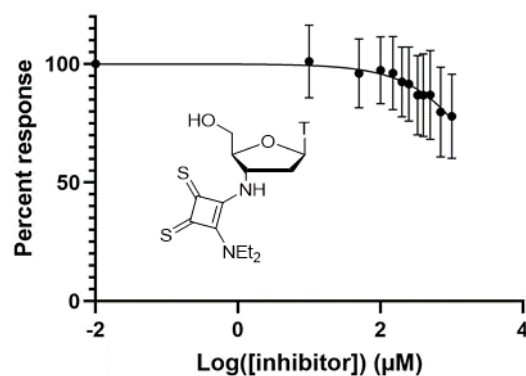
IC₅₀ of *N*-hydroxysquaramide **46** – n.d.



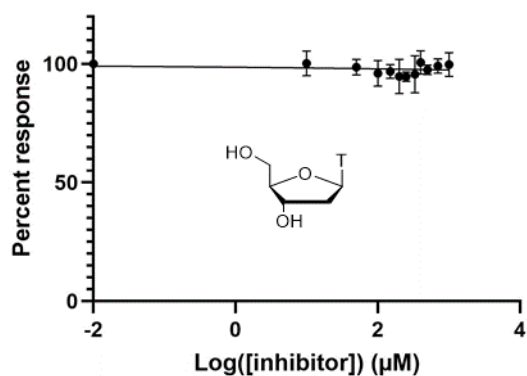
IC₅₀ of squaric acid **47** – 456 μM
(95% CI: 427.1-487.7 μM)



IC₅₀ of 5'-thiosquaramide **57** – 238 μM
(95% CI: 218-259 μM)

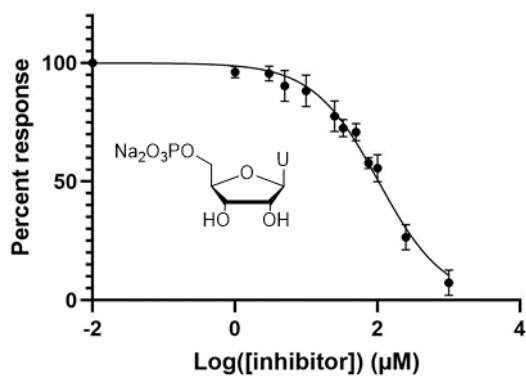


IC₅₀ of 3'-thiosquaramide **61** – n.d.

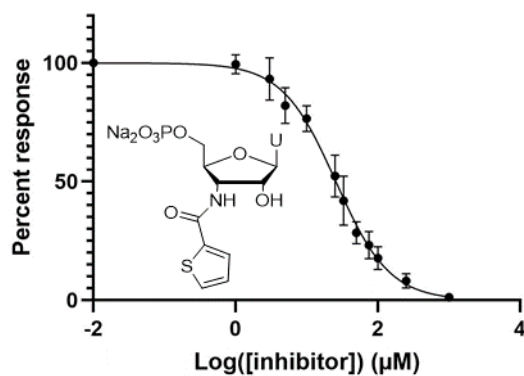


IC₅₀ of thymidine – n.d.

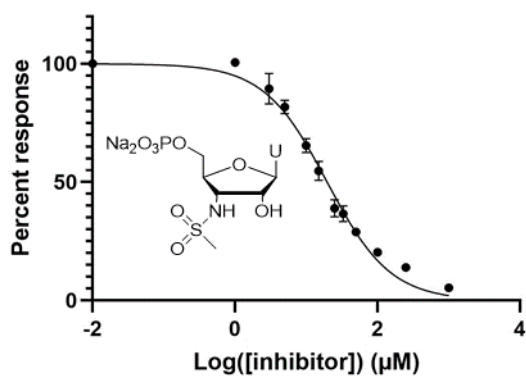
Figure 6.1. Dose-response curves obtained for squaramide-bearing nucleosides from real-time fluorescence assays. Error bars are \pm one standard deviation.



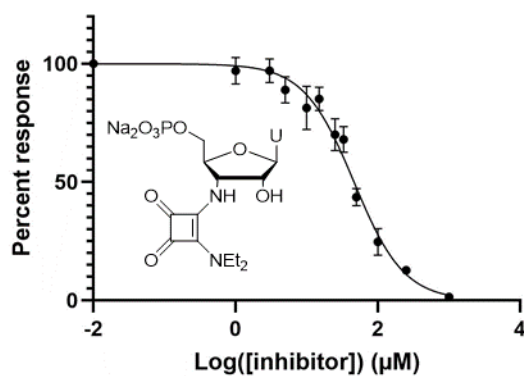
IC₅₀ of uridine 5'-phosphate **70** – 103.7
 μM (95% CI: 93.2 – 115.4 μM)



IC₅₀ of thiophene derivative **77** – 25.7
 μM (95% CI: 23.5 – 28.2 μM)

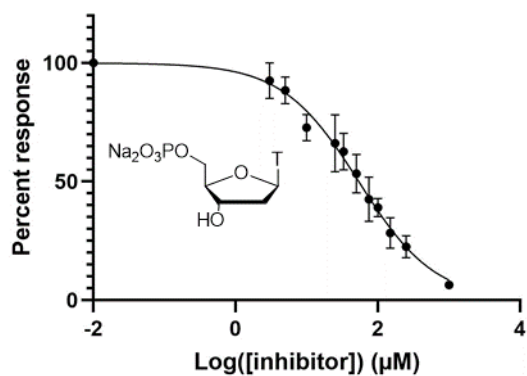


IC₅₀ of sulfonamide **73** – 19.7 μM
 (95% CI: 18.4 – 21.0 μM)

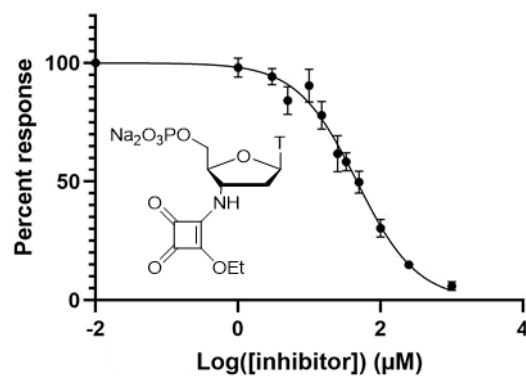


IC₅₀ of squaramide **81** – 46.3 μM
 (95% CI: 42.2 – 50.7 μM)

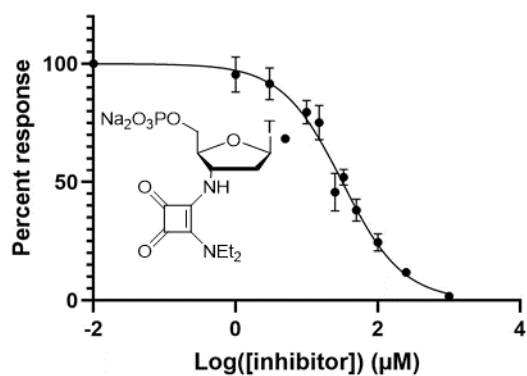
Figure 6.2. Dose-response curves obtained for phosphorylated uridine derivatives from real-time fluorescence assays. Error bars are ± one standard deviation based on six replicates.



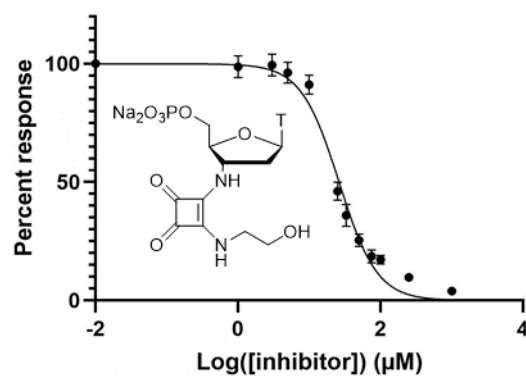
IC₅₀ of thymidine 5'-phosphate **85** – 54.0 μM (95% CI: 48.5 – 60.0 μM)



IC₅₀ of squaryl monoamide **87** – 46.7 μM (95% CI: 42.8 – 51.0 μM)



IC₅₀ of squaramide **89** – 31.7 μM (95% CI: 28.8 – 34.9 μM)



IC₅₀ of squaramide **91** – 26.2 μM (95% CI: 24.6 – 28.0 μM)

Figure 6.3. Dose-response curves obtained from real-time fluorescence assays. Error bars are ± one standard deviation based on six replicates.

6.3.3. UV-vis titrations

UV-vis absorption spectra were recorded using a Varian Cary 50 spectrophotometer. A spectroscopic window of 600 – 200 nm was used for all spectra. Baseline correction from blank solvent was used for all spectra. Spectra were measured in a 1 cm quartz cuvette at room temperature.

Solutions of modified nucleosides in DMSO (spectroscopic grade), or in water in the case of squaric acid **47**, were made up to *ca.* 2.5 – 4 mM and diluted to *ca.* 10 μ M in MeCN (spectroscopic grade) for the titrations. Solutions of zinc salts were made up in MeCN (spectroscopic grade) and aliquoted into the cuvette such that the volume in the cuvette did not increase by more than 10% in total during the titration. The data obtained were fitted to trial models of metal-ligand binding through global nonlinear regression analysis using the ReactLab Equilibria software (Jplus Consulting Pty Ltd., East Fremantle, WA, Australia).

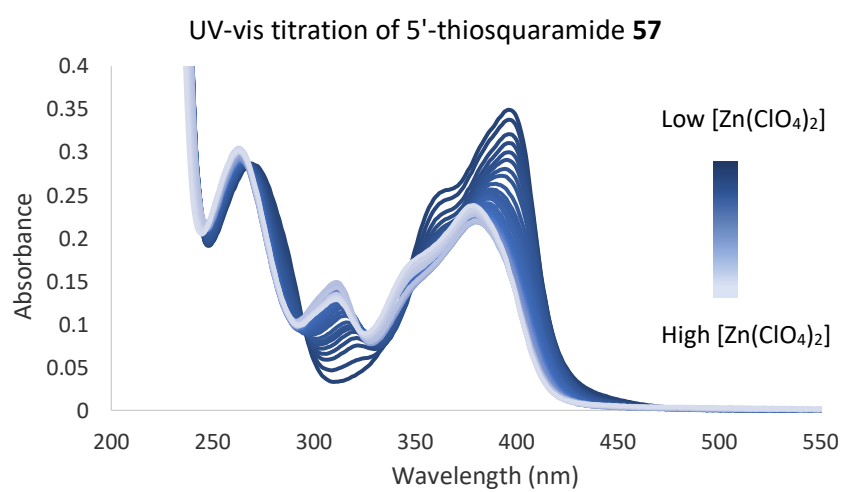
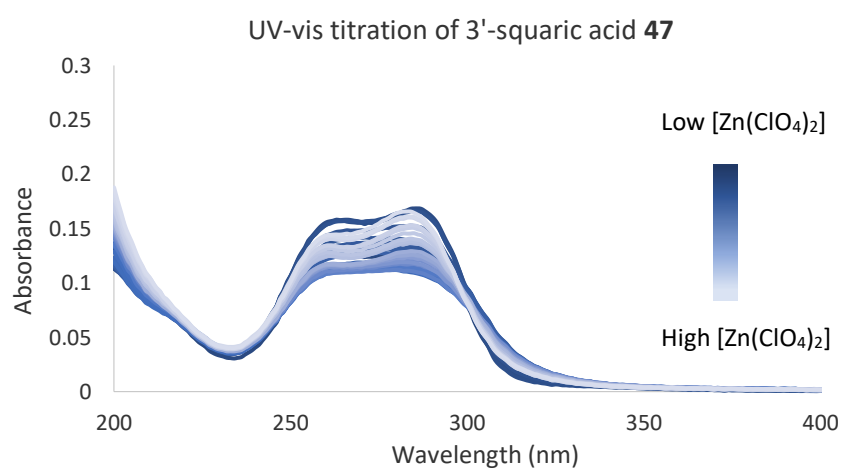
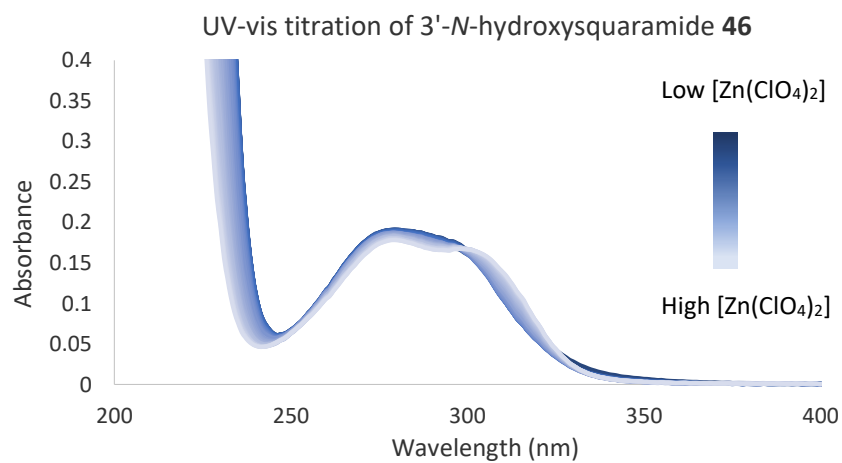


Figure 6.4. UV-vis titration data for titrations of compounds **46**, **47** and **57** with $\text{Zn}(\text{ClO}_4)_2$ in MeCN.

6.3.4. Parallel artificial membrane permeability assay (PAMPA)

The PAMPA assay was performed using a 96-well MultiScreen Filter Plate (Merck), with underdrain removed, as the donor plate, and a 96-well MultiScreen Transport Receiver Plate (Merck) as the acceptor plate, following the manufacturer's protocol.³⁴⁹ Two known drug compounds, carbamazepine and furosemide were used as controls. Solutions of squaramides **46**, **47**, **52**, **57**, **61**, and **63**, and carbamazepine and furosemide (500 μ M) in PBS buffer (pH 7.4) containing 5% DMSO were prepared. PBS buffer (pH 7.4) containing 5% DMSO (300 μ L) was added to each well of the acceptor plate. A solution of lecithin in dodecane (5 μ L, 1% w/v) was added onto the filter at the bottom of donor well to form an artificial membrane. The drug solutions (150 μ L, 500 μ M) were immediately added to each well of the donor plate. The donor plate was then placed into the acceptor plate and incubated at room temperature for 16 h. After the incubation, a sample of each donor well solution (100 μ L) and of each acceptor well solution (250 μ L) were transferred into a UV-star 96-well plate (Greiner Bio-one). Solutions of each compound at the equilibrium concentration expected in both the donor and acceptor wells if the compounds are permeable were made up in PBS (pH 7.4) containing 5% DMSO and samples of the equilibrium solution (250 μ L) were also transferred to the UV-star 96-well plate (Greiner Bio-one). Absorbance was measured from 250-500 nm using a SpectraMax i3x microplate reader. Concentrations in each well were determined using calibration curves for each compound. The effective permeability P_e was calculated using the following equation:

$$P_e = -\ln(1 - r) \left(\frac{V_D V_A}{(V_D + V_A) A t} \right) \quad \text{where} \quad r = \frac{[\text{drug}]_{\text{acceptor}}}{[\text{drug}]_{\text{equilibrium}}}$$

V_D = volume of donor well, 0.15 cm³

V_A = volume of acceptor well, 0.30 cm³

A = area of the filter, 0.3 cm²

t = incubation time, 57 600 s

Each compound was tested in this way in quadruplicate, and values reported are mean \pm one standard deviation. The results obtained for carbamazepine and furosemide, $\log P_e = -5.3 \pm 0.07$ and $\log P_e = -7.1 \pm 0.007$, respectively, were in good agreement with previously reported values.³⁴⁹

References

1. Lindahl, T. & Nyberg, B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610–3618 (1972).
2. Lindahl, T. & Nyberg, B. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* **13**, 3405–3410 (1974).
3. Shen, J. C., Rideout, W. M. & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972–6 (1994).
4. Rydberg, B. & Lindahl, T. Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *Embo J.* **1**, 211–216 (1982).
5. Fink, S. P., Reddy, G. R. & Marnett, L. J. Mutagenicity in *Escherichia coli* of the major DNA adduct derived from the endogenous mutagen malondialdehyde. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8652–8657 (1997).
6. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, 85–98 (2008).
7. Ravanat, J.-L. & Douki, T. UV and ionizing radiations induced DNA damage, differences and similarities. *Radiat. Phys. Chem.* **128**, 92–102 (2016).
8. Wogan, G. N., Hecht, S. S., Felton, J. S., Conney, A. H. & Loeb, L. A. Environmental and chemical carcinogenesis. *Semin. Cancer Biol.* **14**, 473–486 (2004).
9. Christmann, M., Verbeek, B., Roos, W. P. & Kaina, B. O6-Methylguanine-DNA methyltransferase (MGMT) in normal tissues and tumors: enzyme activity, promoter methylation and immunohistochemistry. *Biochim. Biophys. Acta - Rev.*

- Cancer* **1816**, 179–190 (2011).
10. Sancar, A., Lindsey-Boltz, L. A., Üsal-Kaçmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73**, 39–85 (2004).
 11. Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
 12. Torgovnick, A. & Schumacher, B. DNA repair mechanisms in cancer development and therapy. *Front. Genet.* **6**, 157 (2015).
 13. Manchado, E., Guillaumot, M. & Malumbres, M. Killing cells by targeting mitosis. *Cell Death and Differentiation* **19**, 369–377 (2012).
 14. Niedernhofer, L. J., Daniels, J. S., Rouzer, C. A., Greene, R. E., Marnett, L. J. & Hancock, B. Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. *J. Biol. Chem.* **278**, 31426–31433 (2003).
 15. Stonez, M. P., Cho, Y. J., Huang, H., Kim, H. Y., Kozekov, I. D., Kozekova, A., Wang, H., Minko, I. G., Lloyd, R. S., Harris, T. M. & Rizzo, C. J. Interstrand DNA cross-links induced by α,β -unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Acc. Chem. Res.* **41**, 793–804 (2008).
 16. Colis, L. C., Raychaudhury, P. & Basu, A. K. Mutational specificity of γ -radiation-induced guanine-thymine and thymine-guanine intrastrand cross-links in mammalian cells and translesion synthesis past the guanine-thymine lesion by human DNA polymerase η . *Biochemistry* **47**, 8070–8079 (2008).
 17. Box, H. C., Budzinski, E. E., Dawidzik, J. D., Wallace, J. C., Evans, M. S. & Gobey, J. S.

- Radiation-induced formation of a crosslink between base moieties of deoxyguanosine and thymidine in deoxygenated solutions of d(CpGpTpA). *Radiat. Res.* **145**, 641–643 (1996).
18. Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer* **11**, 467–480 (2011).
 19. Rycenga, H. B. & Long, D. T. The evolving role of DNA inter-strand crosslinks in chemotherapy. *Curr. Opin. Pharmacol.* **41**, 20–26 (2018).
 20. Guainazzi, A. & Schärer, O. D. Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy. *Cell. Mol. Life Sci.* **67**, 3683–3697 (2010).
 21. Clauson, C., Schärer, O. D. & Niedernhofer, L. Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012732 (2013).
 22. Henriques, J. A. P. & Moustacchi, E. Isolation and characterization of pso mutants sensitive to photo-addition of psoralen derivatives in *Saccharomyces cerevisiae*. *Genetics* **95**, 273–288 (1980).
 23. Haase, E., Riehl, D., Mack, M. & Brendel, M. Molecular cloning of SNM1, a yeast gene responsible for a specific step in the repair of cross-linked DNA. *MGG Mol. Gen. Genet.* **218**, 64–71 (1989).
 24. Brendel, M., Bonatto, D., Strauss, M., Revers, L. F., Pungartnik, C., Saffi, J. & Henriques, J. A. P. Role of PSO genes in repair of DNA damage of *Saccharomyces cerevisiae*. *Mutat. Res.* **544**, 179–193 (2003).

25. Dominski, Z. Nucleases of the metallo- β -lactamase family and their role in DNA and RNA metabolism. *Critical Reviews in Biochemistry and Molecular Biology* **42**, 67–93 (2007).
26. Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L. & Tong, L. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953–956 (2006).
27. Brzezniak, L. K., Bijata, M., Szczesny, R. J. & Stepien, P. P. Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol.* **8**, (2011).
28. Hazrati, A., Ramis-Castellort, M., Sarkar, S., Barber, L. J., Schofield, C. J., Hartley, J. A. & McHugh, P. J. Human SNM1A suppresses the DNA repair defects of yeast *pso2* mutants. *DNA Repair (Amst)*. **7**, 230–238 (2008).
29. Nagase, T., Miyajima, N., Tanaka, A., Sazuka, T., Seki, N., Sato, S., Tabata, S., Ishikawa, K.-I., Kawarabayasi, Y., Kotani, H. & Nomura, N. Prediction of the coding sequences of unidentified human genes. III. The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* **2**, 37–43 (1995).
30. Demuth, I. & Digweed, M. Genomic organization of a potential human DNA-crosslink repair gene, KIAA0086. *Mutat. Res.* **409**, 11–16 (1998).
31. Dronkert, M. L. G., de Wit, J., Boeve, M., Vasconcelos, M. L., van Steeg, H., Tan, T. L. R., Hoeijmakers, J. H. J. & Kanaar, R. Disruption of mouse SNM1 causes increased sensitivity to the DNA interstrand cross-linking agent mitomycin C. *Mol. Cell. Biol.*

- 20**, 4553–4561 (2000).
32. Zhang, J. & Walter, J. C. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst)*. **19**, 135–142 (2014).
 33. Sengerová, B., Wang, A. T. & McHugh, P. J. Orchestrating the nucleases involved in DNA interstrand cross-link (ICL) repair. *Cell Cycle* **10**, 3999–4008 (2011).
 34. Iyama, T., Lee, S. Y., Berquist, B. R., Gileadi, O., Bohr, V. A., Seidman, M. M., McHugh, P. J. & Wilson, D. M. I. CSB interacts with SNM1A and promotes DNA interstrand crosslink processing. *Nucleic Acids Res.* **43**, 247–258 (2015).
 35. Jones, M. J. K. & Huang, T. T. The Fanconi anemia pathway in replication stress and DNA crosslink repair. *Cell. Mol. Life Sci.* **69**, 3963–3974 (2012).
 36. Sengerová, B., Allerston, C. K., Abu, M., Lee, S. Y., Hartley, J., Kiakos, K., Schofield, C. J., Hartley, J. A., Gileadi, O. & McHugh, P. J. Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J. Biol. Chem.* **287**, 26254–26267 (2012).
 37. Zhang, J., Dewar, J. M., Budzowska, M., Motnenko, A., Cohn, M. A. & Walter, J. C. DNA interstrand cross-link repair requires replication-fork convergence. *Nat. Struct. Mol. Biol.* **22**, 242–247 (2015).
 38. Wang, A. T., Sengerová, B., Cattell, E., Inagawa, T., Hartley, J. M., Kiakos, K., Burgess-Brown, N. A., Swift, L. P., Enzlin, J. H., Schofield, C. J., Gileadi, O., Hartley, J. A. & McHugh, P. J. Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev.* **25**, 1859–1870 (2011).
 39. Yang, K., Moldovan, G.-L. & D'Andrea, A. D. RAD18-dependent recruitment of

- SNM1A to DNA repair complexes by a ubiquitin-binding zinc finger. *J. Biol. Chem.* **285**, 19085–19091 (2010).
40. Ishiai, M., Kimura, M., Namikoshi, K., Yamazoe, M., Yamamoto, K., Arakawa, H., Agematsu, K., Matsushita, N., Takeda, S., Buerstedde, J.-M. & Takata, M. DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. *Mol. Cell. Biol.* **24**, 10733–10741 (2004).
41. Sarangi, P. & Zhao, X. SUMO-mediated regulation of DNA damage repair and responses. *Trends in Biochemical Sciences* **40**, 233–242 (2015).
42. Richie, C. T., Peterson, C., Lu, T., Hittelman, W. N., Carpenter, P. B. & Legerski, R. J. hSnm1 colocalizes and physically associates with 53BP1 before and after DNA damage. *Mol. Cell. Biol.* **22**, 8635–8647 (2002).
43. Harper, J. W. & Elledge, S. J. The DNA damage response: ten years after. *Molecular Cell* **28**, 739–745 (2007).
44. Abdullah, U. B., McGouran, J. F., Brolih, S., Ptchelkine, D., El-Sagheer, A. H., Brown, T. & McHugh, P. J. RPA activates the XPF-ERCC1 endonuclease to initiate processing of DNA interstrand crosslinks. *EMBO J.* **36**, 2047–2060 (2017).
45. Ho, T. V., Guainazzi, A., Derkunt, S. B., Enoiu, M. & Schärer, O. D. Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases. *Nucleic Acids Res.* **39**, 7455–7464 (2011).
46. Minko, I. G., Harbut, M. B., Kozekov, I. D., Kozekova, A., Jakobs, P. M., Olson, S. B., Moses, R. E., Harris, T. M., Rizzo, C. J. & Lloyd, R. S. Role for DNA polymerase κ in the processing of N2-N2-guanine interstrand cross-links. *J. Biol. Chem.* **283**, 17075–

- 17082 (2008).
47. Baddock, H. T., Yosaatmadja, Y., Newman, J. A., Schofield, C. J., Gileadi, O. & McHugh, P. J. The SNM1A DNA repair nuclease. *DNA Repair (Amst)*. **95**, 102941 (2020).
 48. De Laat, W. L., Appeldoorn, E., Jaspers, N. G. J. & Hoeijmakers, J. H. J. DNA structural elements required for ERCC1-XPF endonuclease activity. *J. Biol. Chem.* **273**, 7835–7842 (1998).
 49. Buzon, B., Grainger, R., Huang, S., Rzaeki, C. & Junop, M. S. Structure-specific endonuclease activity of SNM1A enables processing of a DNA interstrand crosslink. *Nucleic Acids Res.* **46**, 9057–9066 (2018).
 50. Tiefenbach, T. & Junop, M. Pso2 (SNM1) is a DNA structure-specific endonuclease. *Nucleic Acids Res.* **40**, 2131–2139 (2012).
 51. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794 (2002).
 52. Hejna, J., Philip, S., Ott, J., Faulkner, C. & Moses, R. The hSNM1 protein is a DNA 5'-exonuclease. *Nucleic Acids Res.* **35**, 6115–6123 (2007).
 53. MacKay, C., Déclais, A. C., Lundin, C., Agostinho, A., Deans, A. J., MacArtney, T. J., Hofmann, K., Gartner, A., West, S. C., Helleday, T., Lilley, D. M. J. & Rouse, J. Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell* **142**, 65–76 (2010).
 54. Zhao, Q., Xue, X., Longerich, S., Sung, P. & Xiong, Y. Structural insights into 5' flap

- DNA unwinding and incision by the human FAN1 dimer. *Nat. Commun.* **5**, 5726 (2014).
55. Pizzolato, J., Mukherjee, S., Schärer, O. D. & Jiricny, J. FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro. *J. Biol. Chem.* **290**, 22602–22611 (2015).
56. Porro, A., Berti, M., Pizzolato, J., Bologna, S., Kaden, S., Saxer, A., Ma, Y., Nagasawa, K., Sartori, A. A. & Jiricny, J. FAN1 interaction with ubiquitylated PCNA alleviates replication stress and preserves genomic integrity independently of BRCA2. *Nat. Commun.* **8**, 1–14 (2017).
57. Fontebasso, Y., Etheridge, T. J., Oliver, A. W., Murray, J. M. & Carr, A. M. The conserved Fanconi anemia nuclease Fan1 and the SUMO E3 ligase Pli1 act in two novel Pso2-independent pathways of DNA interstrand crosslink repair in yeast. *DNA Repair (Amst)*. **12**, 1011–1023 (2013).
58. Thongthip, S., Bellani, M., Gregg, S. Q., Sridhar, S., Conti, B. A., Chen, Y., Seidman, M. M. & Smogorzewska, A. Fan1 deficiency results in DNA interstrand cross-link repair defects, enhanced tissue karyomegaly, and organ dysfunction. *Genes Dev.* **30**, 645–659 (2016).
59. Enoiu, M., Jiricny, J. & Schärer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res.* **40**, 8953–8964 (2012).
60. Spivak, G. & Ganesan, A. K. The complex choreography of transcription-coupled repair. *DNA Repair (Amst)*. **19**, 64–70 (2014).

61. Xie, W., Ling, T., Zhou, Y., Feng, W., Zhu, Q., Stunnenberg, H. G., Grummt, I. & Tao, W. The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8161–8166 (2012).
62. Kato, N., Kawasoe, Y., Williams, H., Coates, E., Roy, U., Shi, Y., Beese, L. S., Schärer, O. D., Yan, H., Gottesman, M. E., Takahashi, T. S. & Gautier, J. Sensing and processing of DNA interstrand crosslinks by the mismatch repair pathway. *Cell Rep.* **21**, 1375–1385 (2017).
63. Hosono, Y., Abe, T., Ishiai, M., Takata, M., Enomoto, T. & Seki, M. The role of SNM1 family nucleases in etoposide-induced apoptosis. *Biochem. Biophys. Res. Commun.* **410**, 568–573 (2011).
64. Akhter, S. & Legerski, R. J. SNM1A acts downstream of ATM to promote the G1 cell cycle checkpoint. *Biochem. Biophys. Res. Commun.* **377**, 236–241 (2008).
65. Akhter, S., Richie, C. T., Deng, J. M., Brey, E., Zhang, X., Patrick, C., Behringer, R. R. & Legerski, R. J. Deficiency in SNM1 abolishes an early mitotic checkpoint induced by spindle stress. *Mol. Cell. Biol.* **24**, 10448–10455 (2004).
66. Giono, L. E. & Manfredi, J. J. The p53 tumor suppressor participates in multiple cell cycle checkpoints. *J. Cell. Physiol.* **209**, 13–20 (2006).
67. Zhang, X., Richie, C. & Legerski, R. J. Translation of hSNM1 is mediated by an internal ribosome entry site that upregulates expression during mitosis. *DNA Repair (Amst)*. **1**, 379–390 (2002).
68. Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A. A., Stern-Ginossar, N.,

- Yakhini, Z., Weinberger, A. & Segal, E. Comparative genetics: Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, (2016).
69. Scolnick, D. M. & Halazonetis, T. D. Chfr defines a mitotic stress checkpoint that delays entry into metaphase. *Nature* **406**, 430–435 (2000).
70. Ahel, I., Ahel, D., Matsusaka, T., Clark, A. J., Pines, J., Boulton, S. J. & West, S. C. Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins. *Nature* **451**, 81–85 (2008).
71. Oberoi, J., Richards, M. W., Crumpler, S., Brown, N., Blagg, J. & Bayliss, R. Structural basis of poly(ADP-ribose) recognition by the multizinc binding domain of Checkpoint with Forkhead-associated and RING domains (CHFR). *J. Biol. Chem.* **285**, 39348–39358 (2010).
72. Barkauskaite, E., Jankevicius, G., Ladurner, A. G., Ahel, I. & Timinszky, G. The recognition and removal of cellular poly(ADP-ribose) signals. *FEBS J.* **280**, 3491–3507 (2013).
73. Michael, S., Travé, G., Ramu, C., Chica, C. & Gibson, T. J. Discovery of candidate KEN-box motifs using cell cycle keyword enrichment combined with native disorder prediction and motif conservation. *Bioinformatics* **24**, 453–457 (2008).
74. Privette, L. M. & Petty, E. M. CHFR: A novel mitotic checkpoint protein and regulator of tumorigenesis. *Transl. Oncol.* **1**, 57–64 (2008).
75. Liu, L., Akhter, S., Bae, J. B., Mukhopadhyay, S. S., Richie, C. T., Liu, X. & Legerski, R. SNMIB/Apollo interacts with astrin and is required for the prophase cell cycle

- checkpoint. *Cell Cycle* **8**, 628–638 (2009).
76. Sudo, T., Ota, Y., Kotani, S., Nakao, M., Takami, Y., Takeda, S. & Saya, H. Activation of Cdh1-dependent APC is required for G1 cell cycle arrest and DNA damage-induced G2 checkpoint in vertebrate cells. *EMBO J.* **20**, 6499–6508 (2001).
77. Digue, L., Orsière, T., De Méo, M., Mattéi, M. G., Depetris, D., Duffaud, F., Favre, R. & Botta, A. Evaluation of the genotoxic activity of paclitaxel by the in vitro micronucleus test in combination with fluorescent in situ hybridization of a DNA centromeric probe and the alkaline single cell gel electrophoresis technique (comet assay) in human T-lymph. *Environ. Mol. Mutagen.* **34**, 269–278 (1999).
78. Callebaut, I., Moshous, D., Mornon, J. & de Villartay, J. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res.* **30**, 3592–3601 (2002).
79. Allerston, C. K., Lee, S. Y., Newman, J. A., Schofield, C. J., McHugh, P. J. & Gileadi, O. The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res.* **43**, 11047–11060 (2015).
80. Colson, P., Pinault, L., Azza, S., Armstrong, N., Chabriere, E., La Scola, B., Pontarotti, P. & Raoult, D. A protein of the metallo-hydrolase/oxidoreductase superfamily with both beta-lactamase and ribonuclease activity is linked with translation in giant viruses. *Sci. Rep.* **10**, (2020).
81. Baddock, H. T., Newman, J. A., Yosaatmadja, Y., Bielinski, M., Schofield, C. J., Gileadi, O. & McHugh, P. J. A phosphate binding pocket is a key determinant of

- exo- versus endo-nucleolytic activity in the SNM1 nuclease family. *Nucleic Acids Res.* **49**, 9294–9309 (2021).
82. Yosaatmadja, Y., Baddock, H. T., Newman, J. A., Bielski, M., Gavard, A. E., Mukhopadhyay, S. M. M., Dannerfjord, A. A., Schofield, C. J., McHugh, P. J. & Gileadi, O. Structural and mechanistic insights into the Artemis endonuclease and strategies for its inhibition. *Nucleic Acids Res.* **49**, 9310–9326 (2021).
83. Cheung-Ong, K., Giaever, G. & Nislow, C. DNA-damaging agents in cancer chemotherapy: serendipity and chemical biology. *Chem. Biol.* **20**, 648–659 (2013).
84. Chu, G. Cellular responses to cisplatin: the roles of DNA-binding proteins and DNA repair. *J. Biol. Chem.* **269**, 787–790 (1994).
85. Bowden, N. A. Nucleotide excision repair: why is it not used to predict response to platinum-based chemotherapy? *Cancer Lett.* **346**, 163–171 (2014).
86. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **8**, 193–204 (2008).
87. Spanswick, V. J., Craddock, C., Sekhar, M., Mahendra, P., Shankaranarayana, P., Hughes, R. G., Hochhauser, D. & Hartley, J. A. Repair of DNA interstrand crosslinks as a mechanism of clinical resistance to melphalan in multiple myeloma. *Blood* **100**, 224–229 (2002).
88. Kauffmann, A., Rosselli, F., Lazar, V., Winnepeninckx, V., Mansuet-Lupo, A., Dessen, P., Van Den Oord, J. J., Spatz, A. & Sarasin, A. High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene* **27**, 565–573 (2008).

89. Herrera, M., Dominguez, G., Garcia, J. M., Peña, C., Jimenez, C., Silva, J., Garcia, V., Gomez, I., Diaz, R., Martin, P. & Bonilla, F. Differences in repair of DNA cross-links between lymphocytes and epithelial tumor cells from colon cancer patients measured in vitro with the comet assay. *Clin. Cancer Res.* **15**, 5466–5472 (2009).
90. Helleday, T. Putting poly (ADP-ribose) polymerase and other DNA repair inhibitors into clinical practice. *Curr. Opin. Oncol.* **25**, 609–614 (2013).
91. Sabharwal, A. & Middleton, M. R. Exploiting the role of O6-methylguanine-DNA-methyltransferase (MGMT) in cancer therapy. *Curr. Opin. Pharmacol.* **6**, 355–363 (2006).
92. Madhusudan, S. & Hickson, I. D. DNA repair inhibition: a selective tumour targeting strategy. *Trends Mol. Med.* **11**, 503–511 (2005).
93. Tumey, L. N., Bom, D., Huck, B., Gleason, E., Wang, J., Silver, D., Brunden, K., Boozer, S., Rundlett, S., Sherf, B., Murphy, S., Dent, T., Leventhal, C., Bailey, A., Harrington, J. & Bennani, Y. L. The identification and optimization of a N-hydroxyurea series of flap endonuclease 1 inhibitors. *Bioorg. Med. Chem. Lett.* **15**, 277–281 (2005).
94. Tumey, L. N., Huck, B., Gleason, E., Wang, J., Silver, D., Brunden, K., Boozer, S., Rundlett, S., Sherf, B., Murphy, S., Bailey, A., Dent, T., Leventhal, C., Harrington, J. & Bennani, Y. L. The identification and optimization of 2,4-diketobutyric acids as flap endonuclease 1 inhibitors. *Bioorg. Med. Chem. Lett.* **14**, 4915–4918 (2004).
95. Rai, G., Vyjayanti, V. N., Dorjsuren, D., Simeonov, A., Jadhav, A., Wilson, D. M. & Maloney, D. J. Synthesis, biological evaluation, and structure-activity relationships

- of a novel class of apurinic/aprimidinic endonuclease 1 inhibitors. *J. Med. Chem.* **55**, 3101–3112 (2012).
96. Manvilla, B. A., Wauchope, O., Seley-Radtke, K. L. & Drohat, A. C. NMR studies reveal an unexpected binding site for a redox inhibitor of AP endonuclease 1. *Biochemistry* **50**, 10540–10549 (2011).
97. Huang, F., Motlekar, N. A., Burgwin, C. M., Napper, A. D., Diamond, S. L. & Mazin, A. V. Identification of specific inhibitors of human RAD51 recombinase using high-throughput screening. *ACS Chem. Biol.* **6**, 628–635 (2011).
98. Kohno, T., Sakiyama, T., Kunitoh, H., Goto, K., Nishiwaki, Y., Saito, D., Hirose, H., Eguchi, T., Yanagitani, N., Saito, R., Sasaki-Matsumura, R., Mimaki, S., Toyama, K., Yamamoto, S., Kuchiba, A., Sobue, T., Ohta, T., Ohki, M. & Yokota, J. Association of polymorphisms in the MTH1 gene with small cell lung carcinoma risk. *Carcinogenesis* **27**, 2448–2454 (2006).
99. Wang, X., Wang, S., Zhou, L., Yu, L. & Zhang, L. A network-pathway based module identification for predicting the prognosis of ovarian cancer patients. *J. Ovarian Res.* **9**, 1–8 (2016).
100. Shahi, R. B., De Brakeleer, S., Caljon, B., Pauwels, I., Bonduelle, M., Joris, S., Fontaine, C., Vanhoeij, M., Van Dooren, S., Teugels, E. & De Grève, J. Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. *BMC Cancer* **19**, 1–11 (2019).
101. Laporte, G. A., Leguisamo, N. M., Gloria, H. de C. e., Azambuja, D. B., Kalil, A. N. &

- Saffi, J. The role of double-strand break repair, translesion synthesis, and interstrand crosslinks in colorectal cancer progression—clinicopathological data and survival. *J. Surg. Oncol.* **121**, 906–916 (2020).
102. DCLRE1A Gene. Available at:
<https://portal.gdc.cancer.gov/genes/ENSG00000198924>. (Accessed: 21st April 2020)
103. Macheret, M. & Halazonetis, T. D. DNA replication stress as a hallmark of cancer. *Annu. Rev. Pathol. Mech. Dis.* **10**, 425–448 (2015).
104. Visconti, R., Della Monica, R. & Grieco, D. Cell cycle checkpoint in cancer: A therapeutically targetable double-edged sword. *Journal of Experimental and Clinical Cancer Research* **35**, 153 (2016).
105. Ubhi, T. & Brown, G. W. Exploiting DNA replication stress for cancer treatment. *Cancer Res.* **79**, 1730–1739 (2019).
106. Tacconi, E. M., Badie, S., Gregoriis, G. De, Reisländer, T., Lai, X., Porru, M., Folio, C., Moore, J., Kopp, A., Torres, J. B., Sneddon, D., Green, M., Dedic, S., Lee, J. W., Batra, A. S., Rueda, O. M., Bruna, A., Leonetti, C., Caldas, C., *et al.* Chlorambucil targets BRCA1/2-deficient tumours and counteracts PARP inhibitor resistance. *EMBO Mol. Med.* **11**, e9982 (2019).
107. Ahkter, S., Richie, C. T., Zhang, N., Behringer, R. R., Zhu, C. & Legerski, R. J. Snm1-deficient mice exhibit accelerated tumorigenesis and susceptibility to infection. *Mol. Cell. Biol.* **25**, 10071–10078 (2005).
108. Hemphill, A. W., Bruun, D., Thrun, L., Akkari, Y., Torimaru, Y., Hejna, K., Jakobs, P.

- M., Hejna, J., Jones, S., Olson, S. B. & Moses, R. E. Mammalian SNM1 is required for genome stability. *Mol. Genet. Metab.* **94**, 38–45 (2008).
109. Ewan M McNeil & David W. Melton. DNA repair endonuclease ERCC1–XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic Acids Res.* **40**, 9990–10004 (2012).
110. Shirota, Y., Stoehlmacher, J., Brabender, J., Xiong, Y. P., Uetake, H., Danenberg, K. D., Groshen, S., Tsao-Wei, D. D., Danenberg, P. V. & Lenz, H. J. ERCC1 and thymidylate synthase mRNA levels predict survival for colorectal cancer patients receiving combination oxaliplatin and fluorouracil chemotherapy. *J. Clin. Oncol.* **19**, 4298–4304 (2001).
111. Dabholkar, M., Vionnet, J., Bostick-Bruton, F., Yu, J. J. & Reed, E. Messenger RNA levels of XPAC and ERCC1 in ovarian cancer tissue correlate with response to platinum-based chemotherapy. *J. Clin. Invest.* **94**, 703–708 (1994).
112. Braun, M. S., Richman, S. D., Quirke, P., Daly, C., Adlard, J. W., Elliott, F., Barrett, J. H., Selby, P., Meade, A. M., Stephens, R. J., Parmar, M. K. B. & Seymour, M. T. Predictive biomarkers of chemotherapy efficacy in colorectal cancer: Results from the UK MRC FOCUS trial. *J. Clin. Oncol.* **26**, 2690–2698 (2008).
113. Arora, S., Kothandapani, A., Tillison, K., Kalman-Maltese, V. & Patrick, S. M. Downregulation of XPF-ERCC1 enhances cisplatin efficacy in cancer cells. *DNA Repair (Amst)*. **9**, 745–753 (2010).
114. McNeil, E. M., Astell, K. R., Ritchie, A. M., Shave, S., Houston, D. R., Bakrania, P., Jones, H. M., Khurana, P., Wallace, C., Chapman, T., Wear, M. A., Walkinshaw, M.

- D., Saxty, B. & Melton, D. W. Inhibition of the ERCC1-XPF structure-specific endonuclease to overcome cancer chemoresistance. *DNA Repair (Amst)*. **31**, 19–28 (2015).
115. Tsodikov, O. V, Ivanov, D., Orelli, B., Staresincic, L., Shoshani, I., Oberman, R., Schärer, O. D., Wagner, G. & Ellenberger, T. Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J*. **26**, 4768–4776 (2007).
116. Barakat, K. H., Torin Huzil, J., Luchko, T., Jordheim, L., Dumontet, C. & Tuszynski, J. Characterization of an inhibitory dynamic pharmacophore for the ERCC1–XPA interaction using a combined molecular dynamics and virtual screening approach. *J. Mol. Graph. Model*. **28**, 113–130 (2009).
117. Barakat, K. H., Jordheim, L. P., Perez-Pineiro, R., Wishart, D., Dumontet, C. & Tuszynski, J. A. Virtual screening and biological evaluation of inhibitors targeting the XPA-ERCC1 interaction. *PLoS One* **7**, e51329 (2012).
118. Jordheim, L. P., Barakat, K. H., Heinrich-Balard, L., Matera, E. L., Cros-Perrial, E., Bouledrak, K., Sabeh, R. El, Perez-Pineiro, R., Wishart, D. S., Cohen, R., Tuszynski, J. & Dumontet, C. Small molecule inhibitors of ERCC1-XPF protein-protein interaction synergize alkylating agents in cancer cells. *Mol. Pharmacol*. **84**, 12–24 (2013).
119. Faridounnia, M., Folkers, G. E. & Boelens, R. Function and interactions of ERCC1-XPF in DNA damage response. *Molecules* **23**, 3205 (2018).
120. Kirschner, K. & Melton, D. W. Multiple roles of the ERCC1-XPF endonuclease in DNA repair and resistance to anticancer drugs. *Anticancer Res*. **30**, 3223–3232

- (2010).
121. Mulderrig, L. & Garaycochea, J. I. XPF-ERCC1 protects liver, kidney and blood homeostasis outside the canonical excision repair pathways. *PLOS Genet.* **16**, e1008555 (2020).
 122. Wu, S. H., Hsiao, Y. T., Chen, J. C., Lin, J. H., Hsu, S. C., Hsia, T. C., Yang, S. T., Hsu, W. H. & Chung, J. G. Bufalin alters gene expressions associated DNA damage, cell cycle, and apoptosis in human lung cancer NCI-H460 cells in vitro. *Molecules* **19**, 6047–6057 (2014).
 123. Lee, S. Y., Brem, J., Pettinati, I., Claridge, T. D. W., Gileadi, O., Schofield, C. J. & McHugh, P. J. Cephalosporins inhibit human metallo β -lactamase fold DNA repair nucleases SNM1A and SNM1B/apollo. *Chem. Commun.* **52**, 6727–6730 (2016).
 124. Buzon, B., Grainger, R. A., Rzaeki, C., York, S., Huang, M. & Junop, M. Identification of bioactive SNM1A inhibitors. *ACS Omega* **6**, 9352–9361 (2021).
 125. Dürr, E.-M., Doherty, W., Lee, S. Y., El-Sagheer, A. H., Shivalingam, A., McHugh, P. J., Brown, T. & McGouran, J. F. Squaramide-based 5'-phosphate replacements bind to the DNA repair exonuclease SNM1A. *ChemistrySelect* **3**, 12824–12829 (2018).
 126. Doherty, W., Dürr, E. M., Baddock, H. T., Lee, S. Y., McHugh, P. J., Brown, T., Senge, M. O., Scanlan, E. M. & McGouran, J. F. A hydroxamic-acid-containing nucleoside inhibits DNA repair nuclease SNM1A. *Org. Biomol. Chem.* **17**, 8094–8105 (2019).
 127. Dürr, E.-M. & McGouran, J. F. Probing the binding requirements of modified nucleosides with the DNA nuclease SNM1A. *Molecules* **26**, 320 (2021).
 128. Laugel, V. Cockayne syndrome: the expanding clinical and mutational spectrum.

- Mech. Ageing Dev.* **134**, 161–170 (2013).
129. Lu, Z., Wang, Q., Jiang, S., Zhang, G. & Ma, Y. Truncation of the unique N-terminal domain improved the thermostability and specific activity of alkaline α -amylase Amy703. *Sci. Rep.* **6**, 1–10 (2016).
130. Ramirez, D. H., Yang, B., D'Souza, A. K., Shen, D. & Woo, C. M. Truncation of the TPR domain of OGT alters substrate and glycosite selection. *Anal. Bioanal. Chem.* **413**, 7385–7399 (2021).
131. Trevino, R. J., Gliubich, F., Berni, R., Cianci, M., Chirgwin, J. M., Zanotti, G. & Horowitz, P. M. NH₂-Terminal sequence truncation decreases the stability of bovine rhodanese, minimally perturbs its crystal structure, and enhances interaction with GroEL under native conditions. *J. Biol. Chem.* **274**, 13938–13947 (1999).
132. Willems, L. I., Overkleeft, H. S. & Van Kasteren, S. I. Current developments in activity-based protein profiling. *Bioconjug. Chem.* **25**, 1181–1191 (2014).
133. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686–4687 (2003).
134. Hoch, D. G., Abegg, D. & Adibekian, A. Cysteine-reactive probes and their use in chemical proteomics. *Chem. Commun.* **54**, 4501–4512 (2018).
135. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M. & Cravatt, B. F. Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10000–10005 (2004).

136. Geurink, P. P., Prely, L. M., Van Der Marel, G. A., Bischoff, R. & Overkleeft, H. S. Photoaffinity labeling in activity-based protein profiling. *Topics in Current Chemistry* **324**, 85–113 (2012).
137. Smith, R. A. G. & Knowles, J. R. Aryldiazirines: Potential reagents for photolabeling of biological receptor sites. *J. Am. Chem. Soc.* **95**, 5072–5073 (1973).
138. Brunner, J., Serin, H. & Richards, F. M. 3-Trifluoromethyl-3-phenyldiazirine: a new carbene generating group for photolabeling reagents. *J. Biol. Chem.* **255**, 3313–3318 (1980).
139. Winnacker, M., Breeger, S., Strasser, R. & Carell, T. Novel diazine-containing DNA photoaffinity probes for the investigation of DNA-protein-interactions. *ChemBioChem* **10**, 109–118 (2009).
140. Tate, J. J., Persinger, J. & Bartholomew, B. Survey of four different photoreactive moieties for DNA photoaffinity labeling of yeast RNA polymerase III transcription complexes. *Nucleic Acids Res.* **26**, 1421–1426 (1998).
141. Galardy, R. E., Craig, L. C. & Printz, M. P. Benzophenone triplet: A new photochemical probe of biological ligand-receptor interactions. *Nat. New Biol.* **242**, 127–128 (1973).
142. Fleet, G. W. J., Porter, R. R. & Knowles, J. R. Affinity labelling of antibodies with aryl nitrene as reactive group. *Nature* **224**, 511–512 (1969).
143. Galmozzi, A., Dominguez, E., Cravatt, B. F. & Saez, E. Application of activity-based protein profiling to study enzyme function in adipocytes. *Methods Enzymol.* **538**, 151–169 (2014).

144. Wang, C., Abegg, D., Dwyer, B. G. & Adibekian, A. Discovery and evaluation of new activity-based probes for serine hydrolases. *ChemBioChem* **20**, 2212–2216 (2019).
145. Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D. H., Bulleit, R. F., *et al.* HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
146. Edgington, L. E. & Bogyo, M. In vivo imaging and biochemical characterization of protease function using fluorescent activity-based probes. *Curr. Protoc. Chem. Biol.* **5**, 25–44 (2013).
147. He, Y., Yu, J., Hu, X., Huang, S., Cai, L., Yang, L., Zhang, H., Jiang, Y., Jia, Y. & Sun, H. An activity-based fluorescent probe and its application for differentiating alkaline phosphatase activity in different cell lines. *Chem. Commun.* **56**, 13323–13326 (2020).
148. Withana, N. P., Garland, M., Verdoes, M., Ofori, L. O., Segal, E. & Bogyo, M. Labeling of active proteases in fresh-frozen tissues by topical application of quenched activity-based probes. *Nat. Protoc.* **11**, 184–191 (2016).
149. Berkers, C. R., Van Leeuwen, F. W. B., Groothuis, T. A., Peperzak, V., Van Tilburg, E. W., Borst, J., Neefjes, J. J. & Ovaa, H. Profiling proteasome activity in tissue with fluorescent probes. *Mol. Pharm.* **4**, 739–748 (2007).
150. Ren, G., Blum, G., Verdoes, M., Liu, H., Syed, S., Edgington, L. E., Gheysens, O., Miao, Z., Jiang, H., Gambhir, S. S., Bogyo, M. & Cheng, Z. Non-invasive imaging of

- cysteine cathepsin activity in solid tumors using a ^{64}Cu -labeled activity-based probe. *PLoS One* **6**, e28029 (2011).
151. Niphakis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341–377 (2014).
 152. Jessani, N., Humphrey, M., McDonald, W. H., Niessen, S., Masuda, K., Gangadharant, B., Yates, J. R., Mueller, B. M. & Cravatt, B. F. Carcinoma and stromal enzyme activity profiles associated with breast tumor growth in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13756–13761 (2004).
 153. Nomura, D. K., Long, J. Z., Niessen, S., Hoover, H. S., Ng, S. W. & Cravatt, B. F. Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell* **140**, 49–61 (2010).
 154. Berkers, C. R., Verdoes, M., Lichtman, E., Fiebiger, E., Kessler, B. M., Anderson, K. C., Ploegh, H. L., Ovaa, H. & Galardy, P. J. Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib. *Nat. Methods* **2**, 357–362 (2005).
 155. Cao, Y., Qiu, T., Kathayat, R. S., Azizi, S. A., Thorne, A. K., Ahn, D., Fukata, Y., Fukata, M., Rice, P. A. & Dickinson, B. C. ABHD10 is an S-depalmitoylase affecting redox homeostasis through peroxiredoxin-5. *Nat. Chem. Biol.* **15**, 1232–1240 (2019).
 156. Cavalli, S., Houben, A. J. S., Albers, H. M. H. G., Van Tilburg, E. W., De Ru, A., Aoki, J., Van Veelen, P., Moolenaar, W. H. & Ovaa, H. Development of an activity-based probe for autotaxin. *ChemBioChem* **11**, 2311–2317 (2010).
 157. Hong, J. A., Choi, N. E., La, Y. K., Nam, H. Y., Seo, J. & Lee, J. Development of a smart activity-based probe to detect subcellular activity of asparaginyl

- endopeptidase in living cells. *Org. Biomol. Chem.* **15**, 8018–8022 (2017).
158. Conole, D., Mondal, M., Majmudar, J. D. & Tate, E. W. Recent developments in cell permeable deubiquitinating enzyme activity-based probes. *Frontiers in Chemistry* **7**, 876 (2019).
159. Sieber, S. A., Niessen, S., Hoover, H. S. & Cravatt, B. F. Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat. Chem. Biol.* **2**, 274–281 (2006).
160. Salisbury, C. M. & Cravatt, B. F. Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1171–1176 (2007).
161. Staub, I. & Sieber, S. A. β -Lactam probes as selective chemical-proteomic tools for the identification and functional characterization of resistance associated enzymes in MRSA. *J. Am. Chem. Soc.* **131**, 6271–6276 (2009).
162. Su, J., Liu, J., Chen, C., Zhang, Y. & Yang, K. Ebsulfur as a potent scaffold for inhibition and labelling of New Delhi metallo- β -lactamase-1 in vitro and in vivo. *Bioorg. Chem.* **84**, 192–201 (2019).
163. Singha, M., Kumar, G., Jain, D., Kumar, G. N., Ray, D., Ghosh, A. S. & Basak, A. Rapid fluorescent-based detection of New Delhi metallo- β -lactamases by photo-cross-linking using conjugates of azidonaphthalimide and zinc(II)-chelating motifs. *ACS Omega* **4**, 10891–10898 (2019).
164. Xie, H., Mire, J., Kong, Y., Chang, M., Hassounah, H. A., Thornton, C. N., Sacchettini, J. C., Cirillo, J. D. & Rao, J. Rapid point-of-care detection of the tuberculosis

- pathogen using a BlaC-specific fluorogenic probe. *Nat. Chem.* **4**, 802–809 (2012).
165. Mao, W., Wang, Y., Qian, X., Xia, L. & Xie, H. A carbapenem-based off–on fluorescent probe for specific detection of metallo- β -lactamase activities. *ChemBioChem* **20**, 511–515 (2019).
166. Patricelli, M. P., Szardenings, A. K., Liyanage, M., Nomanbhoy, T. K., Wu, M., Weissig, H., Aban, A., Chun, D., Tanner, S. & Kozarich, J. W. Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **46**, 350–358 (2007).
167. Patricelli, M. P., Nomanbhoy, T. K., Wu, J., Brown, H., Zhou, D., Zhang, J., Jagannathan, S., Aban, A., Okerberg, E., Herring, C., Nordin, B., Weissig, H., Yang, Q., Lee, J. D., Gray, N. S. & Kozarich, J. W. In situ kinase profiling reveals functionally relevant properties of native kinases. *Chem. Biol.* **18**, 699–710 (2011).
168. Grace Villamor, J., Kaschani, F., Colby, T., Oeljeklaus, J., Zhao, D., Kaiser, M., Patricelli, M. P. & Van Der Hoorn, R. A. L. Profiling protein kinases and other ATP binding proteins in arabidopsis using Acyl-ATP probes. *Mol. Cell. proteomics* **12**, 2481–2496 (2013).
169. Sadler, N. C., Angel, T. E., Lewis, M. P., Pederson, L. A. M., Chauvigné-Hines, L. M., Wiedner, S. D., Zink, E. M., Smith, R. D. & Wright, A. T. Activity-based protein profiling reveals mitochondrial oxidative enzyme impairment and restoration in diet-induced obese mice. *PLoS One* **7**, e47996 (2012).
170. Stoddard, E. G., Volk, R. F., Carson, J. P., Ljungberg, C. M., Murphree, T. A., Smith, J. N., Sadler, N. C., Shukla, A. K., Ansong, C. & Wright, A. T. Multifunctional activity-

- based protein profiling of the developing lung. *J. Proteome Res.* **17**, 2623–2634 (2018).
171. Ratcliffe, S. J., Yi, T. & Khandekar, S. S. Synthesis and characterization of 5'-p-fluorosulfonylbenzoyl-2'(or 3')-(biotinyl)adenosine as an activity-based probe for protein kinases. *J. Biomol. Screen.* **12**, 126–132 (2007).
172. Tomich, J. M., Marti, C. & Colman, R. F. Modification of two essential cysteines in rabbit muscle pyruvate kinase by the guanine nucleotide analogue 5'-[p-(fluorosulfonyl)benzoyl]guanosine. *Biochemistry* **20**, 6711–6720 (1981).
173. Hanouille, X., Van Damme, J., Staes, A., Martens, L., Goethals, M., Vandekerckhove, J. & Gevaert, K. A new functional, chemical proteomics technology to identify purine nucleotide binding sites in complex proteomes. *J. Proteome Res.* **5**, 3438–3445 (2006).
174. Manvar, D., Singh, K. & Pandey, V. N. Affinity labeling of hepatitis C virus replicase with a nucleotide analogue: Identification of binding site. *Biochemistry* **52**, 432–444 (2013).
175. Zhang, X., Marchand, C., Pommier, Y. & Burke, T. R. Design and synthesis of photoactivatable aryl diketo acid-containing HIV-1 integrase inhibitors as potential affinity probes. *Bioorg. Med. Chem. Lett.* **14**, 1205–1207 (2004).
176. Sechi, M., Carta, F., Sannia, L., Dallochio, R., Dessì, A., Al-Safi, R. I. & Neamati, N. Design, synthesis, molecular modeling, and anti-HIV-1 integrase activity of a series of photoactivatable diketo acid-containing inhibitors as affinity probes. *Antiviral Res.* **81**, 267–276 (2009).

177. Zhao, X. Z., Semenova, E. A., Liao, C., Nicklaus, M., Pommier, Y. & Burke, T. R. Biotinylated biphenyl ketone-containing 2,4-dioxobutanoic acids designed as HIV-1 integrase photoaffinity ligands. *Bioorg. Med. Chem.* **14**, 7816–7825 (2006).
178. Kim, D., Jetson, R. R. & Krusemark, C. J. A DNA-assisted immunoassay for enzyme activity via a DNA-linked, activity-based probe. *Chem. Commun.* **53**, 9474–9477 (2017).
179. Zhu, G. & Lippard, S. J. Photoaffinity labeling reveals nuclear proteins that uniquely recognize cisplatin-DNA interstrand cross-links. *Biochemistry* **48**, 4916–4925 (2009).
180. Winnacker, M., Welzmler, V., Strasser, R. & Carell, T. Development of a DNA photoaffinity probe for the analysis of 8-OxodG-binding proteins in a human proteome. *ChemBioChem* **11**, 1345–1349 (2010).
181. Maltseva, E. A., Rechkunova, N. I., Gillet, L. C., Petruseva, I. O., Schärer, O. D. & Lavrik, O. I. Crosslinking of the NER damage recognition proteins XPC-HR23B, XPA and RPA to photoreactive probes that mimic DNA damages. *Biochim. Biophys. Acta* **1770**, 781–789 (2007).
182. Jeong, H. S., Hayashi, G. & Okamoto, A. Diazirine photocrosslinking recruits activated FTO demethylase complexes for specific N6-methyladenosine recognition. *ACS Chem. Biol.* **10**, 1450–1455 (2015).
183. Macdonald, D., Perrier, H., Liu, S., Laliberte, F., Rasori, R., Robichaud, A., Masson, P. & Huang, Z. Hunting the emesis and efficacy targets of PDE4 inhibitors: Identification of the photoaffinity probe 8-(3-azidophenyl)-6-[(4-iodo-1H-1-

- imidazolyl)methyl]quinoline (APIIMQ). *J. Med. Chem.* **43**, 3820–3823 (2000).
184. Picq, M., Huang, Y., Lagarde, M., Doutheau, A. & Nemoz, G. Synthesis of photoreactive phosphatidic acid analogues displaying activatory properties on cyclic AMP-phosphodiesterases. Photoaffinity labeling of an isoform of phosphodiesterase. *J. Med. Chem.* **45**, 1678–1685 (2002).
185. Schülke, J. P., McAllister, L. A., Geoghegan, K. F., Parikh, V., Chappie, T. A., Verhoest, P. R., Schmidt, C. J., Johnson, D. S. & Brandon, N. J. Chemoproteomics demonstrates target engagement and exquisite selectivity of the clinical phosphodiesterase 10a inhibitor mp-10 in its native environment. *ACS Chem. Biol.* **9**, 2823–2832 (2014).
186. Hernandez, L. I., Ozalp, V. C. & Hernandez, F. J. Nuclease activity as a specific biomarker for breast cancer. *Chem. Commun.* **52**, 12346–12349 (2016).
187. Matsumoto, N., Toga, T., Hayashi, R., Sugasawa, K., Katayanagi, K., Ide, H., Kuraoka, I. & Iwai, S. Fluorescent probes for the analysis of DNA strand scission in base excision repair. *Nucleic Acids Res.* **38**, e101 (2010).
188. Kladova, O. A., Iakovlev, D. A., Groisman, R., Ishchenko, A. A., Saparbaev, M. K., Fedorova, O. S. & Kuznetsov, N. A. An assay for the activity of base excision repair enzymes in cellular extracts using fluorescent DNA probes. *Biochemistry* **85**, 480–489 (2020).
189. Balian, A., Gonzalez, J. G., Bastida, N., Akhtar, K. T. K., Borsa, B. A. & Hernandez, F. J. Kinetic screening of nuclease activity using nucleic acid probes. *J. Vis. Exp.* **153**, e60005 (2019).

190. Wilson, D. L. & Kool, E. T. Fluorescent probes of DNA repair. *ACS Chemical Biology* **13**, 1721–1733 (2018).
191. He, H. Z., Chan, W. I., Mak, T. Y., Liu, L. J., Wang, M., Chan, D. S. H., Ma, D. L. & Leung, C. H. Detection of 3'→5' exonuclease activity using a metal-based luminescent switch-on probe. *Methods* **64**, 218–223 (2013).
192. Sato, S. & Takenaka, S. Highly sensitive nuclease assays based on chemically modified DNA or RNA. *Sensors* **14**, 12437–12450 (2014).
193. Kasai, Y., Sato, K., Utsumi, S. & Ichikawa, S. Improvement of SNA_r reaction rate by an electron-withdrawing group in the crosslinking of DNA cytosine-5 methyltransferase by a covalent oligodeoxyribonucleotide inhibitor. *ChemBioChem* **19**, 1866–1872 (2018).
194. Jacobsen, J. A., Major Jourden, J. L., Miller, M. T. & Cohen, S. M. To bind zinc or not to bind zinc: An examination of innovative approaches to improved metalloproteinase inhibition. *Biochim. Biophys. Acta - Mol. Cell Res.* **1803**, 72–94 (2010).
195. Zhang, L., Zhang, J., Jiang, Q., Zhang, L. & Song, W. Zinc binding groups for histone deacetylase inhibitors. *J. Enzyme Inhib. Med. Chem.* **33**, 714–721 (2018).
196. Muri, E., Nieto, M., Sindelar, R. & Williamson, J. Hydroxamic acids as pharmacological agents. *Curr. Med. Chem.* **9**, 1631–1653 (2012).
197. Chiu, Y.-H., Gabriel, G. J. & Canary, J. W. Ternary ligand–zinc–hydroxamate complexes. *Inorg. Chem.* **44**, 40–44 (2005).
198. Wagner, F. F., Weyiwier, M., Lewis, M. C. & Holson, E. B. Small molecule inhibitors

- of zinc-dependent histone deacetylases. *Neurotherapeutics* **10**, 589–604 (2013).
199. Cross, J. B., Duca, J. S., Kaminski, J. J. & Madison, V. S. The active site of a zinc-dependent metalloproteinase influences the computed pKa of ligands coordinated to the catalytic zinc ion. *J. Am. Chem. Soc.* **124**, 11004–11007 (2002).
200. Babine, R. E. & Bender, S. L. Molecular recognition of protein–ligand complexes: applications to drug design. *Chem. Rev.* **97**, 1359–1472 (1997).
201. Ducháčková, L. & Roithová, J. The interaction of zinc(II) and hydroxamic acids and a metal-triggered Lossen rearrangement. *Chem. Eur. J.* **15**, 13399–13405 (2009).
202. Izquierdo-Martin, M. & Stein, R. L. Mechanistic studies on the inhibition of thermolysin by a peptide hydroxamic acid. *J. Am. Chem. Soc.* **114**, 325–331 (1992).
203. Citarella, A., Moi, D., Pinzi, L., Bonanni, D. & Rastelli, G. Hydroxamic acid derivatives: from synthetic strategies to medicinal chemistry applications. *ACS Omega* **6**, 21843–21849 (2021).
204. Day, J. A. & Cohen, S. M. Investigating the selectivity of metalloenzyme inhibitors. *J. Med. Chem.* **56**, 7997–8007 (2013).
205. O’Brien, E. C., Farkas, E., Gil, M. J., Fitzgerald, D., Castineras, A. & Nolan, K. B. Metal complexes of salicylhydroxamic acid (H₂Sha), anthranilic hydroxamic acid and benzohydroxamic acid. Crystal and molecular structure of [Cu(phen)₂(Cl)]Cl · H₂Sha, a model for a peroxidase-inhibitor complex. *J. Inorg. Biochem.* **79**, 47–51 (2000).
206. Augé, F., Hornebeck, W., Decarme, M. & Laronze, J.-Y. Improved gelatinase a selectivity by novel zinc binding groups containing galardin derivatives. *Bioorg.*

- Med. Chem. Lett.* **13**, 1783–1786 (2003).
207. Wang, Y., Stowe, R. L., Pinello, C. E., Tian, G., Madoux, F., Li, D., Zhao, L. Y., Li, J. L., Wang, Y., Wang, Y., Ma, H., Hodder, P., Roush, W. R. & Liao, D. Identification of histone deacetylase inhibitors with benzoylhydrazide scaffold that selectively inhibit class I histone deacetylases. *Chem. Biol.* **22**, 273–284 (2015).
208. Whittaker, M., Floyd, C. D., Brown, P. & Gearing, A. J. H. Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem. Rev.* **99**, 2735–2776 (1999).
209. Kawai, K. & Nagata, N. Metal-ligand interactions: An analysis of zinc binding groups using the Protein Data Bank. *Eur. J. Med. Chem.* **51**, 271–276 (2012).
210. Li, W., Li, J., Wu, Y., Wu, J., Hotchandani, R., Cunningham, K., McFadyen, I., Bard, J., Morgan, P., Schlerman, F., Xu, X., Tam, S., Goldman, S. J., Williams, C., Sypek, J. & Mansour, T. S. A selective matrix metalloprotease 12 inhibitor for potential treatment of chronic obstructive pulmonary disease (COPD): discovery of (S)-2-(8-(Methoxycarbonylamino)dibenzo[b,d]furan-3-sulfonamido)-3-methylbutanoic acid (MMP408). *J. Med. Chem.* **52**, 1799–1802 (2009).
211. Krauze, A. V., Myrehaug, S. D., Chang, M. G., Holdford, D. J., Smith, S., Shih, J., Tofilon, P. J., Fine, H. A. & Camphausen, K. A phase 2 study of concurrent radiation therapy, temozolomide, and the histone deacetylase inhibitor valproic acid for patients with glioblastoma. *Int. J. Radiat. Oncol. Biol. Phys.* **92**, 986–992 (2015).
212. Patchett, A. A., Harris, E., Tristram, E. W., Wyvratt, M. J., Wu, M. T., Taub, D., Peterson, E. R., Ikeler, T. J., Ten Broeke, J., Payne, L. G., Ondeyka, D. L., Thorsett, E.

- D., Greenlee, W. J., Lohr, N. S., Hoffsommer, R. D., Joshua, H., Ruyle, W. V., Rothrock, J. W., Aster, S. D., *et al.* A new class of angiotensin-converting enzyme inhibitors. *Nature* **288**, 280–283 (1980).
213. Attenni, B., Ontoria, J. M., Cruz, J. C., Rowley, M., Schultz-Fademrecht, C., Steinkühler, C. & Jones, P. Histone deacetylase inhibitors with a primary amide zinc binding group display antitumor activity in xenograft model. *Bioorg. Med. Chem. Lett.* **19**, 3081–3084 (2009).
214. Von Reedern, E. G., Grams, F., Brandstetter, H. & Moroder, L. Design and synthesis of malonic acid-based inhibitors of human neutrophil collagenase (MMP8). *J. Med. Chem.* **41**, 339–345 (1998).
215. Miller, M. J., Anderson, K. S., Braccolino, D. S., Cleary, D. G., Gruys, K. J., Han, C. Y., Lin, K. C., Pansegrau, P. D., Ream, J. E., Douglas Sammons, R. & Sikorski, J. A. EPSP synthase inhibitor design II. The importance of the 3-phosphate group for ligand binding at the shikimate-3-phosphate site & the identification of 3-malonate ethers as novel 3-phosphate mimics. *Bioorg. Med. Chem. Lett.* **3**, 1435–1440 (1993).
216. Desvergnès, S., Courtiol-Legourd, S., Daher, R., Dabrowski, M., Salmon, L. & Therisod, M. Synthesis and evaluation of malonate-based inhibitors of phosphosugar-metabolizing enzymes: Class II fructose-1,6-bis-phosphate aldolases, type I phosphomannose isomerase, and phosphoglucose isomerase. *Bioorg. Med. Chem.* **20**, 1511–1520 (2012).
217. Stewart, A. O., Bhatia, P. A., Martin, J. G., Summers, J. B., Rodrigues, K. E., Martin,

- M. B., Holms, J. H., Moore, J. L., Craig, R. A., Kolasa, T., Ratajczyk, J. D., Mazdiyasi, H., Kerdesky, F. A. J., DeNinno, S. L., Maki, R. G., Bouska, J. B., Young, P. R., Lanni, C., Bell, R. L., *et al.* Structure–activity relationships of N-hydroxyurea 5-lipoxygenase inhibitors. *J. Med. Chem.* **40**, 1955–1968 (1997).
218. Campestre, C., Agamennone, M., Tortorella, P., Prezioso, S., Biasone, A., Gavuzzo, E., Pochetti, G., Mazza, F., Hiller, O., Tschesche, H., Consalvi, V. & Gallina, C. N-Hydroxyurea as zinc binding group in matrix metalloproteinase inhibition: Mode of binding in a complex with MMP-8. *Bioorg. Med. Chem. Lett.* **16**, 20–24 (2006).
219. Temperini, C., Innocenti, A., Scozzafava, A. & Supuran, C. T. N-Hydroxyurea - A versatile zinc binding function in the design of metalloenzyme inhibitors. *Bioorg. Med. Chem. Lett.* **16**, 4316–4320 (2006).
220. Higgin, J. J., Yakovlev, G. I., Mitkevich, V. A., Makarov, A. A. & Raines, R. T. Zinc(II)-mediated inhibition of a ribonuclease by an N-hydroxyurea nucleotide. *Bioorg. Med. Chem. Lett.* **13**, 409–412 (2003).
221. Pace, N. J. & Weerapana, E. Zinc-binding cysteines: diverse functions and structural motifs. *Biomolecules* **4**, 419–434 (2014).
222. Maret, W. Zinc and sulfur: a critical biological partnership. *Biochemistry* **43**, 3301–3309 (2004).
223. Redelinghuys, P., Nchinda, A. T. & Sturrock, E. D. Development of domain-selective angiotensin I-converting enzyme inhibitors. *Ann. N. Y. Acad. Sci.* **1056**, 160–175 (2005).
224. Ondetti, M., Rubin, B. & Cushman, D. Design of specific inhibitors of angiotensin-

- converting enzyme: new class of orally active antihypertensive agents. *Science* **196**, 441–444 (1977).
225. Klingler, F. M., Wichelhaus, T. A., Frank, D., Cuesta-Bernal, J., El-Delik, J., Müller, H. F., Sjuts, H., Göttig, S., Koenigs, A., Pos, K. M., Pogoryelov, D. & Proschak, E. Approved drugs containing thiols as inhibitors of metallo- β -lactamases: Strategy to combat multidrug-resistant bacteria. *J. Med. Chem.* **58**, 3626–3630 (2015).
226. Suzuki, T., Nagano, Y., Kouketsu, A., Matsuura, A., Maruyama, S., Kurotaki, M., Nakagawa, H. & Miyata, N. Novel inhibitors of human histone deacetylases: design, synthesis, enzyme inhibition, and cancer cell growth inhibition of SAHA-based non-hydroxamates. *J. Med. Chem.* **48**, 1019–1032 (2005).
227. Hurst, D. R., Schwartz, M. A., Jin, Y., Ghaffari, M. A., Kozarekar, P., Cao, J. & Sang, Q. X. A. Inhibition of enzyme activity of and cell-mediated substrate cleavage by membrane type 1 matrix metalloproteinase by newly developed mercaptosulphide inhibitors. *Biochem. J.* **392**, 527–536 (2005).
228. Suzuki, T., Matsuura, A., Kouketsu, A., Nakagawa, H. & Miyata, N. Identification of a potent non-hydroxamate histone deacetylase inhibitor by mechanism-based drug design. *Bioorg. Med. Chem. Lett.* **15**, 331–335 (2005).
229. Ferraris, D. V., Majer, P., Ni, C., Slusher, C. E., Rais, R., Wu, Y., Wozniak, K. M., Alt, J., Rojas, C., Slusher, B. S. & Tsukamoto, T. δ -Thiolactones as prodrugs of thiol-based glutamate carboxypeptidase II (GCPII) inhibitors. *J. Med. Chem.* **57**, 243–247 (2014).
230. Grant, C., Rahman, F., Piekarz, R., Peer, C., Frye, R., Robey, R. W., Gardner, E. R.,

- Figg, W. D. & Bates, S. E. Romidepsin: A new therapy for cutaneous T-cell lymphoma and a potential therapy for solid tumors. *Expert Review of Anticancer Therapy* **10**, 997–1008 (2010).
231. Vesci, L., Bernasconi, E., Milazzo, F. M., De Santis, R., Gaudio, E., Kwee, I., Rinaldi, A., Pace, S., Carollo, V., Giannini, G. & Bertoni, F. Preclinical antitumor activity of ST7612AA1: A new oral thiol-based histone deacetylase (HDAC) inhibitor. *Oncotarget* **6**, 5735–5748 (2015).
232. Suzuki, T., Kouketsu, A., Itoh, Y., Hisakawa, S., Maeda, S., Yoshida, M., Nakagawa, H. & Miyata, N. Highly potent and selective histone deacetylase 6 inhibitors designed based on a small-molecular substrate. *J. Med. Chem.* **49**, 4809–4812 (2006).
233. Tanakit, A., Rouffet, M., Martin, D. P. & Cohen, S. M. Investigating chelating sulfonamides and their use in metalloproteinase inhibitors. *Dalt. Trans.* **41**, 6507–6515 (2012).
234. Alterio, V., Di Fiore, A., D'Ambrosio, K., Supuran, C. T. & De Simone, G. Multiple binding modes of inhibitors to carbonic anhydrases: How to design specific drugs targeting 15 different isoforms? *Chem. Rev.* **112**, 4421–4468 (2012).
235. Supuran, C. T. Structure-based drug discovery of carbonic anhydrase inhibitors. *J. Enzyme Inhib. Med. Chem.* **27**, 759–772 (2012).
236. Koike, T., Kimura, E., Nakamura, I., Hashimoto, Y. & Shiro, M. The first anionic sulfonamide-binding zinc(II) complexes with a macrocyclic triamine: chemical verification of the sulfonamide inhibition of carbonic anhydrase. *J. Am. Chem. Soc.*

- 114**, 7338–7345 (1992).
237. Huang, P., Ramphal, J., Wei, J., Liang, C., Jallal, B., McMahon, G. & Tang, C. Structure-based design and discovery of novel inhibitors of protein tyrosine phosphatases. *Bioorg. Med. Chem.* **11**, 1835–1849 (2003).
238. Maryanoff, B. E., McComsey, D. F., Costanzo, M. J., Hochman, C., Smith-Swintosky, V. & Shank, R. P. Comparison of sulfamate and sulfamide groups for the inhibition of carbonic anhydrase-II by using topiramate as a structural platform. *J. Med. Chem.* **48**, 1941–1947 (2005).
239. Wahhab, A., Smil, D., Ajamian, A., Allan, M., Chantigny, Y., Therrien, E., Nguyen, N., Manku, S., Leit, S., Rahil, J., Petschner, A. J., Lu, A. H., Nicolescu, A., Lefebvre, S., Montcalm, S., Fournel, M., Yan, T. P., Li, Z., Besterman, J. M., *et al.* Sulfamides as novel histone deacetylase inhibitors. *Bioorg. Med. Chem. Lett.* **19**, 336–340 (2009).
240. Sang, Q. X. A., Jia, M. C., Schwartz, M. A., Jaye, M. C., Kleinman, H. K., Ghaffari, M. A. & Luo, Y. L. New thiol and sulfodiimine metalloproteinase inhibitors and their effect on human microvascular endothelial cell growth. *Biochem. Biophys. Res. Commun.* **274**, 780–786 (2000).
241. AM, C., RS, A. & DW, C. Structural comparison of sulfodiimine and sulfonamide inhibitors in their complexes with zinc enzymes. *J. Biol. Chem.* **267**, 19192–19197 (1992).
242. Matziari, M., Beau, F., Cuniasse, P., Dive, V. & Yiotakis, A. Evaluation of P1'-diversified phosphinic peptides leads to the development of highly selective inhibitors of MMP-11. *J. Med. Chem.* **47**, 325–336 (2004).

243. Vassiliou, S., Mucha, A., Cuniasse, P., Georgiadis, D., Lucet-Levannier, K., Beau, F., Kannan, R., Murphy, G., Knäuper, V., Rio, M. C., Basset, P., Yiotakis, A. & Dive, V. Phosphinic pseudo-tripeptides as potent inhibitors of matrix metalloproteinases: A structure-activity study. *J. Med. Chem.* **42**, 2610–2620 (1999).
244. Duchin, K. L., Waclawski, A. P., Tu, J. I., Manning, J., Frantz, M. & Willard, D. A. Pharmacokinetics, safety, and pharmacologic effects of fosinopril sodium, an angiotensin-converting enzyme inhibitor in healthy subjects. *J. Clin. Pharmacol.* **31**, 58–64 (1991).
245. Biasone, A., Tortorella, P., Campestre, C., Agamennone, M., Prezioso, S., Chiappini, M., Nuti, E., Carelli, P., Rossello, A., Mazza, F. & Gallina, C. α -Biphenylsulfonlamino 2-methylpropyl phosphonates: Enantioselective synthesis and selective inhibition of MMPs. *Bioorg. Med. Chem.* **15**, 791–799 (2007).
246. Bird, J., Harper, G. P., Hughes, I., Hunter, D. J., Karran, E. H., Markwell, R. E., Miles-Williams, A. J., Rahman, S. S. & Ward, R. W. Inhibitors of human collagenase: dipeptide mimetics with lactam and azalactam moieties at the P2' P3' position. *Bioorg. Med. Chem. Lett.* **5**, 2593–2598 (1995).
247. Temperini, C., Innocenti, A., Guerri, A., Scozzafava, A., Rusconi, S. & Supuran, C. T. Phosph(on)ate as a zinc-binding group in metalloenzyme inhibitors: X-ray crystal structure of the antiviral drug foscarnet complexed to human carbonic anhydrase I. *Bioorg. Med. Chem. Lett.* **17**, 2210–2215 (2007).
248. Nocentini, A., Gratteri, P. & Supuran, C. T. Phosphorus versus sulfur: discovery of benzenephosphonamidates as versatile sulfonamide-mimic chemotypes acting as

- carbonic anhydrase inhibitors. *Chem. - A Eur. J.* **25**, 1188–1192 (2019).
249. Alissa, S. A., Alghulikah, H. A., Alothman, Z. A., Osman, S. M., Del Prete, S., Capasso, C., Nocentini, A. & Supuran, C. T. Phosphoramidates are the first phosphorus-based zinc binding motif to show inhibition of β -class carbonic anhydrases from bacteria, fungi, and protozoa. *J. Enzyme Inhib. Med. Chem.* **35**, 59–64 (2020).
250. Jacobsen, N. E. & Bartlett, P. A. A phosphoramidate dipeptide analogue as an inhibitor of carboxypeptidase A. *J. Am. Chem. Soc.* **103**, 654–657 (1981).
251. Bartlett, P. A. & Marlowe, C. K. Evaluation of intrinsic binding energy from a hydrogen bonding group in an enzyme inhibitor. *Science* **235**, 569–571 (1987).
252. Mucha, A., Grembecka, J., Cierpicki, T. & Kafarski, P. Hydrolysis of the phosphoramidate bond in phosphono dipeptide analogues— the influence of the nature of the N-terminal functional group. *Eur. J. Org. Chem.* **2003**, 4797–4803 (2003).
253. Breuer, E., Salomon, C. J., Katz, Y., Chen, W., Lu, S., Röschenthaler, G. V., Hadar, R. & Reich, R. Carbamoylphosphonates, a new class of in vivo active matrix metalloproteinase inhibitors. 1. Alkyl- and cycloalkylcarbamoylphosphonic acids. *J. Med. Chem.* **47**, 2826–2832 (2004).
254. Reich, R., Katz, Y., Hadar, R. & Breuer, E. Carbamoylphosphonate matrix metalloproteinase inhibitors 3: In vivo evaluation of cyclopentylcarbamoylphosphonic acid in experimental metastasis and angiogenesis. *Clin. Cancer Res.* **11**, 3925–3929 (2005).
255. Hoffman, A., Qadri, B., Frant, J., Katz, Y., Bhusare, S. R., Breuer, E., Hadar, R. &

- Reich, R. Carbamoylphosphonate matrix metalloproteinase inhibitors 6: cis-2-aminocyclohexylcarbamoylphosphonic acid, a novel orally active antimetastatic matrix metalloproteinase-2 selective inhibitor-synthesis and pharmacodynamic and pharmacokinetic analysis. *J. Med. Chem.* **51**, 1406–1414 (2008).
256. Farkas, E., Katz, Y., Bhusare, S., Reich, R., Rösenthaller, G. V., Königsmann, M. & Breuer, E. Carbamoylphosphonate-based matrix metalloproteinase inhibitor metal complexes: Solution studies and stability constants. Towards a zinc-selective binding group. *J. Biol. Inorg. Chem.* **9**, 307–315 (2004).
257. Grzywa, R. & Oleksyszyn, J. First synthesis of α -aminoalkyl-(N-substituted)thiocarbamoyl-phosphinates: Inhibitors of aminopeptidase N (APN/CD13) with the new zinc-binding group. *Bioorg. Med. Chem. Lett.* **18**, 3734–3736 (2008).
258. Jones, P., Altamura, S., Chakravarty, P. K., Cecchetti, O., Francesco, R. De, Gallinari, P., Ingenito, R., Meinke, P. T., Petrocchi, A., Rowley, M., Scarpelli, R., Serafini, S. & Steinkühler, C. A series of novel, potent, and selective histone deacetylase inhibitors. *Bioorg. Med. Chem. Lett.* **16**, 5948–5952 (2006).
259. Nielsen, T. K., Hildmann, C., Dickmanns, A., Schwienhorst, A. & Ficner, R. Crystal structure of a bacterial class 2 histone deacetylase homologue. *J. Mol. Biol.* **354**, 107–120 (2005).
260. Madsen, A. S., Kristensen, H. M. E., Lanz, G. & Olsen, C. A. The effect of various zinc binding groups on inhibition of histone deacetylases 1-11. *ChemMedChem* **9**, 614–626 (2014).

261. Ignatyev, I. S., Montejo, M., Rodríguez Ortega, P. G. & González, J. J. L. Quantum chemical study of silanediols as metal binding groups for metalloprotease inhibitors. *J. Mol. Model.* **19**, 1819–1834 (2013).
262. Kraker, A. J., Mizzen, C. A., Hartl, B. G., Miin, J., Allis, C. D. & Merriman, R. L. Modulation of histone acetylation by [4-(acetylamino)-N-(2-amino-phenyl) benzamide] in HCT-8 colon carcinoma. *Mol. Cancer Ther.* **2**, 401–408 (2003).
263. Suzuki, T., Ando, T., Tsuchiya, K., Fukazawa, N., Saito, A., Mariko, Y., Yamashita, T. & Nakanishi, O. Synthesis and histone deacetylase inhibitory activity of new benzamide derivatives. *J. Med. Chem.* **42**, 3001–3003 (1999).
264. Dong, M., Ning, Z. Q., Xing, P. Y., Xu, J. L., Cao, H. X., Dou, G. F., Meng, Z. Y., Shi, Y. K., Lu, X. P. & Feng, F. Y. Phase I study of chidamide (CS055/HBI-8000), a new histone deacetylase inhibitor, in patients with advanced solid tumors and lymphomas. *Cancer Chemother. Pharmacol.* **69**, 1413–1422 (2012).
265. Limban, C., Nuță, D. C., Chiriță, C., Negreș, S., Arsene, A. L., Goumenou, M., Karakitsios, S. P., Tsatsakis, A. M. & Sarigiannis, D. A. The use of structural alerts to avoid the toxicity of pharmaceuticals. *Toxicology Reports* **5**, 943–953 (2018).
266. Li, Y. & Woster, P. M. Discovery of a new class of histone deacetylase inhibitors with a novel zinc binding group. *Medchemcomm* **6**, 613–618 (2015).
267. Maola, K., Wilbs, J., Touati, J., Sabisz, M., Kong, X. D., Baumann, A., Deyle, K. & Heinis, C. Engineered peptide macrocycles can inhibit matrix metalloproteinases with high selectivity. *Angew. Chem. Int. Ed.* **58**, 11801–11805 (2019).
268. Nishino, N., Yoshikawa, D., Watanabe, L. A., Kato, T., Jose, B., Komatsu, Y., Sumida,

- Y. & Yoshida, M. Synthesis and histone deacetylase inhibitory activity of cyclic tetrapeptides containing a retrohydroxamate as zinc ligand. *Bioorg. Med. Chem. Lett.* **14**, 2427–2431 (2004).
269. Jacobsen, F. E., Lewis, J. A. & Cohen, S. M. A new role for old ligands: Discerning chelators for zinc metalloproteinases. *J. Am. Chem. Soc.* **128**, 3156–3157 (2006).
270. Chen, A. Y., Adamek, R. N., Dick, B. L., Credille, C. V., Morrison, C. N. & Cohen, S. M. Targeting metalloenzymes for therapeutic intervention. *Chem. Rev.* **119**, 1323–1455 (2019).
271. Akama, T., Baker, S. J., Zhang, Y. K., Hernandez, V., Zhou, H., Sanders, V., Freund, Y., Kimura, R., Maples, K. R. & Plattner, J. J. Discovery and structure-activity study of a novel benzoxaborole anti-inflammatory agent (AN2728) for the potential topical treatment of psoriasis and atopic dermatitis. *Bioorg. Med. Chem. Lett.* **19**, 2129–2132 (2009).
272. Marchetti, L. A., Kumawat, L. K., Mao, N., Stephens, J. C. & Elmes, R. B. P. The versatility of squaramides: from supramolecular chemistry to chemical biology. *Chem* **5**, 1398–1485 (2019).
273. Quiñonero, D., Frontera, A., Ballester, P. & Deyà, P. M. A theoretical study of aromaticity in squaramide and oxocarbons. *Tetrahedron Lett.* **41**, 2001–2005 (2000).
274. Malerich, J. P., Hagihara, K. & Rawal, V. H. Chiral squaramide derivatives are excellent hydrogen bond donor catalysts. *J. Am. Chem. Soc.* **130**, 14416–14417 (2008).

275. Zhao, B., Li, J. & Du, D. Squaramide-catalyzed asymmetric reactions. *Chem. Rec.* **17**, 994–1018 (2017).
276. Prohens, R., Tomàs, S., Morey, J., Deyà, P. M., Ballester, P. & Costa, A. Squaramido-based receptors: Molecular recognition of carboxylate anions in highly competitive media. *Tetrahedron Lett.* **39**, 1063–1066 (1998).
277. Garau, C., Frontera, A., Ballester, P., Quiñonero, D., Costa, A. & Deyà, P. M. A theoretical ab initio study of the capacity of several binding units for the molecular recognition of anions. *Eur. J. Org. Chem.* **2005**, 179–183 (2005).
278. Niewiadomski, S., Beebejaun, Z., Denton, H., Smith, T. K., Morris, R. J. & Wagner, G. K. Rationally designed squaryldiamides - A novel class of sugar-nucleotide mimics? *Org. Biomol. Chem.* **8**, 3488–3499 (2010).
279. Tomàs, S., Prohens, R., Vega, M., Rotger, M. C., Deyà, P. M., Ballester, P. & Costa, A. Squaramido-based receptors: Design, synthesis, and application to the recognition of tetraalkylammonium compounds. *J. Org. Chem.* **61**, 9394–9401 (1996).
280. Rotger, M. C., Piña, M. N., Frontera, A., Martorell, G., Ballester, P., Deyà, P. M. & Costa, A. Conformational preferences and self-template macrocyclization of squaramide-based foldable modules. *J. Org. Chem.* **69**, 2302–2308 (2004).
281. Ian Storer, R., Aciro, C. & Jones, L. H. Squaramides: Physical properties, synthesis and applications. *Chem. Soc. Rev.* **40**, 2330–2346 (2011).
282. Tomàs, S., Rotger, M. C., González, J. F., Deyà, P. M., Ballester, P. & Costa, A. Squaramide-based receptors: Synthesis and application to the recognition of

- polyalkyl ammonium salts. *Tetrahedron Lett.* **36**, 2523–2526 (1995).
283. Quiñonero, D., Frontera, A., Suñer, G. A., Morey, J., Costa, A., Ballester, P. & Deyà, P. M. Squaramide as a binding unit in molecular recognition. *Chem. Phys. Lett.* **326**, 247–254 (2000).
284. Onaran, M. B., Comeau, A. B. & Seto, C. T. Squaric acid-based peptidic inhibitors of matrix metalloprotease-1. *J. Org. Chem.* **70**, 10792–10802 (2005).
285. Charton, J., Déprez, B. P. & Déprez-Poulain, R. F. Synthesis of a 200-member library of squaric acid N-hydroxylamide amides. *Bioorg. Med. Chem. Lett.* **18**, 4968–4971 (2008).
286. Schaeffer, H. F. Squaric acid: Reactions with certain metals. *Microchem. J.* **17**, 443–455 (1972).
287. Liu, H., Tomooka, C. S. & Moore, H. W. An efficient general synthesis of squarate esters. *Synth. Commun.* **27**, 2177–2180 (1997).
288. Tantry, S. J., Markad, S. D., Shinde, V., Bhat, J., Balakrishnan, G., Gupta, A. K., Ambady, A., Raichurkar, A., Kedari, C., Sharma, S., Mudugal, N. V., Narayan, A., Naveen Kumar, C. N., Nanduri, R., Bharath, S., Reddy, J., Panduga, V., Prabhakar, K. R., Kandaswamy, K., *et al.* Discovery of imidazo[1,2-a]pyridine ethers and squaramides as selective and potent inhibitors of mycobacterial adenosine triphosphate (ATP) synthesis. *J. Med. Chem.* **60**, 1379–1399 (2017).
289. Kumawat, L. K., Abogunrin, A. A., Kickham, M., Pardeshi, J., Fenelon, O., Schroeder, M. & Elmes, R. B. P. Squaramide-naphthalimide conjugates as ‘turn-on’ fluorescent sensors for bromide through an aggregation-disaggregation approach. *Front.*

- Chem.* **7**, 354 (2019).
290. Wurm, F., Steinbach, T. & Klok, H. A. One-pot squaric acid diester mediated aqueous protein conjugation. *Chem. Commun.* **49**, 7815–7817 (2013).
291. Lefeber, D. J., Kamerling, J. P. & Vliegthart, J. F. G. Synthesis of *Streptococcus pneumoniae* type 3 neoglycoproteins varying in oligosaccharide chain length, loading and carrier protein. *Chem. Eur. J.* **7**, 4411–4421 (2001).
292. Ohara, K., Takeda, Y., Daikoku, S., Hachisu, M., Seko, A. & Ito, Y. Profiling aglycon-recognizing sites of UDP-glucose:glycoprotein glucosyltransferase by means of squarate-mediated labeling. *Biochemistry* **54**, 4909–4917 (2015).
293. Kamath, V. P., Diedrich, P. & Hindsgaul, O. Use of diethyl squarate for the coupling of oligosaccharide amines to carrier proteins and characterization of the resulting neoglycoproteins by MALDI-TOF mass spectrometry. *Glycoconj. J.* **13**, 315–319 (1996).
294. Yan, H., Aguilar, A. L. & Zhao, Y. Preparation of carbohydrate-oligonucleotide conjugates using the squarate spacer. *Bioorg. Med. Chem. Lett.* **17**, 6535–6538 (2007).
295. Zhao, Y., Tram, K. & Yan, H. Synthesis and characterization of mannosylated oligoribonucleotides. *Carbohydr. Res.* **344**, 2137–2143 (2009).
296. Ivancová, I., Pohl, R., Hubálek, M. & Hocek, M. Squaramate-modified nucleotides and DNA for specific cross-linking with lysine-containing peptides and proteins. *Angew. Chem. Int. Ed.* **58**, 13345–13348 (2019).
297. Fonvielle, M., Sakkas, N., Iannazzo, L., Le Fournis, C., Patin, D., Mengin-Lecreulx, D.,

- El-Sagheer, A., Braud, E., Cardon, S., Brown, T., Arthur, M. & Etheve-Quelquejeu, M. Electrophilic RNA for peptidyl-RNA synthesis and site-specific cross-linking with tRNA-binding enzymes. *Angew. Chem. Int. Ed.* **55**, 13553–13557 (2016).
298. Rudd, S. E., Roselt, P., Cullinane, C., Hicks, R. J. & Donnelly, P. S. A desferrioxamine B squaramide ester for the incorporation of zirconium-89 into antibodies. *Chem. Commun.* **52**, 11889–11892 (2016).
299. Martínez, L., Sampedro, A., Sanna, E., Costa, A. & Rotger, C. Synthesis and conformational studies of peptido-squaramide foldable modules: A new class of turn-mimetic compounds. *Org. Biomol. Chem.* **10**, 1914–1921 (2012).
300. Martínez, L., Martorell, G., Sampedro, Á., Ballester, P., Costa, A. & Rotger, C. Hydrogen bonded squaramide-based foldable module induces both β - And α -turns in hairpin structures of α -peptides in water. *Org. Lett.* **17**, 2980–2983 (2015).
301. Martínez-Crespo, L., Escudero-Adán, E. C., Costa, A. & Rotger, C. The Role of N-methyl squaramides in a hydrogen-bonding strategy to fold peptidomimetic compounds. *Chem. - A Eur. J.* **24**, 17802–17813 (2018).
302. Sampedro, A., Villalonga-Planells, R., Vega, M., Ramis, G., Fernández De Mattos, S., Villalonga, P., Costa, A. & Rotger, C. Cell uptake and localization studies of squaramide based fluorescent probes. *Bioconjug. Chem.* **25**, 1537–1546 (2014).
303. Fernández-Moreira, V., Alegre-Requena, J. V., Herrera, R. P., Marzo, I. & Gimeno, M. C. Synthesis of luminescent squaramide monoesters: Cytotoxicity and cell imaging studies in HeLa cells. *RSC Adv.* **6**, 14171–14177 (2016).
304. Silverberg, N. B., Lim, J. K., Paller, A. S. & Mancini, A. J. Squaric acid immunotherapy

- for warts in children. *J. Am. Acad. Dermatol.* **42**, 803–808 (2000).
305. Micali, G., Cicero, R. L., Nasca, M. R. & Sapuppo, A. Treatment of alopecia areata with squaric acid dibutylester. *Int. J. Dermatol.* **35**, 52–56 (2007).
306. Kinney, W. A., Abou-Gharbia, M., Garrison, D. T., Schmid, J., Kowal, D. M., Bramlett, D. R., Miller, T. L., Tasse, R. P., Zaleska, M. M. & Moyer, J. A. Design and synthesis of [2-(8,9-dioxo-2,6-diazabicyclo[5.2.0]non-1(7)-en-2-yl)-ethyl]phosphonic acid (EAA-090), a potent N-methyl-D-aspartate antagonist, via the use of 3-cyclobutene-1,2-dione as an achiral α -amino acid bioisostere. *J. Med. Chem.* **41**, 236–246 (1998).
307. Study evaluating EAA-090 in adult outpatients with neuropathic pain associated with diabetic neuropathy. Available at:
<https://clinicaltrials.gov/ct2/show/NCT00073034>. (Accessed: 30th April 2020)
308. Long-term study of the effects of navarixin (SCH 527123, MK-7123) in participants with moderate to severe COPD (MK-7123-019). Available at:
<https://clinicaltrials.gov/ct2/show/NCT01006616>. (Accessed: 30th April 2020)
309. Efficacy and safety study of navarixin (MK-7123) in combination with pembrolizumab (MK-3475) in adults with selected advanced/metastatic solid tumors (MK-7123-034). Available at:
<https://clinicaltrials.gov/ct2/show/NCT03473925>. (Accessed: 30th April 2020)
310. Lee, C.-W., Cao, H., Ichiyama, K. & Rana, T. M. Design and synthesis of a novel peptidomimetic inhibitor of HIV-1 Tat–TAR interactions: Squaryldiamide as a new potential bioisostere of unsubstituted guanidine. *Bioorg. Med. Chem. Lett.* **15**,

- 4243–4246 (2005).
311. Kim, C. U. & Misco, P. F. A facile synthesis of 1-hydroxy-2-phosphonocyclobutenedione. *Tetrahedron Lett.* **33**, 3961–3962 (1992).
312. Ishida, T., Shinada, T. & Ohfuné, Y. Synthesis of novel amino squaric acids via addition of dianion enolates derived from N-Boc amino acid esters. *Tetrahedron Lett.* **46**, 311–314 (2005).
313. Shinada, T., Ishida, T., Hayashi, K. ich, Yoshida, Y., Shigeri, Y. & Ohfuné, Y. Synthesis of leucine-enkephalin analogs containing α -amino squaric acid. *Tetrahedron Lett.* **48**, 7614–7617 (2007).
314. Maeda, K., Kiniwa, Y. I., Ohfuné, Y., Ishiguro, S., Suzuki, K., Murata, K., Matsuda, H. & Shinada, T. Solid phase synthesis of α -amino squaric acid-containing peptides. *RSC Adv.* **4**, 50639–50643 (2014).
315. Chan, P. C. M., Roon, R. J., Koerner, J. F., Taylor, N. J. & Honek, J. F. A 3-amino-4-hydroxy-3-cyclobutene-1,2-dione-containing glutamate analogue exhibiting high affinity to excitatory amino acid receptors. *J. Med. Chem.* **38**, 4433–4438 (1995).
316. Seio, K., Miyashita, T., Sato, K. & Sekine, M. Synthesis and properties of new nucleotide analogues possessing squaramide moieties as new phosphate isosters. *Eur. J. Org. Chem.* **2005**, 5163–5170 (2005).
317. Xie, J., Comeau, A. B. & Seto, C. T. Squaric acids: a new motif for designing inhibitors of protein tyrosine phosphatases. *Org. Lett.* **6**, 83–86 (2004).
318. Saha, A., Panda, S., Paul, S. & Manna, D. Phosphate bioisostere containing amphiphiles: a novel class of squaramide-based lipids. *Chem. Commun.* **52**, 9438–

- 9441 (2016).
319. Glüsenkamp, K.-H., Drosdziok, W., Eberle, G., Jähde, E. & Rajewsky, M. F. Squaric acid diethylester: a simple and convenient coupling reagent. *Zeitschrift für Naturforsch. C* **46**, 498–501 (1991).
320. Zhang, Y., Jumppanen, M., Maksimainen, M. M., Auno, S., Awol, Z., Ghemtio, L., Venkannagari, H., Lehtiö, L., Yli-Kauhaluoma, J., Xhaard, H. & Boije af Gennäs, G. Adenosine analogs bearing phosphate isosteres as human MDO1 ligands. *Bioorg. Med. Chem.* **26**, 1588–1597 (2018).
321. Soukarieh, F., Nowicki, M. W., Bastide, A., Pöyry, T., Jones, C., Dudek, K., Patwardhan, G., Meullenet, F., Oldham, N. J., Walkinshaw, M. D., Willis, A. E. & Fischer, P. M. Design of nucleotide-mimetic and non-nucleotide inhibitors of the translation initiation factor eIF4E: Synthesis, structural and functional characterisation. *Eur. J. Med. Chem.* **124**, 200–217 (2016).
322. Sato, K., Seio, K. & Sekine, M. Squaryl group as a new mimic of phosphate group in modified oligodeoxynucleotides: synthesis and properties of new oligodeoxynucleotide analogues containing an internucleotidic squaryldiamide linkage. *J. Am. Chem. Soc.* **124**, 12715–12724 (2002).
323. Sato, K., Tawarada, R., Seio, K. & Sekine, M. Synthesis and structural properties of new oligodeoxynucleotide analogues containing a 2',5'-internucleotidic squaryldiamide linkage capable of formation of a Watson-Crick base pair with adenine and a Wobble base pair with guanine at the 3'-downstream junction site. *Eur. J. Org. Chem.* **2004**, 2142–2150 (2004).

324. Shivalingam, A., Taemaitree, L., El-Sagheer, A. H. & Brown, T. Squaramides and ureas: a flexible approach to polymerase-compatible nucleic acid assembly. *Angew. Chem.* **132**, 11513–11519 (2020).
325. Lu, M., Lu, Q. Bin & Honek, J. F. Squarate-based carbocyclic nucleosides: Syntheses, computational analyses and anticancer/antiviral evaluation. *Bioorg. Med. Chem. Lett.* **27**, 282–287 (2017).
326. Rombola, M. & Rawal, V. H. Dicyclopentyl dithiosquarate as an intermediate for the synthesis of thiosquaramides. *Org. Lett.* **20**, 514–517 (2018).
327. Busschaert, N., Elmes, R. B. P., Czech, D. D., Wu, X., Kirby, I. L., Peck, E. M., Hendzel, K. D., Shaw, S. K., Chan, B., Smith, B. D., Jolliffe, K. A. & Gale, P. A. Thiosquaramides: pH switchable anion transporters. *Chem. Sci.* **5**, 3617–3626 (2014).
328. Ho, J., Zwicker, V. E., Yuen, K. K. Y. & Jolliffe, K. A. Quantum chemical prediction of equilibrium acidities of ureas, deltamides, squaramides, and croconamides. *J. Org. Chem.* **82**, 10732–10736 (2017).
329. Lu, T. & Wheeler, S. E. Origin of the superior performance of (thio)squaramides over (thio)ureas in organocatalysis. *Chem. - A Eur. J.* **19**, 15141–15147 (2013).
330. Rombola, M., Sumaria, C. S., Montgomery, T. D. & Rawal, V. H. Development of chiral, bifunctional thiosquaramides: enantioselective Michael additions of barbituric acids to nitroalkenes. *J. Am. Chem. Soc.* **139**, 5297–5300 (2017).
331. Yang, M., Chen, C., Yi, X., Li, Y., Wu, X., Li, Q. & Ban, S. Thiosquaramide-catalysed asymmetric double Michael addition of 2-(3:H)-furanones to nitroolefines. *Org.*

- Biomol. Chem.* **17**, 2883–2886 (2019).
332. Nagy, S., Dargó, G., Kisszékelyi, P., Fehér, Z., Simon, A., Barabás, J., Höltzl, T., Mátravölgyi, B., Kárpáti, L., Drahos, L., Huszthy, P. & Kupai, J. New enantiopure binaphthyl-cinchona thiosquaramides: synthesis and application for enantioselective organocatalysis. *New J. Chem.* **43**, 5948–5959 (2019).
333. Elmes, R. B. P., Busschaert, N., Czech, D. D., Gale, P. A. & Jolliffe, K. A. pH Switchable anion transport by an oxothiosquaramide. *Chem. Commun.* **51**, 10107–10110 (2015).
334. Howe, E. N. W., Busschaert, N., Wu, X., Berry, S. N., Ho, J., Light, M. E., Czech, D. D., Klein, H. A., Kitchen, J. A. & Gale, P. A. pH-Regulated nonelectrogenic anion transport by phenylthiosemicarbazones. *J. Am. Chem. Soc.* **138**, 8301–8308 (2016).
335. Kawana, M. & Kuzuhara, H. General method for the synthesis of 2'-azido-2',3'-dideoxynucleosides by the use of [1,2]-hydride shift and β -elimination reactions. *J. Chem. Soc., Perkin Trans. 1* **0**, 469–478 (1992).
336. Hakimelahi, G. H., Proba, Z. A. & Ogilvie, K. K. New catalysts and procedures for the dimethoxytritylation and selective silylation of ribonucleosides. *Can. J. Chem.* **60**, 1106–1113 (1982).
337. Bege, M., Bereczki, I., Herczeg, M., Kicsák, M., Eszenyi, D., Herczegh, P. & Borbás, A. A low-temperature, photoinduced thiol–ene click reaction: a mild and efficient method for the synthesis of sugar-modified nucleosides. *Org. Biomol. Chem.* **15**, 9226–9233 (2017).
338. Rao, B. G. Recent developments in the design of specific matrix metalloproteinase

- inhibitors aided by structural and computational studies. *Current Pharmaceutical Design* **11**, 295–322 (2005).
339. Kotikam, V. & Rozners, E. Concurrent hydrogenation of three functional groups enables synthesis of C3'-homologated nucleoside amino acids. *Org. Lett.* **19**, 4122–4125 (2017).
340. Kojima, N., Szabo, I. E. & Bruice, T. C. Synthesis of ribonucleic guanidine: replacement of the negative phosphodiester linkages of RNA with positive guanidinium linkages. *Tetrahedron* **58**, 867–879 (2002).
341. Ogawa, A., Tanaka, M., Sasaki, T. & Matsuda, A. Nucleosides and Nucleotides. 180. Synthesis and antitumor activity of nucleosides that have a hydroxylamino group instead of a hydroxyl group at the 2'- or 3'-position of the sugar moiety. *J. Med. Chem.* **41**, 5094–5107 (1998).
342. Robins, M. J., Samano, V. & Johnson, M. D. Periodinane oxidation, selective primary deprotection, and remarkably stereoselective reduction of tert-butyldimethylsilyl-protected ribonucleosides. Synthesis of 9-(β -D-xylofuranosyl)adenine or 3'-deuterioadenosine from adenosine. *J. Org. Chem.* **55**, 410–412 (1990).
343. Gündüzalp, A. B. & Erk, B. Copper(II) and zinc(II) complexes of thiophene/furan carboxamides: Synthesis, structure and properties. *Russ. J. Inorg. Chem.* **55**, 1094–1102 (2010).
344. Jones, P., Bottomley, M. J., Carfí, A., Cecchetti, O., Ferrigno, F., Lo Surdo, P., Ontoria, J. M., Rowley, M., Scarpelli, R., Schultz-Fademrecht, C. & Steinkühler, C. 2-

- Trifluoroacetylthiophenes, a novel series of potent and selective class II histone deacetylase inhibitors. *Bioorg. Med. Chem. Lett.* **18**, 3456–3461 (2008).
345. Ontoria, J. M., Altamura, S., Di Marco, A., Ferrigno, F., Laufer, R., Muraglia, E., Palumbi, M. C., Rowley, M., Scarpelli, R., Schultz-Fademrecht, C., Serafini, S., Steinkühler, C. & Jones, P. Identification of novel, selective, and stable inhibitors of class II histone deacetylases. Validation studies of the inhibition of the enzymatic activity of HDAC4 by small molecules as a novel approach for cancer therapy. *J. Med. Chem.* **52**, 6782–6789 (2009).
346. Coyle, J. D. The photochemistry of thiocarbonyl compounds. *Tetrahedron* **41**, 5393–5425 (1985).
347. Maciejewski, A. & Steer, R. P. The photophysics, physical photochemistry, and related spectroscopy of thiocarbonyls. *Chem. Rev.* **93**, 67–98 (1993).
348. Solans, X., Aguiló, M., Gleizes, A., Faus, J., Miguel Julve, L. & Verdaguer, M. Coordination modes of the squarate ligand: syntheses and crystal structures of six copper(II) squarate complexes. *Inorg. Chem.* **29**, 775–784 (1990).
349. Schmidt, D. & Lynch, J. Millipore Corporation Application Note. Lit. No. AN1728EN00 (2003).
350. Moreno, S., Brunner, M., Delazer, I., Rieder, D., Lusser, A. & Micura, R. Synthesis of 4-thiouridines with prodrug functionalization for RNA metabolic labeling. *RSC Chem. Biol.* **3**, 447–455 (2022).
351. Wiemer, A. J. Metabolic efficacy of phosphate prodrugs and the remdesivir paradigm. *ACS Pharmacol. Transl. Sci.* **3**, 613–626 (2020).

352. Sinokrot, H., Smerat, T., Najjar, A. & Karaman, R. Advanced prodrug strategies in nucleoside and non-nucleoside antiviral agents: a review of the recent five years. *Molecules* **22**, 1736 (2017).
353. Ray, A. S. & Hostetler, K. Y. Application of kinase bypass strategies to nucleoside antivirals. *Antiviral Res.* **92**, 277–291 (2011).
354. Sofia, M. J. Nucleotide prodrugs for the treatment of HCV infection. *Adv. Pharmacol.* **67**, 39–73 (2013).
355. Bannwarth, W. & Trzeciak, A. A simple and effective chemical phosphorylation procedure for biomolecules. *Helv. Chim. Acta* **70**, 175–186 (1987).
356. Stigliani, J. L. & Bernardes-Génisson, V. New insights into the chemical behavior of S-oxide derivatives of thiocarbonyl-containing antitubercular drugs and the influence on their mechanisms of action and toxicity. *Ann. Pharm. Françaises* **77**, 126–135 (2019).
357. Berney, M., Doherty, W., Jauslin, W. T., T Manoj, M., Dürr, E.-M. & McGouran, J. F. Synthesis and evaluation of squaramide and thiosquaramide inhibitors of the DNA repair enzyme SNM1A. *Bioorg. Med. Chem.* **46**, 116369 (2021).
358. Sheppard, T. L. & Breslow, R. C. Selective binding of RNA, but not DNA, by complementary 2',5'-linked DNA. *J. Am. Chem. Soc.* **118**, 9810–9811 (1996).
359. Wasner, M., Arion, D., Borkow, G., Noronha, A., Uddin, A. H., Parniak, M. A. & Damha, M. J. Physicochemical and biochemical properties of 2',5'-linked RNA and 2',5'-RNA:3',5'-RNA "hybrid" duplexes. *Biochemistry* **37**, 7478–7486 (1998).
360. Kandimalla, E., Manning, A., Zhao, Q., Shaw, D. R., Byrn, R. A., Sasisekharan, V. &

- Agrawal, S. Mixed backbone antisense oligonucleotides: design, biochemical and biological properties of oligonucleotides containing 2'-5'-ribo- and 3'-5'-deoxyribonucleotide segments. *Nucleic Acids Res.* **25**, 370–378 (1997).
361. Habibian, M., Harikrishna, S., Fakhoury, J., Barton, M., Ageely, E. A., Cencic, R., Fakh, H. H., Katolik, A., Takahashi, M., Rossi, J., Pelletier, J., Gagnon, K. T., Pradeepkumar, P. I. & Damha, M. J. Effect of 2'-5'/3'-5' phosphodiester linkage heterogeneity on RNA interference. *Nucleic Acids Res.* **48**, 4643–4657 (2020).
362. Sinha, S., Kim, P. H. & Switzer, C. 2',5'-Linked DNA is a template for polymerase-directed DNA synthesis. *J. Am. Chem. Soc.* **126**, 40–41 (2004).
363. Engelhart, A. E., Powner, M. W. & Szostak, J. W. Functional RNAs exhibit tolerance for non-heritable 2'–5' versus 3'–5' backbone heterogeneity. *Nat. Chem.* **5**, 390–394 (2013).
364. Prakash, T. P., Kraynack, B., Baker, B. F., Swayze, E. E. & Bhat, B. RNA interference by 2',5'-linked nucleic acid duplexes in mammalian cells. *Bioorg. Med. Chem. Lett.* **16**, 3238–3240 (2006).
365. Myers, A. G., Gin, D. Y. & Rogers, D. H. Synthetic studies of the tunicamycin antibiotics. preparation of (+)-tunicaminyuracil, (+)-tunicamycin-V, and 5'-epi-tunicamycin-V. *J. Am. Chem. Soc.* **116**, 4697–4718 (1994).
366. Pankiewicz, K. W. & Watanabe, K. A. Nucleosides. CXLIV. Some reactions of 2'-O-triflyl-2,3'-anhydroxylosyluracil with nucleophilic reagents. synthesis of 2'-chloro-2',3'-dideoxyuridinene. Studies directed toward the synthesis of 2'-deoxy-2'-substituted arabino nucleosides. *Chem. Pharm. Bull. (Tokyo)*. **35**, 4498–4502

- (1987).
367. Samano, V. & Robins, M. J. Nucleic acid related compounds. 60. Mild periodinane oxidation of protected nucleosides to give 2' - and 3'-ketonucleosides. The first isolation of a purine 2'-deoxy-3'-ketonucleoside derivative. *J. Org. Chem.* **55**, 5186–5188 (1990).
368. Nomura, M., Shuto, S., Tanaka, M., Sasaki, T., Mori, S., Shigeta, S. & Matsuda, A. Nucleosides and Nucleotides. 185. Synthesis and biological activities of 4'α-C-branched-chain sugar pyrimidine nucleosides. *J. Med. Chem.* **42**, 2901–2908 (1999).
369. De, S., Groaz, E. & Herdewijn, P. Tailoring peptide–nucleotide conjugates (PNCs) for nucleotide delivery in bacterial cells. *Eur. J. Org. Chem.* **2014**, 2322–2348 (2014).

