Development of Chemical Tools for Ten-Eleven Translocation Enzymes



Klemensas Šimelis

Exeter College
University of Oxford
Hilary term, 2022

A Thesis Submitted to the Board of the Faculty of Physical Sciences of the University of Oxford in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

Abstract

DNA methylation is one of the three epigenetic mechanisms regulating gene expression and is tightly regulated by DNA methyltransferase and demethylation mechanism activity. TET enzymes are key players in the active demethylation pathway and act in an O_2, Fe(II), and 2-oxoglutarate (2-OG)-dependent manner to sequentially oxidise 5-methylcytosine (^{5m}C) to 5-hydroxymethylcytosine (^{5hm}C), 5-formylcytosine (^{5f}C), and 5-carboxycytosine (^{5ca}C). While ^{5f}C and ^{5ca}C are prime targets for demethylation, all ^{5m}C derivatives are also epigenetic marks with distinct functions. Despite the importance of TETs in cellular processes and the prevalence of TET mutations in haematopoietic malignancies, *in vitro* characterisation and chemical tool development efforts have been surprisingly limited. The work described herein thus aims to advance biochemical TET characterisation and expand the chemical toolkit for TET probing.

In Chapter 2, the relation between substrate oxidation and 2-OG decarboxylation was investigated using a readily obtainable model TET homologue from *Naegleria gruberi*, NgTet1. Kinetic analysis by ^{1}H NMR revealed an unusually high 2-OG decarboxylation rate in the absence of substrate, further enhanced by both substrate and "non-substrate" DNA. TET inhibitors suppressed substrate and 2-OG turnover without discrimination. The results highlight the poor regulation of 2-OG consumption by NgTet1 and suggest that 2-OG turnover may be modulated by small molecules and oligonucleotides *in cellulo*.

Chapter 3 describes the development of chemical tools based on the TET cyclic peptide inhibitor TiP1. TiP1 demonstrated nanomolar potency and >30-fold selectivity for human TET1 (hTET1) over a panel of 2-OG dioxygenases *in vitro*, and SAR studies identified distinct interactions between TIP1 and hTET paralogues, permitting further selectivity optimisation. TiP1 functionalisation with non-obstructive handles and photoreactive moieties did not impact potency or selectivity. The work establishes TiP1 as the most potent and selective hTET1 inhibitor to date and supports the future utility of TiP1 derivatives as potential antibody mimetics and photo-affinity probes for studying TETs in a cellular context.

In Chapter 4, screening of novel small molecule scaffolds for TET inhibition identified 5-carboxy-2-thiouracils as low micromolar inhibitors of hTET2 with up to 50-fold selectivity over hTET1/3 and other 2-OG dioxygenases *in vitro*. Carboxylic acid analogues were inactive in cells, whereas methyl ester prodrugs inhibited hTETs *in cellulo* in a non-specific manner with $EC_{50} \approx 10$ μM. The novel class of cell-active pan-hTET inhibitors expands the chemical toolkit for TETs; furthermore, addressing the paralogue selectivity loss observed *in cellulo* may provide hTET2-selective inhibitors.

Chapter 5 describes work towards the structural characterisation of interactions between hTETs and TET activity modulators. Crystallisation efforts were initially hindered by the challenging production of a crystallisable hTET2 construct in *E. coli*, but extensive optimisation of expression and purification protocols furnished hTET2 in high purity and tractable yield. Previously reported crystallisation conditions and broad screens yielded promising crystalline precipitate, but further condition optimisation is required to obtain macrocrystals.

Overall, the work presented herein provides insight into the mechanistic understanding of the TET enzymes, describes progress towards the development of novel chemical tools for TET function probing, and details an advancement towards structure-guided design of TET inhibitors. The results lay the foundation for further characterisation of an important class of enzymes involved in epigenetic regulation in health and disease.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Akane Kawamura, without whom none of this work would have materialised. Akane has been a constant source of support, guidance, and encouragement throughout my DPhil, and continues to inspire me with her boundless energy, incredible time management, and knowledge of all things scientific. I am also very grateful to Professor Chris Schofield for his time, the stream of brilliant scientific ideas, and suggestions for collaborations with members of the Schofield group.

Dr Roman Belle and Eidarus (Eddy) Salah are the unsung heroes of this work. Roman, who has been my *de facto* day-to-day supervisor in the lab, has had to suffer through my myriad questions and a non-insignificant number of failed AlphaScreen experiments while I was getting to grips with every aspect of biochemistry. Support from Eddy has been absolutely indispensable for all recombinant protein production in terms of his experience, resources (hTET1 and hTET3!), and for his enthusiasm pushing the project forward.

I am indebted to many other colleagues that I've had the pleasure of working with throughout my DPhil, including Dr Anthony (Tony) Tumber for his kind assistance using the RapidFire instrument, Dr Oliver (Ollie) Coleman for his help with molecular biology and the SPOT synthesiser, Dr Nader Amin and Elisabete Pires for their help with NMR and MS, respectively, Dr Tobias (Toby) John for teaching me the joys of Western blotting, Dr Tom McAllister for keen insights into experimental design and the scientific approach in general, Dr Adam Hardy for the understated day-to-day lab support, and every other member of the Kawamura group, Schofield group, and CRL staff.

Extra thanks go to Akane, Roman, Raphael, and Tim for proof-reading (parts of) this work. Pleasingly, their detailed feedback has revealed an extremely well-behaved piece of written work.

I would be remiss if I didn't thank the many friends that have made the last five years a truly enjoyable experience. Roman, Tim, Raph, Toby, Jame, Ziggy, Freddie, Jos, and Will (lovingly referred to as "the boys") have all been great sources of personal support, digression, and discussions spanning a rather eclectic range of topics. I am forever thankful to Miglė for always being there for me during a tough first year and providing a little piece of Lithuania in Oxford. Thanks to Roman, Giorgia, Toby, Amelia, Freddie, Ziggy, Tika, Raph, Xiao, Patrick, Karina, and the rest of the LG2 crew for tolerating me throughout the years and making the lab a pleasant environment to work in. Special thanks to Cynthia for experimentally determining the tensile strength of a Pipetman L ejector button, and the LibertyBlue for demonstrating that flogging a dead horse long enough does eventually make it start working again. Finally, I would like to thank the Tesco Meal Deal for providing a steady source of nourishment throughout my degree.

Esu be galo dėkingas savo draugams Lietuvoje ir svetur. Ypač Žygimantui, Mariui, Martynui, ir Godai, kurie buvo, yra, ir, tikiuosi, bus vieni artimiausių man žmonių. Galiausiai, šis iššūkis būtų buvęs neįmanomas be nuolatinio palaikymo iš mano šeimos, kuri visada nekantriai laukė mano atostogų Lietuvoje, ir visus šiuos metus buvo namai, į kuriuos žinojau visados galėsiąs sugrįžti. Esu ypač dėkingas tėčiui Aleksandrui, kurio meilė, pasiaukojimas, rūpestis, palaikymas mokantis užsienyje, ir nuolatinis raginimas tobulėti neabejotinai buvo svarbūs veiksniai, nulėmę ne tik mano sprendimą siekti daktaro laipsnio, bet ir padėję atsidurti Oksforde.

Declaration and contribution list

I declare that the work discussed and figures presented in this thesis are entirely my own, unless a reference to a person or a published source has been made, with the exception of the following work:

Chapter 2:

Cofactor concentrations for AlphaScreen and MALDI-TOF MS assays were optimised previously by Dr Roman Belle (University of Oxford).

Chapter 3:

RaPID selection for hTET1-binding cyclic peptides and analysis of sequencing results was carried out by Kosuke Nishio (University of Tokyo).

A pET-28b plasmid containing an open reading frame coding for hTET2 catalytic domain with an insertion-deletion, with an N–terminal His_6-FLAG-[thrombin cleavage site]-SUMO (ubiquitin-like protein Smt3) tag was kindly gifted by Prof. Yanhui Xu (State Key Laboratory of Genetic Engineering, Fudan University, China).

Screening of TiP1–3 against KDM enzymes was carried out by Dr Anthony Tumber (University of Oxford).

Recombinant hTET1 and hTET3 enzymes were kindly produced by Eidarus Salah (University of Oxford).

All cell-based assays were carried out by Dr Hilal Sarac (Newcastle University).

RaPID selection for hTET2-binding cyclic peptides and analysis of sequencing results was carried out by Dr Tom McAllister (Newcastle University).

Chapter 4:

All compounds, excluding IOX1, were synthesised by the group of Assoc. Prof. Dante Rotili (Sapienza University of Rome).

All cell-based assays were carried out by Dr Hilal Sarac (Newcastle University).

Chapter 5:

TEV protease was kindly provided by Dr Patrick Rabe; Ulp3 protease was kindly provided by Eidarus Salah; SenP2 protease was kindly provided by Dr Zhihong Zhang (University of Oxford).

Broad screen crystallisation trials using tag-free hTET2 were carried out by Dr Arnaud Basle (Newcastle Structural Biology Laboratory).

Contents

Abstract ..... **i**

Acknowledgements ..... **ii**

Declaration and contribution list ..... **iv**

Contents ..... **vi**

Abbreviations ..... **x**

Chapter 1 Introduction **1**

1.1 Epigenetics 1

1.2 Epigenetic modifications to chromatin 3

1.2.1 Chromatin structure 3

1.2.2 DNA methylation by DNA methyltransferases 5

1.2.3 5-Methylcytosine oxidation by Ten-Eleven Translocation enzymes 7

1.2.4 DNA demethylation 8

1.2.5 Other epigenetic nucleobase modifications 10

1.3 Biochemical characterisation of Ten-Eleven Translocation enzymes 12

1.3.1 Shared features of the Fe(II)/2-OG-dependent dioxygenase superfamily 12

1.3.2 Uncoupled 2-OG decarboxylation and L-ascorbate dependence 14

1.3.3 Sequential ^{5m}C oxidation 16

1.3.4 Substrate specificity 17

1.4 Structural characterisation of Ten-Eleven Translocation enzymes 20

1.4.1 2-OG-dependent TETs 20

1.4.2 L-Ascorbate-dependent *C. reinhardtii* TET homologue, CMD1 24

1.5 Biological significance of Ten-Eleven Translocation enzymes 27

1.5.1 Function of ^{5m}C oxidation products 27

1.5.2 TET function in development 28

1.5.3 Aberrant TET function in cancers 29

1.6 Inhibitors of Ten-Eleven Translocation enzymes 31

1.6.1 Endogenous inhibitors 31

1.6.2 Fe(II)-chelating 2-OG mimetics 32

1.6.3 2-OG-non-competitive Fe(II) chelators 35

1.6.4 Cytosine mimetics 36

1.6.5 Cyclic peptide inhibitors 37

1.7 Thesis objectives 38

Chapter 2 NgTet1 biochemical characterisation and inhibitor screening using DNA oxidation- and 2-OG decarboxylation-based detection methods **40**

2.1 Introduction 40
2.1.1 Methods for TET catalytic activity detection 40
2.2 Aims 44
2.3 Results 46
2.3.1 Production of *Naegleria* Tet-like dioxygenase NgTet1 46
2.3.2 Selection of catalytic detection methods 49
2.3.3 NgTet1 catalytic activity detection using AlphaScreen 49
2.3.4 NgTet1 catalytic activity detection using MALDI-TOF MS 54
2.3.5 NgTet1 catalytic activity detection using Succinate-Glo™ 57
2.3.6 NgTet1 catalytic activity detection using ^{1}H NMR spectroscopy 60
2.3.7 NgTet1 kinetic parameter determination by ^{1}H NMR 64
2.3.8 Characterisation of TET inhibitors against NgTet1 68
2.4 Discussion 73
2.5 Conclusions and future work 79
Chapter 3 Development of cyclic peptide inhibitors for human TET enzymes 81
3.1 Introduction 81
3.2 Aims 84
3.3 Results 85
3.3.1 Characterisation of cyclic peptide TET inhibitors TiP1–3 85
3.3.2 Characterisation of additional cyclic peptide sequences from the RaPID selection against hTET1 91
3.3.3 Characterisation of TiP1–3 and KS01–13 against hTET2 using mass spectrometry 94
3.3.4 Evaluation of TiP1 cellular activity 98
3.3.5 Characterisation of TiP1 alanine variants 101
3.3.6 Synthesis of TiP1-cell-penetrating peptide (CPP) conjugates 103
3.3.7 Evaluation of TiP1-CPP conjugate inhibitory activity using AlphaScreen 108
3.3.8 Evaluation of TiP1-CPP conjugate cellular activity 110
3.3.9 Investigation of TiP1 binding site 112
3.3.10 Characterisation of cyclic peptides from RaPID selection against hTET2 117
3.4 Discussion 120
3.5 Conclusions and future work 124
Chapter 4 Development of small molecule inhibitors for human TET enzymes 126
4.1 Introduction 126
4.2 Aims 128
4.3 Results 129

4.3.1 Identification of hTET2 inhibitor scaffolds 129
4.3.2 Structural features and synthesis of *S*–substituted 5-carboxy-2-thiouracils 133
4.3.3 Scaffold modification of *S*–benzyl 5-carboxy-2-thiouracil 134
4.3.4 *S*–alkyl 5-carboxy-2-thiouracil analogues 136
4.3.5 *S*–(aryl)alkyl 5-carboxy-2-thiouracil analogues 138
4.3.6 *S*–unsubstituted (thione) analogues 141
4.3.7 hTET paralogue selectivity profiling 143
4.3.8 Cellular activity evaluation for selected 5-carboxylic acid analogues 147
4.3.9 Evaluation of ester prodrug cellular activity 148
4.4 Discussion 152
4.5 Conclusions and future work 156
Chapter 5 Production and crystallisation of human TET2 truncated catalytic domain, hTET2$_{CD\Delta LCI}$ 158
5.1 Introduction 158
5.2 Aims 159
5.3 Results 160
5.3.1 Production of SUMO-tagged hTET2$_{CD\Delta LCI}$ 160
5.3.2 SUMO-tagged hTET2$_{CD\Delta LCI}$ crystallisation trials 164
5.3.3 Optimisation of hTET2$_{CD\Delta LCI}$ expression 165
5.3.4 Optimisation of hTET2 purification 178
5.3.5 Crystallisation of cleaved hTET2 188
5.4 Discussion 190
5.5 Conclusions and future work 195
Chapter 6 Summary and conclusions 197
Chapter 7 Materials and methods 205
7.1 General information 205
7.2 Peptide synthesis 205
7.3 General molecular biology methods 209
7.4 Restriction cloning of pCOLD I 211
7.5 Ligation-independent cloning (LIC): general protocol 213
7.6 General microbiology methods 216
7.7 Recombinant protein expression 220
7.8 Recombinant protein purification 222
7.9 Ni Sepharose resin IMAC for expression trials 229
7.10 Western Blot 231
7.11 SPOT peptide array 232
7.12 2-Oxoglutarate (2-OG) decarboxylation monitoring by ^{1}H NMR 235

7.13 ^{5m}C oxidation monitoring by Amplified Luminescent Proximity Homogenous Assay (AlphaScreen) 238
7.14 Succinate-Glo™ bioluminescent succinate detection assay 243
7.15 ^{5m}C oxidation monitoring using solid-phase extraction-mass spectrometry (SPE-MS) 245
7.16 ^{5m}C oxidation monitoring using Matrix-Assisted Laser Desorption/Ionisation-Time of Flight Mass Spectrometry (MALDI-TOF MS) 249
7.17 Photoreactive probe cross-linking 251
7.18 Protein LC-MS 253
7.19 Peptide LC-MS 254
7.20 Protein crystallography 255
Chapter 8 References 257
Appendix 1 DNA oligomer information ..... a
Appendix 2 Peptide purity and accurate mass information ..... b
Appendix 3 Small molecule inhibitor numbering ..... d
Appendix 4 Protein construct information ..... e

Abbreviations

^{1m}A	*N*1–Methyladenine (as a ribonucleotide)
2,4-PDCA	2,4-Pyridinedicarboxylic acid
2,5-DHAP	2,5-Dihydroxyacetophenone
2-OG	2-Oxoglutarate
3-HPA	3-Hydroxypicolinic acid
^{5}C	Unmodified cytosine (5–*H*) as a deoxyribonucleotide **Note**: All nucleobases and their modifications, including, but not limited to, ^{5}C, ^{5m}C, ^{5hm}C, ^{5f}C, ^{5ca}C, refer to their corresponding deoxyribonucleotides in DNA oligomers unless otherwise mentioned
^{5ca}C	5-Carboxycytosine
^{5ca}U	5-Carboxyluracil
^{5e}C	5-Ethylcytosine
^{5et}C	5-Ethynylcytosine
^{5f}C	5-Formylcytosine
^{5f}U	5-Formyluracil
^{5gm}C	5-Glyceryl-methylcytosine
^{5hm}C	5-Hydroxymethylcytosine
^{5hm}U	5-Hydroxymethyluracil
^{5m}C	5-Methylcytosine
^{5ox}C	Oxidised 5-methylcytosine derivatives
^{5v}C	5-Vinylcytosine
^{5x}C	Modified cytosine
^{6m}A	*N*6–Methyladenine
A260/A280	UV light absorbance ratio (260/280 nm)
AID	Activation-induced deaminase
ALKBH3	Human ALKB homologue 3
AlphaScreen	Amplified luminescent proximity homogeneous assay screen
AML	Acute myeloid leukaemia
Amp	Ampicillin
APOBEC	Apolipoprotein B mRNA-editing enzyme complex
APS	Ammonium persulfate
ATP	Adenosine triphosphate
Base J	β-D-Glucopyranosyloxymethyluracil
BC	Before CXXC
BER	Base excision repair
Bno	5-*n*-Butyl-5-nonyl ester
Boc	*t*-Butyloxycarbonyl protecting group
bp	Base pair
Bpa	4-Benzoyl-L-phenylalanine
BSA	Bovine serum albumin
C. reinhardtii	*Chlamydomonas reinhardtii*
Camp	Chloramphenicol
CD	Catalytic domain

cDNA	Complementary DNA
$hTET2_{CD\Delta LCI}$	hTET2 catalytic domain without low complexity insert
CHES	*N*–Cyclohexyl-2-aminoethanesulfonic acid
CMD1	^{5m}C-modifying *C. reinhardtii* TET homologue
CP	Cyclic peptide
CpG	Cytidine phosphate guanosine
CPP	Cell-penetrating peptide
CTAB	Cetrimonium (cetyltrimethylammonium) bromide
CV	Column volume
DAC	Dibasic ammonium citrate
DAPI	4',6-Diamidino-2-phenylindole
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DIC	*N*,*N*'–Diisopropylcarbodiimide
DIPEA	*N*,*N*–Diisopropylethylamine
DMF	*N*,*N*–Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DnaK	*E. coli* Hsp70 molecular chaperone
DNMT	DNA methyltransferase
dox	Doxycycline
DSBH	Double-stranded β-helix
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EC_{50}	Half-maximal effective concentration
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Epag	Eltrombopag
FA	Formic acid
FH	Fumarate hydratase (fumarase)
FIH	Factor inhibiting HIF
FLAG tag	Polypeptide with sequence DYKDDDDK
Fmoc	Fluorenylmethoxycarbonyl protecting group
FT	Flow-through
GFP	Green fluorescent protein
GST	Glutathione *S*–transferase
$H3K9me_2$	Histone H3 N^{ε}–dimethyllysine 9
$H3K9me_3$	Histone H3 N^{ε}–trimethyllysine 9
HDAC	Histone deacetylase
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HIF PHD	Hypoxia-inducible factor prolyl hydroxylase
hNTH1	Human endonuclease III homologue
HOBt	Hydroxybenzotriazole
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry

HRP	Horseradish peroxidase
hTET	Human TET
IC_{50}	Half-maximal inhibitory concentration
IDH	Isocitrate dehydrogenase
IEX	Ion-exchange chromatography
IMAC	Immobilised metal affinity chromatography
IOX1	8-Hydroxy-5-quinolinecarboxylic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVTT	*In vitro* transcription/translation
Kan	Kanamycin
KDM	Histone lysine demethylase
K_M^{app}	Apparent K_M
KO	Gene knockout
LB	Luria-Bertani medium
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LIC	Ligation independent cloning
MALDI-TOF MS	Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry
MAP	Methionine aminopeptidase
MBHA	4-Methylbenzhydrylamine
MBP	Maltose-binding protein (Chapter 5)
MBP	^{5m}CpG-binding protein (Chapter 1)
MES	2-Ethanesulfonic acid
mESC	Mouse embryonic stem cell
MLL	Mixed-lineage leukaemia
Mpe	β-3-Methylpent-3-yl ester
mRNA	Messenger RNA
mTet	Murine Tet
MW	Molecular weight
MWCO	Molecular weight cut-off
n.d.	Not determined
n.i.	No inhibition
N6AMT1	*N*6–Adenine-specific DNA methyltransferase 1
NgTet1	*Naegleria* Tet-like dioxygenase
n_H	Hill coefficient
NME	N–terminal methionine excision
NMP	*N*–Methyl-2-pyrrolidone
NMR	Nuclear magnetic resonance
NOG	*N*–oxalylglycine
OAA	Octylammonium acetate
OD_{600}	Optical density at 600 nm wavelength
Oxyma	Ethyl cyanohydroxyiminoacetate
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline with 0.1% v/v Tween® 20
PCR	Polymerase chain reaction
PDF	Peptide deformylase

pEC_{50}	The negative logarithm of the EC_{50}
PEG 2000 MME	Polyethylene glycol, average molecular weight 2000 g/mol, monomethyl ether
PEG 300–2000	Polyethylene glycol, average molecular weight 300–2000 g/mol
PGC	Primordial germ cell
PHD	Prolyl 4-hydroxylase
pI	Protein isoelectric point
pIC_{50}	The negative logarithm of the IC_{50}
PLG	6-Phosphogluconolactonase
PTM	Protein post-translational modification
Q-TOF	Quadrupole time-of-flight
R8 tag	Polypeptide with sequence RRRRRRRR
RaPID	Random nonstandard peptide integrated discovery
ROS	Reactive oxygen species
rpm	Revolutions per minute
RRID	Research Resource Identifier
RT	Room temperature
S/N ratio	Signal-to-noise ratio
SAH	*S*–Adenosylhomocysteine
SAM	*S*–Adenosylmethionine
SAR	Structure-activity relationship
SARS-CoV	Severe acute respiratory syndrome-coronavirus
SD	Standard deviation
SDH	Succinate dehydrogenase
SDR	Succinate detection reagent
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEM	Standard error of the mean
SenP2	SUMO-specific peptidase 2
SIDD	Stress-induced duplex destabilisation regions
SMUG1	Single-stranded mono-functional uracil DNA *N*–glycosylase
SPE-MS	Solid-phase extraction-mass spectrometry
SPPS	Solid-phase peptide synthesis
ssDNA	Single-stranded DNA
SUMO	Ubiquitin-like protein Smt3
T	Thymine
T4 β-GT	T4 phage β-glucosyltransferase
TAE	Tris, acetic acid, and EDTA buffer
TAT tag	HIV Tat protein-derived polypeptide with sequence YGRKKRRQRRR
TauD	Taurine dioxygenase
TB	Terrific Broth medium
TBAF	Tetra-*n*-butylammonium fluoride
TCA	Tricarboxylic acid
TCEP	Tris(2-carboxyethyl)phosphine
TDG	Thymine DNA glycosylase

TEMED	Tetramethylethylenediamine
TET	Ten-Eleven Translocation
TEV	Tobacco Etch Virus
TF	Transcription factor
TFA	Trifluoroacetic acid
TfdA	2,4-Dichlorophenoxyacetate dioxygenase
THFA	Tetrahydrofolic acid
TIPS	Triisopropyl silane
TLC	Thin-layer chromatography
Tris	Tris(hydroxymethyl)aminomethane
Tris-d_{11}	Tris(hydroxymethyl-d_3)amino-d_2-methane
tRNA	Transfer RNA
Trx	Thioredoxin
TSP-d_4	3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid
UDP	Uridine diphosphate
v/v	Volume by volume
V_{max}^{app}	Apparent V_{max}
VHL	Von Hippel–Lindau protein
w/v	Weight by volume
XPS	X-ray photoemission spectroscopy
α-CHCA	α-Cyano-4-hydroxycinnamic acid

Chapter 1
Introduction

1.1 Epigenetics

Complex multicellular eukaryotic organisms all begin from a single embryonic stem cell – a cell that is both extremely proliferative and completely unspecialised. Any subsequent diploid cells produced *via* mitosis contain virtually identical genetic information, and so should produce virtually identical cells. However, cellular differentiation during embryonic development leads to the occurrence of a considerable number of phenotypically distinct cell types. According to the central dogma of molecular biology proposed by Francis Crick, sequential information can only be transferred from nucleic acids to proteins, but not in reverse[1]. Since cell phenotype is dictated by the gene products known as proteins, it follows that determination of cell type must occur at the level of gene expression. Gene expression is dependent on DNA accessibility and chromatin structure, regulated by a complex web of chemical modifications, which must be introduced and removed in a tightly controlled spatiotemporal manner to maintain a co-operative multicellular system[2].

The study of chemical modifications to chromatin (DNA and associated histone proteins) that do not involve any alteration to the underlying DNA sequence is termed epigenetics (*ἐπι* - "over, outside of" + *genetics*), initially coined in the 1940s by C. H. Waddington[3]. While epigenetic modifications are typically considered transient, they are not only mitotically and/or meiotically heritable, but can also be inherited trans-generationally through genomic imprinting[4,5]. Epigenetic control is essential for normal development and allows cells to respond and adapt to exogenous stimuli[6]. However, erroneous epigenetic marks are responsible for a plethora of diseases – ranging from cardiovascular disease to cancer – which can occur at any time during the life cycle[7]. The major epigenetic control pathways identified

to date are DNA modification, post-translational modification of histone proteins, and functional non-coding RNA regulation[8]. It is important to note that these pathways do not function in isolation, and epigenetic crosstalk exists between DNA modifications and other chromatin components.

1.2 Epigenetic modifications to chromatin

1.2.1 Chromatin structure

Human genomic DNA, present in full in every cell, is approximately 2 meters in length, and is packaged neatly into chromatin – a macromolecular complex of DNA and proteins – in order to regulate gene expression and DNA replication, and to avoid DNA damage[9]. The fundamental subunit of chromatin is the nucleosome – a small segment of DNA wrapped around a protein octamer composed of two copies of four core histone proteins – H2A, H2B, H3, and H4 (Figure 1.1)[10]. Histone octamer formation begins with the heterodimerisation of H2A with H2B, and H3 with H4. Two H3/H4 dimers self-associate to form a H4/H3:H3/H4 tetramer, which combines with two H2A/H2B heterodimers to finally form the histone octamer complex. Positively charged histones form electrostatic interactions with the negatively charged DNA, enabling the octamers to act as molecular "spools". Approximately 147 base pairs (bp) of DNA can wrap around each octamer, with a DNA linker connecting the nucleosomes together[11]. The length of linker DNA can vary from 10 bp to 80 bp, depending on tissue type and species[12]. Nucleosome structure is stabilised by linker histone H1, which binds to both the nucleosome and the linker DNA. Additionally, histone H1 plays a crucial role in higher-order chromatin architecture by enabling further compaction of nucleosomes through the formation of irregular chromatin fibre[13]. Histones also contain highly basic N–terminal tails, which can protrude from their own nucleosomes to make contact with adjacent nucleosomes[14], and can be decorated with over 15 post-translational modifications (PTMs) with distinct biological functions[15,16].

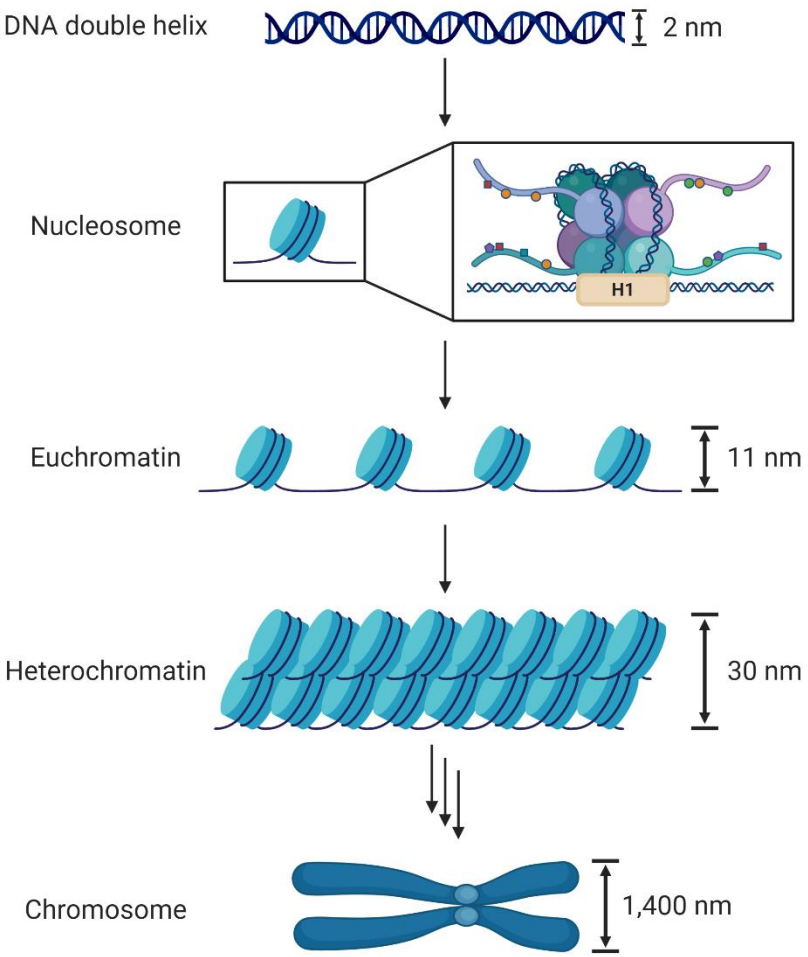


Figure 1.1 | Simplified schematic of DNA packaging. DNA double helix is wrapped around histone octamers and capped with linker histone H1 to form nucleosomes. Nucleosomes and linker DNA form the loosely packed euchromatin, which can be condensed into heterochromatin. Further packing results in DNA organisation into chromosomes. Coloured shapes on histone tails represent potential post-translational modifications. Adapted from literature[12]; created with BioRender.com.

In addition to the tightly controlled compression of genetic material, chromatin also enables spatiotemporal regulation of gene transcription[17]. There are two forms of chromatin: euchromatin, which is an open form, and heterochromatin, which is a densely packed form[18]. The dense packing in heterochromatin prohibits translational machinery from accessing genetic information, effectively silencing the associated genes. The interchange between euchromatin and heterochromatin is regulated by a host of proteins that introduce and remove chemical modifications on DNA and histones.

1.2.2 DNA methylation by DNA methyltransferases

DNA methylation at the C5 position of cytosine to form 5-methylcytosine† (^{5m}C) is the most prevalent epigenetic DNA modification in human tissues, corresponding to around 4.0–4.5% of all cytosines[19]. DNA methylation is generally associated with transcriptional repression, particularly in promoter regions[20], and operates *via* two main pathways: (i) obstruction of DNA-binding proteins that act as or recruit transcriptional activators and factors, and (ii) recruitment of ^{5m}CpG-binding protein (MBP) family proteins, which in turn recruit transcriptional corepressor complexes, ultimately leading to deacetylated heterochromatin (Figure 1.2)[21]. Members of the MBP family, including ^{5m}CpG binding protein 2 (MeCP2), contain a strongly conserved 70–75 amino acid ^{5m}CpG binding domain (MBD) that recognises symmetrically methylated CpG dinucleotides and discriminates hemimethylated and non-methylated CpG dinucleotides[22]. MBPs bound to symmetrically methylated DNA recruit transcriptional corepressors, such as histone deacetylase (HDAC) complexes, and proteins involved in chromatin compaction[23–25], resulting in gene silencing.

† Nucleobase modifications, including ^{5m}C, ^{5hm}C, ^{5f}C, ^{5ca}C, and others, refer to modifications of deoxyribonucleotides in DNA oligomers unless otherwise mentioned.

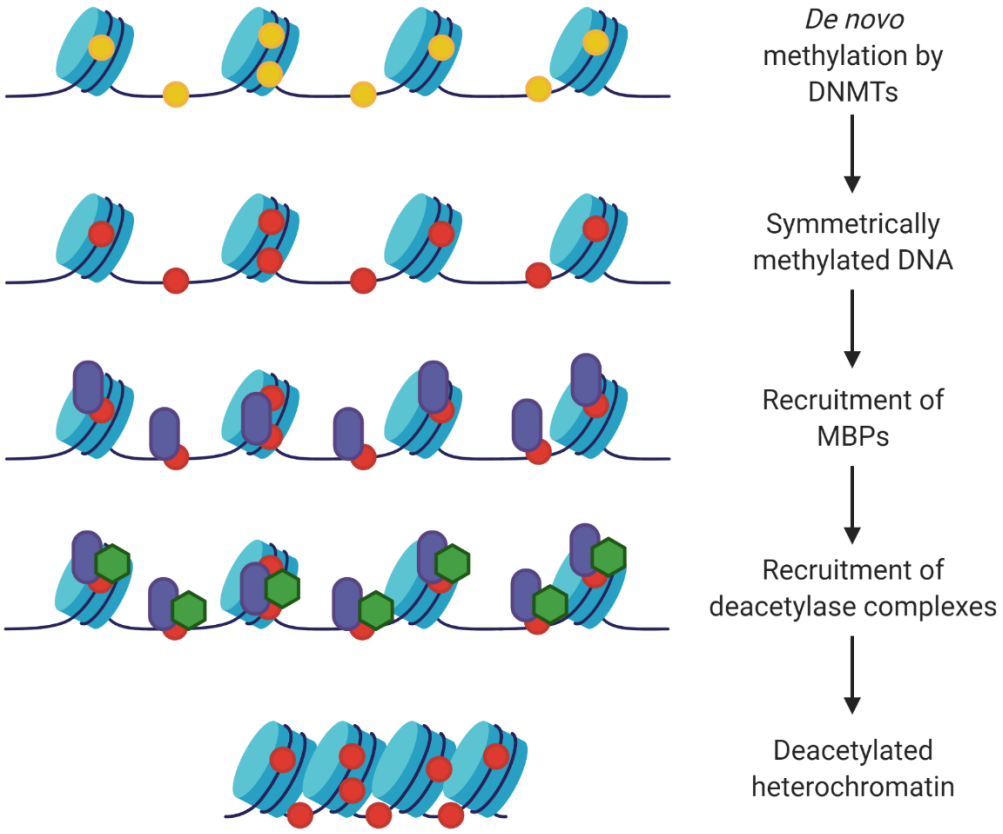
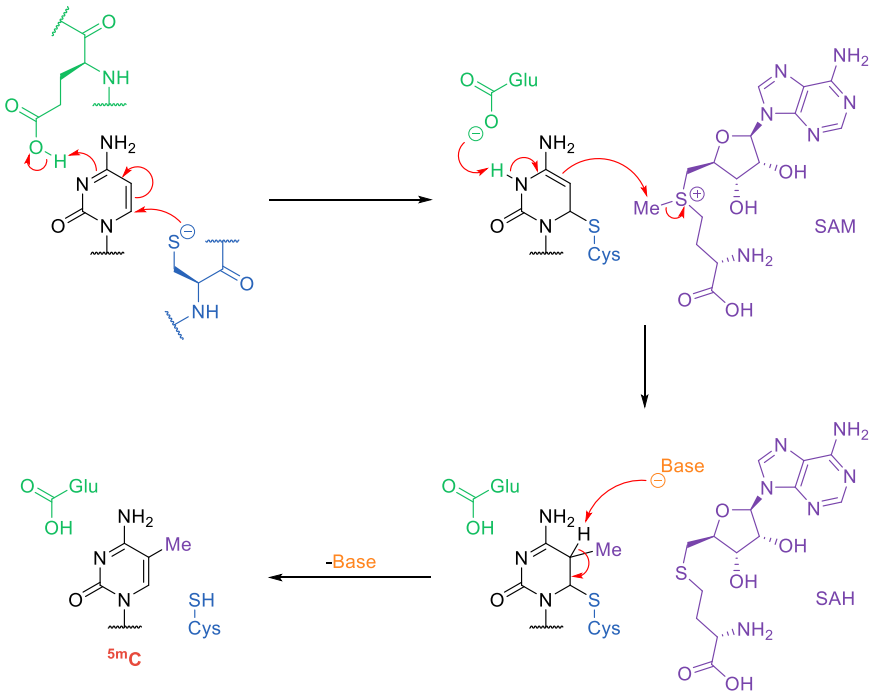


Figure 1.2 | Schematic representation of d*e novo* DNA methylation-mediated gene silencing. Symmetric methylation of DNA by DNMTs recruits MBPs, which recruit histone deacetylase complexes; deacetylation of chromatin results in condensation to heterochromatin. Figure adapted from literature[21]; created with biorender.com.

^{5m}C is introduced by a class of enzymes known as DNA methyltransferases (DNMTs), which utilise *S*–adenosylmethionine (SAM) as a methyl donor[26,27]. The catalytic mechanism[28,29] begins with a base flipping in order to situate the target cytosine in the DNMT active site. A conserved Glu residue facilitates a nucleophilic attack from the Cys side chain thiol to C6 of cytosine, resulting in the formation of an enamine (Scheme 1.1). A methyl group is then transferred from SAM onto C5, followed by a deprotonation on C5 by a DNMT-provided base to form ^{5m}C and *S*–adenosylhomocysteine (SAH). The Cys thiol is released, and the Glu residue is regenerated for the next catalytic cycle. There are three cytosine DNMTs in humans: DNMT1 preserves methylation states by methylating newly synthesised non-methylated strands on hemimethylated double-stranded DNA, thus retaining a symmetric ^{5m}CpG dinucleotide enrichment[29], while DNMT3A and DNMT3B are *de novo* methyltransferases and are involved in the methylation of new sites[30].



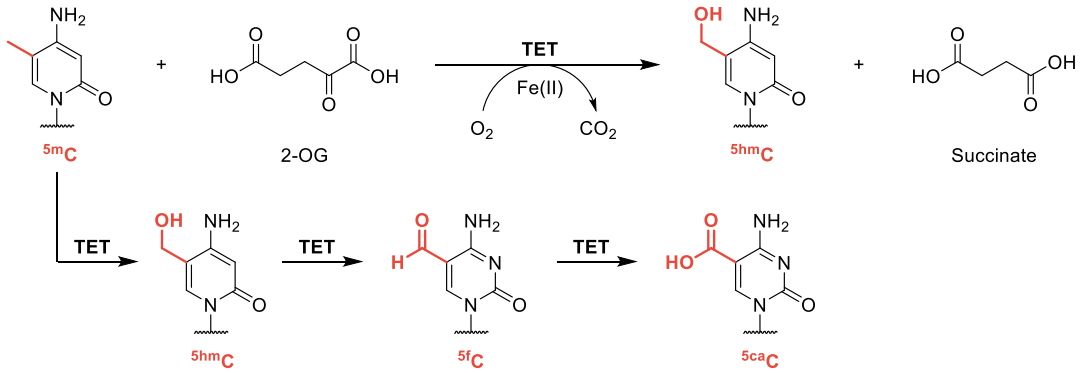
Scheme 1.1 | The catalytic mechanism of DNMT, highlighting the formation of ^{5m}C using SAM as a cofactor[28,29]. Glutamine and cysteine residues are part of the DNMT active site. **SAM**, *S*–adenosylmethionine; **SAH**, *S*–adenosylhomocysteine.

While cytosine methylation is mostly associated with transcriptional silencing, further studies have revealed the effect to be context-dependent, and multiple genes have been reported for which cytosine methylation in fact promotes expression[31,32]. Although most classes of transcription factors (TFs) do not bind to methylated DNA, TFs such as homeodomain, POU, and NFAT proteins bind preferentially to ^{5m}CpG[33]. Furthermore, mounting evidence suggests that TF binding could dictate the methylation state of neighbouring DNA by recruiting DNA-modifying enzymes[34]. In mammals, ^{5m}C plays a crucial role in the maintenance of cellular functions, genomic stability, and processes such as X-chromosome inactivation and genomic imprinting[35]. Consequently, aberrant cytosine methylation levels are frequently a hallmark of human cancers[36].

1.2.3 5-Methylcytosine oxidation by Ten-Eleven Translocation enzymes

Ten-Eleven Translocation (TET) enzymes, named after a chromosomal translocation common in acute myeloid leukaemia (AML)[37,38], were reported in 2009 as ^{5m}C hydroxylases that utilised Fe(II), 2-OG, and O_2 as cofactors to oxidise the 5-methyl position of ^{5m}C, forming

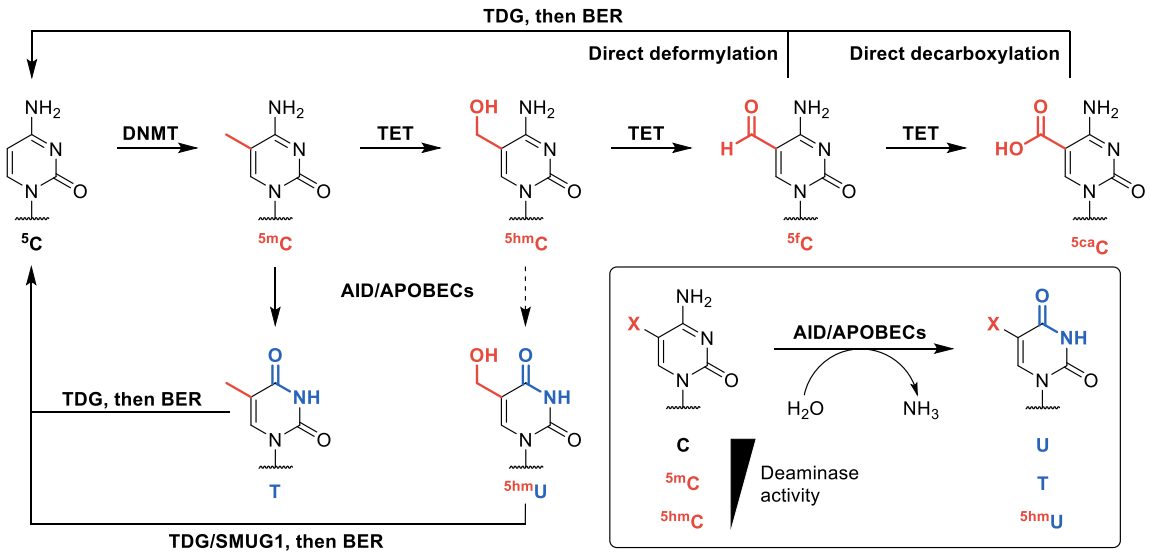
5-hydroxymethylcytosine (^{5hm}C, Scheme 1.2)[39]. Further studies on the enzymes found that TETs could also generate 5-formylcytosine (^{5f}C) and 5-carboxycytosine (^{5ca}C) *via* subsequent oxidation of ^{5hm}C (Scheme 1.2)[40,41]. The discovery of TET function prompted investigations into the potential biological significance of ^{5m}C oxidation products (^{5ox}C), revealing distinct roles for each modification (see section 1.5.1).



Scheme 1.2 | Catalytic pathway for generation of ^{5m}C oxidation products by the Fe(II)/2-OG-dependent TET dioxygenases. TET, Ten-Eleven Translocation enzymes; **^{5hm}C**, 5-hydroxymethylcytosine; **^{5f}C**, 5-formylcytosine; **^{5ca}C**, 5-carboxycytosine.

1.2.4 DNA demethylation

DNA demethylation can be carried out either passively or actively. Passive DNA demethylation takes place during successive rounds of DNA replication in the absence of functional DNA methylation maintenance mechanisms, such as following inhibition or dysfunction of DNMT1. By contrast, active DNA demethylation requires enzymatic intervention to modify or remove the methyl group of ^{5m}C, with a number of demethylation pathways proposed for mammalian systems (Scheme 1.3). Initial reports on active demethylation suggested that it proceeds through deamination of ^{5m}C to thymine (T) by the AID/APOBEC (activation-induced deaminase/apolipoprotein B mRNA-editing enzyme complex) family of cytidine deaminases[42,43], with the consequent T•G mismatch facilitating thymine excision by thymine DNA glycosylase (TDG)[44]. The abasic site produced by TDG is then replaced with unmodified cytosine (^{5}C) *via* base excision repair (BER) mechanisms.



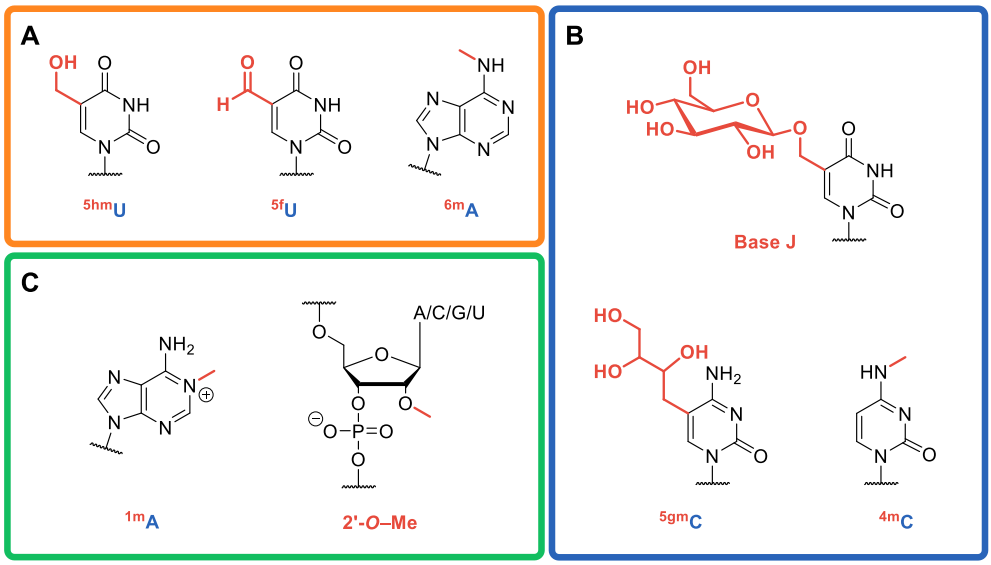
Scheme 1.3 | Schematic representation of active cytosine demethylation pathways. **^{5}C**, unmodified cytosine; **^{5hm}U**, 5-hydroxymethyluracil; **DNMT**, DNA methyltransferase; **AID**, activation-induced deaminase; **APOBEC**, apolipoprotein B mRNA-editing enzyme complex; **TDG**, thymine DNA glycosylase; **SMUG1**, single-stranded mono-functional uracil DNA *N*–glycosylase; **BER**, base excision repair.

Following the discovery that TET proteins can oxidise ^{5m}C up to ^{5ca}C[39,40], TDG was found to rapidly excise both ^{5f}C and ^{5ca}C[45,46], providing an alternative active demethylation pathway. Additionally, Apobec1 has been reported to support ^{5hm}C demethylation in mouse brain *via* deamination to 5-hydroxymethyluracil (^{5hm}U)[47], with the corresponding ^{5hm}U•G mismatch being a known substrate for both TDG[44] and single-stranded mono-functional uracil DNA *N*–glycosylase (SMUG1)[48]. However, AID/APOBEC-mediated deamination of ^{5hm}C could not be reproduced *in vitro* or in HEK 293T cells overexpressing AID/APOBEC family members[49], challenging the plausibility of the proposed pathway in humans.

Direct deformylation of ^{5f}C *via C–C* bond cleavage has been observed *in vivo*[50], but the enzyme responsible for this transformation has not been identified to date. Similarly, evidence for direct decarboxylation of ^{5ca}C in cells has been reported[51], with one *in vitro* study observing decarboxylase activity by the mammalian DNMT3B/DNMT3L complex[52]. Direct cytosine regeneration mechanisms (deformylation and decarboxylation) are ideal for DNA demethylation, as they bypass mutagenic intermediates and potential replication errors that can occur during BER[53].

1.2.5 Other epigenetic nucleobase modifications

While cytosine modifications are the best characterised epigenetic DNA marks, a number of additional nucleobase modifications have been identified in the genomes of mammals and lower species (Scheme 1.4).



Scheme 1.4 | Examples of other epigenetic nucleobase modifications. A, DNA modifications found in mammals; **B**, DNA modifications found in lower eukaryotes, prokaryotes, and archaea; **C**, RNA modifications found in mammals. **^{5f}U**, 5-formyluracil; **^{6m}A**, *N*6–methyladenine; **base J**, β-D-glucopyranosyloxymethyluracil; **^{5gm}C**, 5-glyceryl-methylcytosine; **^{4m}C**, 4-methylcytosine; **^{1m}A**, *N*1–methyladenine; **2'-*O*–Me**, 2'-*O*–ribosemethylation.

Oxidised thymine (T) derivatives, ^{5hm}U and ^{5f}U (5-formyluracil), are typically formed as a result of DNA damage induced by reactive oxygen species (ROS)[54,55], although a TET homologue, *Naegleria* Tet-like dioxygenase (NgTet1), was found to sequentially oxidise T *in vitro*[56]. Thymine can also reportedly be oxidised to ^{5hm}U by murine Tet1 (mTet1) in mouse embryonic stem cells (mESCs)[57], although the activity could not be reproduced *in vitro*[56]. Specific glycosylases of oxidised thymine derivatives have been identified as SMUG1[48] for ^{5hm}U and human endonuclease III homologue (hNTH1)[58] for ^{5f}U. ^{5hm}U levels are elevated in the blood of breast cancer patients[59], although it is also recognised by some reader proteins, suggesting a distinct epigenetic function[57]. Conversely, ^{5f}U is an oxidative DNA stress marker implicated in DNA replication stalling, base misincorporation, and interference with DNA/RNA metabolism[60–62].

Methylation of adenine in DNA to form *N*6–methyladenine (^{6m}A) has been identified in mESCs[63] and, more recently, in the human genome[64]. The modification is purportedly deposited by the *N*6–adenine-specific DNA methyltransferase 1 (N6AMT1), and removed by the Fe(II)/2-OG-dependent demethylase ALKBH1[64]. The presence of DNA ^{6m}A has been shown *in vitro* to substantially reduce the affinity of SATB1, a crucial chromatin organiser, for stress-induced duplex destabilisation regions (SIDD)[65], highlighting its regulatory role through control of DNA secondary structure; accordingly, decreased ^{6m}A levels reportedly promote tumorigenesis in mouse models[64].

A number of DNA modifications absent from mammalian genomes have been identified, with some examples of particular interest discussed below (Scheme 1.4 B). β-D-Glucopyranosyloxymethyluracil (base J) present in kinetoplastids is historically important as the first modified nucleobase identified in eukaryotic DNA[66], subsequently setting off the hunt for other DNA modifications. Conjugation of a glyceryl moiety to the methyl group on ^{5m}C to form 5-glyceryl-methylcytosine (^{5gm}C) has been reported in green algae *C. reinhardtii*[67]; the modification is deposited by an Fe(II)- and L-ascorbate-dependent TET homologue CMD1, demonstrating the first enzymatic 2-OG-independent functionalisation of the 5-methyl group on ^{5m}C. Finally, 4-methylcytosine (^{4m}C), long considered to be exclusive to prokaryotes and archaea[68], has very recently been detected in eukaryotic organisms[69], highlighting the ongoing identification of novel epigenetic marks.

While epigenetics are mostly concerned with DNA modifications, more than 100 epitranscriptomic RNA modifications have been reported to date, including analogous ^{5m}C and ^{6m}A modifications, and RNA-specific marks such as *N*1–methyladenine (^{1m}A), 2'-*O*–ribosemethylation, and others[70] (Scheme 1.4 C).

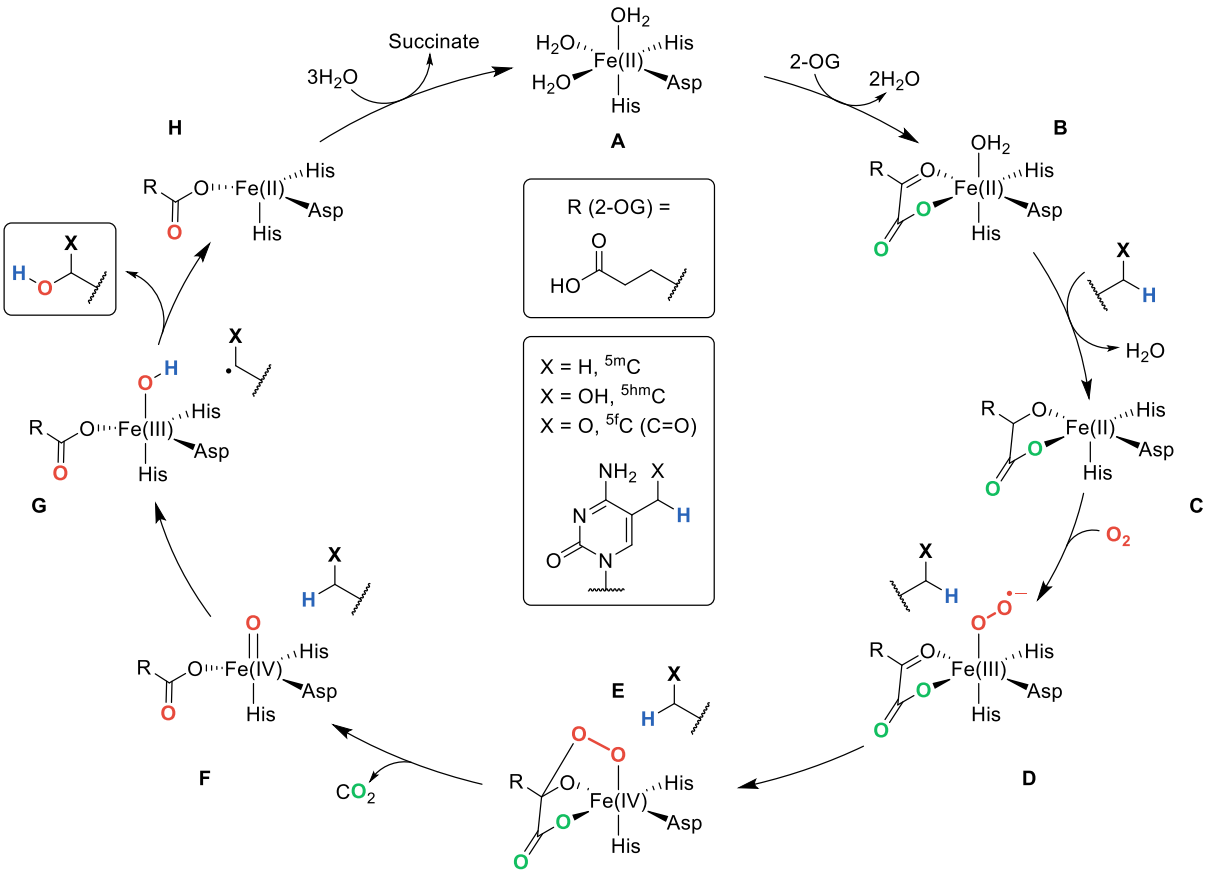
1.3 Biochemical characterisation of Ten-Eleven Translocation enzymes

1.3.1 Shared features of the Fe(II)/2-OG-dependent dioxygenase superfamily

Fe(II)/2-OG-dependent dioxygenases are a large superfamily of enzymes that perform diverse oxidative transformations with important biological roles, including hydroxylation, halogenation, ring formation, and others[71]. More than 60 members of the superfamily have been identified in humans to date[72], with substrate hydroxylation, typically at an unactivated carbon centre, being the dominant mode of reactivity. 2-OG-dependent hydroxylases exhibit absolute dependence on Fe(II), 2-OG, and O_2, and recognise a wide array of substrates with distinct biological functions. For example, asparaginyl hydroxylase factor inhibiting HIF (FIH) regulates oxidative metabolism by controlling hypoxia-inducible factor (HIF) transcriptional activity[73]; prolyl hydroxylases alter protein conformational stability, prime HIF for proteasomal degradation, and contribute to other pathophysiological processes[74]; DNA repair enzymes AlkB directly remove alkyl adducts from nucleobases by oxidative dealkylation[75]; finally, histone lysine demethylases (KDMs), one of the largest groups of 2-OG hydroxylases and also perhaps the closest to TETs in terms of their biological function, epigenetically regulate chromatin structure and gene expression by removing methylation marks from N–terminal histone tails[72].

Nearly all members of the 2-OG dioxygenase superfamily share a number of structural features, including a double-stranded β-helix (DSBH) fold and conserved sequence motifs necessary for iron and cofactor binding (see section 1.4.1). Owing to the similarities of key structural elements, all 2-OG-dependent hydroxylases, including TETs, are believed to follow a consensus catalytic mechanism[71] (Scheme 1.5). Briefly, 2-OG replaces two water molecules to chelate to the iron centre in a bidentate manner (**B**), followed by substrate binding to the

active site (**C**), and oxygen coordination to the metal centre (**D**). The resulting Fe(III)-superoxo (**D**) species attacks the iron-bound 2-OG *via* a peroxohemiketal bicyclic intermediate (**E**) to release CO_2, producing a reactive high-valent Fe(IV)-oxo complex (**F**). This species consequently abstracts a hydrogen atom from the substrate, producing a reactive radical intermediate (**G**). Hydroxyl radical rebound yields the oxidised product and a reduced Fe(II) complex (**H**). Finally, succinate is released and three water molecules fill the vacant coordination sites on Fe(II), completing the catalytic cycle (**A**).



Scheme 1.5 | Proposed consensus mechanism of the oxidative reactions catalysed by Fe(II)/2-OG-dependent hydroxylases, including TETs. For clarity, only the catalytic triad specific to TET enzymes (HxD…H) and only canonical substrates of TETs are shown. Adapted from literature[71].

Despite low sequence identities, 2-OG-dependent hydroxylases share conserved sequence motifs, structural elements, and partially overlapping biological roles. Histone lysine methylation, regulated in part by KDM activity, is an important mechanism of epigenetic control complementary to DNA methylation: lysine methylation marks can occur at multiple sites on different histones and influence the overall structure of chromatin and/or binding of

effector molecules in a distinct manner, with each position and methylation state associated with a discrete function. Furthermore, crosstalk between different modifications *via* positive or negative influence between them introduces an additional layer of complexity and epigenetic fine-tuning[76]. As demethylation of both histone lysines and cytosines is carried out by 2-OG hydroxylases (KDM and TET families, respectively), development of specific 2-OG hydroxylase chemical probes, particularly in the case of 2-OG mimetic inhibitors, must ensure a sufficient degree of inter-family selectivity. Non-selective inhibition of multiple 2-OG dioxygenases could lead to off-target effects in cells and *in vivo*, producing undesirable phenotypes and precluding interrogation of protein of interest biological functions.

1.3.2 Uncoupled 2-OG decarboxylation and L-ascorbate dependence

By its nature, turnover of substrate by Fe(II)/2-OG-dependent enzymes is stoichiometrically linked to 2-OG consumption, as oxidative decarboxylation of 2-OG is a requisite for the formation of the reactive Fe(IV)-oxo complex. However, in some cases, 2-OG turnover can take place without the associated substrate modification, resulting in "uncoupled" 2-OG decarboxylation. Prolyl 4-hydroxylase (PHD) and lysyl hydroxylase were shown to consume 2-OG in the absence of substrate at rates corresponding to 1.3–6.5% of rates at saturating substrate conditions[77–79]. Interestingly, total 2-OG consumption in the absence of substrate reportedly approached 50% relative to substrate-containing conditions for hTET1 and human DNA-repair enzyme ALKBH3[80], suggesting that 2-OG-dependent DNA hydroxylases may be more predisposed towards uncoupled 2-OG decarboxylation. Uncoupled 2-OG turnover can also take place in the presence of substrate, the rate of which seemingly depends on the nature of the substrate itself: the proportion of uncoupled turnover by histone lysine demethylase KDM4E reportedly depended on the substrate used (25 ± 1% and 16 ± 1% for histone H3 N^{ε}–trimethyllysine 9, H3K9me$_3$, and N^{ε}–dimethyllysine 9, H3K9me$_2$, respectively)[81]. In a more striking example, truncation of an oligonucleotide substrate

recognised by the *E. coli* DNA-repair enzyme AlkB resulted in a substantial increase in the proportion of uncoupled turnover (<5% to approx. 80%)[82], demonstrating that utilisation of poorly-recognised substrates can lead to artificially elevated 2-OG uncoupling levels *in vitro*.

2-OG-dependent enzymes often exhibit a requirement for L-ascorbate in order to retain maximal activity, which may be a consequence of the observed uncoupled 2-OG turnover. Both PHD and lysyl hydroxylase were shown to oxidise substrates in the absence of ascorbate at a high initial rate, but the reaction rate decreased rapidly (down to 8% activity within 30 sec) with a simultaneous enrichment of enzyme-bound iron to Fe(III)[78,83]; in the case of PHD, subsequent addition of ascorbate or Fe(II) partially recovered the turnover rate[83]. Thus, ascorbate may act as an alternative electron sink in the uncoupled catalytic cycle, reducing the oxidised iron intermediates and returning the enzyme-metal complex to its active form[83]. Consumption of ascorbate is sub-stoichiometric in coupled reactions where the native substrate acts as the oxygen acceptor[84], but appears to tend towards stoichiometric levels with respect to 2-OG decarboxylation in the absence of substrate[85], supporting the role of ascorbate as an oxygen acceptor required to complete the catalytic cycle in the absence of preferred substrate. Treatment of stem cells with ascorbate promoted TET-mediated DNA demethylation[86,87], demonstrating that ascorbate dependence extends to TET enzymes in cellular settings.

While the presence of ascorbate can supress enzyme inactivation due to stalling at the Fe(III) complex, uncoupled 2-OG decarboxylation can also lead to largely irreversible inactivation *via* self-hydroxylation. 2-OG dioxygenases, including 2,4-dichlorophenoxyacetate dioxygenase (TfdA)[88], taurine dioxygenase (TauD)[89], and *E. coli* and human AlkB homologues[90,91], reportedly auto-hydroxylate Trp, Tyr, or Ala residues proximal to the iron centre in the absence of substrate. Consequently, coordination to the metal centre by the hydroxylated

aromatic residues or disruption of 2-OG binding by Ala-OH abrogates further catalytic activity. TET enzyme propensity for auto-hydroxylation has not been determined to date, but the reported high levels of uncoupled 2-OG turnover[80] may suggest a degree of resistance to oxidative damage.

1.3.3 Sequential ^{5m}C oxidation

The multifunctional nature of TETs, where the product of one oxidation reaction is accepted as a substrate for a subsequent oxidation, presents two possible sequential oxidation mechanisms: (i) distributive oxidation, where the oxidised product is released following the completion of a single catalytic cycle, or (ii) iterative oxidation, where one molecule of substrate is oxidised multiple times in succession following a single substrate binding event, with conflicting evidence in support of both modes reported to date. Tamanaha *et al*.[92] argued for a distributive model due to the observed accumulation of intermediate oxidation products ^{5hm}C and ^{5f}C, while Crawford *et al*.[93] proposed an iterative mode of oxidation based on 5-[$^{13}C^{2}H_3$]-C labelling experiments which revealed concomitant depletion of ^{5hm}C and enrichment of ^{5f}C/^{5ca}C. However, the latter study did not address the factors that would control the oxidation state at which the substrate is released, as, in the case of a fully iterative mechanism, only ^{5ca}C accumulation should be observed. The crystal structures of hTET2[94] and NgTet1[95] in complex with ^{5m}C-containing oligonucleotides show that 2-OG is localised deep within the enzyme active site (see section 1.4.1), suggesting that cofactor recharging without prior substrate dissociation would be challenging, although crystal structures may fail to capture the structural flexibility of the enzyme-substrate complex in solution. Interestingly, Crawford *et al*.[93] did not observe any kinetic isotope effects when using deuterium-labelled ^{5m}C despite substantial changes to reaction rates and regioselectivity reported for hTET2 and other Fe(II)/2-OG-dependent hydroxylases with analogously deuterated substrates[96–98].

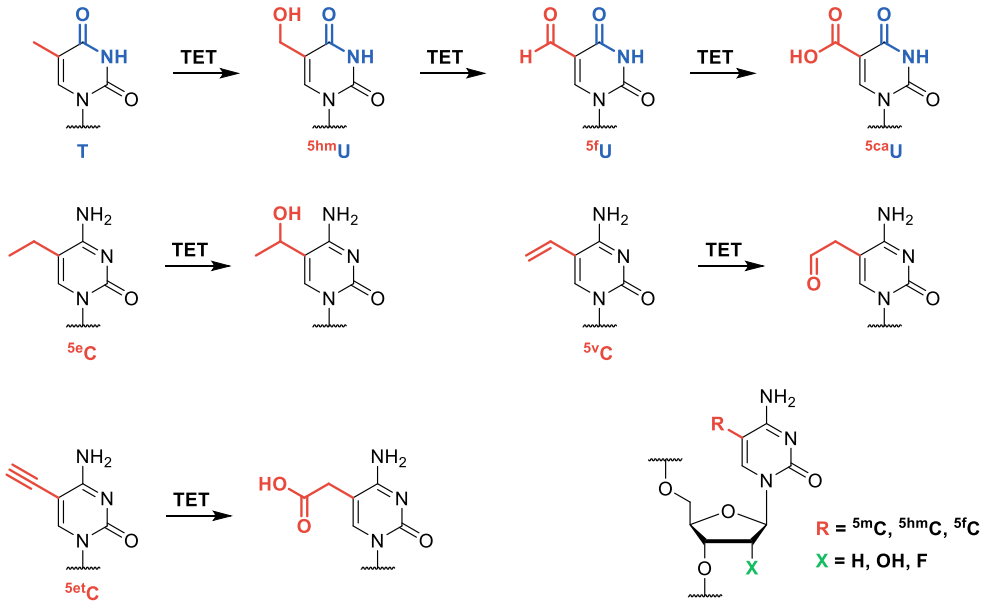
1.3.4 Substrate specificity

1.3.4.1 ^{5m}C oxidation state

TET enzymes oxidise ^{5m}C to ^{5ca}C in a sequential fashion, but the efficiencies of the three reactions differ substantially. Kinetic studies on hTETs have shown that deoxynucleotide ^{5m}C exhibits the highest rate of turnover, with significant reduction in turnover rate observed for subsequent oxidation events on ^{5hm}C and ^{5f}C substrates (up to 7-fold and 13-fold, respectively)[40,96,99]. The retardation of subsequent oxidation turnover rate is, however, not due to differences in substrate binding affinity, as hTET2 bound to oligonucleotides containing ^{5}C, ^{5m}C, ^{5hm}C, or ^{5f}C with comparable affinities (K_d = 0.75 ± 0.33 μM)[96]. Structural insight into hTET2 complexed with ^{5m}C and ^{5ox}C revealed that substrate preference likely stems from conformational differences between ^{5m}C and ^{5ox}C derivatives within the active site (see section 1.4.1)[96].

1.3.4.2 Alternative nucleotide substrates

The canonical function of TET enzymes is the oxidation of ^{5m}C, but TETs have been shown to be relatively promiscuous in terms of their substrate specificity, with a number of alternative nucleotide substrates identified (Scheme 1.6). mTet1 reportedly oxidised thymine to ^{5hm}U *in cellulo*[57], but the observed activity could not be reproduced *in vitro*[56]. Conversely, NgTet1 was shown to sequentially oxidise T up to 5-carboxyuracil (^{5ca}U) *in vitro*, although thymine hydroxylase activity was minor (12-fold lower) compared to activity on ^{5m}C substrates[56]. The relatively spacious active site of TETs facilitates single oxidation of 5-ethylcytosine (^{5e}C)[100], 5-vinylcytosine (^{5v}C)[101], and 5-ethynylcytosine (^{5et}C)[101], albeit at substantially reduced rates. TETs have also been shown to tolerate modifications to the deoxyribose sugar: a ribonucleotide analogue of ^{5m}C (2'-(R)-OH)[102,103], as well as a fluorinated analogue (2'-(R)-F)[104] were accepted as substrates, although turnover rates were reduced compared to the deoxynucleotide analogue.



Scheme 1.6 | Non-canonical substrates recognised by TET enzymes. ^{5ca}**U**, 5-carboxyuracil; ^{5e}**C**, 5-ethylcytosine; ^{5v}**C**, 5-vinylcytosine; ^{5et}**C**, 5-ethynylcytosine.

1.3.4.3 Oligonucleotide flanking sequence

The context of the modified nucleotide (i.e., the flanking sequence) can also affect enzyme activity. The nucleotide 3' downstream of the modified C (^{5x}C, +1 position) has been shown to exhibit the most pronounced effect on catalytic activity, although flanking nucleotides in positions −3 to +2 (relative to ^{5x}C) also affected activity[56,105], in agreement with the range of base pairs found to interact with hTET2 *in crystallo*[94,96]. Deviation from the preferred ^{5x}CpG dinucleotide resulted in a substantial (>75%) reduction in relative activity, with the least preferred ^{5x}CpT dinucleotide showing only marginal turnover levels[94,99,105]. Although position −1 and the more distal flanking positions did not exhibit strong effects individually, the cumulative impact was substantial enough to allow ^{5x}CpA or ^{5x}CpC substrates with preferred flanking sequences to be oxidised at similar or faster rates than ^{5x}CpG substrates with disfavoured flanking sequences[105]. Interestingly, the specific flanking sequence preferences were different between TET1 and TET2, and were also distinct for ^{5m}C and ^{5hm}C substrates[105], supporting the distinct functions of TET paralogues and suggesting a potential mechanism for the sequence-specific ^{5m}C oxidation product accumulation profiles in genomic DNA.

1.3.4.4 Oligonucleotide hybridisation

TET enzymes recognise ^{5m}C nucleotides in both single-stranded (ss) DNA and RNA[102,103]. While a complementary DNA strand is not required for activity, double-stranded (ds) DNA substrates are oxidised at a similar or faster rate than single-stranded oligonucleotides; conversely, hybridisation of ssRNA with complementary DNA reduces activity, whereas dsRNA substrates are not recognised by TETs[102,103]. In DNA sequences containing ^{5x}CpG dinucleotides, the methylation state of the complementary strand does not seem to affect turnover rate[56], consistent with structural observations indicating no protein interactions with the C5 position of the complementary cytosine nucleotide[94,95].

1.4 Structural characterisation of Ten-Eleven Translocation enzymes

1.4.1 2-OG-dependent TETs

TETs found in metazoa are large, multidomain proteins with a C–terminal catalytic domain (CD) containing a double-stranded β-helix (DSBH) fold, characteristic of Fe(II)/2-OG-dependent dioxygenases (Figure 1.3 A)[106]. In metazoan TETs, the DSBH domain is split in two by a large low complexity insert (LCI) with no assigned function to date[107], and is typically augmented by secondary structural elements or additional domains that distinguish different 2-OG dioxygenase families; in TETs, the N–terminal end of the CD is decorated with a Cys-rich region which flanks the DSBH domain, helping stabilise segments involved in DNA recognition and catalysis[94]. TETs also contain a positively charged Arg- and Lys-rich cleft which promotes interactions between the enzyme and the negatively charged oligonucleotide phosphate backbone.

In humans (and other jawed vertebrates), gene duplication or triplication events led to the emergence of three TET paralogues (hTET1, hTET2, hTET3)[108]. While the paralogues share high sequence identity within the catalytic domain, the C–termini are more differentiated, suggesting divergent biological roles[109]: TET1s contain a BC domain (“before CXXC”) situated at the outset of the N–terminus, which has been shown to promote global chromatin binding[110], while hTET2s and hTET3s contain a number of conserved sequence motifs at the N–terminal end and within the LCI with functions unidentified to date (not shown in Figure 1.3 A).

The N–termini of TET1/3 contain CXXC DNA-binding domains, which have been shown to bind ^{5x}CpG with paralogue-dependent specificities[111,112]. CXXC is absent in TET2s due to a chromosomal inversion that occurred during evolution; however, the resulting stand-alone

CXXC4 domain (IDAX) can reportedly interact with the CD of hTET2 to reconstitute a functional CXXC-TET complex[113]. CXXC domains recruit TETs to promoters and CpG islands, and have been implicated in degradation priming[107], chromatin association[110], and genomic imprinting removal during development[110].

The large, multidomain character of metazoan TETs makes them a challenging target for *in vitro* studies requiring recombinant protein production. Importantly, TET catalytic activity is not dependent on CXXC, allowing construction of truncated constructs retaining catalytic activity for *in vitro* studies (Figure 1.3 B). Furthermore, an additional truncation of $hTET2_{CD}$ *via* a deletion-insertion in the LCI region afforded a reportedly crystallisable, catalytically active construct ($hTET2_{CD\Delta LCI}$, Figure 1.3 C)[94]. Although the LCI is not essential for TET activity, the crystal structure of $hTET2_{CD\Delta LCI}$ in complex with ^{5m}C-containing DNA revealed that parts of the LCI region may potentially interact with the major groove of the DNA substrate, suggesting a role in TET activity or specificity control[114].

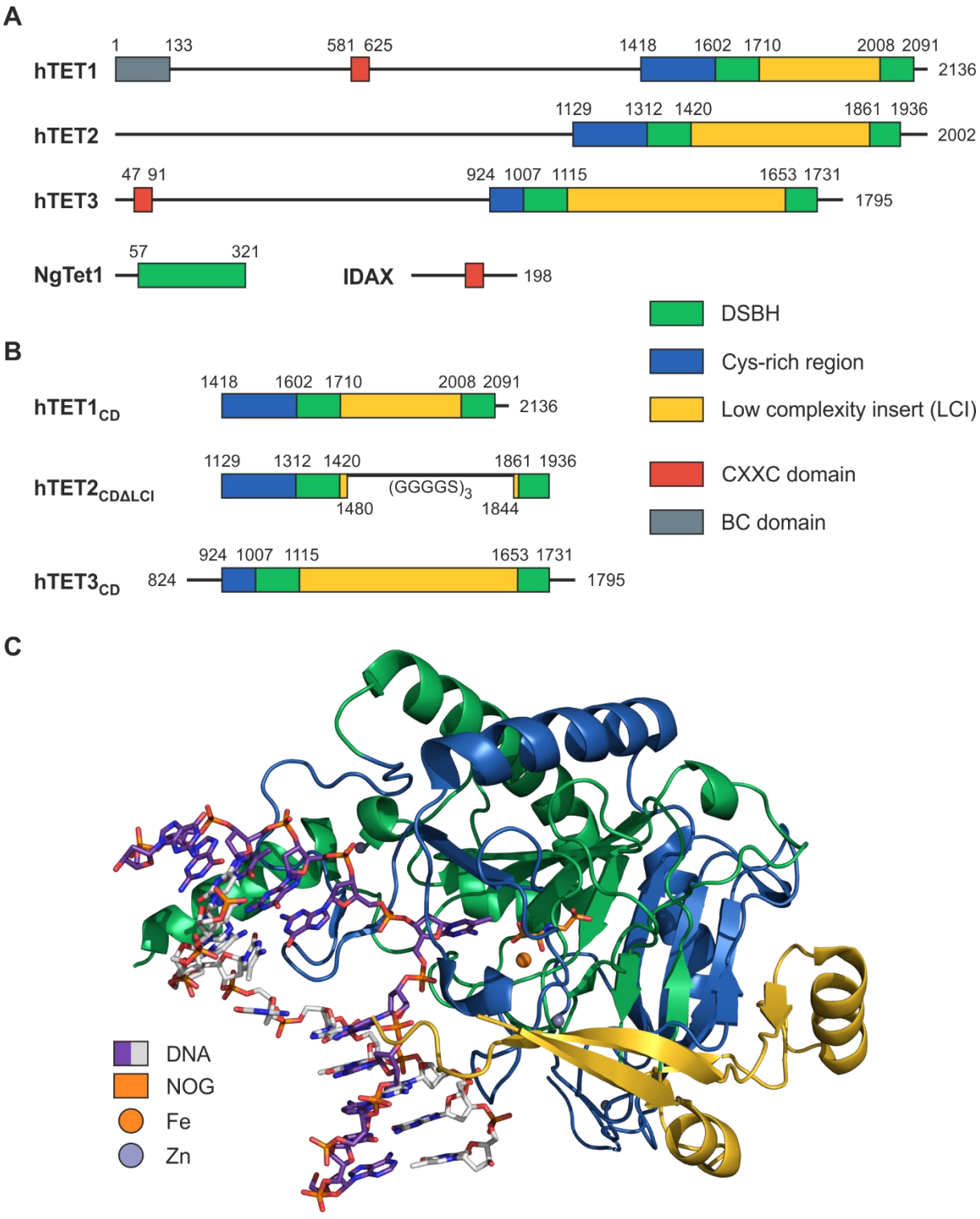


Figure 1.3 | Structural information for selected TET enzymes. A, domain architectures of full-length hTET paralogues, NgTET1, and IDAX (CXXC4); **B**, hTET constructs used in this work; **C**, crystal structure of hTET2$_{CD\Delta LCI}$ in complex with Fe(II), NOG, and ^{5m}C-containing DNA, highlighting the different domains (PDB ID: 4NM6)[94]. **DSBH**, double-stranded β-helix; **CD**, catalytic domain; **BC domain**, "before CXXC" domain. hTET2 structural elements in **C** are coloured according to the colour scheme used in **A** and **B**; substrate and cofactor colouring: violet and grey, DNA strand; orange, NOG; rust, Fe; grey, Zn. Adapted from literature[115].

While lower eukaryotic TET homologues such as NgTet1 typically show simpler architectures (Figure 1.3 A) and poor sequence overlap compared to metazoan TETs[107], active site feature

comparison of hTET2$_{CD\Delta LCI}$ and NgTet1 complexed with ^{5x}C-containing oligonucleotides revealed substantial structural similarities (Figure 1.4), making NgTet1 an attractive model enzyme for TET function investigations *in vitro*. hTET2$_{CD\Delta LCI}$ and NgTet1 both contain a catalytic Fe(II) metal ion coordinated by a 2-His-1-carboxylate facial triad (HxD…H). Fe(II) completes the octahedral binding geometry by coordinating 2-OG (substituted with *N*–oxalylglycine, NOG, *in crystallo*), with additional 2-OG binding stabilisation provided by H–bonding to nearby hydrophilic residues, and a water molecule, which is likely displaced by O_2 during catalysis. Correspondingly, mutation of the residues interacting with Fe(II) and 2-OG leads to partial or complete loss of catalytic activity[94,95].

Although residues interacting with the substrate are less conserved across the species, three analogous interactions remain: (i) π– π stacking between the pyrimidine ring of ^{5x}C and an aromatic residue (omitted for clarity in Figure 1.4), (ii) H–bonding between H–bond acceptor pyrimidine N3 and a third active site His, and (iii) H–bonding between exocyclic H–bond donor pyrimidine N4 and an Asn/Glu residue. Conversely, distinct modes of H–bonding with the exocyclic pyrimidine O2 are observed for the two enzymes: NgTet1 donates a H–bond from N147 to O2 in both ^{5m}C and ^{5hm}C (Figure 1.4 B), while coordination of O2 in hTET2$_{CD\Delta LCI}$ seems to depend on the oxidation state of the C5 substituent. A fourth active site His residue adopts different conformations in ^{5m}C/^{5hm}C/^{5f}C complexes, and seemingly only participates in H–bonding with the pyrimidine O2 of ^{5hm}C (Figure 1.4 A). ^{5hm}C and ^{5f}C are potentially further stabilised in hTET2$_{CD\Delta LCI}$ by a water-mediated H–bond between T1393 and the exocyclic pyrimidine N4. While the observed structural differences may suggest potential differentiation with regards to reaction kinetics and substrate specificity, biochemical characterisation of the two TETs revealed highly comparable rate constants for turnover of ^{5m}C, ^{5hm}C, and ^{5f}C[56,96].

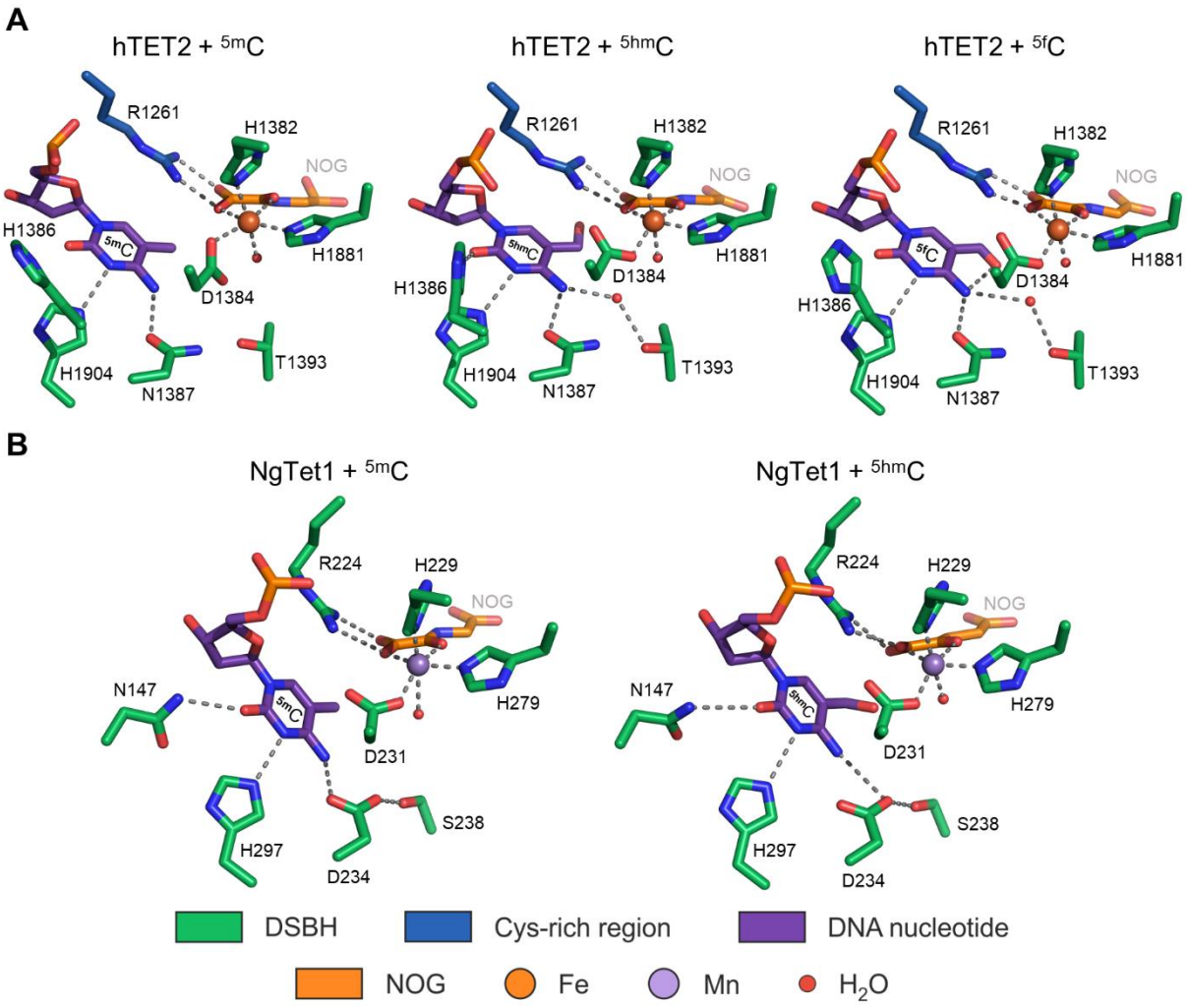


Figure 1.4 | Active site structures of hTET2$_{CD\Delta LCI}$ and NgTet1. A, hTET2$_{CD\Delta LCI}$ complexes with Fe(II), NOG, and DNA oligomers containing ^{5m}C, ^{5hm}C, or ^{5f}C (PDB IDs: 4NM6, 5DEU, and 5D9Y, respectively)[94,96]; **B**, NgTet1 complexes with Mn(II), NOG, and DNA oligomers containing ^{5m}C or ^{5hm}C (PDB IDs: 4LT5 and 5CG8, respectively)[95,99]. Residues involved in interactions with DNA substrate and cofactors are shown in stick representation (π–π stacking residues Y1902 for hTET2 and F295 for NgTet1 not shown for clarity); metal-ligand and H–bonding interactions are shown as dashed lines. Structural elements are represented by specific colours: pastel blue, Cys-rich domain; green, DSBH; violet, targeted DNA substrate nucleotide; orange, NOG. Heteroatom colouring: red, O; blue, N; orange, S; rust, Fe; lilac, Mn. Waters are represented as red spheres.

1.4.2 L-Ascorbate-dependent *C. reinhardtii* TET homologue, CMD1

Although a number of TET homologues have been computationally identified in metazoa, fungi, and prokaryotes[107], the putative homologues are typically predicted to exhibit Fe(II)/2-OG-dependent dioxygenase activity, with to date biochemically characterised TETs supporting the predictions[115]. Conversely, a ^{5m}C-modifying TET homologue isolated from green alga *Chlamydomonas reinhardtii*, CMD1, has been reported to glycerylate ^{5m}C *via* direct C–C bond formation with no apparent requirement for 2-OG, instead exhibiting absolute dependence on Fe(II) and L-ascorbate[67].

Despite its distinct function, CMD1 exhibits some commonalities with ^{5m}C-oxidising TETs, as it forms a DSBH fold containing the canonical Fe(II) coordinating motif (HxD…H) in the active site (Figure 1.5)[116]. Furthermore, the binding of ^{5m}C is stabilised in a fashion analogous to other TETs: the ^{5m}C pyrimidine engages in π–π stacking with an aromatic residue and a number of H–bonding interactions are present between pyrimidine heteroatoms (O2, N3, and exocyclic N4) and hydrophilic residues (Figure 1.5 B). The main source of structural divergence in the active site of CMD1 comes from the distinct coordination of the essential cofactor: unlike the bidentate 2-OG chelation to Fe(II) observed in the 2-OG-dependent hTET2 and NgTet1, L-ascorbate only forms a monodentate interaction with the metal centre (Figure 1.5 C); furthermore, there is virtually no overlap in H–bonding interactions with the cofactors between the two enzymes.

The unique structural and functional features of CMD1 highlight the breadth of TET homologue divergence across different species, and provide evidence that L-ascorbate can perform functions beyond the typically described role in active site metal reduction.

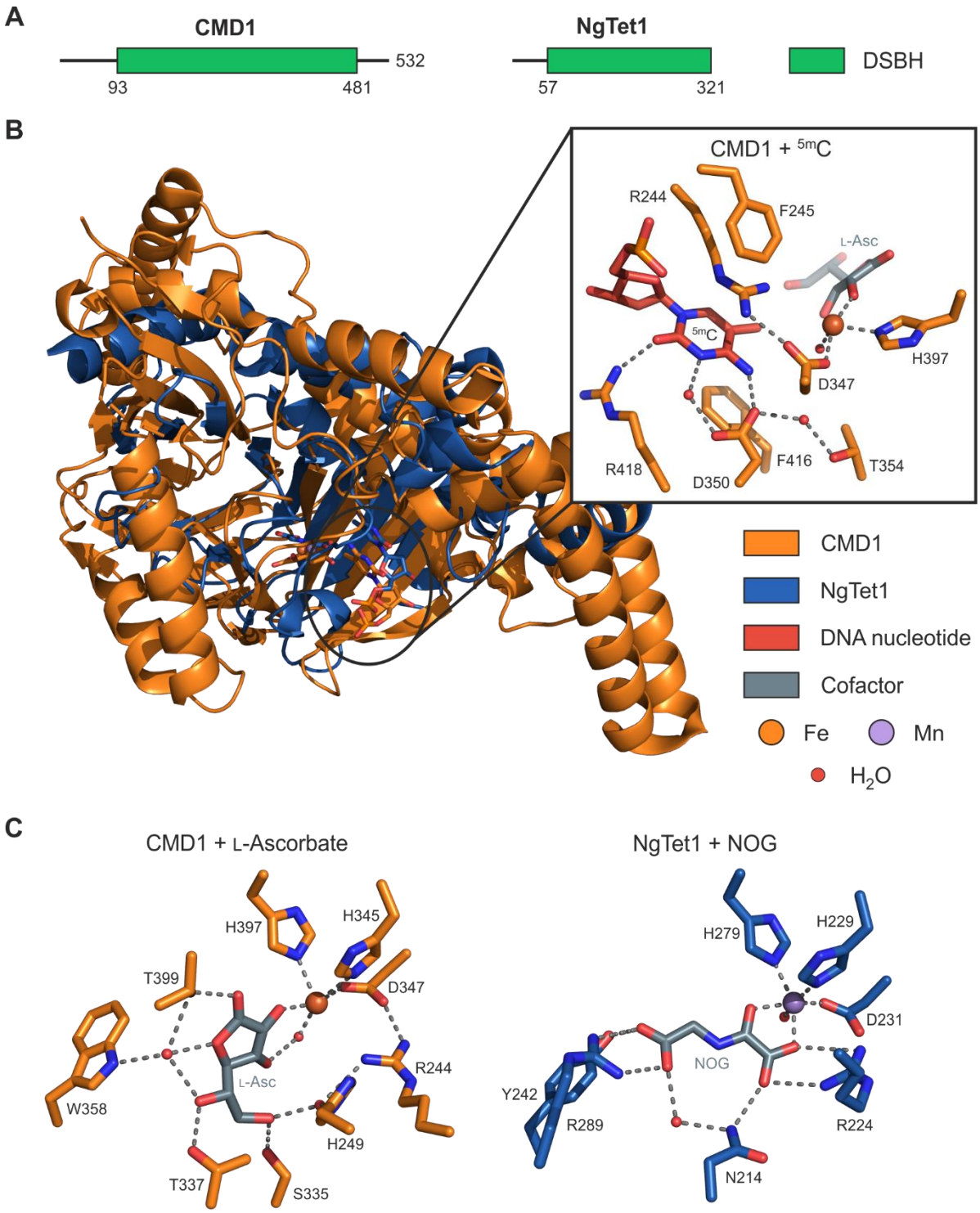


Figure 1.5 | Structural information for CMD1. A, domain architectures of CMD1 and NgTet1; **B**, overlap of CMD1 and NgTet1 (inset: CMD1 interactions with ^{5m}C substrate); **C**, interactions of CMD1 and NgTet1 with their corresponding cofactors. Residues involved in interactions with DNA substrate and cofactors are shown in stick representation; in **B** inset, H345 is omitted for clarity; metal-ligand and H–bonding interactions are shown as dashed lines. Structural elements are represented by specific colours: orange, CMD1; pastel blue, NgTet1; red, targeted DNA substrate nucleotide; grey, cofactor (L-ascorbate or NOG). Heteroatom colouring: red, O; blue, N; orange, S; rust, Fe; lilac, Mn. Waters are represented as red spheres. PDB IDs: 7CY8 (CMD1)[116], 4LT5 (NgTet1)[99].

1.5 Biological significance of Ten-Eleven Translocation enzymes

1.5.1 Function of ^{5m}C oxidation products

^{5m}C oxidation products are established as intermediates towards cytosine demethylation, but each ^{5m}C oxidation state (^{5hm}C, ^{5f}C, and ^{5ca}C) also functions as an epigenetic mark with distinct function and enrichment patterns. ^{5hm}C abundance is significantly lower than that of ^{5m}C, with prevalence ranging from 0.7% of all cytosines in central nervous system (CNS) tissues down to 0.03% in other tissues[19,117,118]. Unlike ^{5m}C, ^{5hm}C distribution is non-uniform, and its content is inversely correlated with the proportion of proliferating cells in a tissue[119], highlighted by decreased ^{5hm}C levels in cancer cells[120,121]. ^{5hm}C inhibits DNMT activity and MBP association, priming the sites for passive demethylation[122–124]. The high abundance of ^{5hm}C in brain tissue and further enrichment during neuronal differentiation suggests a function as an active epigenetic mark in mammalian neuronal development[125], consistent with ^{5hm}C-mediated changes to DNA duplex stability and DNA-protein interactions observed *in vitro*[126,127]. Furthermore, ^{5hm}C is enriched in gene bodies of actively transcribed genes and at poised and active enhancers, and a number of reader proteins for ^{5hm}C have been identified[128,129].

The prevalence of ^{5f}C, a stable[130,131], minimally mutagenic[132] nucleobase with functional roles that go beyond being a demethylation intermediate, is between 0.002–0.02% of all cytosines, 10- to 100-fold lower than that of ^{5hm}C[40,41], although ^{5f}C can reach levels comparable to ^{5hm}C at specific genomic loci[133]. Despite the effect of ^{5f}C on global DNA structure being a matter of debate[134,135], ^{5f}C has been shown to generate local structural changes: ^{5f}C presence increases DNA flexibility by reducing ^{5f}C:G base-pairing stability[136], leading to enhanced nucleosome mechanical stability[137]. The reactive aldehyde moiety of ^{5f}C also enables the formation of reversible DNA-protein crosslinks by forming Schiff bases with primary amines on histones[138],

with ^{5f}C–histone H3 conjugates in mESCs linked to decay in transcription elongation[139]. Accumulation of ^{5f}C at poised enhancers in TDG-deficient mESCs enhances association with transcriptional coactivator histone acetyltransferase p300[140], with a number of other ^{5f}C binders identified, including transcriptional regulators, chromatin regulators, and DNA repair factors[129,141].

The reported prevalence of ^{5ca}C is the lowest of all ^{5ox}C modifications at around 0.0003% of all cytosines, 10- to 100-fold lower than ^{5f}C[40]. However, ^{5ca}C levels may be underestimated due to ^{5ca}C resistance to certain nucleases commonly used in sample preparation for nucleoside/nucleotide MS, with up to a 10-fold increase in ^{5ca}C abundance reported when using alternative DNA hydrolysis methods[142]. ^{5ca}C is a prime target for active demethylation mechanisms, although a number of other cellular functions have been identified to date. Similarly to ^{5f}C, ^{5ca}C reduces the rate and substrate specificity of transcription enzymes[143,144] which can have a significant effect on transcriptional fidelity, alternative splicing, and other transcription processes[145,146], but epigenomic mapping revealed little overlap between ^{5f}C- and ^{5ca}C-enriched positions, supporting distinct regulatory roles for the two modifications[147]. The modification is highly enriched at active transcriptional start sites and enhancers in mESCs[148], and ^{5ca}C levels are also elevated in human breast cancers and gliomas[149]; together with the observation that the CXXC domain of hTET3 binds preferentially to ^{5ca}CpG[111], the findings suggest that ^{5ca}C may play a role in gene activation.

1.5.2 TET function in development

Although TET paralogues contain highly conserved active sites, TETs are functionally non-redundant, likely as a result of varying flanking sequence preferences, expression levels, and recruitment mechanisms. TETs have been shown to be differentially expressed throughout development in mice: TET3 plays a key role in global DNA demethylation for epigenetic reprogramming during pre-implantation development[150], whereas TET1 is

responsible for genomic imprinting erasure in primordial germ cells (PGCs)[151]. Both TET1 and TET2 are part of the pluripotency regulatory circuit in mESCs *via* active regulation of ^{5hm}C levels with distinct gene target preferences[152]; following differentiation, TET3-mediated ^{5hm}C accumulation has been proven essential in the correct development of neural tissues[153,154]. Furthermore, TET-mediated oxidation of ^{5m}C to ^{5hm}C in RNA was shown to play a key role in pluripotency factor downregulation to enable differentiation[155]. Although TET hydroxylase activity is essential for development, a number of TET protein functions are in fact catalytic activity-independent: TETs facilitate transcription repression by recruiting histone deacetylation-mediating proteins[156,157] and have been reported to directly control modifications on specific histone lysine positions[158]. In another example, a catalytically inactive variant of Tet3 was shown to partially reverse the developmental defects caused by Tet3 knockout (KO) in frog[159], demonstrating that TET functions are not limited to ^{5m}C hydroxylase activity.

1.5.3 Aberrant TET function in cancers

Aberrant DNA methylation has been established as a hallmark of cancer[160], suggesting that active demethylation mechanisms may contribute to cancer development. Indeed, *TET2* mutations are prevalent in haematopoietic malignancies[161], while decreased TET protein and ^{5hm}C levels are also closely associated with human breast, liver, lung, pancreatic, and prostate cancers[162]. Inactivating mutations in *TET2* are the most common lesions in myeloid lineage malignancies[163], although it was the fusion of *TET1* and mixed-lineage leukaemia (*MLL*) genes in acute myeloid leukaemia (AML) that first led to the discovery of the TET family[37]. Oncogenic transformations can stem not only from loss-of-function *TET* mutations, but also mutually exclusive mutations of TET activity regulators. For example, neomorphic mutations of isocitrate dehydrogenases *IDH1* and *IDH2* lead to the accumulation of 2-hydroxyglutarate (2-HG)[164], which is a 2-OG-competitive TET inhibitor[165], resulting in leukaemia-associated

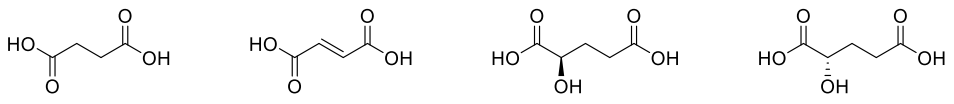
phenotypes; similarly, mutations to the gene coding for Wilms' tumour protein (WT1), a transcription factor that binds to the CD of hTET2 and recruits hTET2 to its target genes, have been shown to result in leukaemia cell proliferation[166]. Interestingly, TET levels were found to be upregulated in hepatoblastomas with an associated enrichment in ^{5hm}C[167], demonstrating that enhanced TET activity can also lead to carcinogenicity.

1.6 Inhibitors of Ten-Eleven Translocation enzymes

TET proteins play key roles in both mammalian development and cancer progression, making them an attractive target for therapeutic intervention. In order to assess the influence of TET catalytic activity modulation on cancer development and probe catalytic activity-independent functions, cell-permeable inhibitors which target specific TET paralogues are required. Although a number of TET inhibition strategies are available, no TET paralogue-specific molecular tools have been identified to date. Selected compounds with reported inhibitory activities against hTETs to date are described below.

1.6.1 Endogenous inhibitors

The catalytic activity of TET enzymes is dependent on 2-OG; as such, it is unsurprising that dysregulation of the tricarboxylic acid (TCA) cycle, of which 2-OG is part of, would likely lead to aberrant TET activity profiles. Indeed, mutations to genes coding for TCA cycle enzymes are highly prevalent in cancers[168]. Loss-of-function gene variants of succinate dehydrogenase (*SDH*) and fumarate hydratase (*FH*) lead to the accumulation of the corresponding TCA cycle intermediates, succinate and fumarate, and, subsequently, arrest of 2-OG production. *SDH*[169] and *FH*[170] mutations have been shown to inhibit HIF prolyl hydroxylases (PHDs), resulting in carcinogenesis: *in vitro* evaluation of succinate and fumarate (Scheme 1.7) revealed inhibitory activities against a number of 2-OG-dependent dioxygenases, with IC_{50} values ranging from 3 µM to 15 mM[72], highlighting the possibility of multiple oncogenic pathways.



Scheme 1.7 | Structures of TCA cycle intermediates (succinate, fumarate) and oncometabolites ((R)/(S)-2-HG).

Accordingly, the two TCA intermediates reportedly inhibit mTet1/2 and hTET1/2 in the micromolar range (IC_{50} = 83–570 µM, Table 1.1), showing that TET function may be adversely affected by both depletion of 2-OG and accumulation of succinate and fumarate.

Table 1.1 | Reported IC_{50} values of TCA cycle intermediates (succinate, fumarate) and oncometabolites ((R)/(S)-2-HG) for mTet1/2 and hTET1/2. Catalytic domain constructs were used for all reported enzymes.

Metabolite	IC_{50} (μM)			
	mTet1	**mTet2**	**hTET1**	**hTET2**
Succinate	540[171]	570[171]	110[172]	138[172]
Fumarate	390[171]	400[171]	83[172]	204[172]
(R)-2-HG	4,000[173]	5,000[173]	1,000[172]	5,300[174] (15[172])
(S)-2-HG	1,000[173]	1,600[173]	417[172]	12,400[174] (13[172])

Neomorphic IDH mutations (section 1.5.3) were shown to lead to reduced levels of ^{5hm}C *in cellulo*[164], while concurrent *in vitro* work by Xu *et al.*[165] identified 2-HG (Scheme 1.7) as a millimolar inhibitor of mTet1/2, with (R)-2-HG being the more potent enantiomer. Both enantiomers of 2-HG have subsequently been closely evaluated as inhibitors of TETs[173,174] with IC_{50} values between 1–12 mM (Table 1.1); interestingly, both studies identified (S)-2-HG as the more potent stereoisomer, although (R)-2-HG is the only enantiomer produced by IDH mutants. Conversely, inhibition studies carried out within our group[172] found 2-HG to be a double-digit micromolar inhibitor of hTET2, with no observed enantiomeric discrimination (hTET2 IC_{50} = 15 and 13 μM for (R)- and (S)-2-HG, respectively). While the *in vitro* 2-HG characterisation results are conflicting, the hypermethylation phenotype observed in mutant *IDH* cells[164] supports 2-HG as a competitive inhibitor of TET enzymes.

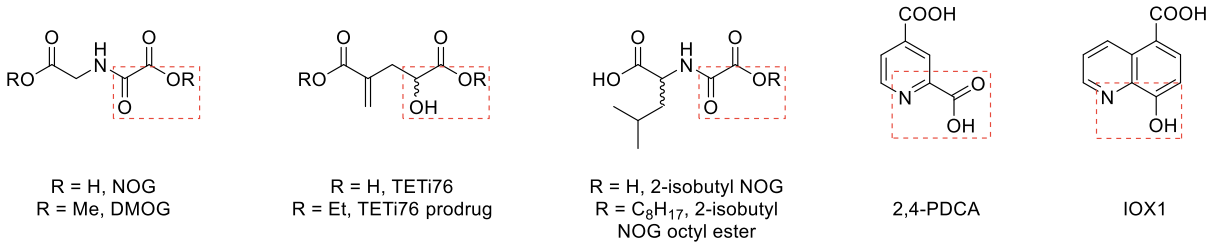
Endogenous inhibitors of 2-OG dioxygenases are unlikely to exhibit chemical probe-like potency and selectivity profiles, but structural derivatisation of the metabolites could lead to more selective analogues.

1.6.2 Fe(II)-chelating 2-OG mimetics

2-OG mimetics preclude the binding and subsequent decarboxylation of 2-OG by occupying the 2-OG binding site and chelating iron *via* bidentate coordination. Competition of 2-OG binding is an extensively utilised approach for 2-OG dioxygenase inhibition, and the number

and structural diversity of 2-OG dioxygenase inhibitors identified to date is extensive[72]; thus, only compounds previously reported as TET inhibitors will be discussed herein.

N–oxalylglycine (NOG, Scheme 1.8)[175] is the closest non-productive analogue of 2-OG, and is commonly used as a general 2-OG dioxygenase inhibitor. NOG is reported to inhibit hTET2 with triple-digit micromolar potency (IC_{50} = 149 μM[174]), and has been used as an inactive 2-OG analogue in crystallographic studies on TETs[94–96,99], confirming its 2-OG-like mode of binding. While NOG exhibits no activity in cells, a cell-penetrating dimethyl ester form of NOG (dimethyl *N*–oxalylglycine, DMOG) is active in cells, and has been shown to non-selectively reduce ^{5hm}C levels in a dose-dependent manner in U2OS cells overexpressing catalytic domains of hTET1–3 (unpublished work by Dr Hilal Sarac).



Scheme 1.8 | Structures of selected 2-OG mimetics reported to inhibit hTETs. Bidentate iron-chelating motifs highlighted in red boxes.

Functionalisation of NOG at the glycine C^{α} has been shown to afford selectivity for either FIH[176] or KDM4s[177] over HIF prolyl hydroxylase PHD2 by exploiting hydrophobic subpockets adjacent to the 2-OG binding site. The approach demonstrated that decoration of non-selective 2-OG mimetics can yield substantially improved selectivity profiles.

Guan *et al*.[178] employed an analogous strategy to develop TETi76, a 4-methenylated 2-HG derivative (Scheme 1.8), which inhibited hTETs with single-digit micromolar potencies (IC_{50} = 1.5, 9.4, and 8.8 μM for hTET1–3, respectively), but was inactive against a panel of 16 other 2-OG dioxygenases. Substitution of 2-hydroxy to 2-oxo resulted in a 5-fold loss of potency[179], suggesting that 4-substituted 2-OG derivatives may still undergo productive

decarboxylation in TETs. Although TETi76 did not exhibit exceptional selectivity among the hTET paralogues (approx. 6-fold selectivity for hTET1), the diethyl ester of TETi76 was shown to induce synthetic lethality in hTET2-deficient cell lines without substantially affecting the viability of closely-related hTET2-proficient cells, illustrating the potential of targeted therapeutic strategies for TET2 mutation-associated diseases[178].

Sudhamalla *et al*.[180] employed complementary steric engineering to develop a "bumped" NOG derivative (2-isobutyl NOG) which selectively inhibited artificial (not found endogenously) hTET variants with sterically expanded active sites, and showed selective cellular activity as the octyl ester. While this approach, termed the "bump-and-hole" tactic[181], offers remarkable selectivity for the particular enzyme variant, the approach is unsuitable for the study of endogenous enzymes and requires prior structural understanding of the protein of interest.

Some of the broadest-spectrum 2-OG dioxygenase inhibitors, 2,4-pyridinedicarboxylate (2,4-PDCA)[182] and 5-carboxy-8-hydroxyquinoline (IOX1)[183], have been shown to attenuate the catalytic activity of hTETs with single-digit micromolar potencies in a largely non-specific manner[172]. Although 2,4-PDCA and IOX1 offer little structural overlap with 2-OG, co-crystallisation with KDM4A revealed that the compounds localise at the 2-OG binding site and interact with the active site metal *via* bidentate chelation in a manner analogous to NOG (Figure 1.6)[184], justifying their classification as 2-OG mimetics herein. Generally, pan-specific inhibitors are of little interest as TET chemical probes due to their non-selective inhibitory profiles, but structural derivatisation of 2,4-PDCA and IOX1 may confer selectivity for a particular set of 2-OG dioxygenases, as exemplified by work on NOG and 2-HG derivatives.

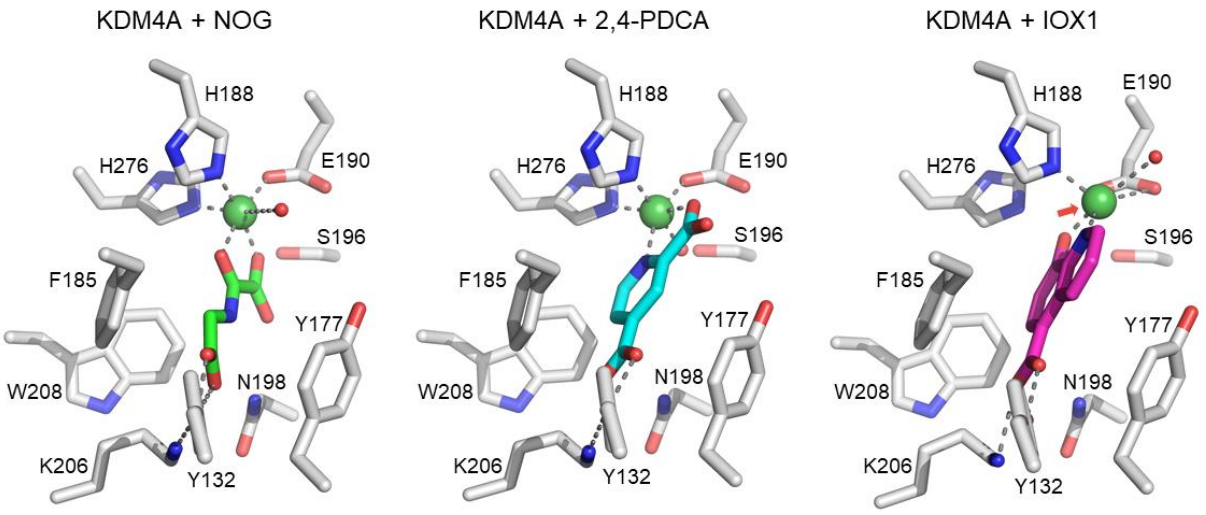


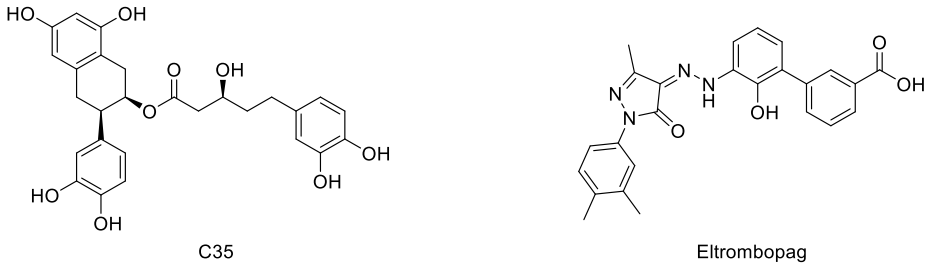
Figure 1.6 | Crystal structures of NOG, 2,4-PDCA, and IOX1 bound to histone lysine demethylase KDM4A. In all structures, the active site Fe(II) is substituted for Ni(II). Red arrow represents IOX1-induced active site metal translocation. Residues involved in interactions with inhibitors are shown in stick representation; metal-ligand and H–bonding interactions are shown as dashed lines. Structural elements are represented by specific colours: light grey, KDM4A; green, NOG; cyan, 2,4-PDCA; magenta, IOX1. Heteroatom colouring: red, O; blue, N; green, Ni. Waters are represented as red spheres. Adapted from literature[184]. PDB IDs: 2OQ6, 2VD7, and 3NJY (left to right)[184].

1.6.3 2-OG-non-competitive Fe(II) chelators

Fe(II) chelators that are non-competitive with respect to 2-OG are a relatively rare class of inhibitors, with only two compounds reported to date. Compound C35 (Scheme 1.9) was identified through a virtual ligand library screening and was found to inhibit hTETs *in vitro* with pan-specific activity (IC_{50} = 1.2, 3.5, and 2.3 µM for hTET1–3, respectively)[185]. The compound purportedly did not affect 2-OG binding, distinguishing it from the broadly available 2-OG mimetics, although more detailed competition experiments are required to validate the claim; molecular modelling suggested C35 partially occupied the DNA substrate binding site, supporting the reported experimental results. C35 was inactive against a panel of 12 2-OG-dependent dioxygenases at 10 µM, and was found to supress formation of ^{5hm}C in cells in a dose-dependent manner, demonstrating the potential utility of the compound as a chemical tool.

Eltrombopag (Epag), an FDA-approved thrombopoietin receptor agonist, has been retrospectively identified as a non-specific TET inhibitor (IC_{50} = 1.0, 1.3, and 1.8 µM for

hTET1–3, respectively)[186]. Molecular docking suggested that Epag inhibits TETs by complexing with iron, in line with previous reports[187], while cofactor saturation experiments revealed non-competitive behaviour towards 2-OG. Epag exhibited cellular activity and was synthetically lethal in TET2-deficient cells[186], in a manner similar to the 2-OG mimetic TETi76[178]. However, Epag has not been profiled against other Fe(II)/2-OG-dependent dioxygenases, and has been recently reported to directly inhibit the pro-apoptotic BCL-2 protein BAX[188], demonstrating poor selectivity for the TET family.



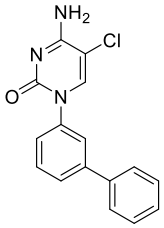
Scheme 1.9 | Structures of 2-OG-non-competitive TET inhibitors C35 (ref. [185]) and eltrombopag (ref. [186]).

1.6.4 Cytosine mimetics

Although substrate mimetics are one of the most common classes of competitive enzyme inhibitors, only one cytosine-based TET inhibitor, Bobcat339, has been reported to date[189] (Scheme 1.10). Bobcat339 is a double-digit micromolar inhibitor of hTETs (IC_{50} = 33 and 73 µM for hTET1 and hTET2, respectively) with excellent selectivity over the cytosine-modifying DNMT3A (<10% inhibition at 500 µM). Computational docking suggested that the 5-chlorocytosine moiety occupies the enzyme active site in a fashion analogous to ^{5m}C, supporting the proposed cytosine mimetic-based mode of action. Bobcat339 demonstrated a degree of cellular activity, as exemplified by a 3-fold reduction of ^{5hm}C levels in HT-22 cells following 10 µM treatment for 24 h.

Importantly, a recent study by Weirath *et al*.[190] suggested that the initially reported inhibitory activity of Bobcat339 is primarily a result of Cu(II) contamination stemming from the use of the Cu(II)-catalysed Chan−Lam coupling reaction to produce the *N*–phenylated compound,

with virtually no inhibitory activity observed when screening samples of stringently purified Bobcat339.

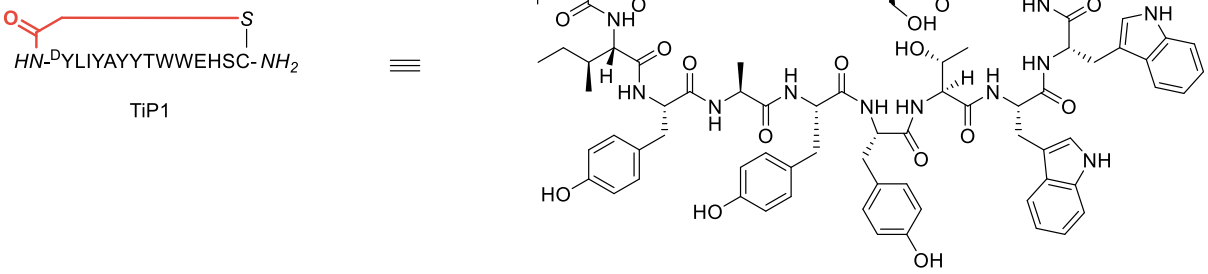


Bobcat339

Scheme 1.10 | Structure of cytosine mimetic Bobcat339 (ref. [189]).

1.6.5 Cyclic peptide inhibitors

Nishio et al.[191] used the mRNA display-based random nonstandard peptide integrated discovery (RaPID) methodology to identify TiP1 (Scheme 1.11), a hTET1-selective thioether macrocyclic peptide inhibitor. While TiP1 inhibited hTET1 with >7-fold selectivity over hTET2 (IC_{50} = 1.48 and >10 μM for hTET1 and hTET2, respectively), selectivity over other 2-OG dioxygenases was not determined. Nevertheless, TiP1 is the most selective inhibitor of a specific hTET paralogue to date, highlighting the difficulty of obtaining selectivity both within the hTET family and the broader 2-OG-dependent dioxygenase superfamily.



TiP1

Scheme 1.11 | Schematic representation and structure of thioether macrocyclic peptide TiP1 identified using the RaPID system[191]. Letters representing individual atoms are italicised in the schematic representation (left) to distinguish from single-letter amino acid codes.

1.7 Thesis objectives

The work described herein aims to contribute towards the expansion of biochemical understanding of TET enzymes and the development of TET chemical probes.

The aim of Chapter 2 is to characterise the biochemical properties of NgTet1 and assess the influence of TET activity modulators on uncoupled 2-OG turnover. Catalytic detection methods based on DNA oxidation and succinate formation will be developed in order to facilitate bimodal evaluation of NgTet1 activity. Using these methods, known TET inhibitors will be screened against NgTet1 for inhibition of substrate turnover and uncoupled 2-OG decarboxylation.

Chemical probes targeting epigenetic proteins are indispensable tools for phenotypic profiling, mechanistic investigations, and target validation[192], but selective TET-targeting molecular tools have not been identified to date. Cyclic peptide TiP1[191] has been recently reported as a hTET1-selective inhibitor, but TiP1 selectivity profiling is lacking, and cellular activity has not been evaluated. Thus, Chapter 3 will focus on the *in vitro* profiling of TiP1 and related peptides against a panel of TET enzymes, followed by efforts to obtain cellular activity with TiP1 and its derivatives. Alongside cyclic peptide inhibitor development, small molecule TET inhibitor development will be explored as well. Identification of a novel hTET2 inhibitor scaffold, followed by SAR studies and subsequent evaluation of selected analogue cellular activity will be described in Chapter 4.

Structural understanding of the interactions between inhibitors and TET enzymes is fundamental to inhibitor mechanism of action determination and streamlined structure-guided inhibitor optimisation, but the 2-OG mimetic NOG is the only inhibitor co-crystallised with TETs to date[94–96,99]. Thus, the aim of Chapter 5 is to develop a platform for the co-crystallisation of inhibitors with the crystallisable construct $hTET2_{CD\Delta LCI}$. Protein

expression and purification will be optimised to allow isolation of sufficient amounts of enzyme, and crystallisation trials with DNA substrate and inhibitors will be carried out.

Chapter 2
NgTet1 biochemical characterisation and inhibitor screening using DNA oxidation- and 2-OG decarboxylation-based detection methods

2.1 Introduction

While the roles of TET enzymes in the control of DNA methylation state, development, and cancer have been firmly established in cells and *in vivo* (see section 1.5), biochemical characterisation of the isolated enzymes has been challenging due to the lack of readily available recombinant TET enzymes and high-throughput activity assays. LC-MS-based activity assays and binding assays have been used to determine the substrate preference of NgTet1[193] and hTET2$_{CD\Delta LCI}$[96] (section 1.3.4), and attempts have been made to identify the mechanism of sequential ^{5m}C oxidation (section 1.3.3), albeit with conflicting results. Because the approach used in the majority of previous work relies on the direct detection of DNA substrate oxidation from ^{5m}C through to ^{5ca}C, little information is reported on the enzymatic decarboxylation of 2-OG and its regulation by catalytic activity modifiers, such as DNA oligomers or inhibitors. Laukka *et al*.[171] used a method detecting the release of $^{14}CO_2$ from 2-oxo[1-^{14}C]glutarate to obtain 2-OG K_M values for murine Tets (mTet1/2) in the presence of DNA substrate; Alves *et al.*[80] described a comparison of total hTET1-mediated succinate production in the presence and absence of substrate using the enzyme-coupled Succinate-Glo™ method; however, there have been no reports on enzymatic 2-OG turnover regulation by endogenous or synthetic TET-binding molecules.

2.1.1 Methods for TET catalytic activity detection

A number of methods have been described for the detection of TET catalytic activity (summarised in Table 2.1). The majority of TET biochemical characterisation work to date has

been carried out using LC-MS/MS methods[94,96,99,193], where DNA oligomers are purified after incubation with TET enzyme, digested to nucleosides, and subjected to LC-MS/MS analysis. While this method enables direct quantification of cytidine nucleoside derivatives (^{5m}C through to ^{5ca}C), the labour- and resource-intensive low-throughput nature of the assay precludes practical collection of larger data sets.

Table 2.1 | Overview of TET catalytic activity detection methods.

	Method	Advantages	Disadvantages
Substrate detection	LC-MS/MS[94,96,99,193]	Simultaneous detection of all ^{5m}C oxidation states; direct read-out	Laborious sample preparation; long analysis time
	MALDI-TOF MS[140,174]	Direct sample analysis without purification; direct read-out; short analysis time	Low sensitivity compared to LC-MS/MS; Impossible to distinguish ^{5hm}C from ^{5f}C without derivatisation; only short oligonucleotides usable
	DNA radiolabelling[194]	Superseded by LC-MS/MS	No molecular identification; laborious sample preparation; low-throughput; requires specialist equipment
	Antibody dot blot[195]	Superseded by ELISA/AlphaScreen	Only one ^{5m}C oxidation state observed; low-throughput
	Commercial TET ELISA kits[189,196]	Remarkably sensitive; high-throughput	Only one ^{5m}C oxidation state observed; expensive; proprietary composition
	AlphaScreen[191]	Remarkably sensitive; high-throughput	Only one ^{5m}C oxidation state observed; expensive; assay component concentration limited by detrimental interactions; only ssDNA usable
Succinate detection	2-OG radiolabelling[171]	Enzyme-free read-out (*cf.* Succinate-Glo™)	Low-throughput; requires specialist equipment
	Succinate-Glo™[80]	High-throughput	Multiple enzyme-coupled read-out; expensive; proprietary composition; low reported linear detection range (up to 15 µM succinate)
	^{1}H NMR[81]	Continuous reaction monitoring, direct read-out	High protein and substrate consumption; unavailable in plate format

MALDI-TOF MS facilitates detection of full-length DNA oligonucleotides without the need for enzymatic digestion and purification of the oxidised ^{5m}C products. Instead, the sample is directly co-crystallised with a matrix of choice and ionised using a high power laser[140,174]. The time required to analyse each sample is substantially shorter compared to LC-MS/MS (seconds *vs* minutes). However, the method suffers from lower sensitivity compared to LC-MS/MS,

reliance on short oligonucleotides for sufficient ionisation, and a lower level of isotopic resolution resulting in the overlap of ^{5hm}C and ^{5f}C peaks.

Two radiolabelling approaches have been reported for TET catalytic activity detection. In work by Ito *et al.*[194], DNA oligomers were fragmented following incubation with TET (MspI), labelled with [γ-^{32}P]ATP, and digested to nucleotides. Separation on a TLC plate and exposure to X-ray film enabled relative abundance determination for ^{5m}C and its oxidation products. DNA radiolabelling suffers from a lack of direct molecular identification and low-throughput due to laborious sample preparation; it has thus been largely superseded by LC-MS/MS methods. Conversely, Laukka *et al.*[171] employed 2-oxo[1-^{14}C]-glutarate as a cofactor to measure the release of $^{14}CO_2$ as a read-out of enzymatic 2-OG decarboxylation. This method, while an interesting option due to it being the only detection method for CO_2 evolution, is disadvantaged by low-throughput as well as the use of radiolabels that require specialised plate readers.

Multiple antibody-based ^{5m}C oxidation product detection methods have been reported for TETs: Szulwach *et al.*[195] described the use of a dot blot assay with an anti-^{5hm}C antibody to obtain relative abundance values for ^{5hm}C; anti-^{5hm}C antibody-based TET ELISA kits are available from multiple manufacturers and have been used by a number of groups[189,196]; and an anti-^{5hm}C antibody detection method using AlphaScreen technology has been developed by our group[191]. Antibody-based methods enable analyte detection down to femtomole quantities (10^{-13}–10^{-15} moles[191]), facilitating use of minute enzyme and substrate amounts. This is of particular importance in cases where isolation of pure catalytically active enzyme is challenging, such as for hTET1 and hTET3. However, as antibodies are specific for a single cytosine modification, only one ^{5m}C oxidation state can be analysed at a time. Commercial ELISA kits are expensive and method optimisation is largely impractical due to the proprietary

nature of reagent composition. AlphaScreen does not permit expedient kinetic parameter determination due to the low concentrations of substrate (considerably below K_M) required to avoid detrimental binary association between AlphaScreen beads and biotinylated substrate.

Succinate-Glo™ is an enzyme-coupled bioluminescent end-point assay used for the detection of succinate[80]. The method uses two proprietary succinate detection reagents (SDR I and SDR II) in succession to produce bioluminescence signal based on the concentration of succinate present in the sample. The main drawbacks of the method are the enzyme-coupled nature of the assay and the low linear detection range (up to 15 µM succinate).

Finally, a report by Hopkinson *et al.*[81] on demethylase activity monitoring using ^{1}H NMR demonstrated that NMR spectroscopy could be used to monitor 2-OG conversion to succinate in real time. While the sensitivity of ^{1}H NMR spectroscopy is worse than that of antibody-based methods or assays such as Succinate-Glo™, ^{1}H NMR is the only method that allows continuous reaction monitoring. Furthermore, there are no restrictions on the concentrations of assay components, supporting ^{1}H NMR as a powerful tool for kinetic parameter determination. The main drawbacks of the method are high consumption of enzyme and substrate, and the single-reaction monitoring nature of NMR.

2.2 Aims

Comprehensive biochemical characterisation of TETs has been hindered by difficulties in preparation of recombinant TET enzymes and a lack of high-throughput assays, with current literature focusing primarily on the studies of substrate preference and the effects of oligonucleotide sequence on catalytic efficiency. Thus, robust medium- to high-throughput assays are required to facilitate further kinetic characterisation, mechanistic studies, and profiling of large inhibitor libraries.

Simultaneous observation of ^{5m}C oxidation and 2-OG turnover would facilitate the determination of TET enzyme efficiency by comparing levels of coupled and uncoupled 2-OG decarboxylation. Characterisation of TET inhibitors using both methods may reveal additional information regarding their mode of action – discrepancy of potency between the different assays may relate to changes in enzymatic 2-OG decarboxylation rates due to a re-organisation of the active site in the presence of the binder. Furthermore, regulation of 2-OG turnover by TET ligands, such as endogenous metabolites and DNA oligomers or exogenous inhibitors, could have implications for the control of 2-OG levels within the context of the citric acid cycle.

Due to the difficulty in obtaining sufficient quantities of pure recombinant mammalian TET enzymes, *Naegleria* Tet-like dioxygenase NgTet1[99] was selected as a model enzyme, encouraged by extensive overlap of key active site residues with hTET2 (section 1.4, Figure 1.4).

Taking into account the catalytic activity detection types, advantages, and drawbacks of the discussed methodologies, AlphaScreen, MALDI-TOF MS, Succinate-Glo™, and ^{1}H NMR methods were selected for assay optimisation and characterisation of NgTet1.

The aims of the work presented in this chapter are as follows:

- Produce catalytically active NgTet1.

- Develop robust TET catalytic activity assays for the simultaneous detection of ^{5m}C oxidation and 2-OG decarboxylation.
- Using the developed assays, determine NgTet1 kinetic parameters for 2-OG in the presence of methylated and non-methylated DNA motifs.
- Characterise known 2-OG dioxygenase inhibitors using orthogonal assays to determine the effect compound mode of inhibition may have on uncoupled 2-OG decarboxylation.

2.3 Results

2.3.1 Production of *Naegleria* Tet-like dioxygenase NgTet1

Production of *Naegleria* Tet-like dioxygenase NgTet1 was carried out following the method reported by Hashimoto *et al.*[99] with minor modifications. Recombinant expression of full-length NgTet1 fused with an N–terminal hexahistidine (His_6) tag and a thrombin cleavage site in a pET28 vector was carried out in *E. coli* BL21(DE3) strain. Transformants were grown according to conditions in Figure 2.1 A. NgTet1 was purified from cell lysate by immobilised metal affinity chromatography using Ni Sepharose® High Performance resin (HisTrap IMAC), followed by size exclusion chromatography using Superdex® 75 resin (SEC S75, Figure 2.1 B). Cell lysis by sonication followed by HisTrap IMAC afforded a single peak by UV which contained mainly the protein of interest by SDS-PAGE (Figure 2.1 C, E). During the subsequent SEC step, a main peak and a "shoulder" were observed in the UV trace, both containing the protein of interest by SDS-PAGE (Figure 2.1 D, E). Protein eluted in the "shoulder" was discarded to ensure sample homogeneity; fractions containing the centre of the main peak were pooled to afford NgTet1 in good yield and purity (4.2 mg/L of cell culture, >90% purity by SDS-PAGE, stored at 15.9–18.6 mg/mL in 20 mM HEPES, 150 mM NaCl, pH 7.5).

LC-MS confirmed the identity of the protein isolate as His_6-tagged NgTet1 with the observed mass corresponding to the loss of N–terminal methionine (Figure 2.1 F, predicted mass: 38,530.5 Da, observed mass: 38,529.9 Da). N–terminal methionine excision (NME) is a well-characterised pathway[197–199], resulting from formylmethionine deformylation by peptide deformylases (PDFs) followed by methionine excision by methionine aminopeptidases (MAPs). MAP catalytic efficiency for N–terminal M▼G sequences has been shown to be remarkably high (>97%)[197], resulting in a near-uniform methionine-excised protein population.

A

Strain	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
BL21(DE3)	2×YT	37 °C	0.78	500 µM	16 °C	19 h

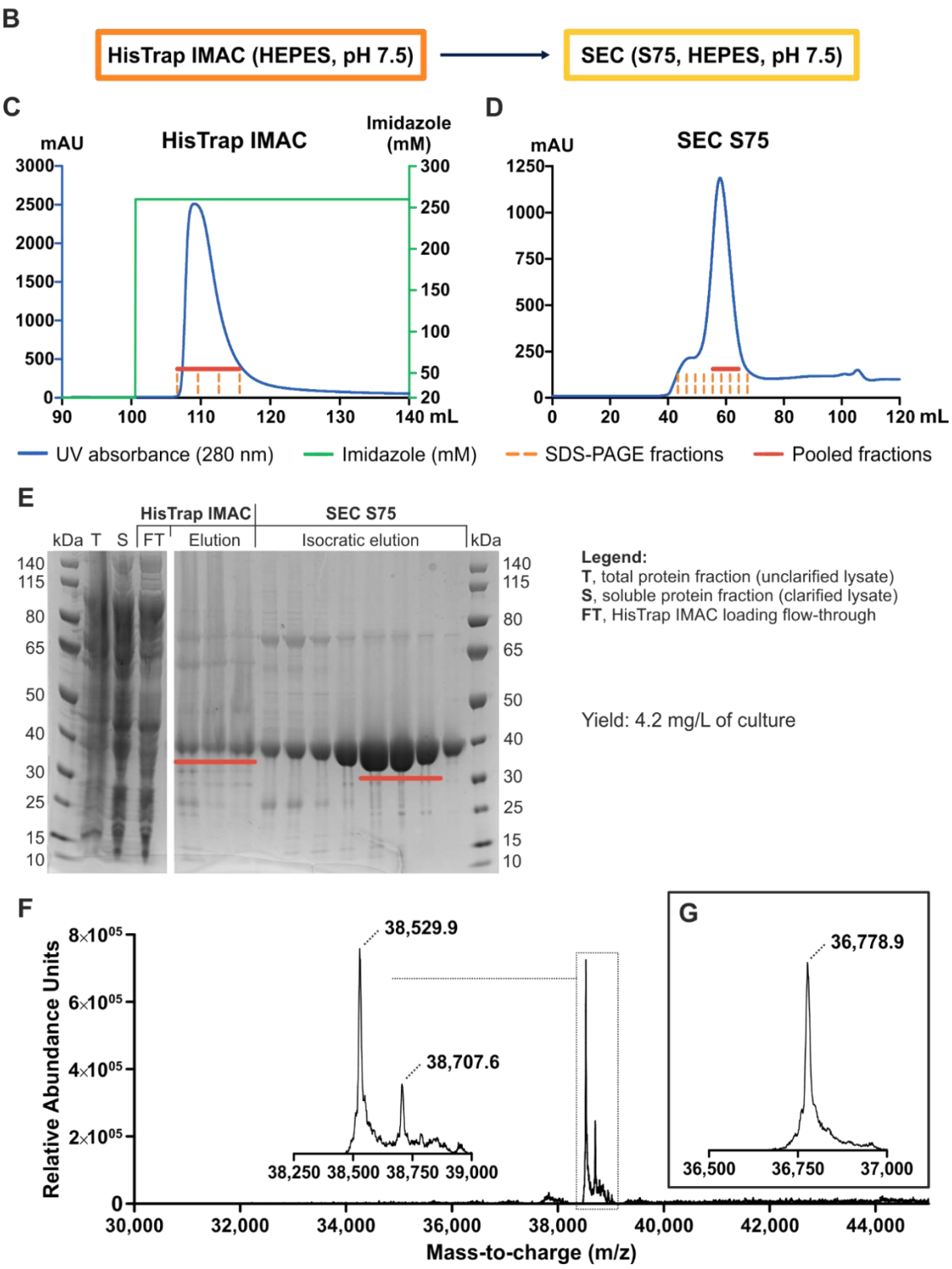
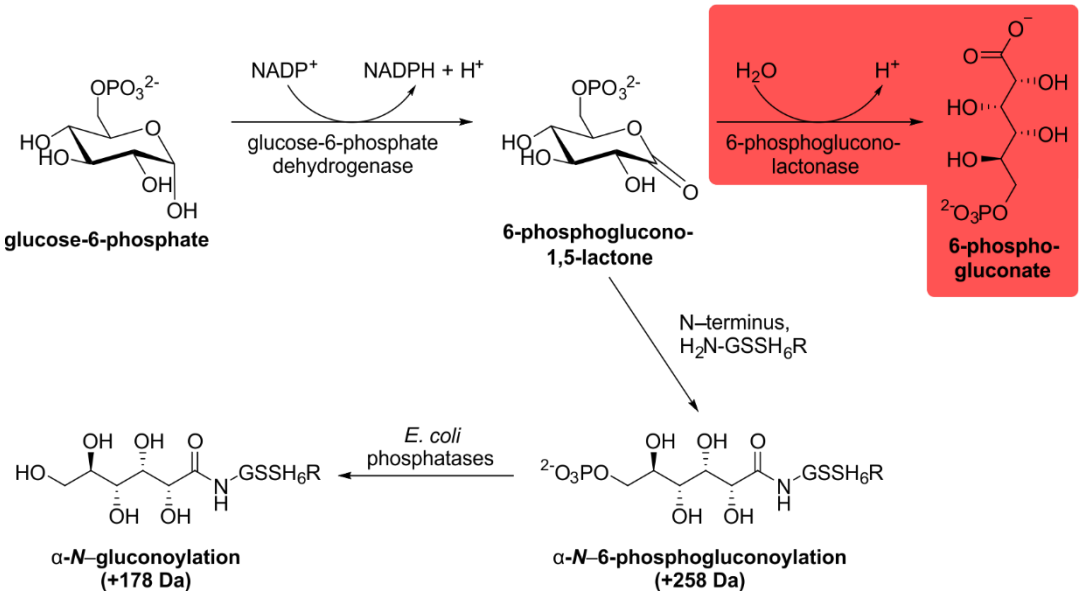


Figure 2.1 | Production of recombinant NgTet1. A, cell growth conditions; **B**, purification workflow; **C**, HisTrap IMAC 280 nm elution profile; **D**, size exclusion (S75) 280 nm elution profile; **E**, SDS-PAGE gel for HisTrap IMAC and size exclusion (S75); **F**, LC-MS analysis of purified NgTet1 (predicted mass for methionine-excised[199] NgTet1: 38,530.5 Da, observed mass: 38,529.9 Da); **G**, LC-MS analysis of purified NgTet1 after thrombin treatment (predicted mass: 36,779.6 Da, observed mass: 36,778.9 Da). Red lines indicating pooled fractions are placed underneath the protein of interest bands in the SDS-PAGE gel.

A +178 Da peak (around 20% of ion intensity) was observed in the mass spectrum, attributed to the spontaneous α-*N*–gluconoylation of the N–terminal Gly preceding the His_6 tag (**H_2N–G**SSH_6R, Scheme 2.1)[200]. Treatment of the purified NgTet1 with thrombin protease confirmed that the modification was situated at the N–terminus (Figure 2.1 G, predicted mass: 36,779.6 Da, observed mass: 36,778.9 Da). *E. coli* BL21(DE3) strains are known to accumulate 6-phosphoglucono-1,5-lactone due to the absence of 6-phosphogluconolactonase (PGL)[201]. In the presence of 6-phosphoglucono-1,5-lactone, α-*N*–6-phosphogluconoylation has been shown to occur spontaneously on the α-amino group, followed by dephosphorylation by *E. coli* phosphatases to produce α-*N*–gluconoylation[200].



Scheme 2.1 | Outline of proposed chemical route for the α-*N*–gluconoylation of N–terminal Gly preceding the His_6 tag[200]. Formation of 6-phosphogluconate (shaded in red) is not possible in *E. coli* BL21(DE3) strains due to the absence of 6-phosphogluconolactonase (PGL), leading to the accumulation of 6-phosphoglucono-1,5-lactone. Letter R in the single amino acid code sequence refers to the remaining NgTet1 construct sequence.

Partial gluconoylation is reported to interfere with protein crystallisation, but no adverse effects on protein folding have been observed[202]. As such, the His_6 tag was not cleaved and the partially gluconoylated protein was used for optimisation of catalytic activity detection methods without further purification.

2.3.2 Selection of catalytic detection methods

In order to carry out the characterisation of NgTet1, robust methods for the detection of both ^{5m}C substrate oxidation and 2-OG decarboxylation to succinate were required. Out of the discussed DNA oxidation and 2-OG decarboxylation detection methods (summarised in Table 2.1), AlphaScreen, MALDI-TOF MS, and Succinate-Glo™ were selected for catalytic activity detection due to the high-throughput nature of the approaches, while ^{1}H NMR was selected due to its capacity for continuous reaction monitoring.

2.3.3 NgTet1 catalytic activity detection using AlphaScreen

In order to measure the catalytic activity of NgTet1, Amplified Luminescent Proximity Homogeneous Assay (AlphaScreen) was explored first, as the method had already been optimised within our group for hTET activity detection[191]. AlphaScreen relies on a proximity-based chemical reaction cascade between photosensitiser-conjugated donor beads and singlet state oxygen ($^{1}O_2$)-activated fluorophore-conjugated acceptor beads (Figure 2.2). Photosensitiser phthalocyanine converts ambient oxygen to the reactive $^{1}O_2$ ($t_{1/2}$ = 4 μsec) upon irradiation at 680 nm. If a donor bead is within 200 nm of an acceptor bead, energy transfer to a 3-dye system (thioxene/anthracene/rubrene) results in light emission at 520–620 nm. AlphaScreen signal is therefore dependent on bead proximity, which is dictated by a chain of interactions. The donor bead is coated with streptavidin which binds tightly to biotinylated single-stranded DNA (bt-ssDNA). The bt-ssDNA strand contains a single ^{5m}C base which can be oxidised by TET enzymes to produce ^{5hm}C[†]. ^{5hm}C is in turn recognised by an anti-^{5hm}C IgG antibody that is bound to the protein A-coated acceptor beads, forming a tether between the two beads. As the interaction chain is dependent on ^{5hm}C level, AlphaScreen signal decreases in relation to enzyme inhibition.

† Nucleobase modifications, including ^{5m}C, ^{5hm}C, ^{5f}C, ^{5ca}C, and others, refer to modifications of deoxyribonucleotides in DNA oligomers unless otherwise mentioned.

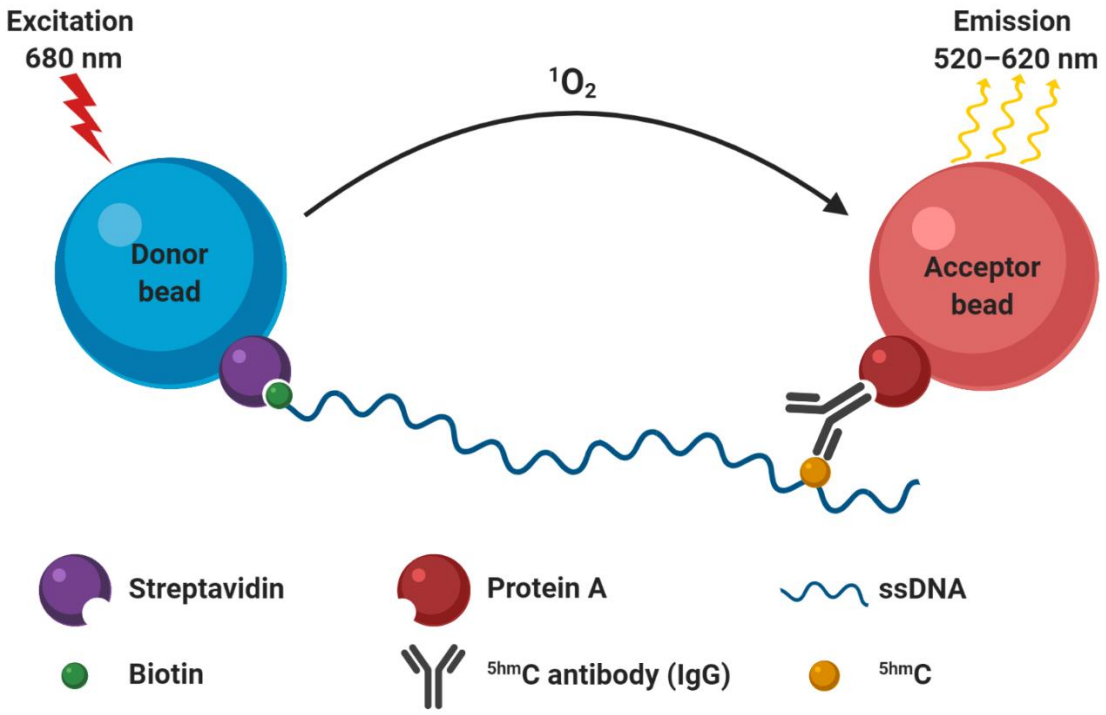


Figure 2.2 | A schematic representation of anti-^{5hm}C antibody-dependent AlphaScreen General IgG (Protein A) assay. Upon irradiation at 680 nm, photosensitiser-coated donor beads produce singlet state oxygen ($^1O_2$), which results in light emission at 520–620 nm by the acceptor beads. The proximity required for luminescence is achieved by employing IgG-based antibodies: ^{5hm}C on the DNA strand is recognised by an anti-^{5hm}C IgG antibody, which is bound to protein A coated acceptor beads; the biotin moiety on the DNA strand binds to streptavidin-coated donor beads. The tethering effect is sufficient to hold the beads in close proximity. AlphaScreen signal decreases in relation to enzyme inhibition. Created with biorender.com.

2.3.3.1 Optimisation of ^{5m}C turnover detection

To evaluate the selectivity of the anti-^{5hm}C antibody and the linear range of the assay, three DNA standard curves were prepared using previously reported DNA oligomers[191]: (i) 0–10 nM of **^{5m}C 32 nt bt-ssDNA**, (ii) 0–10 nM of **^{5hm}C 32 nt bt-ssDNA**, and (iii) 0–100% of **^{5hm}C 32 nt bt-ssDNA** in **^{5m}C 32 nt bt-ssDNA** at a 10 nM total DNA concentration (DNA oligomer sequence information tabulated in Appendix 1). Concentrations of DNA oligomers above 10 nM were not explored due to saturation of the luminescence detector at higher analyte levels.

Standard curve (i) containing **^{5m}C 32 nt bt-ssDNA** produced minimal levels of luminescence (Figure 2.3), while standard curve (ii) containing **^{5hm}C 32 nt bt-ssDNA** produced a linear increase in signal (R^2 = 0.99) throughout the full concentration range, confirming the antibody selectivity for ^{5hm}C over ^{5m}C. To simulate enzymatic conversion of ^{5m}C to ^{5hm}C, curve (iii) was prepared where **^{5hm}C 32 nt bt-ssDNA** was titrated into **^{5m}C 32 nt bt-ssDNA** at a 10 nM total

DNA concentration. Curve (iii) showed a ^{5hm}C concentration-dependent increase in signal, but the linear response range was decreased compared to (ii). While 0–2 nM of **^{5hm}C 32 nt bt-ssDNA** in **^{5m}C 32 nt bt-ssDNA** produced a highly linear response (R^2 = 0.98), further increase in **^{5hm}C 32 nt bt-ssDNA** concentration resulted in loss of response linearity.

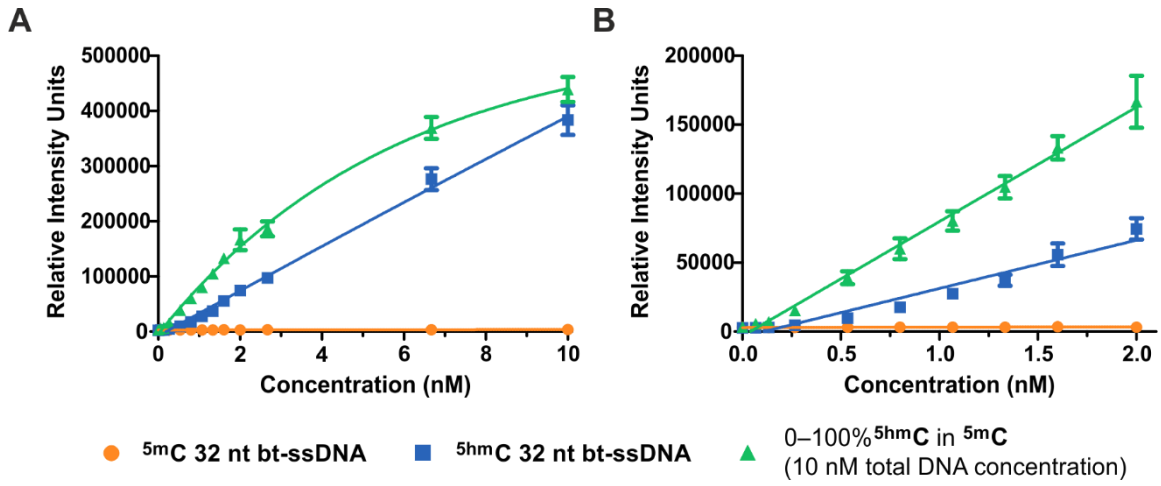


Figure 2.3 | DNA standard curves for TET catalytic activity AlphaScreen assay. A, standard curves (0–10 nM); **B**, partial range (0–2 nM) of the standard curves. **^{5m}C 32 nt bt-ssDNA** produced low signal. **^{5hm}C 32 nt bt-ssDNA** produced a linear increase in signal (R^2 = 0.99) throughout the full concentration range. The titration of **^{5hm}C 32 nt bt-ssDNA** into **^{5m}C 32 nt bt-ssDNA** did not afford a full linear range of detection, but was linear within the 0–2 nM region (R^2 = 0.98). The concentrations refer to concentrations in a 10 µL reaction volume. Data shown as the mean of 4 experimental replicates (n = 4; mean ± SD).

In order to increase the luminescence signal and improve the signal-to-noise (S/N) ratio at ^{5hm}C concentrations below 1 nM, the concentration of anti-^{5hm}C antibody in the final assay mixture was optimised (Figure 2.4). The antibody dilution series (1:4,000–1:24,000) contained **^{5m}C 32 nt bt-ssDNA** (8.5 nM) and **^{5hm}C 32 nt bt ssDNA** (1.5 nM) to emulate signal stemming from TET-mediated ^{5m}C oxidation to ^{5hm}C. The 1:8,000 dilution was selected as the optimal condition for future experiments, as it improved the S/N ratio without reducing the linear response range (data not shown).

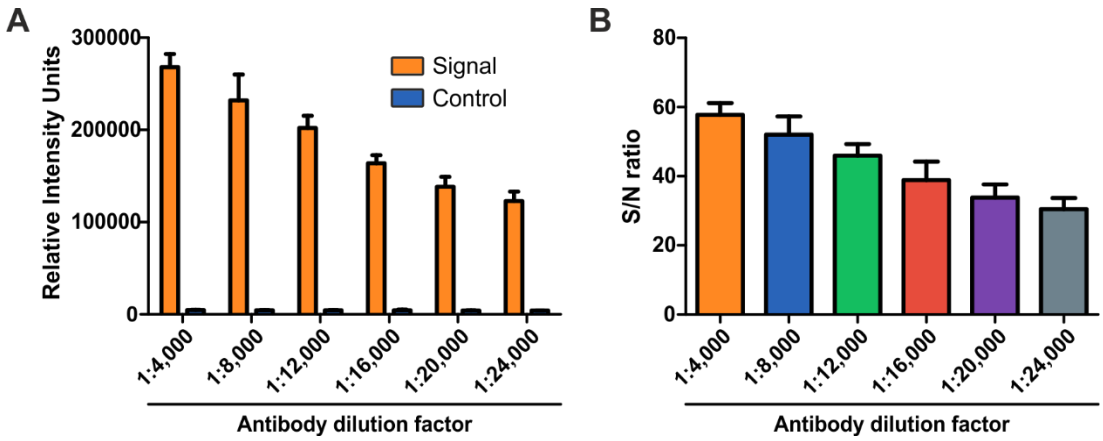


Figure 2.4 | Anti-^{5hm}C antibody concentration optimisation for TET catalytic activity AlphaScreen assay. A, signal and no **^{5hm}C 32 nt bt-ssDNA** control for each antibody dilution; **B**, signal-to-noise (S/N) ratio for each antibody dilution. *x* axis labels refer to the final anti-^{5hm}C antibody dilution factor in a 20 µL final volume, which contained 4.25 nM **^{5m}C 32 nt bt-ssDNA**, 0.75 nM **^{5hm}C 32 nt bt-ssDNA**, 50 µM sodium L-ascorbate, 5 µM $(NH_4)_2Fe(SO_4)_2$, and 5 µM disodium 2-OG. S/N ratio, signal-to-noise ratio. Data shown as the mean of 14 experimental replicates (n = 14; mean ± SD).

2.3.3.2 NgTet1 catalytic activity detection

With the optimised AlphaScreen assay conditions in hand, the catalytic activity of recombinant NgTet1 enzyme was assessed. A time course was carried out using previously reported cofactor concentrations[191] to investigate NgTet1 activity in the presence of DNA substrate **^{5m}C 32 nt bt-ssDNA** (10 nM) and cofactors sodium L-ascorbate (100 µM), $(NH_4)_2Fe(SO_4)_2$ (10 µM), and disodium 2-OG (10 µM). The catalytic activity of NgTet1 was substantially lower than that reported for hTETs[191], requiring at least 200 nM of NgTet1 to produce appreciable levels of luminescence (Figure 2.5 A). To quantify the turnover of ^{5m}C by NgTet1, the experiment was normalised against a standard curve (**^{5hm}C 32 nt bt-ssDNA** in **^{5m}C 32 nt bt-ssDNA** at a 10 nM total DNA concentration), revealing that turnover did not exceed 20% of the ^{5m}C substrate after 30 min (Figure 2.5 B).

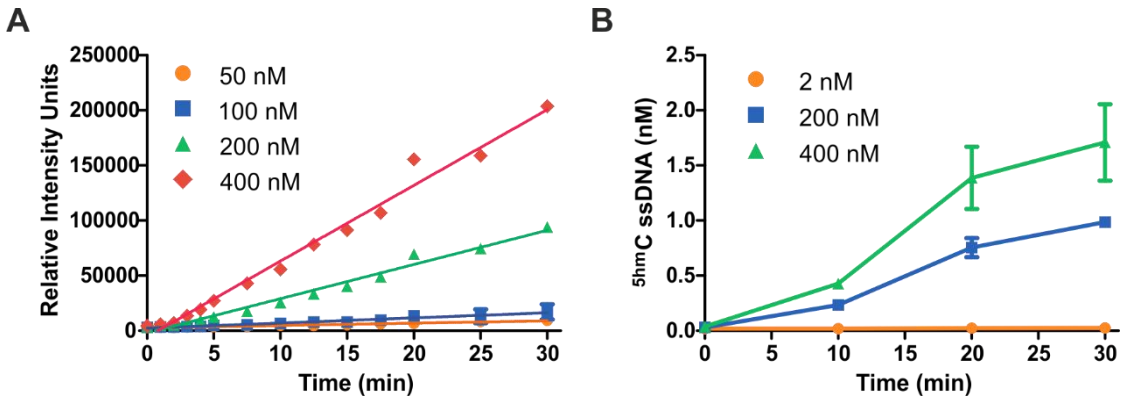


Figure 2.5 | NgTet1 catalytic activity detection using AlphaScreen. A, detailed time course at different NgTet1 concentrations; **B**, normalised time course for selected NgTet1 concentrations. Conditions: 2–400 nM NgTet1, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. Data shown as the mean of 1–3 experimental replicates (n = 1–3; mean ± SD).

Detection of NgTet1 catalytic activity using AlphaScreen was encouraging; however, the high concentration of NgTet1 required for appreciable ^{5m}C turnover (20-fold excess over substrate **^{5m}C 32 nt bt-ssDNA**) challenged the quality of the recombinantly produced NgTet1, as well as the compatibility of the protein with the assay conditions. Firstly, the recombinant enzyme could be partially unfolded, resulting in a much lower concentration of the functional enzyme species. Secondly, NgTet1 could be an inherently less efficient hydroxylase compared to hTETs, as no information regarding the relative activity of NgTet1 and its human counterpart hTETs is available. Thirdly, although NgTet1 has been shown to be active on both ssDNA and dsDNA[193], no kinetic data comparing ssDNA and dsDNA substrates is available. NgTet1 may therefore have a much lower affinity for single-stranded DNA oligomers and/or 2-OG when compared to hTETs, resulting in a low apparent activity.

While AlphaScreen has proven to be a powerful tool for detecting femtomole quantities of ^{5hm}C, the anti-^{5hm}C antibody-based approach neglects the subsequent oxidation products ^{5f}C and ^{5ca}C. As such, care must be taken to ensure all assays are carried out within the linear range of ^{5hm}C product formation, while a sufficient amount of unreacted ^{5m}C is still remaining.

2.3.4 NgTet1 catalytic activity detection using MALDI-TOF MS

Only biotinylated single-stranded oligonucleotides could be used as substrates in AlphaScreen, as both assay repeatability and the signal-to-noise ratio were reported to be substantially worse when using double-stranded substrates[172]. In order to investigate double-stranded DNA (dsDNA) substrates, a MALDI-TOF MS-based method for TET catalytic activity detection was optimised, providing an orthogonal condition set for ^{5m}C oxidation detection. Assays were carried out following the method reported by Sudhamalla *et al*.[174] with minor modifications.

2.3.4.1 Optimisation of DNA detection

DNA oligomers used in AlphaScreen (**^{5m}C/^{5hm}C 32 nt bt-ssDNA**) were found to produce insufficient levels of ionisation by MALDI-TOF MS (results not shown). Conversely, a shorter, self-complementary double-stranded **^{5m}C 12 bp dsDNA** oligomer (DNA oligomer sequence information tabulated in Appendix 1), reportedly used for co-crystallisation with hTET2[94], produced much better signal and could be observed by MALDI-TOF MS at concentrations down to 1 μM.

2.3.4.2 NgTet1 catalytic activity detection

A time course was carried out with NgTet1 (2 μM) in the presence of **^{5m}C 12 bp dsDNA** (2 μM) and cofactors sodium L-ascorbate (500 μM), $(NH_4)_2Fe(SO_4)_2$ (50 μM), and 2-OG (50 μM), demonstrating a time-dependent decrease in relative ^{5m}C substrate abundance, concomitant with the appearance of an approximately +16 Da peak (Figure 2.6 A, B). Although mass spectrometry methods generally benefit from the simultaneous analysis of all analytes present in the sample, the small mass difference (−2 Da) between ^{5hm}C and ^{5f}C combined with the poor isotopic resolution of MALDI-TOF MS prohibited differentiation of the two oxidation products (Figure 2.6 B). As such, catalytic activity was measured by the relative consumption of ^{5m}C, rather than the appearance of oxidation product peaks.

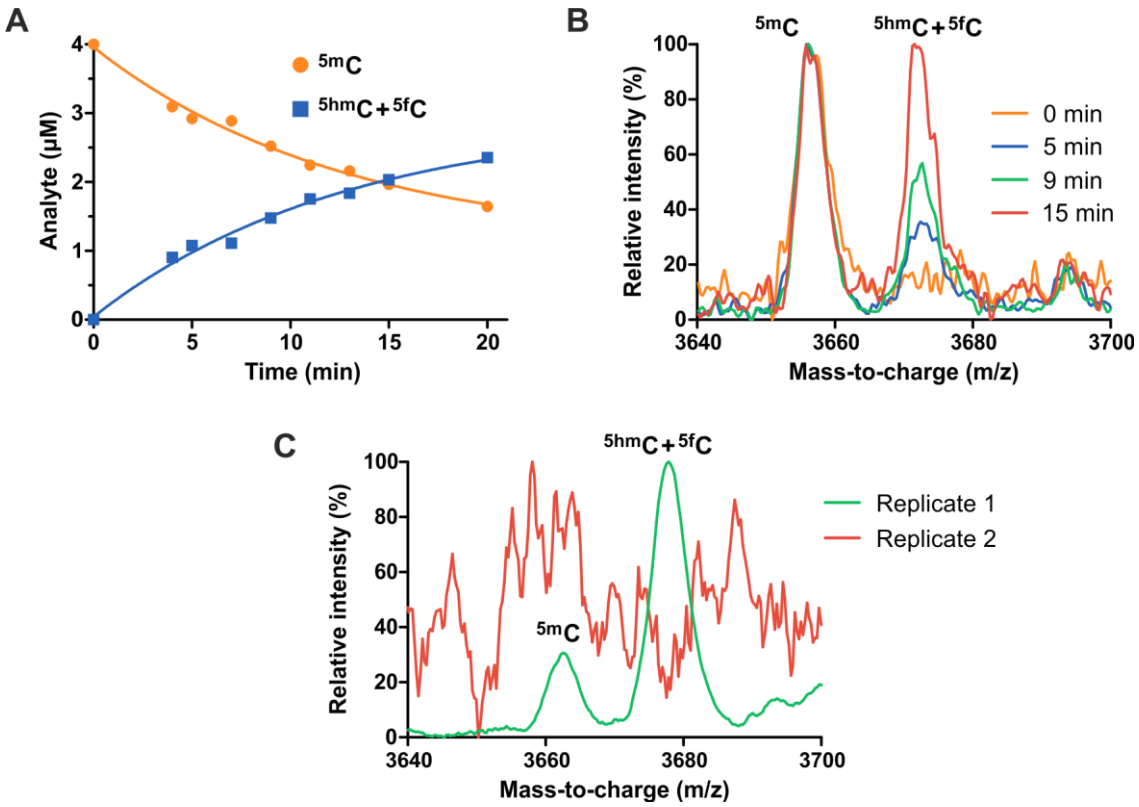


Figure 2.6 | NgTet1 catalytic activity detection using MALDI-TOF MS. A, a time course of NgTet1-mediated DNA substrate consumption; **B**, representative mass spectra from time course **A**; **C**, spectra highlighting ionisation inconsistency in multiple replicates of a single sample. Conditions (10 μL reaction volume): 2 μM NgTet1, 2 μM **^{5m}C 12 bp dsDNA**, 500 μM sodium L-ascorbate, 50 μM $(NH_4)_2Fe(SO_4)_2$, and 50 μM disodium 2-OG. Data from single replicates shown.

NgTet1-catalysed ^{5m}C oxidation was more efficient in the MALDI-TOF MS assay (over 50% turnover after 20 min with 1:2 NgTet1/^{5m}C stoichiometry) compared to the AlphaScreen assay (under 20% turnover after 30 min with 40:1 NgTet1/^{5m}C stoichiometry), demonstrating much higher NgTet1 catalytic activity than initially observed using AlphaScreen. The marked difference in substrate turnover levels between the two assays suggested that the low activity observed by AlphaScreen was likely not due to protein misfolding. Previously, Pais *et al.*[193] used single-digit micromolar concentrations of DNA substrate and 2:1 NgTet1/^{5m}C stoichiometry to measure the NgTet1-catalysed oxidation of ^{5m}C in ssDNA, suggesting that NgTet1 turnover of ssDNA may not be efficient. While the K_M values are unknown, a comparison of the concentrations of ^{5m}C substrate and 2-OG used in MALDI-TOF MS and AlphaScreen assays (^{5m}C: 4 μM *vs* 10 nM; 2-OG: 50 μM *vs* 10 μM) further imply that the DNA substrate and 2-OG concentrations used in AlphaScreen may be below those required for

optimal catalytic activity. Although NgTet1 time course datasets could be successfully collected using MALDI-TOF MS, data collection was hindered by poor repeatability even when analysing replicates of a single sample (Figure 2.6 C), as the ionisation of DNA oligomers is hugely dependent on the homogeneous co-crystallisation of matrix and analyte. Co-crystallisation was especially hindered by the presence of DMSO (0.1–1% v/v), which is often used as a vehicle for inhibitor screening, reducing the applicability of the assay (data not shown).

2.3.4.3 Differentiation of ^{5m}C oxidation products ^{5hm}C and ^{5f}C

In an effort to resolve the ^{5hm}C and ^{5f}C mass peaks, ^{5hm}C derivatisation was carried out based on the method reported by Szwagierczak *et al*.[203]: **^{5m}C 12 bp dsDNA** was enzymatically oxidised, and the resulting oxidation product mixture was treated with T4 phage β-glucosyltransferase (T4 β-GT) and uridine diphosphate (UDP)-glucose at 37 °C (Figure 2.7). A UDP-glucose concentration-dependent increase in glucosylated ^{5hm}C (^{5ghm}C) could be observed after 2 h by MALDI-TOF MS, but the experiment suffered from poor ionisation; mass spectra shown in Figure 2.7 were obtained only after sample desalting and reverse-phase C_{18} purification (ZipTip), whereas analysis of crude reaction mixture samples produced no ionisation for the DNA oligomers (data not shown).

In order to validate ^{5hm}C glucosylation as an accurate method for separating ^{5hm}C and ^{5f}C analytes, a substantial optimisation effort would be required to establish that complete glucosylation of ^{5hm}C can be achieved over a range of ^{5hm}C concentrations. Furthermore, enzyme-coupled product derivatisation would introduce an additional source of error into an otherwise direct measurement of TET-mediated ^{5m}C oxidation. Poor consistency of sample ionisation coupled with the need for sample purification considerably reduced the throughput

of the assay, making it impractical for higher-throughput experiments such as kinetic parameter determination or inhibitor characterisation.

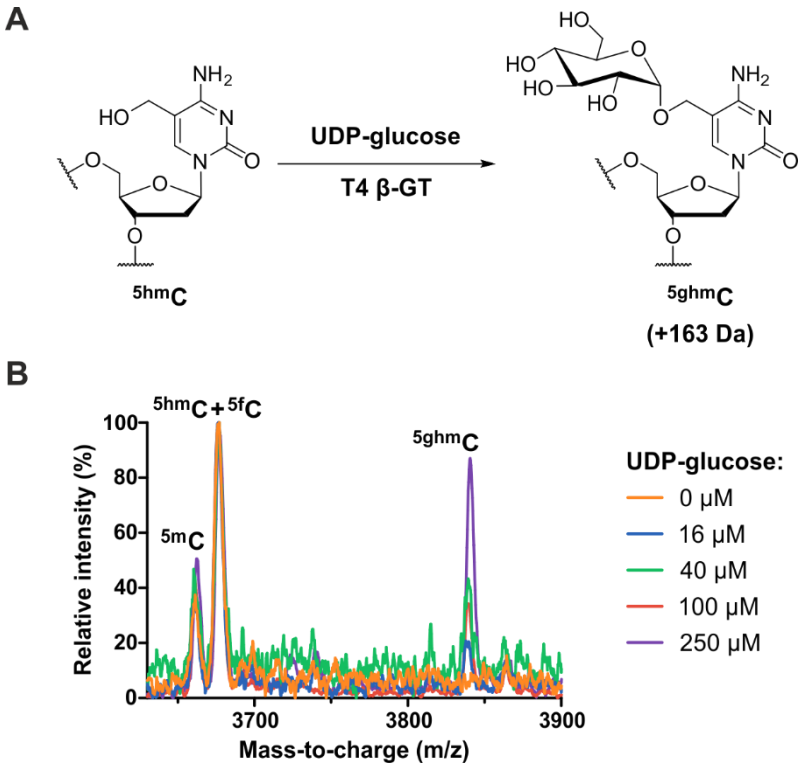
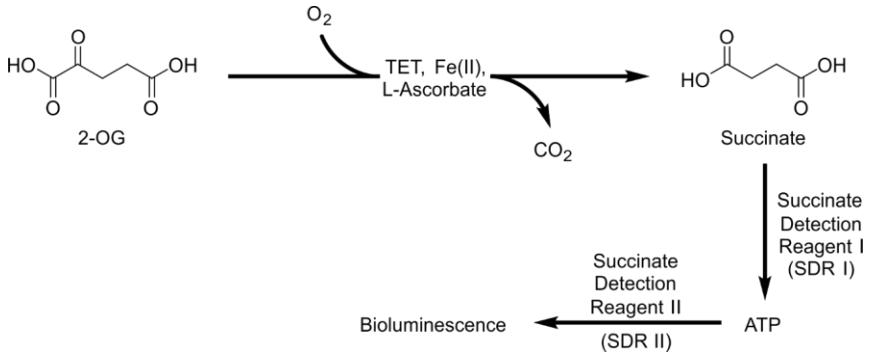


Figure 2.7 | T4 β-GT-mediated derivatisation of ^{5hm}C using UDP-glucose. A, scheme showing glucosylation of ^{5hm}C by T4 β-GT using UDP-glucose; **B**, ^{5hm}C derivatisation carried out on an enzymatically produced **^{5m}C 12 bp dsDNA** oxidation product mixture using T4 β-GT (0.1 U/μL) and UDP-glucose (0–250 μM) at 37 °C for 2 h.

2.3.5 NgTet1 catalytic activity detection using Succinate-Glo™

Next, to study NgTet1 activity as a function of 2-OG decarboxylation, the Succinate-Glo™ succinate detection assay was explored. Succinate-Glo™ is an enzyme-coupled end-point assay which detects 2-OG turnover to succinate in 2 steps (Scheme 2.2): (i) conversion of succinate to ATP by Succinate Detection Reagent I (SDR I), and (ii) conversion of ATP to bioluminescence through a luciferase/luciferin reaction using Succinate Detection Reagent II (SDR II). While oxidative 2-OG decarboxylation by 2-OG dioxygenases is typically coupled to substrate turnover, direct correlation of 2-OG consumption with substrate turnover is inaccurate due to the varying degrees of substrate-uncoupled enzymatic 2-OG decarboxylation observed within the superfamily (section 1.3.2).



Scheme 2.2 | A schematic representation of the Succinate-Glo™ bioluminescent succinate detection assay. Enzymatic turnover of 2-OG to succinate is detected *via* a 2-step transformation: succinate is converted to ATP by Succinate Detection Reagent I (SDR I), and ATP is then converted to bioluminescence by Succinate Detection Reagent II (SDR II).

2.3.5.1 Optimisation of 2-OG turnover detection

A succinate standard curve was generated to ascertain the compatibility of SDR I and SDR II reagents with the enzymatic reaction mixture components, and identify the linear response range of the assay (Figure 2.8 A); the assay retained linearity (R^2 = 0.99) within the previously reported[80] range (0–15 µM). Optimisation of enzyme concentration showed that NgTet1 produced sufficient 2-OG turnover at 1 µM (Figure 2.8 B). Subsequent optimisation of ^{5m}C concentration revealed that the enzyme operated at near-maximal velocity in the presence of 5 µM of **^{5m}C 12 bp dsDNA** (Figure 2.8 C, D). NgTet1 showed high affinity for ^{5m}C DNA (K_M^{app} = 0.23 ± 0.04 µM), in line with the values reported for hTET2 (K_M = 0.48 ± 0.19 µM)[96], and turned over 2-OG at an apparent maximal rate (V_{max}^{app}) of 0.26 ± 0.01 µM/min. In order to determine the level of NgTet1-mediated uncoupled 2-OG decarboxylation, a time course was carried out with NgTet1 (1 µM) in the absence of DNA and in the presence of non-methylated (**^{5}C**) and methylated (**^{5m}C**) **12 bp dsDNA** (Figure 2.8 E). While the initial succinate production rate was faster in the presence of methylated DNA (v_0 = 0.139, 0.080, and 0.081 µM/min for ^{5m}C, ^{5}C, and no DNA, respectively), linear NgTet1 activity was retained longer with no DNA or **^{5}C 12 bp dsDNA**, resulting in similar levels of succinate production after 60 min**.** The convergence of succinate levels in all three conditions is highly unusual, as rates

2-OG decarboxylation rates were over 20-fold slower in the absence of substrate for other 2-OG dioxygenases[78,79].

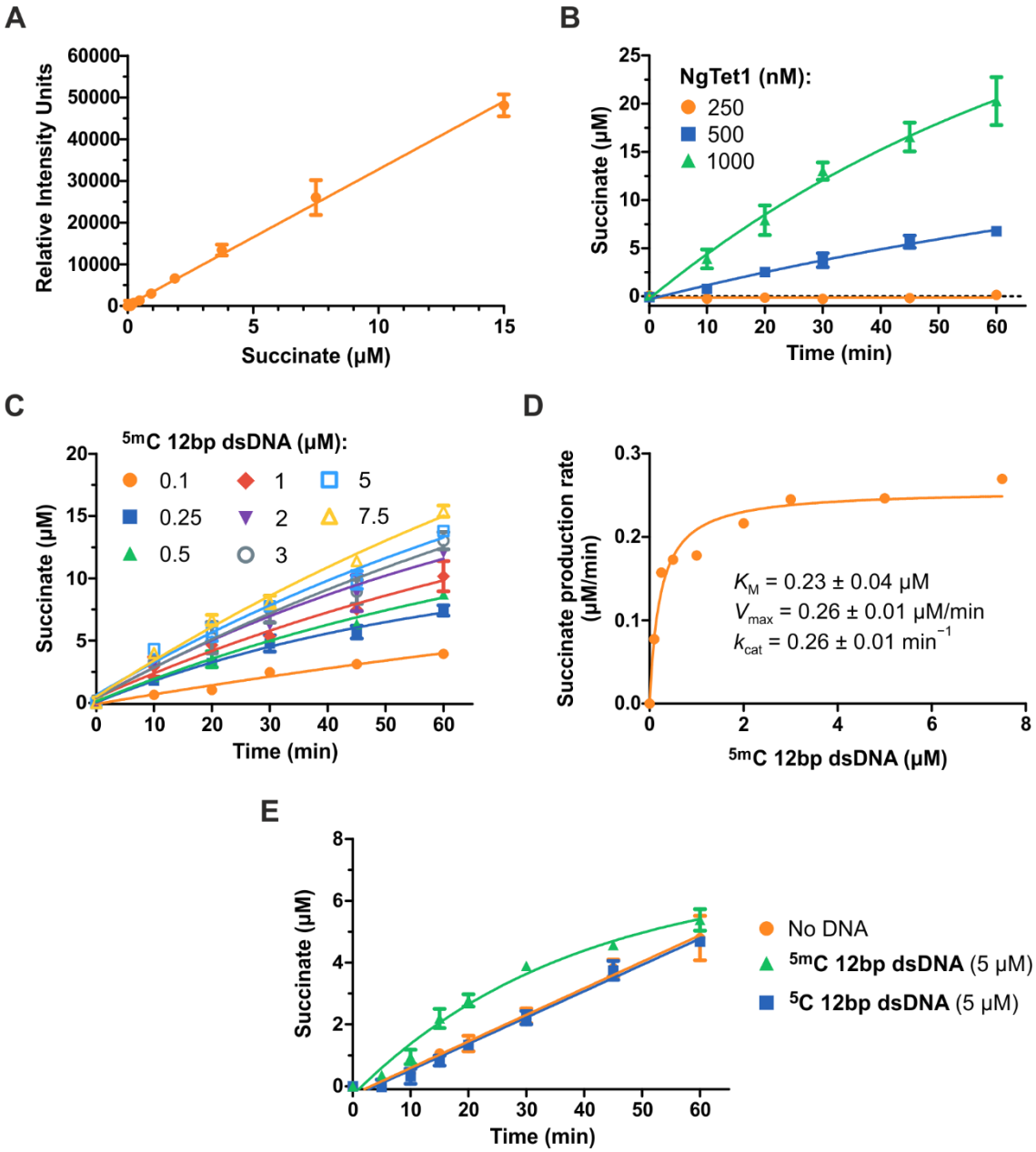


Figure 2.8 | NgTet1 catalytic activity detection using Succinate-Glo™. A, succinate standard curve; **B**, optimisation of NgTet1 concentration in the presence of 5 μM **^{5m}C 12 bp dsDNA**; **C**, optimisation of DNA substrate **^{5m}C 12 bp dsDNA** for optimal NgTet1 activity at 1 μM; **D**, Michaelis-Menten substrate saturation curve re-plotted using data from **C**; **E**, effect of cytosine methylation on 2-OG turnover. Conditions: 0.25–1 μM NgTet1, 0.1–7.5 μM **^{5}C / ^{5m}C 12 bp dsDNA**, 1 mM sodium L-ascorbate, 100 μM $(NH_4)_2Fe(SO_4)_2$, 100 μM disodium 2-OG. Data shown as the mean of 2–4 experimental replicates (n = 2–4; mean ± SD).

In order to elucidate the underlying reasons for the apparently comparable levels of succinate production regardless of substrate presence, comparison of kinetic parameters for ^{5}C/^{5m}C DNA oligomers and 2-OG would be required. However, the narrow linear range of detection

(up to 15 µM of succinate) and end-point read-out method did not allow facile collection of kinetic information, prompting investigation into ^{1}H NMR as an alternative succinate detection method.

2.3.6 NgTet1 catalytic activity detection using ^{1}H NMR spectroscopy

The previously discussed report by Hopkinson *et al*.[81] provided a foundation upon which conditions could be optimised for the detection of TET-mediated 2-OG decarboxylation using ^{1}H NMR spectroscopy. ^{1}H NMR is a powerful tool for catalytic activity detection, as it enables direct and quantitative observation of both 2-OG consumption and succinate production; the method is also the only approach to date that offers continuous detection of the analytes of interest.

Method optimisation began with a ^{1}H NMR analysis of all assay components to ensure the analyte signals (2-OG and succinate) did not overlap with signals from other components. A triplet signal corresponding to 2-OG and a singlet corresponding to succinate were free of overlapping signals, enabling *in situ* quantification of 2-OG turnover (Figure 2.9).

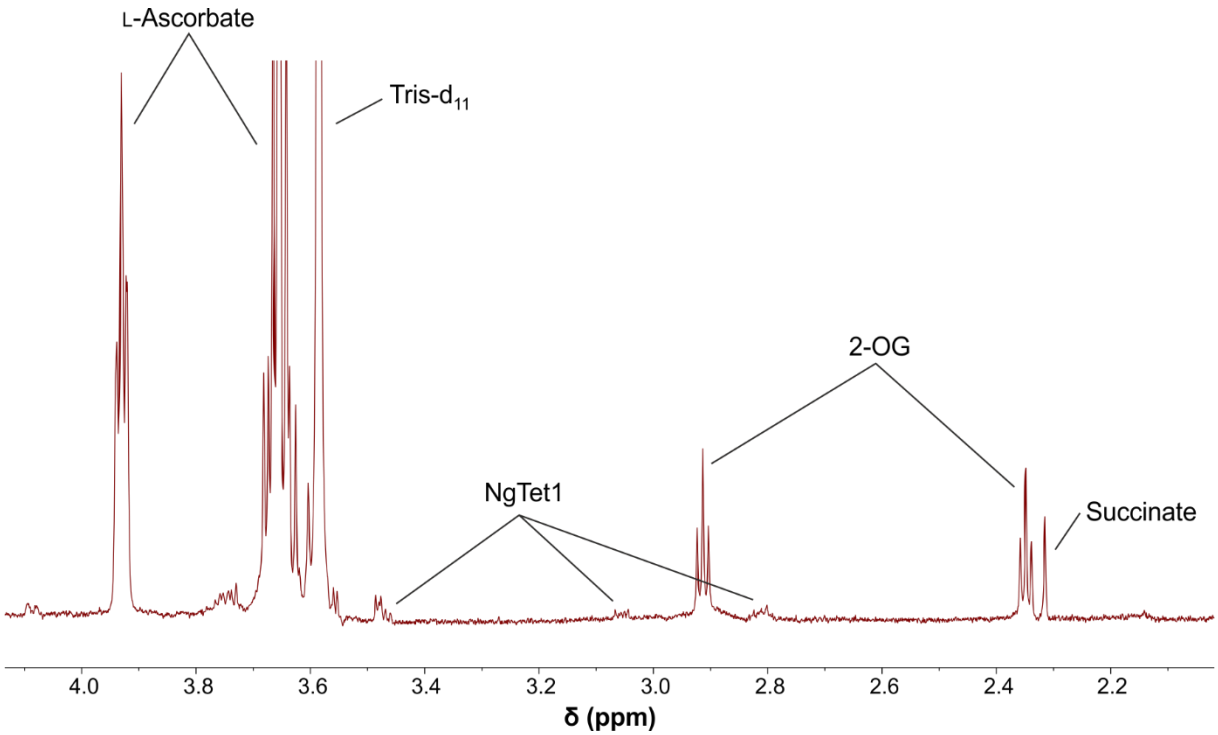


Figure 2.9 | NgTet1 catalytic turnover assay component ^{1}H NMR trace analysis. Components: 5 µM NgTet1, 2 mM sodium L-ascorbate, 100 µM $(NH_4)_2Fe(SO_4)_2$, 175 µM disodium 2-OG, 25 µM disodium succinate, 50 mM Tris-d_{11}, pH 7.5.

2-OG has been shown to undergo non-enzymatic oxidative decarboxylation in the presence of L-ascorbate[204]. In the presence of O_2, L-ascorbate can react to produce dehydroascorbate (DHA) and H_2O_2, which can then react with 2-OG to form succinate and CO_2. In order to evaluate the non-enzymatic succinate production, a number of assay component mixtures was prepared and analysed using ^{1}H NMR after 1 h and 8 h (Table 2.2). Mixture **2** (2-OG and L-ascorbate in the absence of iron) produced substantial levels of succinate in a time-dependent fashion, consistent with other reports[204]. Fe(II) suppressed L-ascorbate-mediated 2-OG decarboxylation to levels observed in the absence of L-ascorbate (mixtures **1** *vs* **3**), with the succinate levels remaining largely unchanged over the course of 8 h. In the absence of either cofactor, only background levels of succinate formation could be observed (mixture **4**). Having shown that non-enzymatic 2-OG decarboxylation is negligible under the assay conditions, mixtures **5**–**8** were prepared to determine the dependence of NgTet1 activity on the presence of cofactors. In the absence of Fe(II), succinate production was comparable to enzyme-free controls (mixture **6**), suggesting that the recombinant NgTet1 does not co-purify

as an active complex with Fe(II). Importantly, L-ascorbate was shown to be essential for NgTet1 activity under substrate-free conditions (mixture **5**), in line with reports on other 2-OG dioxygenases (section 1.3.2).

Table 2.2 | Evaluation of non-enzymatic oxidative 2-OG decarboxylation under ^{1}H NMR assay conditions. 2-OG was present in all mixtures. ^{1}H NMR spectra were collected 1 h and 8 h after mixture preparation. Spectra for mixtures containing NgTet1 were collected after 1 h only. Conditions: 5 µM NgTet1, 2 mM sodium L-ascorbate, 100 µM $(NH_4)_2Fe(SO_4)_2$, 500 µM disodium 2-OG, 50 mM Tris-d_{11}, pH 7.5. N.d., not determined. Data from single replicates shown.

Mixture	L-Ascorbate	Fe(II)	NgTet1	Succinate, µM (% turnover)	
				After 1 h	After 8 h
1*	+	+	-	2.0 (0.4%)	3.3 (0.7%)
2*	+	-	-	6.1 (1.2%)	14.6 (2.9%)
3*	-	+	-	2.6 (0.5%)	3.0 (0.6%)
4*	-	-	-	0.3 (0.1%)	0.4 (0.1%)
5†	-	+	+	7.3 (1.5%)	n.d.
6†	+	-	+	3.7 (0.8%)	n.d.
7†	-	-	+	0 (0%)	n.d.
8†	+	+	+	35.4 (8.8%)‡	n.d.

**1,024 scans were collected.*
†32 scans were collected.
‡Data was collected in the presence of 400 µM 2-OG after 25 min.

Having shown that L-ascorbate and Fe(II) effect minimal non-enzymatic 2-OG decarboxylation and are essential for NgTet1 activity, buffer pH and NaCl concentration were adjusted for optimal NgTet1 activity. The enzyme favoured neutral conditions (pH 7.0) with a rapid decline in activity in more basic conditions (Figure 2.10 A). While neutral pH provided optimal NgTet1 activity, the slightly more basic pH 7.5 was selected for future experiments in favour of a longer linear enzyme activity window. The addition of salt seemingly had little effect on the reaction rate, with the reaction plateauing at a slightly higher succinate conversion level in the absence of salt (Figure 2.10 B). Salt-free buffer was used for further experiments, as sample conductivity has been shown to inversely correlate with the sensitivity of cryogenic probes used for spectra recording[205].

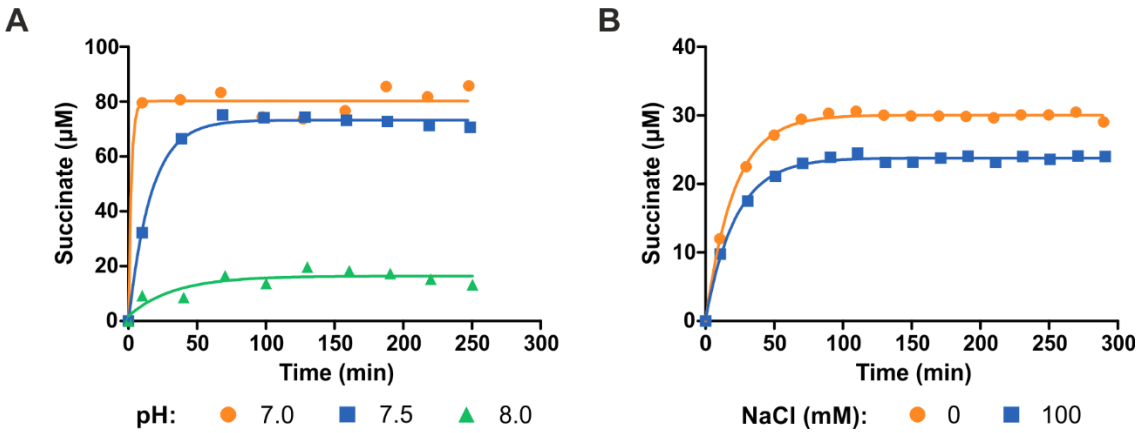


Figure 2.10 | Buffer optimisation for NgTet1 catalytic activity. A, optimisation of buffer pH (conditions: 10 μM NgTet1, 2 mM L-ascorbate, 200 μM $(NH_4)_2Fe(SO_4)_2$, 200 μM disodium 2-OG); **B**, optimisation of NaCl concentration (conditions: 5 μM NgTet1, 2 mM sodium L-ascorbate, 100 μM $(NH_4)_2Fe(SO_4)_2$, 100 μM disodium 2-OG, pH 7.5). Data from single replicates shown.

Following buffer component optimisation, the concentration of NgTet1 was optimised (Figure 2.11). While the enzyme linear activity window was comparable for both 2 μM and 5 μM of NgTet1, 5 μM was selected to allow for turnover margin in future experiments where activity may be partially suppressed, such as inhibitor screening.

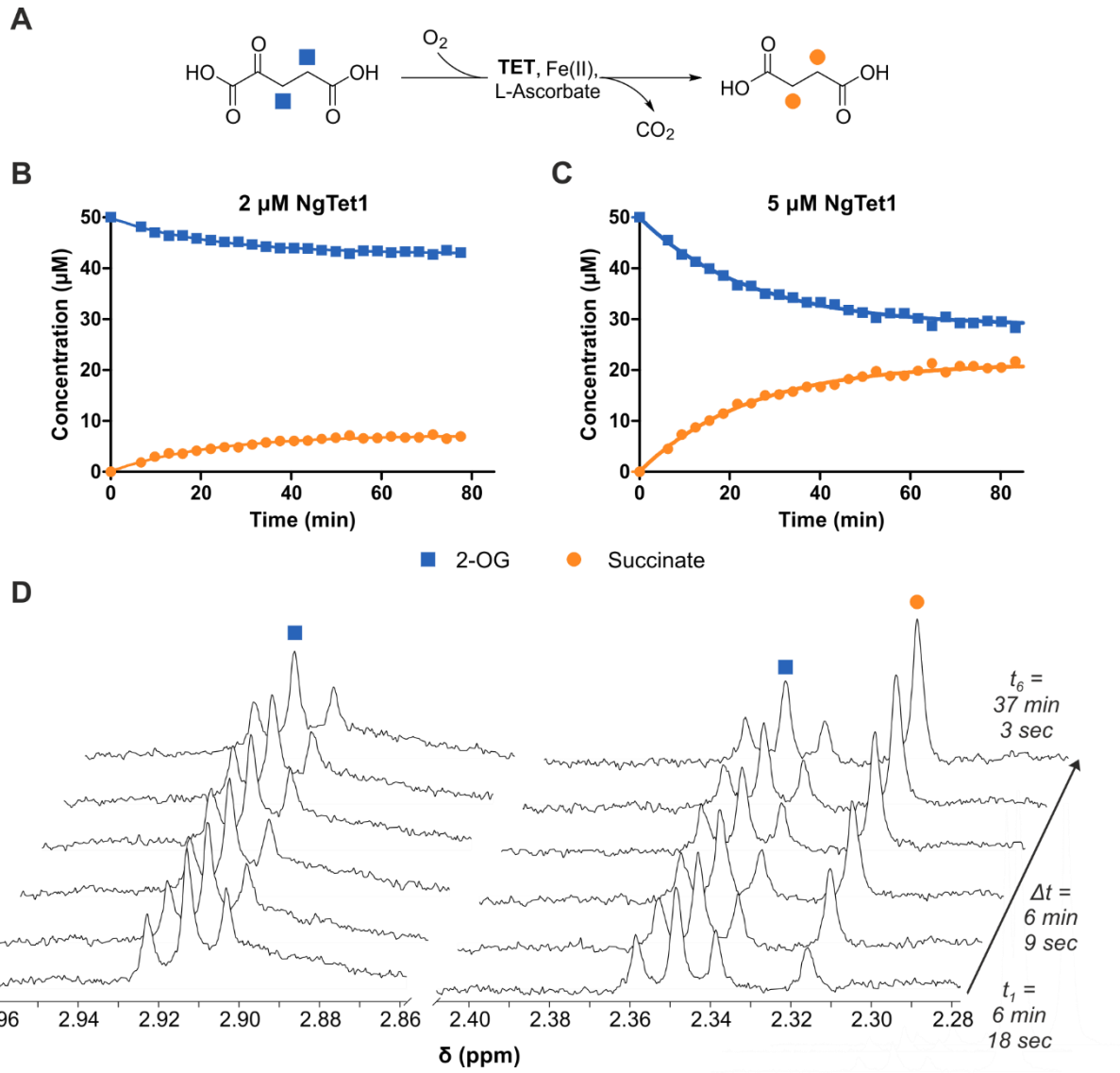


Figure 2.11 | Optimisation of NgTet1 concentration. A, enzymatic 2-OG decarboxylation reaction scheme; **B**, 2-OG turnover time course with 2 μM NgTet1; **C**, 2-OG turnover time course with 5 μM NgTet1; **D**, selected ^{1}H NMR spectra from time course **C**. Conditions: 2/5 μM NgTet1, 2 mM sodium L-ascorbate, 100 μM $(NH_4)_2Fe(SO_4)_2$, 50 μM disodium 2-OG. Data from single replicates shown.

2.3.7 NgTet1 kinetic parameter determination by ^{1}H NMR

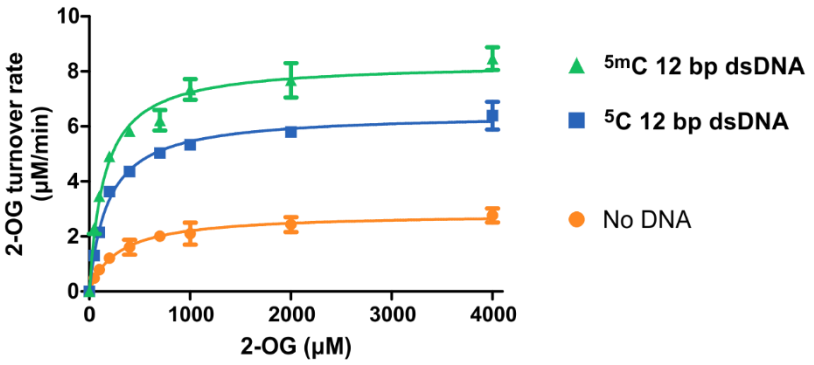
With the optimised assay conditions in hand, kinetics of 2-OG turnover were investigated. Michaelis-Menten kinetics data were collected for 2-OG in the presence and absence of **^{5m}C 12 bp dsDNA** substrate (Figure 2.12). The concentration of substrate dsDNA was limited to 20 μM (40 μM total concentration of ^{5m}C), as analysis of samples containing higher levels of dsDNA was complicated by the overlap of DNA, 2-OG, and succinate signals (data not shown). Because the K_M value of **^{5m}C 12 bp dsDNA** had not been unambiguously determined, apparent values (K_M^{app} and V_{max}^{app}) are reported for kinetic parameters of 2-OG. Addition of methylated

DNA resulted in a nearly 2-fold increase in 2-OG affinity (K_M^{app} = 284 and 150 μM without and with **^{5m}C** dsDNA, respectively) and a nearly 3-fold increase in the maximum reaction rate (V_{max}^{app} = 2.83 and 8.31 μM/min without and with **^{5m}C** dsDNA, respectively). However, K_M^{app} for 2-OG in the presence of substrate DNA was around 3-fold lower than values obtained for mTet1/2 using a 2-OG radiolabelling assay[171] (K_M = 55, 60, and 150 μM for mTet1, mTet2, and NgTet1, respectively). The apparent 2-OG decarboxylase catalytic efficiency of NgTet1, k_{cat}^{app}/K_M^{app}, was increased 5.5-fold by the addition of substrate (0.00201 to 0.0111 $min^{-1}\mu M^{-1}$), indicating that 2-OG decarboxylation rate is regulated by binding to DNA.

Next, the rate-enhancing effect of substrate DNA was compared to that of non-substrate non-methylated DNA (**^{5}C 12 bp dsDNA**)‡, as $hTET2_{CD\Delta LCI}$ was reported to bind unmodified DNA with similar affinity to DNA containing modified cytosine (K_d = 0.54 and 0.99 μM for ^{5}C and ^{5m}C 18 bp dsDNA, respectively)[96]. Addition of non-substrate DNA improved NgTet1 affinity for 2-OG (K_M^{app} = 284 and 187 μM without and with **^{5}C** dsDNA, respectively) and increased the maximum reaction rate (V_{max}^{app} = 2.83 and 6.47 μM/min without and with **^{5}C** dsDNA, respectively), although to a lesser degree than substrate DNA. **^{5}C** DNA presence improved the catalytic efficiency 3.5-fold (2.01×10^{-3} to 6.95×10^{-3} $min^{-1}\mu M^{-1}$), although no productive 2-OG turnover should be possible, indicating that non-substrate oligonucleotides stabilise the enzyme active site and enable higher efficiency of uncoupled 2-OG decarboxylation.

‡ ^{5}C refers to an unmodified deoxycytidine monophosphate deoxynucleotide in a DNA oligomer.

A



B

DNA oligomer	K_m^{app} (2-OG, μM)	V_{max}^{app} (μM/min)	K_{cat}^{app} (min^{-1})	Catalytic efficiency (k_{cat}^{app} / K_M^{app}, min^{-1} μM^{-1})
● No DNA	284 ± 42	2.83 ± 0.12	0.57 ± 0.02	2.01×10^{-3}
▲ **^{5m}C 12 bp dsDNA**	150 ± 16	8.31 ± 0.21	1.66 ± 0.04	11.1×10^{-3}
■ **^{5}C 12 bp dsDNA**	187 ± 14	6.47 ± 0.12	1.30 ± 0.02	6.95×10^{-3}

Figure 2.12 | NgTet1 kinetic parameter determination for 2-OG. A, Michaelis-Menten 2-OG saturation curves in the presence of no DNA, non-methylated DNA (**^{5}C**), and methylated DNA (**^{5m}C**); **B**, tabulated kinetic parameters obtained from curves in **A**. Conditions: 5 μM NgTet1, 0/20 μM **12 bp dsDNA**, 2 mM sodium L-ascorbate, 100 μM $(NH4)_2Fe(SO_4)_2$, 50–4,000 μM disodium 2-OG. Data are shown as the mean of two independent replicates (n = 2; mean ± SD).

Further support for this is provided by studies on PHDs, where addition of competitive substrate inhibitors was found to increase the rate of enzymatic uncoupled 2-OG decarboxylation[77,206]. Correspondingly, the decarboxylase efficiency difference between methylated and non-methylated dsDNA was only 1.6-fold (11.1×10^{-3} and 6.95×10^{-3} $min^{-1}\mu M^{-1}$, respectively). While similar results were observed using Succinate-Glo™ (Figure 2.8 E), there was a possibility that the ^{5m}C present in methylated DNA was not efficiently oxidised under the experimental conditions, as different buffering agent and pH were used for ^{1}H NMR compared to AlphaScreen or MALDI-TOF MS assays. To determine if ^{5m}C oxidation was taking place, a time course was carried out using ^{1}H NMR and MALDI-TOF MS methods simultaneously (Figure 2.13). MALDI-TOF MS analysis confirmed that **^{5}C 12 bp dsDNA** is not modified by NgTet1 (Figure 2.13 B, D), while **^{5m}C 12 bp dsDNA** is oxidised by NgTet1 in a time-dependent manner but does not undergo spontaneous oxidation in the absence of enzyme (Figure 2.13 B, C, E). Simultaneous analysis of catalytic activity showed that, after 10 min, approximately 65 μM of succinate was produced, but only 26 μM of ^{5m}C substrate had been

consumed. As discussed previously in section 2.3.4, MALDI-TOF MS does not enable differentiation between ^{5hm}C and ^{5f}C; as such, the ^{5hm}C product may have undergone secondary oxidation to ^{5f}C, amounting to an overall turnover of up to 52 µM of DNA substrate. However, a substantial degree of ^{5f}C formation is unlikely due to the presence of significant levels of ^{5m}C (14 µM, 35% of starting concentration), suggesting that up to 60% of succinate production was not coupled to oxidation of ^{5m}C. The observation indicates that, for NgTet1, uncoupled 2-OG decarboxylation is present to a high degree regardless of the availability of methylated cytosine substrate.

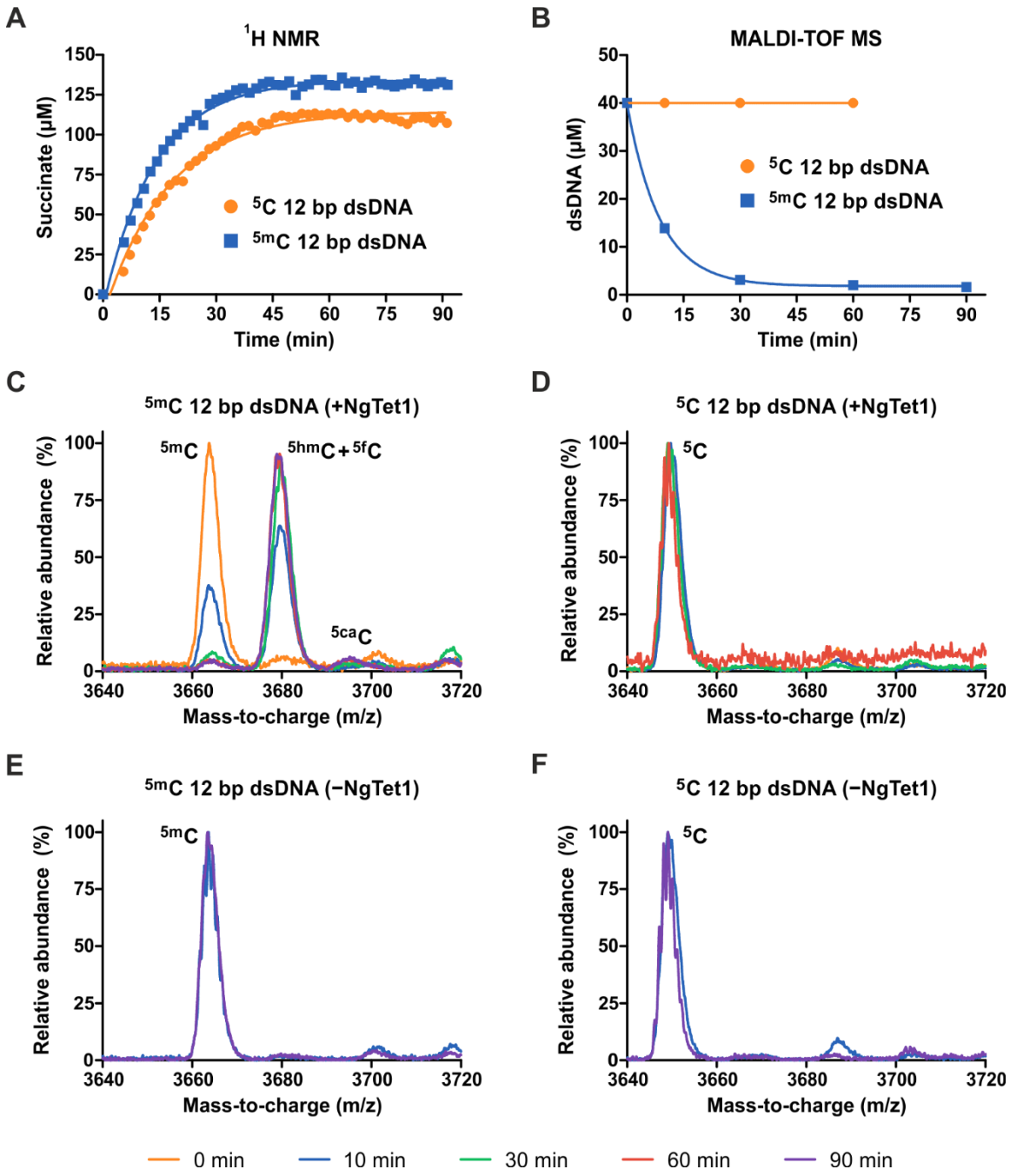


Figure 2.13 | Tandem ^{1}H NMR/MALDI-TOF MS assay for the simultaneous analysis of 2-OG and ^{5m}C turnover. **A**, ^{1}H NMR time course of 2-OG turnover in the presence of **^{5}C/^{5m}C 12 bp dsDNA**; **B**, MALDI-TOF MS time course of NgTet1-mediated DNA oxidation; **C**, **D**, MALDI-TOF mass spectra of DNA time courses in **B**; **E**, **F**, MALDI-TOF mass spectra of DNA at 10 min and 90 min time points in the absence of NgTet1. Conditions: 5 μM NgTet1, 20 μM dsDNA, 2 mM sodium L-ascorbate, 100 μM $(NH_4)_2Fe(SO_4)_2$, 200 μM disodium 2-OG. Data from single replicates shown.

2.3.8 Characterisation of TET inhibitors against NgTet1

Bimodal catalytic activity detection, facilitated by the development of DNA substrate and 2-OG turnover detection assays, enabled the dual characterisation of TET inhibitors. Three 2-OG mimetics (NOG[207], IOX1[182], and 2,4-PDCA[208]) were selected for the screening along with a recently reported hTET1-selective thioether macrocyclic peptide TiP1[191] (section 1.6, Figure

2.14 A). The compounds were first characterised as inhibitors using AlphaScreen (Figure 2.14 B, C), stopping the reactions within the linear range of ^{5hm}C product formation. 2,4-PDCA was found to be the most potent inhibitor under the experimental conditions with IC_{50} = 3.2 µM. IOX1 and TiP1 inhibited NgTet1 with identical double-digit micromolar potency (IC_{50} = 26 µM), while NOG showed 8-fold lower potency in comparison (IC_{50} = 200 µM).

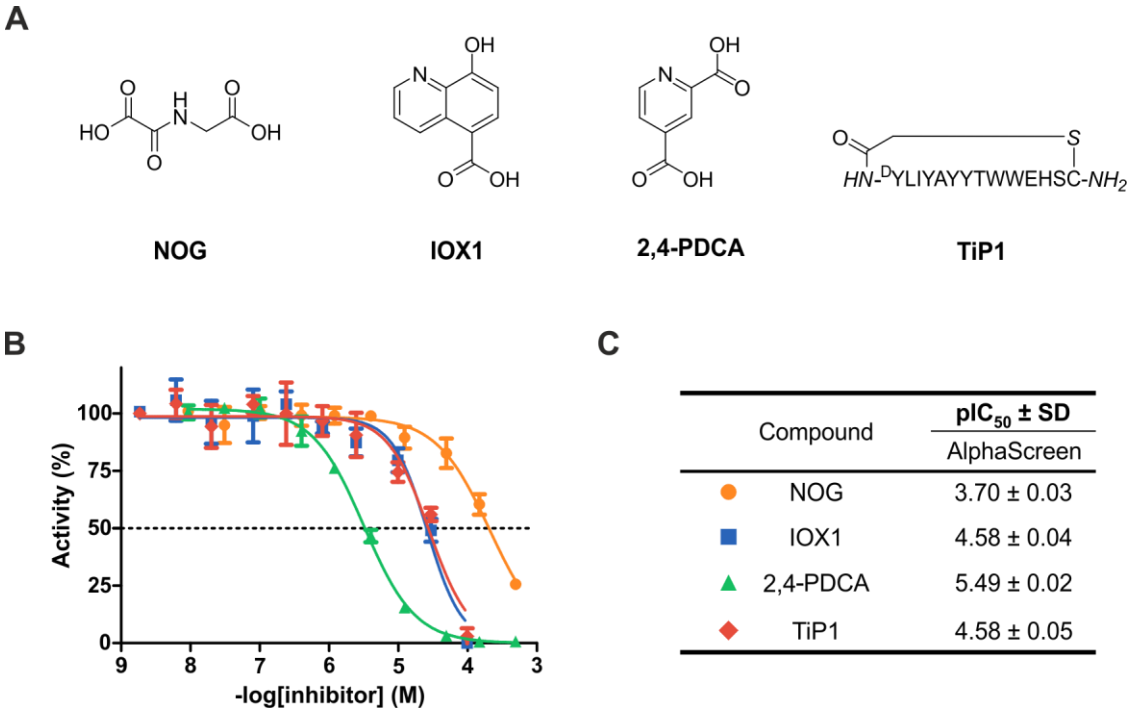
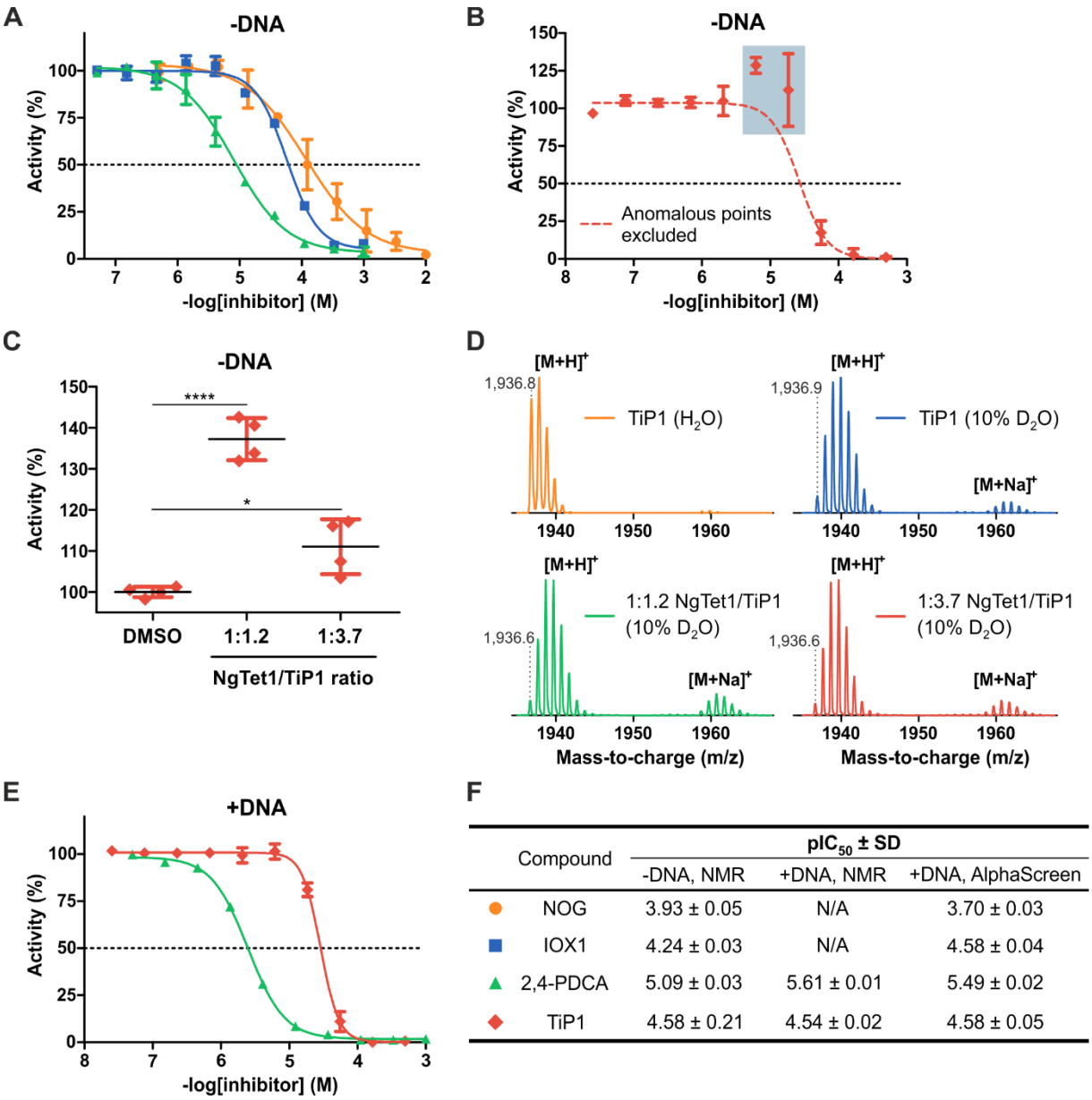


Figure 2.14 | Characterisation of NgTet1 inhibitors using AlphaScreen. A, structures of selected inhibitors (letters representing individual atoms in TiP1 are italicised to distinguish from single-letter amino acid codes); **B**, inhibitor dose-response curves against NgTet1; **C**, tabulated pIC_{50} values from dose-response curves. Conditions: 400 nM NgTet1, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, 10 µM disodium 2-OG. Data shown as the mean of 3 experimental replicates (n = 3; mean ± SD). Tabulated data shown as the mean of 3 independent replicates (n = 3; mean ± SD).

Next, compound inhibitory activity on 2-OG decarboxylation was tested using ^{1}H NMR (Figure 2.15). The concentration of 2-OG was fixed at 50 µM to produce data more closely comparable to results from AlphaScreen, which uses 10 µM 2-OG. The reactions were stopped within the linear range of succinate formation using a mixture of NOG and Zn(II); NOG was found to be a poor inhibitor of uncoupled 2-OG turnover (approx. 10 mM required for complete inhibition), so the quenching mixture was supplemented with the ^{1}H NMR-silent Zn(II). Uncoupled 2-OG decarboxylation inhibition potencies were largely comparable to values obtained using AlphaScreen (Figure 2.15 A, B, E), with less than a 3-fold difference in potencies observed

across the board (max. ΔpIC_{50} = ± 0.4). Unexpectedly, an increase in 2-OG decarboxylation activity was observed when approaching 1:1 stoichiometry between TiP1 and NgTet1. A focused repeat of the 2 concentrations of interest (1:1.2 and 1:3.7 ratios of NgTet1/TiP1) confirmed that TiP1 increased the rate of uncoupled 2-OG turnover in a statistically significant manner (Figure 2.15 C). Activity in the presence of TiP1 (6.17 µM, 1:1.2 ratio of NgTet1/TiP1) was significantly higher than the normalised control activity (137 ± 5% *vs* 100 ± 1%) according to an unpaired *t* test; $t(6) = 14.04$, $p \leq 0.0001$. The level of activity enhancement decreased with increasing TiP1 concentration, allowing a classical sigmoidal IC_{50} curve fit when the two anomalous data points were excluded from the fit. Subsequent TiP1 analysis was carried out using MALDI-TOF MS to ensure that the observed catalytic activity enhancement was not due to oxidation of TiP1 in a substrate-like fashion. No change in the peptide mass could be observed after reaction with NgTet1 (1:1.2 or 1:3.7 ratio of NgTet1/TiP1), indicating that the peptide is not recognised as a substrate (Figure 2.15 D). The +2 Da shift of the major TiP1 mass peak was attributed to partial hydrogen-deuterium exchange of the peptide amide backbone protons enabled by the presence of D_2O in the sample (10% v/v)[209].

Two compounds were selected for further 2-OG turnover inhibitor characterisation in the presence of methylated DNA substrate. 2,4-PDCA was selected as the most potent 2-OG mimetic, while TiP1 was chosen due to its agonistic effect at near-stoichiometric ratios with NgTet1 (Figure 2.15 E). 2,4-PDCA showed higher 2-OG turnover inhibition in the presence of DNA (IC_{50} = 2.5 and 8.1 µM with and without **^{5m}C** dsDNA, respectively), which correlated very closely to the ^{5m}C oxidation inhibition potency obtained using AlphaScreen (IC_{50} = 3.2 µM). TiP1 inhibited NgTet1 with nearly identical potency across ^1H NMR and AlphaScreen assays (IC_{50} = 29 and 26 µM, respectively) but did not enhance 2-OG turnover activity at near-stoichiometric ratios in the presence of substrate DNA.



F

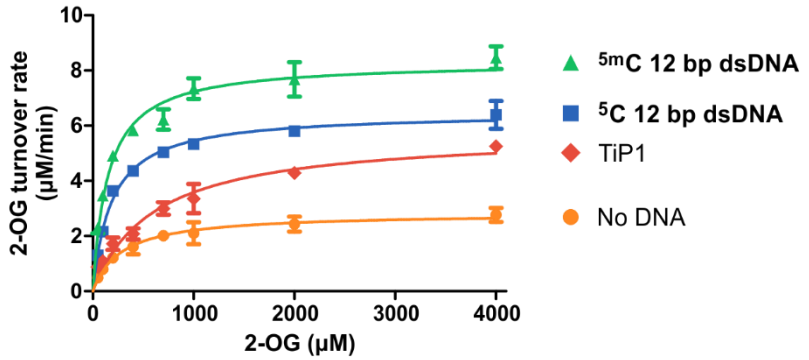
Compound	pIC_{50} ± SD		
	-DNA, NMR	+DNA, NMR	+DNA, AlphaScreen
NOG	3.93 ± 0.05	N/A	3.70 ± 0.03
IOX1	4.24 ± 0.03	N/A	4.58 ± 0.04
2,4-PDCA	5.09 ± 0.03	5.61 ± 0.01	5.49 ± 0.02
TiP1	4.58 ± 0.21	4.54 ± 0.02	4.58 ± 0.05

Figure 2.15 | Characterisation of NgTet1 inhibitors using ^{1}H NMR. A, **B**, dose-response curves for 2-OG turnover inhibition without DNA; **C**, TiP1-mediated NgTet1 activation without DNA; **D**, MALDI-TOF MS spectra of TiP1 in H_2O, 10% D_2O, and after reaction with NgTet1; **E**, dose-response curves for 2-OG turnover inhibition with **^{5m}C** dsDNA; **F**, tabulated inhibition data. Conditions: 5 μM NgTet1, 0/20 μM **^{5m}C 12 bp dsDNA**, 2 mM sodium L-ascorbate, 100 μM $(NH_4)_2Fe(SO_4)_2$, 50 μM disodium 2-OG. Data shown as the mean of 2 independent replicates (n = 2; mean ± SD). Statistical significance of NgTet1 activation was calculated using an unpaired *t* test (4 experimental replicates); ****, p ≤ 0.0001; *, p = 0.0173.

To determine if TiP1 enhanced the catalytic efficiency of NgTet1-mediated 2-OG decarboxylation, a 2-OG saturation curve was obtained in the presence of 1.2 equivalents of TiP1 (6.17 μM). Michaelis-Menten analysis (Figure 2.16) revealed that while 2-OG V_{max}^{app} increased from 2.83 to 5.76 μM/min, 2-OG affinity was concurrently decreased 2-fold in the presence of TiP1 (K_M^{app} = 284 μM to 606 μM). Consequently, the catalytic efficiency remained

virtually unchanged between no DNA and TiP1 conditions (0.00201 and 0.00190 $min^{-1} \cdot \mu M^{-1}$, respectively).

A



B

DNA oligomer	K_m^{app} (2-OG, μM)	V_{max}^{app} (μM/min)	K_{cat}^{app} (min^{-1})	Catalytic efficiency (k_{cat}^{app} / K_M^{app}, min^{-1} μM^{-1})
● No DNA	284 ± 42	2.83 ± 0.12	0.57 ± 0.02	2.01×10^{-3}
▲ **^{5m}C 12 bp dsDNA**	150 ± 16	8.31 ± 0.21	1.66 ± 0.04	11.1×10^{-3}
■ **^{5}C 12 bp dsDNA**	187 ± 14	6.47 ± 0.12	1.30 ± 0.02	6.95×10^{-3}
◆ TiP1	606 ± 94	5.76 ± 0.31	1.15 ± 0.06	1.90×10^{-3}

Figure 2.16 | Extended NgTet1 kinetic parameter determination for 2-OG. A, Michaelis-Menten 2-OG saturation curves for no DNA, non-methylated DNA (20 μM), methylated DNA (20 μM), and near-stoichiometric TiP1 (6.17 μM, 1:1.2 NgTet1/TiP1 ratio) conditions; **B**, tabulated kinetic parameters obtained from curves in **A**. Conditions: 5 μM NgTet1, 2 mM sodium L-ascorbate, 100 μM $(NH_4)_2Fe(SO_4)_2$, 50–4,000 μM disodium 2-OG. Data are shown as the mean of two independent replicates (n = 2; mean ± SD).

2.4 Discussion

Recombinant NgTet1 was produced in good yield and purity (4.2 mg/L of cell culture, >90% purity by SDS-PAGE), although sample homogeneity was partially perturbed by the presence of α-*N*–gluconoylation, a known post-translational modification in *E. coli* BL21(DE3) strains[200].

Two catalytic activity detection assays were developed to characterise the DNA hydroxylase activity of TET dioxygenases. MALDI-TOF MS assay enabled the validation of NgTet1 activity through direct detection of ^{5m}C oxidation in dsDNA, but higher-order oxidation states ^{5hm}C and ^{5f}C could not be resolved due to the broadness and poor isotopic resolution of the peaks observed by MALDI-TOF MS. ^{5hm}C derivatisation *via* T4 β-GT-mediated glucosylation produced encouraging results, but the assay was ultimately found unsuitable for the desired applications due to poor repeatability, laborious sample purification, and incompatibility with the commonly used inhibitor vehicle DMSO. AlphaScreen provided a repeatable high-throughput method for detection of the first ^{5m}C oxidation product, ^{5hm}C, and was used to characterise NgTet1 inhibition by known 2-OG dioxygenase inhibitors (NOG, IOX1, 2,4-PDCA) and cyclic peptide TiP1. All four compounds inhibited NgTet1 in the micromolar range, with 2,4-PDCA being the most potent with IC_{50} = 3.2 μM, though generally lower potencies were observed than those reported for other Fe(II)/2-OG-dependent dioxygenases[182,191,207,208]. Data obtained using AlphaScreen was initially considered unreliable due to the assay conditions requiring high excess of enzyme in relation to substrate (20:1 ratio of NgTet1/^{5m}C), but the results were subsequently supported by inhibition data collected using ^{1}H NMR spectroscopy.

With a functional ^{5m}C oxidation detection assay in hand, focus shifted towards development of assays detecting 2-OG decarboxylation. Succinate-Glo™ succinate detection assay was successfully used to obtain initial data on NgTet1-mediated uncoupled 2-OG decarboxylation. Comparison of NgTet1-mediated succinate production without DNA and in the presence of **^{5}C**

or **^{5m}C 12 bp dsDNA** revealed that, although the initial reaction rate was nearly 2-fold higher for ^{5m}C than for no DNA/^{5}C, the window for linear enzyme activity was narrower, resulting in approximately equimolar succinate production after 60 min. These results are in stark contrast with data reported for other 2-OG dioxygenases: a 2-fold difference in total succinate production between reactions with and without substrate was reported for hTET1[80], while dioxygenases such as procollagen-lysine 5-dioxygenase and PHD were reported to catalyse uncoupled 2-OG decarboxylation at rates less than 7% of those observed in the presence of saturating substrate concentrations (section 1.3.2). Due to the narrow linear range of detection (up to 15 µM of succinate), Succinate-Glo™ was found unsuitable for subsequent work on kinetic parameter determination for 2-OG. ^{1}H NMR spectroscopy proved to be a robust and reproducible method for succinate detection, as well as being the only assay currently capable of continuous TET activity monitoring. NgTet1 kinetic parameters for 2-OG were determined using ^{1}H NMR, revealing that 2-OG decarboxylation efficiency increased in the following order: no DNA << **^{5}C 12 bp dsDNA** < **^{5m}C 12 bp dsDNA**. The presence of dsDNA increased enzyme affinity for 2-OG by nearly 2-fold, while cytosine methylation contributed to a further 20% increase in affinity (K_M = 284, 187, and 150 µM for no DNA, **^{5}C**, and **^{5m}C 12 bp dsDNA**, respectively). Although the majority of 2-OG-dependent dioxygenases show substantially lower uncoupled 2-OG decarboxylation levels, similar results have been reported for *E. coli* DNA repair enzyme AlkB, where the rates in the absence of substrate and the presence of a methylated DNA substrate were relatively comparable (1 ± 0.2 and 2.76 ± 0.2 pmol/sec, respectively)[210]. Interestingly, while addition of the non-methylated **^{5}C 12 bp dsDNA** increased the 2-OG decarboxylation rate of NgTet1, an analogous effect was not observed in the case of AlkB[210].

Although the differences in experimental conditions prohibit direct K_M value comparisons, a broad range of K_M values (0.3–110 µM) has been reported for 2-OG dioxygenases, including

TETs, lysine demethylases (KDMs), hypoxia-inducible factor prolyl hydroxylases (HIF PHDs), and factor inhibiting HIF (FIH, Table 2.3). It is worth noting that, while 2-OG K_M values in the presence of substrate-saturating conditions are available for a wide range of dioxygenases, 2-OG affinities in the absence of substrate are rarely evaluated.

Table 2.3 | A comparison of 2-OG kinetics results for selected 2-OG dioxygenases. Substrate concentration used was more than 2-fold above the K_M unless otherwise mentioned. Variation of substrate is reported in brackets following K_M^{app} values. N.d., not determined; n.i., no inhibition.

Family	Enzyme	K_M^{app} (2-OG, μM)	K_i^{app} (2-OG, mM)
TET	NgTet1*	150 ± 16†	n.i. at 4
	NgTet1*	284 ± 42‡	n.i. at 4
	mTet1[171]	55 ± 20	n.d.
	mTet2[171]	60 ± 15	n.d.
	TET2[174]	15.7 ± 2.8	n.d.
	TET3[174]	24.2 ± 2.2	n.d.
	AlkB[210]	approx. 100	approx. 0.6
KDM	KDM2A[211]	12.5 ± 1.4	n.d.
	KDM3A[212]	22 ± 10	n.d.
	KDM3A[211]	5.3 ± 2.2	n.d.
	KDM4A[213]	15.5 ± 5.5	11.5 ± 1.5
	KDM4C[213]	17 ± 5	3.85 ± 0.45
	KDM4E[213]	29 ± 8	11.5 ± 0.5
	KDM5B[214]	2.6 ± 0.6	n.d.
	KDM5C[211]	41.7 ± 6.3	n.d.
	KDM6B[211]	5.4 ± 0.4	n.d.
	KDM7B[211]	16.6 ± 1.8	n.d.
	FIH[215]	110 ± 20 (HIF-1α $CAD_{35\text{-}mer}$)	0.9 ± 0.4
		19 ± 6 (HIF-2α $CAD_{35\text{-}mer}$)	approx. 1.0
HIF	PHD2[215]	13 ± 2 (HIF-1α CODD)	n.d.
		30 ± 9 (HIF-1α NODD)	n.d.
		9 ± 2 (HIF-2α CODD)	n.d.
		17 ± 5 (HIF-2α NODD)	n.d.
	PHD2[216]	0.343 ± 0.028	n.d.
	PHD[77]	28**	n.d.

**This work.*
†Substrate K_M not determined.
‡No substrate (uncoupled turnover only).
***Both with and without substrate.*

Although NgTet1 affinity for 2-OG was the lowest amongst the characterised 2-OG dioxygenases, the K_M value reported herein may be affected by the choice of oligonucleotide substrate. FIH affinity for 2-OG changed considerably depending on the substrate (2-OG K_M = 110 and 19 μM with HIF-1α $CAD_{35\text{-}mer}$ and HIF-2α $CAD_{35\text{-}mer}$ substrates, respectively), and

^{5m}C 12 bp dsDNA has not been validated as the optimal substrate for NgTet1. Interestingly, affinity for 2-OG within the TET family improves across species, with NgTet1 having the poorest 2-OG affinity (150 μM) and hTETs having the highest 2-OG affinity (approx. 20 μM), suggesting an evolutionary advantage for high 2-OG affinity, such as maintaining appropriate TET activity in environments with reduced 2-OG availability.

A tandem ^{1}H NMR/MALDI-TOF MS assay confirmed that ^{5m}C was hydroxylated only in the presence of NgTet1, while ^{5}C remained inert under the reaction conditions. Crucially, a high degree of uncoupled 2-OG turnover (up to 60% of total succinate production) was observed within the window of linear ^{5hm}C product formation, highlighting the poor 2-OG decarboxylation coupling efficiency of NgTet1. However, this observation may be the consequence of using suboptimal oligonucleotide substrate sequences rather than an inherent property of the protein itself, as the uncoupled 2-OG turnover proportion was shown to fluctuate with different substrates for KDM4E and AlkB (section 1.3.2). Variation of the flanking sequence around the ^{5m}CpG site could potentially identify a DNA substrate in the presence of which the proportion of uncoupled turnover ("optimal" substrate) is minimised, highlighting the relation between coupled 2-OG turnover efficiency and substrate preference. Alternatively, a comparison of ^{5m}C and 2-OG turnover by NgTet1 and a TET analogue from a higher species (murine or human) under identical conditions could provide evidence of inherent differences in 2-OG turnover coupling efficiency between enzyme homologues. It is unclear whether uncoupled 2-OG decarboxylation in the presence of substrate is prevalent in other TET dioxygenases such as hTETs or mTets, as simultaneous data on ^{5m}C and 2-OG turnover is not available.

NOG, IOX1, 2,4-PDCA, and TiP1 were screened for inhibitory activity against NgTet1 using AlphaScreen and ^{1}H NMR spectroscopy. 2-OG decarboxylation in the absence of DNA was

inhibited with potencies comparable to those obtained for ^{5m}C substrate oxidation inhibition. Inhibitor ranking was also largely unchanged between AlphaScreen and ^{1}H NMR assays, with the most potent inhibitor 2,4-PDCA losing potency (IC_{50} = 3.2 μM and 8.1 μM for AlphaScreen and ^{1}H NMR without DNA, respectively), and the least potent inhibitor NOG gaining potency (IC_{50} = 200 μM and 117 μM for AlphaScreen and ^{1}H NMR without DNA, respectively). Succinate production inhibition potency of 2,4-PDCA was rescued to levels observed using AlphaScreen in the presence of **^{5m}C 12 bp dsDNA** (IC_{50} = 2.5 μM and 3.2 μM for ^{1}H NMR with ^{5m}C DNA and AlphaScreen, respectively), suggesting that dsDNA binding to NgTet1 may prompt a degree of active site reorganisation.

While the potencies of 2-OG mimetic iron chelators (NOG, IOX1, and 2,4-PDCA) varied between characterisation methods, TiP1 potency remained virtually unchanged. TiP1 was found to have an agonistic effect on 2-OG turnover when present in near-stoichiometric ratios to NgTet1 (1:1.2 NgTet1/TiP1) in the absence of DNA, increasing NgTet1 activity by nearly 40% in a statistically significant manner. No enhancing effect could be observed in the presence of **^{5m}C 12 bp dsDNA**, potentially due to competition for the nucleotide substrate binding site. A similar effect has been reported for PHDs in the presence of substrate competitors such as type II poly(L-proline) and other proline-containing oligomers, with the authors suggesting all substrate competitors may have an enhancing effect on uncoupled 2-OG decarboxylation[77,206]. Addition of poly(L-proline) increased total uncoupled 2-OG decarboxylation by up to 66%, while no prolyl hydroxylation was observed[77]; analogously, no modification to TiP1 could be observed by MALDI-TOF MS, eliminating the possibility of enzymatic TiP1 modification, and supporting TiP1 as a substrate competitor. Unlike poly(L-proline), the activity enhancing effect of TiP1 was dependent on its stoichiometry with NgTet1, with higher TiP1 concentrations inhibiting succinate production in a dose-dependent manner. This observation, along with the steep Hill slopes (n_H) of TiP1 dose-response curves

obtained from ^{1}H NMR experiments (n_H = 3.2 and 1.6 for TiP1 and 2,4-PDCA in the presence of ^{5m}C, respectively), suggests that TiP1 may inhibit NgTet1 *via* aggregation. Aggregation-based inhibitors are often characterised by steep Hill slopes, promiscuous inhibition, and sensitivity to detergents[217–219]. Further work on TiP1 mode of action is required in order to validate this hypothesis, however, as TiP1 has shown selectivity for hTET1 over hTET2[191], as well as no loss in NgTet1 inhibitory potency in the presence of detergent (Figure 2.15).

2.5 Conclusions and future work

Naegleria Tet-like dioxygenase NgTet1 was produced recombinantly and characterised using AlphaScreen, MALDI-TOF MS, Succinate-Glo™, and ^{1}H NMR assays. Kinetic parameter determination for 2-OG using ^{1}H NMR showed that NgTet1 catalytic activity was improved 3.5-fold by the addition of non-methylated oligonucleotide, with a further 1.6-fold improvement observed upon introducing cytosine methylation. Substantial levels of uncoupled 2-OG decarboxylation (up to 60% of total 2-OG consumption) were observed within the window of linear ^{5hm}C formation, demonstrating that NgTet1 2-OG turnover remains uncoupled to a large degree even in the presence of substrate. Further work should focus on a broader exploration of assay conditions (pH and temperature, in particular) and the flanking sequence around the ^{5m}CpG site of substrate DNA to determine if the proportion of uncoupled 2-OG turnover can be regulated by these factors. Additional work is also required to determine if this observation extends to the rest of the TET enzyme family, particularly hTETs, and to cellular contexts.

The fate of the highly reactive Fe(IV)-oxo species formed during 2-OG consumption also warrants further investigation; while a number of Fe(II)/2-OG-dependent dioxygenases have been reported to perform auto-hydroxylation in the absence of substrate (section 1.3.2), no evidence of TET auto-hydroxylation has been reported as of yet. Concurrently, L-ascorbate may be acting as an alternative electron sink in the uncoupled catalytic cycle, reducing the Fe(IV)-oxo species and returning the enzyme to its active form[220], exemplified by the stoichiometric dependence of PHD on L-ascorbate in the absence of substrate (section 1.3.2). The ^{1}H NMR assay could be used to further explore the role of L-ascorbate, with a particular focus on changes to kinetic parameters for L-ascorbate in the presence of substrate competitors or alternative reducing agents, such as tetrahydrofolic acid (THFA)[221] or dithiothreitol (DTT)[222].

Quantification of reactive oxygen species (ROS) may be an alternative approach to investigating temporal changes in ROS levels during uncoupled 2-OG decarboxylation. A number of ROS detection methods have been reported, including fluorescence-, chemiluminescence-, and spectroscopy-based methods[223]. ROS detection *via* oxidation of luminol to 3-aminophthalate could be a promising approach, as luminol is responsive to multiple radical species, including $O_2^{\bullet-}$, $HO^{\bullet}$, and H_2O_2, enabling simultaneous detection of a number of ROS[223]. Care should be taken while developing the method, as L-ascorbate, apparently essential for NgTet1 activity (Table 2.2), is known to scavenge a wide array of ROS[224,225]. Furthermore, hydrogen peroxide-specific methods should be avoided, as Fe(II) is known to react with H_2O_2 to form hydroxyl and hydroxyperoxyl radicals ($HO^{\bullet}$ and $HOO^{\bullet}$, respectively) *via* Fenton chemistry[226].

Characterisation of known dioxygenase inhibitors (NOG, IOX1, 2,4-PDCA, and TiP1) revealed that potency of 2-OG mimetics (NOG, IOX1, 2,4-PDCA) varied up to 3-fold across activity detection methods and conditions (AlphaScreen and ^{1}H NMR; with and without **^{5m}C 12 bp dsDNA**); conversely, TiP1 potency was virtually unchanged across the conditions screened. At near-stoichiometric ratios with NgTet1, TiP1 enhanced NgTet1 catalytic activity in the absence of DNA but not when methylated DNA substrate was present, suggesting substrate-competitive mode of inhibition based on previous studies on PHDs[77,206]. Characterisation of other TET inhibitors that are outside the 2-OG mimetic family, such as C35 and eltrombopag (section 1.6)[189], could provide further insight into the potential effect of compound mode of inhibition on uncoupled 2-OG decarboxylation.

Chapter 3
Development of cyclic peptide inhibitors for human TET enzymes

3.1 Introduction

Chemical tools targeting epigenetic proteins are indispensable for phenotypic profiling, mechanistic investigations, and target validation[192], but the desired combination of selectivity, potency, and cellular activity is difficult to achieve for enzyme families with high structural similarity. The unavailability of selective TET activity-modulating molecules has hampered studies reliant on temporal enzyme inhibition.

Cyclic peptide (CP) probes are excellent molecular tools due to their very favourable selectivity profiles compared to small molecule (≤ 1,000 Da) modulators[227]. Furthermore, identification of potent CP ligands does not require the level of effort typically associated with small molecule inhibitor development, as high-diversity screening ("display") techniques enable rapid identification of target protein binders *via* simultaneous screening of libraries consisting of up to 10^{15} unique peptide sequences[228]. One such technique, the mRNA display-based random nonstandard peptide integrated discovery (RaPID) system (Figure 3.1), has found extensive use in the identification of *de novo* cyclic peptide binders for a broad range of proteins[228,229]. The mRNA display selection process starts with a partially randomised DNA library, which is transcribed to RNA using *in vitro* transcription/translation (IVTT) and linked to puromycin, a non-hydrolysable tyrosyl-tRNA analogue. Following translation of the RNA sequence, puromycin covalently attaches to the translated peptide *via* the α-amino group due to the peptidyl transferase activity of the ribosome, "barcoding" each peptide with its corresponding mRNA sequence. Reprogramming of the AUG start codon to *N*–chloroacetylated D-Tyr enables spontaneous intramolecular cyclisation with the side chain

of a downstream Cys residue to yield mRNA-tagged cyclic peptides. Subsequent reverse transcription generates complementary DNA (cDNA) for each mRNA tag, and the resulting library can be panned against an immobilised protein target of interest. Recovery of target-bound peptides is followed by PCR amplification of the DNA barcodes to generate a binder-enriched library. Subsequent iterations of the selection should, in principle, return a DNA library enriched for sequences coding for peptides that bind to the target with high affinity.

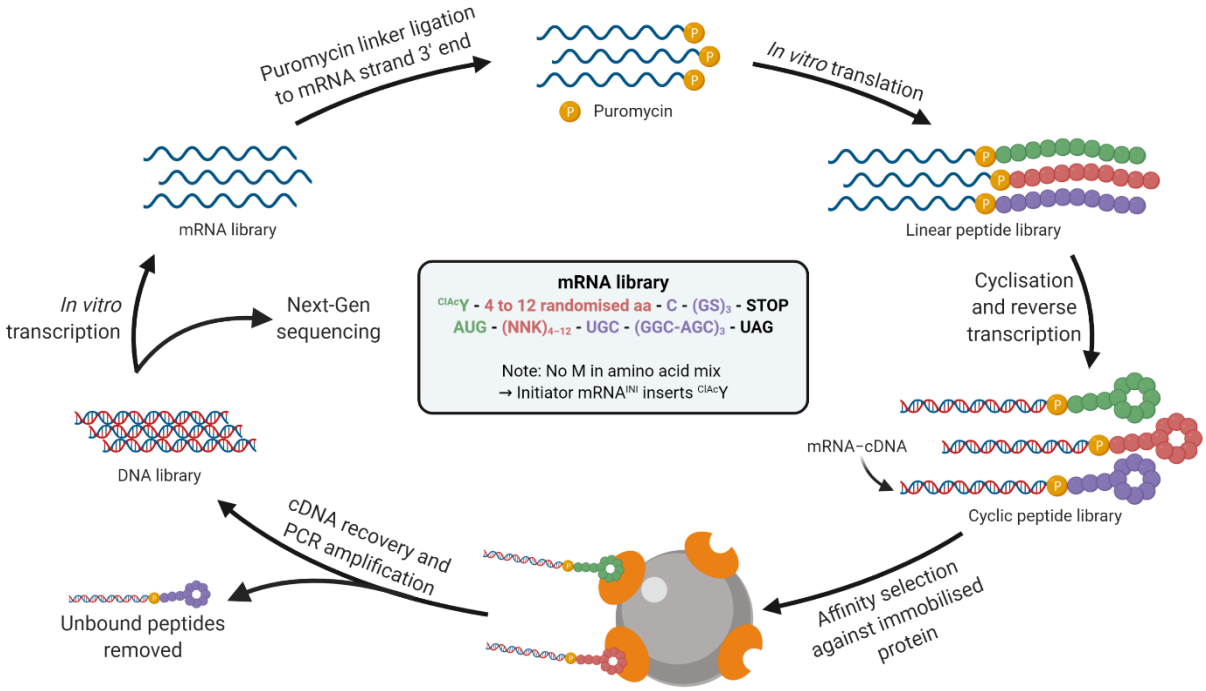


Figure 3.1 | Schematic representation of the mRNA display-based RaPID system. NNK codons represent randomised codons where N = A/C/G/T and K = G/T. Created using biorender.com.

While mRNA display facilitates the simultaneous screening of a large and diverse library of peptides, the main drawback of this approach is that library enrichment is affinity-driven, rather than inhibition-driven. Furthermore, highly lipophilic sequences can be enriched due to non-specific binding to the target. As such, peptide sequence enrichment does not directly correlate with inhibitory potency, and hits must be synthesised and validated as inhibitors using orthogonal methods[230].

The RaPID system was recently used to identify potent hTET1-targeting cyclic peptide inhibitors including TiP1[191], which was selective for hTET1 over hTET2 (section 1.6.5). TiP1 inhibitory potencies and kinetic binding constants were evaluated for hTET1/2, but no further characterisation has been reported.

3.2 Aims

The reported characterisation of TiP1 is insufficient to assess the utility of TiP1 in cellular systems. Characterisation of TiP1 against a broader panel of TETs and other enzymes from the Fe(II)/2-OG-dependent dioxygenase superfamily is required to establish the selectivity profile of the peptide; subsequent evaluation of TiP1 cellular activity would then determine its potential as a chemical tool. Finally, complementary development of hTET2 CP inhibitors could further expand the molecular TET activity modulator toolbox.

The aims of the work presented in this chapter are as follows:

- Characterise the reported CP hTET1 inhibitors TiP1–3[191] and additional peptide sequence hits identified from the hTET1 RaPID selection (KS01–13) against a panel of TETs and 2-OG dioxygenases.
- Evaluate the cellular activity of TiP1 and TiP1 conjugates with cell-penetrating peptides.
- Investigate the binding mode of TiP1 to TET enzymes.
- Evaluate *in vitro* and cellular activity of the novel CP hTET2 inhibitors KTT1–6 selected using mRNA display against hTET2.

3.3 Results

3.3.1 Characterisation of cyclic peptide TET inhibitors TiP1–3

In 2018, Nishio *et al*.[191] reported the use of the RaPID system to identify thioether macrocyclic peptide inhibitors of hTET1 (TiP1–3, all peptide sequences tabulated in Appendix 2). As the peptides were only evaluated for hTET1/2 inhibition, screening against a broader panel of TET enzymes and other Fe(II)/2-OG-dependent dioxygenases was desired in order to investigate the selectivity profiles of the compounds.

TiP1–3 were synthesised using Fmoc solid-phase peptide synthesis (SPPS) and purified using reverse-phase HPLC. TiP1 and TiP2 were obtained in high purity (>99.0% by LC-UV, 220 nm; full peptide characterisation information available in Appendix 2), while isolation of TiP3 proved more difficult (91.9% by LC-UV, 220 nm).

As large-scale recombinant production of catalytically active hTET1/3 was challenging (personal communication with Eidarus Salah), evaluation of peptide inhibitory potencies was carried out using AlphaScreen, which enabled detection of TET catalytic activity with femtomole amounts of enzyme (10^{-14}–10^{-13} moles[191]). This facilitated initial assay optimisation using purchased recombinant hTET1 when the enzyme was not readily producible in-house. The anti-^{5hm}C antibody-based AlphaScreen method used for TET catalytic activity detection in Chapter 2 constitutes a signal-increase assay, producing elevation in signal intensity in relation to the accumulation of analyte of interest (^{5hm}C)[†]. However, as a complex string of interactions between multiple binding partners is required in order to produce the signal (Figure 2.2), perturbation of any of these interactions by an external modulator would lead to a decrease in signal intensity. This signal loss would be practically indistinguishable

† Nucleobase modifications, including ^{5m}C, ^{5hm}C, ^{5f}C, ^{5ca}C, and others, refer to modifications of deoxyribonucleotides in DNA oligomers unless otherwise mentioned.

from one stemming from true enzyme inhibition, resulting in false positive inhibition results. Thus, it is essential that appropriate assay interference controls are in place to differentiate enzyme inhibitors from AlphaScreen signal inhibitors. In order to identify false positive hits, two types of AlphaScreen interference assays were carried out. Firstly, an antibody-independent AlphaScreen TruHits interference assay was used to identify potential singlet oxygen quenchers, colour quenchers, light scatterers, biotin mimetics, and acceptor bead competitors. Laterally, an anti-^{5hm}C antibody-dependent AlphaScreen interference assay, containing 8.5 nM of substrate (^{5m}C) and 1.5 nM of analyte (^{5hm}C) to simulate enzymatic turnover (as in Figure 2.3), was used to identify antibody-associated AlphaScreen signal quenchers.

A small molecule TET inhibitor control, IOX1 (section 1.6.2), exhibited no interference with AlphaScreen signal (Table 3.1), while cyclic peptide TiP1 partially interfered with the assays at the highest peptide concentrations (66.1% and 48.4% signal at 100 µM for TruHits and antibody interference, respectively). Conversely, TiP2 and TiP3 were shown to be potent signal suppressors in the antibody-based interference assay (pIC_{50} = 5.45 and 5.48, respectively), necessitating caution when interpreting all subsequent enzyme inhibition results. The >10-fold difference in potencies between the two types of interference assays suggested that the peptides bound non-specifically either at the interface between protein A immobilised on acceptor beads and the IgG-based anti-^{5hm}C antibody, or the ^{5hm}C recognition domain of the antibody.

Table 3.1 | Tabulated AlphaScreen interference pEC_{50} values for reported cyclic peptide inhibitors TiP1–3. Values reported as pEC_{50} ± SD (M). Values reported as percentage signal (%) at 100 µM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 µM). Data reported as the mean of 3 experimental replicates (n = 3; mean ± SD).

Assay type	pEC_{50} ± SD (M)			
	IOX1	**TiP1**	**TiP2**	**TiP3**
TruHits interference	n.i.	66.1 ± 3.4%	66.7 ± 2.3%	72.3 ± 1.4%
Anti-^{5hm}C antibody interference	n.i.	48.4 ± 3.4%	5.45 ± 0.13	5.48 ± 0.05

Prior to screening for TET inhibition, linear enzymatic activity windows within which inhibition data could be collected were determined for hTET1$_{CD}$ (0–30 min, R^2 = 0.97), SUMO-tagged (0–10 min, R^2 = 0.97) and cleaved (tag-free) hTET2$_{CD\Delta LCI}$ (0–12 min, R^2 = 0.99), hTET3$_{CD}$ (0–10 min, R^2 = 0.99), murine Tet1$_{CD}$ (mTet1$_{CD}$, 0–9 min, R^2 = 0.99), and NgTet1 (0–25 min, R^2 = 1.00) using AlphaScreen (Figure 3.2). For brevity, construct information (CD, CDΔLCI) will be omitted from hereon.

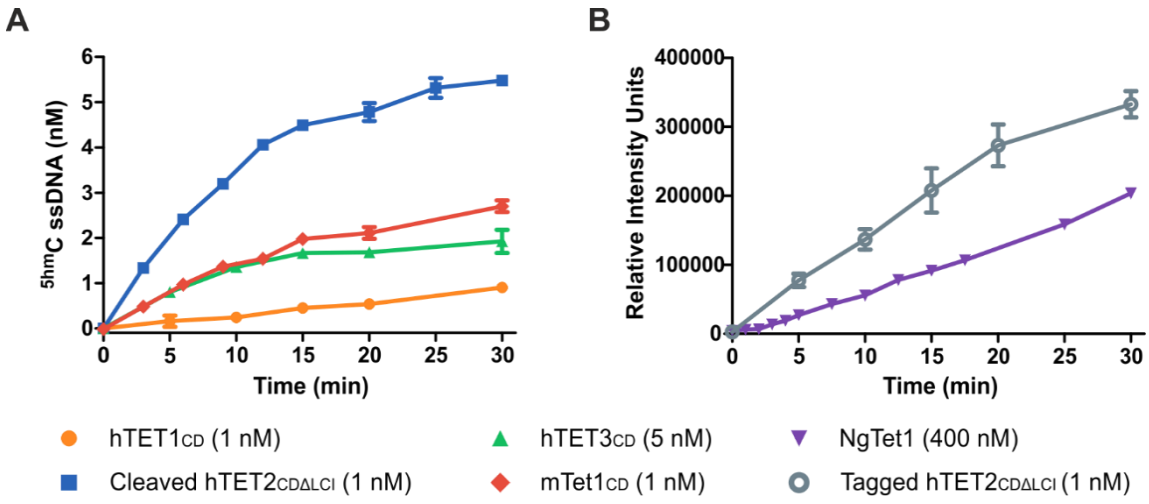


Figure 3.2 | Determination of linear enzymatic activity windows for TET enzymes using AlphaScreen. Enzyme concentrations: 10 nM for hTET1$_{CD}$, 1 nM for cleaved and tagged hTET2$_{CD\Delta LCI}$, 5 nM for hTET3$_{CD}$, 1 nM for mTet1$_{CD}$, and 400 nM for NgTet1. Conditions: TET enzyme, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. Data shown as the mean of 1–3 experimental replicates (n = 1–3; mean ± SD).

Following the determination of linear enzymatic activity windows, peptides TiP1–3 were screened against a panel of five TET enzymes (Figure 3.3, Table 3.2), including hTET1–3, mTet1, and NgTet1. The observed hTET1 inhibitory potencies for both TiP1 and IOX1 (IC_{50} = 200 and 191 nM, respectively) were over 5-fold higher than the reported[191] values (IC_{50} = 1.48 and 1.10 µM, respectively), although virtually identical assay conditions were used in both cases. The discrepancy in potency may be due to batch-to-batch variability of the hTET1 enzyme.

Importantly, TiP1 was >30-fold selective over hTET2/3, demonstrating the selectivity advantage conferred by the use of cyclic peptide scaffolds. TiP1 potency against mTet1 was lower than against hTET1 (pIC_{50} = 6.70 and 5.22 against hTET1 and mTet1, respectively), and decreased further against NgTet1 (pIC_{50} = 4.69), highlighting the structural differences

between TET homologues. Evaluation of an *N*–acetylated linear version of TiP1 (TiP1.L) revealed a >100-fold loss of potency against hTET1 compared to TiP1 (Figure 3.3 B) and a marked reduction in potency against hTET2/3 (Figure 3.3 C, D), demonstrating that the conformational rigidity induced by peptide cyclisation is crucial for inhibitory activity.

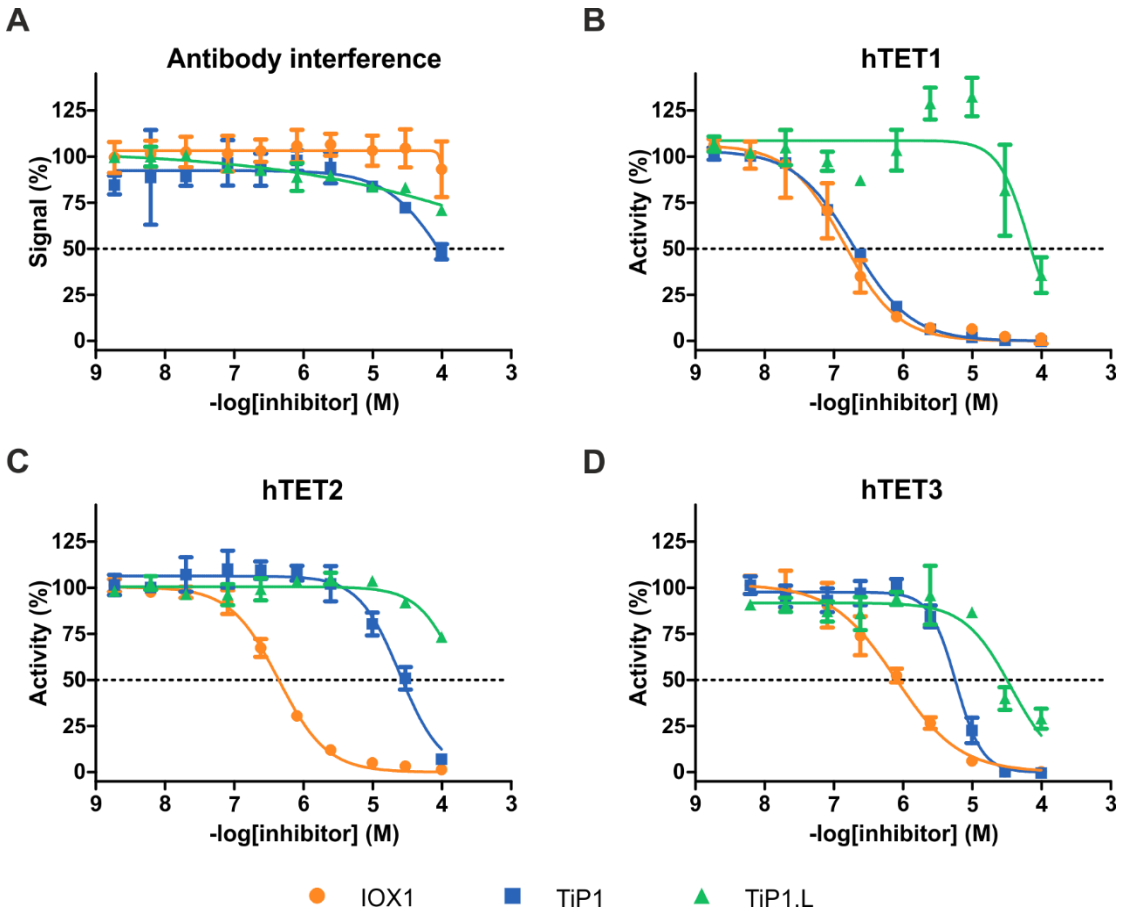


Figure 3.3 | Representative AlphaScreen dose-response curves for IOX1, TiP1, and TiP1.L, highlighting the importance of cyclisation-induced conformational rigidity. A, anti-^{5hm}C antibody interference; **B**, hTET1 inhibition; **C**, hTET2 inhibition; **D**, hTET3 inhibition. Conditions: TET enzyme, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. Data shown as the mean of 2–3 experimental replicates (n = 2–3; mean ± SD).

IC_{50} values obtained using both SUMO-tagged and cleaved hTET2 were largely comparable (data not shown), suggesting that the two enzyme forms could be used interchangeably, at least within the context of cyclic peptide and 2-OG mimetic inhibitor screening.

TiP2 and TiP3 inhibited hTET1 with sub-micromolar potencies, similarly to TiP1 (pIC_{50} = 6.76, and 6.10 for TiP2 and TiP3, respectively, Table 3.2). However, evaluation against the remainder of the TET panel was not possible due to the single-digit micromolar AlphaScreen

interference potencies of the two peptides (pEC_{50} = 5.45 and 5.48 for TiP2 and TiP3, respectively).

Table 3.2 | Tabulated AlphaScreen interference and enzyme inhibition pIC_{50} values for reported cyclic peptide inhibitors TiP1–3 against a panel of TET enzymes. Enzyme inhibition data likely obscured by AlphaScreen interference is shaded in grey. Values reported as pIC_{50} ± SD (M). Values reported as percentage signal (%) at 100 µM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 µM); n.d., not determined. Data for AlphaScreen antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for enzyme inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

		pIC_{50} ± SD (M)				
		IOX1	**TiP1**	**TiP1.L***	**TiP2**	**TiP3**
Antibody interference		n.i.	48.4 ± 3.4%	71.0 ± 0.4%	5.45 ± 0.13	5.48 ± 0.05
Enzyme inhibition	hTET1	6.72 ± 0.66	6.70 ± 0.05	4.15 ± 0.27	6.76 ± 0.23	6.10 ± 0.20
	hTET2	6.23 ± 0.09	4.84 ± 0.48	78.9 ± 9.2%	5.49 ± 0.02	5.52 ± 0.07
	hTET3	5.94 ± 0.13	5.18 ± 0.05	4.51 ± 0.05	5.62 ± 0.06	5.57 ± 0.04
	mTet1	6.37 ± 0.20	5.22 ± 0.10	n.d.	5.46 ± 0.06	5.39 ± 0.20
	NgTet1	4.73 ± 0.21	4.69 ± 0.09	n.d.	5.49 ± 0.07	5.57 ± 0.08

**Data for antibody interference reported as the mean of 2 experimental replicates (n = 2; mean ± SD).*

TiP2/3 dose-response curves for assay interference and hTET2/3 inhibition showed extensive overlap (Figure 3.4), obscuring potential enzymatic inhibition (Table 3.2, values in grey) and precluding the accurate evaluation of TiP2/3 selectivity profiles.

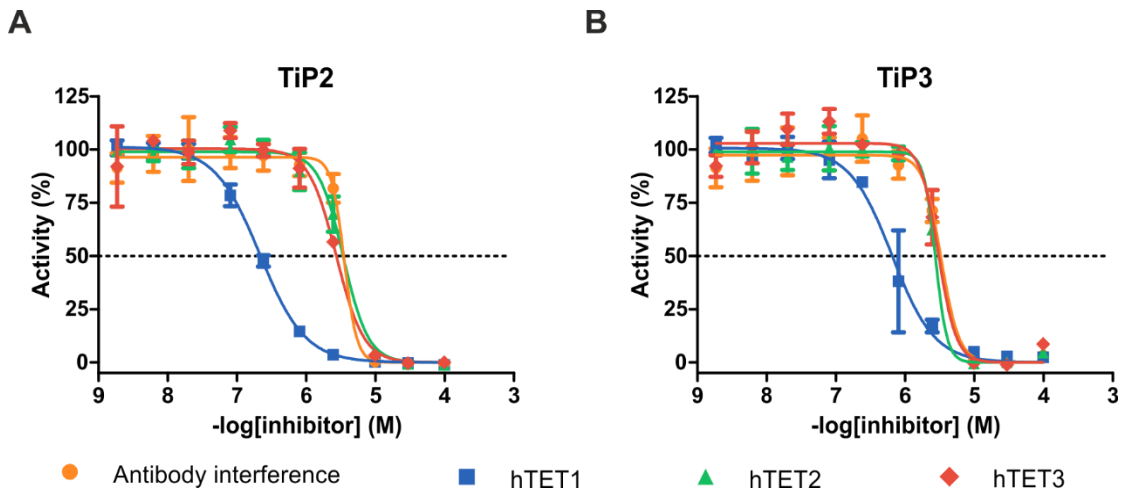
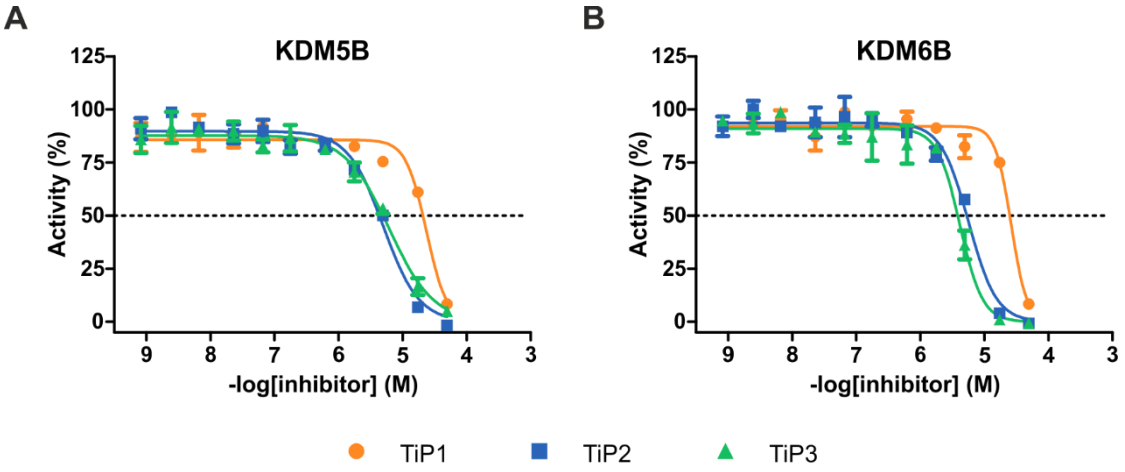


Figure 3.4 | Representative AlphaScreen dose-response curves for TiP2 and TiP3, highlighting strong assay interference. A, TiP2 inhibition; **B**, TiP3 inhibition. Conditions: TET enzyme, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. Data shown as the mean of 2–3 experimental replicates (n = 2–3; mean ± SD).

In order to assess TiP1–3 selectivity for hTET1 over the broader Fe(II)/2-OG-dependent dioxygenase superfamily, the peptides were screened against histone lysine demethylases

KDM5B and KDM6B using AlphaScreen (Figure 3.5). TiP1 was >100-fold less potent against KDM5B/6B compared to hTET1 (IC_{50} = 23, 26, and 0.20 µM, respectively), demonstrating excellent selectivity within the superfamily. The pIC_{50} values of TiP2 and TiP3 for KDMs (pIC_{50} = approx. 5.3) bore high similarity to AlphaScreen interference potencies obtained using the anti-^{5hm}C antibody (pEC_{50} = 5.45 and 5.48, respectively, Table 3.2). While the similarity in pIC_{50}/pEC_{50} values may be circumstantial, this observation could suggest that, rather than binding non-specifically to the anti-^{5hm}C antibody, TiP2 and TiP3 may bind at the interface between protein A immobilised on acceptor beads and IgG-based antibodies.



C

	pIC_{50} ± SD (M)		
	TiP1	TiP2	TiP3
KDM5B	4.63 ± 0.03	5.30 ± 0.04	5.21 ± 0.04
KDM6B	4.58 ± 0.03	5.24 ± 0.03	5.38 ± 0.03
hTET1	6.70 ± 0.05	6.76 ± 0.23	6.10 ± 0.20
Antibody interference	48.4 ± 3.4%	5.45 ± 0.13	5.48 ± 0.05

Figure 3.5 | Cross-screening of TiP1–3 against histone lysine demethylases KDM5B and KDM6B using AlphaScreen. A, KDM5B inhibition; **B**, KDM6B inhibition; **C**, tabulated AlphaScreen pIC_{50} values for KDM5B/6B inhibition, along with antibody interference and hTET1 inhibition (from Table 3.2). KDM5B assay conditions: 0.5 nM KDM5B, 100 nM H3(1-21)K4Me_3-GGK-Biotin, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 5 µM disodium 2-OG; 20 min reaction time. KDM6B assay conditions: 0.1 nM KDM6B, 60 nM H3(21-44)K27Me_3-GGK-Biotin, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG; 10 min reaction time. Data for KDM inhibition shown as the mean of 2 experimental replicates (n = 2; mean ± SD); data for AlphaScreen antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for hTET1 inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD). KDM inhibition experiments were carried out by Dr Anthony Tumber.

3.3.2 Characterisation of additional cyclic peptide sequences from the RaPID selection against hTET1

Characterisation of reported cyclic peptide inhibitor TiP1 revealed a promising selectivity profile, whereas TiP2/3 were potent quenchers of AlphaScreen signal, precluding broader selectivity profiling. Although TiP1 was the most-enriched sequence in the RaPID selection against hTET1[191], sequence enrichment does not necessarily correlate with inhibitory potency (section 3.1). In order to investigate alternative (and potentially superior) CP sequences, sequencing results from the RaPID selection against hTET1 were obtained from Prof. Suga[191] (personal communication), and the peptide sequences were aligned based on identity (Figure 3.6). 13 additional peptides (KS01–13) were selected for synthesis and characterisation against the TET enzyme panel; the selection of peptides largely contained TiP1–3 analogues, with two sequences from unrelated families (KS03, KS10). Sequences containing internal cysteines were avoided due to multiple possible cyclisation variants (loop *vs* lasso cyclisation) and potential issues with side chain thiol oxidation; consequently, Met15 present in KS05 was substituted with Leu (M15L).

Cyclic peptides KS01–13 were synthesised using SPPS and screened for assay interference using both TruHits and anti-^{5hm}C antibody-dependent interference assays, followed by evaluation of inhibitory activity against the TET enzyme panel. Unfortunately, the majority of selected peptides interfered potently with AlphaScreen, complicating inhibitor characterisation (Table 3.3, data shaded in grey). Interestingly, out of three 14-mer peptides with high sequence similarity (TiP1, KS02, and KS08), only one (KS02) was shown to be a potent AlphaScreen signal quencher, suggesting that assay interference is effected by distinct structural features, rather than intrinsic properties such as hydrophobicity. Identification of such structural features could thus prevent subsequent time-consuming preparation and screening of peptides interfering with the AlphaScreen signal; however, due to the small

sample size and heterogeneity of evaluated sequences, features implicated in assay interference could not be unambiguously determined.

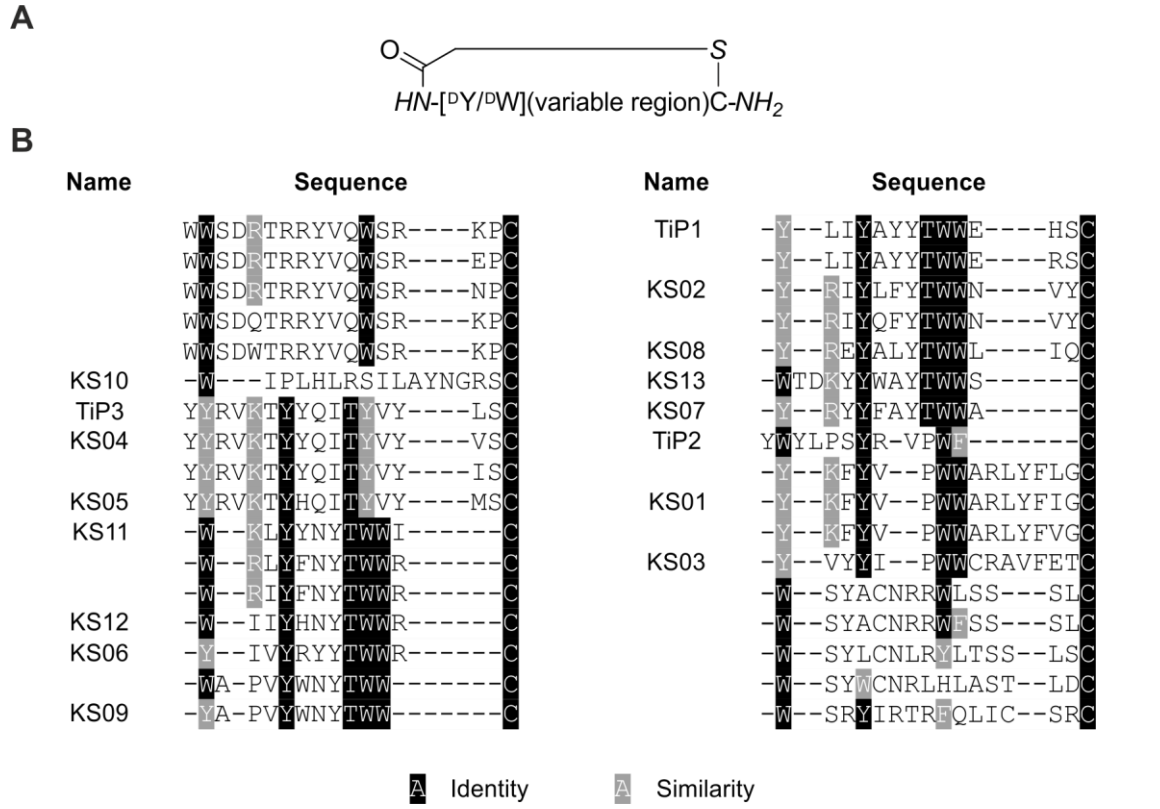


Figure 3.6 | Identity alignment of sequences from the RaPID selection against hTET1. A, schematic representation of the thioether macrocyclic peptide structure, highlighting the variable region (letters representing individual atoms are italicised to distinguish from single-letter amino acid codes); **B**, identity alignment of sequences. N–terminal residues (Tyr or Trp) reported in the sequences are in the D-configuration. C–terminal cysteine is required for the thioether cyclisation step. Black shading indicates residue identity; grey shading indicates residue similarity.

Peptides KS03, KS08, and KS09 were weak AlphaScreen signal quenchers, presenting a sufficient window for enzyme inhibition studies. KS03 exhibited the lowest TET2 inhibition across the board and showed poor selectivity within the hTET family (pIC_{50} = 4.35, <4, and 4.50 for hTET1–3, respectively). KS09 was >10-fold selective for hTET1 over hTET2/3, but showed 10-fold lower potency against hTET1 compared to TiP1 (hTET1 pIC_{50} = 5.75 and 6.70 for KS09 and TiP1, respectively). The inhibitory profile of KS08 was largely comparable to that of TiP1, a result in accordance with the high sequence similarity between the two peptides (Figure 3.6).

Table 3.3 | Tabulated AlphaScreen interference and enzyme inhibition pIC_{50} values for TiP1 and KS01–13 against a panel of TET enzymes. Enzyme inhibition data likely obscured by AlphaScreen interference is shaded in grey. Values reported as pIC_{50} ± SD (M). Values reported as percentage signal (%) at 100 μM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 μM). Data for AlphaScreen antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for enzyme inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

	pIC_{50} ± SD (M)						
Entry	**Interference**		**Enzyme inhibition**				
	TruHits	Antibody	hTET1	hTET2	hTET3	mTet1	NgTet1
IOX1	n.i.	n.i.	6.72 ± 0.66	6.23 ± 0.09	5.94 ± 0.13	6.37 ± 0.20	4.73 ± 0.21
TiP1	66.1 ± 3.4%	48.4 ± 3.4%	6.70 ± 0.05	4.84 ± 0.48	5.18 ± 0.05	5.22 ± 0.10	4.69 ± 0.09
KS01*	4.31 ± 0.12	6.08 ± 0.02	6.62 ± 0.02	5.95 ± 0.02	5.99 ± 0.03	5.96 ± 0.02	5.95 ± 0.04
KS02	73.2 ± 2.6%	5.72 ± 0.03	5.94 ± 0.09	5.59 ± 0.13	5.62 ± 0.07	5.47 ± 0.31	5.73 ± 0.09
KS03†	74.7 ± 2.8%	64.4 ± 7.4%	4.35 ± 0.29	80.0 ± 7.0%	4.50 ± 0.53	4.41 ± 0.37	4.27 ± 0.11
KS04	64.3 ± 2.6%	5.29 ± 0.06	5.87 ± 0.11	5.11 ± 0.10	5.22 ± 0.04	5.06 ± 0.05	5.36 ± 0.04
KS05*	53.3 ± 1%	5.94 ± 0.03	6.11 ± 0.04	5.89 ± 0.02	6.30 ± 0.12	5.87 ± 0.03	5.85 ± 0.03
KS06	4.46 ± 0.10	5.66 ± 0.04	6.57 ± 0.06	5.66 ± 0.05	5.86 ± 0.04	5.75 ± 0.24	5.73 ± 0.06
KS07	4.12 ± 0.044	5.52 ± 0.03	6.33 ± 0.08	5.52 ± 0.03	5.56 ± 0.03	5.44 ± 0.04	5.53 ± 0.05
KS08	88.6 ± 1.3%	76.5 ± 6.1%	6.69 ± 0.29	4.38 ± 0.39	4.31 ± 0.36	4.34 ± 0.19	4.29 ± 0.13
KS09	83.9 ± 1.8%	75.7 ± 7.6%	5.73 ± 0.26	4.17 ± 0.30	4.33 ± 0.16	4.26 ± 0.40	4.05 ± 0.23
KS10	68.9 ± 0.5%	4.51 ± 0.04	6.16 ± 0.07	4.65 ± 0.05	4.74 ± 0.09	4.56 ± 0.12	4.84 ± 0.08
KS11*	4.68 ± 0.06	6.06 ± 0.02	6.22 ± 0.02	6.03 ± 0.03	6.16 ± 0.05	5.98 ± 0.04	6.08 ± 0.03
KS12*	4.98 ± 0.08	6.08 ± 0.02	6.44 ± 0.02	6.08 ± 0.04	6.20 ± 0.06	6.03 ± 0.03	6.15 ± 0.03
KS13	4.2 ± 0.09	5.17 ± 0.07	6.10 ± 0.16	5.41 ± 0.16	5.45 ± 0.09	5.36 ± 0.42	5.49 ± 0.25

**Data for enzyme inhibition reported as the mean of 3 experimental replicates (n = 3; mean ± SD).*

†*Data for hTET2 and hTET3 inhibition reported as the mean of 2 independent replicates (n = 2; mean ± SD).*

3.3.3 Characterisation of TiP1–3 and KS01–13 against hTET2 using mass spectrometry

In order to validate the results obtained using AlphaScreen, an orthogonal characterisation method was employed. A mass spectrometry-based assay using the RapidFire365 solid-phase extraction instrument (SPE-MS) was optimised based on previous work by Dr Roman Belle[172]. Self-complementary DNA oligomer **^{5m}C 12 bp dsDNA** (Appendix 1) was used as the substrate, enabling direct detection of full-length oligonucleotides containing ^{5m}C and its oxidation products; 1 µM of the double-stranded oligonucleotide produced sufficient signal strength using SPE-MS, and resulted in a 2 µM final concentration of the ^{5m}C substrate.

Analysis of ^{5m}C oxidation product (^{5ox}C) accumulation profiles was complicated by the overlap of peaks corresponding to ^{5hm}C and ^{5f}C (−2 Da difference), necessitating deconvolution of the two sets of peaks based on the relative isotopic distribution of ^{5m}C and ^{5hm}C (Figure 3.7 A). As such, the linear activity window of SUMO-tagged hTET2 for inhibitor evaluation was selected within the linear range of the ^{5hm}C product formation, at conditions where subsequent ^{5ox}C formation was minimal (300 nM hTET2, 0–12 min, R^2 = 0.90, Figure 3.7 B). The linear activity window for hTET1 was not determined, as inhibition data on hTET1 could not be collected using SPE-MS due to insufficient enzyme material.

Interestingly, analysis of hTET1 and SUMO-tagged hTET2 catalytic activity using SPE-MS demonstrated a rapid formation of ^{5f}C and revealed differences in ^{5ox}C accumulation profiles: hTET1 maintained a consistent level of ^{5hm}C, whereas hTET2 depleted ^{5hm}C levels from 10 min onwards; furthermore, hTET2 produced ^{5f}C and ^{5ca}C more rapidly than hTET1 (Figure 3.7 C, D).

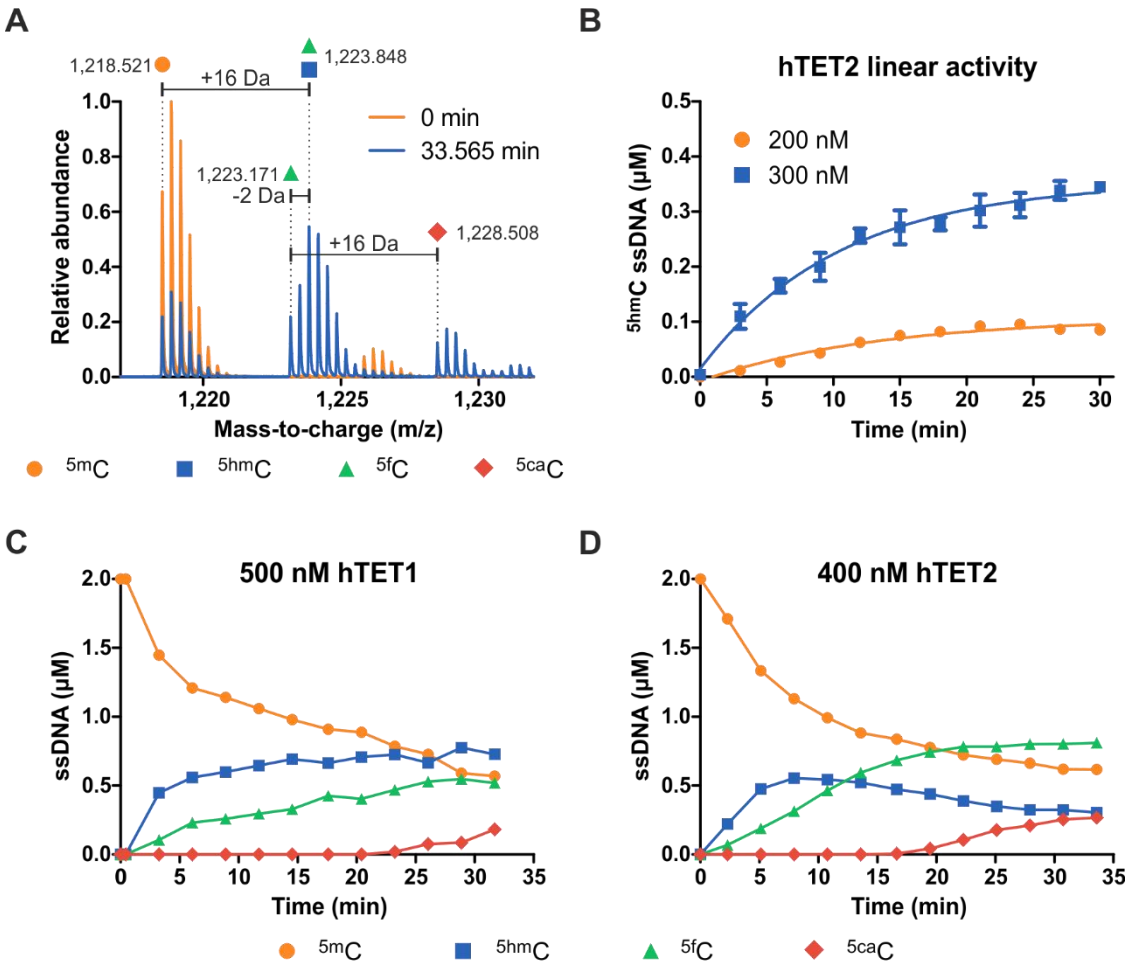


Figure 3.7 | Characterisation of hTET1 and hTET2 catalytic activity using SPE-MS. A, analysis of representative profile mass spectra from **D**, with oligonucleotides observed in the $[M-3H]^{3-}$ charge state; **B**, linear enzymatic activity window determination for SUMO-tagged hTET2 using end-point reaction sampling (300 nM hTET2, 0–12 min, $R^2 = 0.90$); **C**, substrate turnover by 500 nM hTET1 using continuous reaction sampling; **D**, substrate turnover by 400 nM hTET2 using continuous reaction sampling. Conditions: TET enzyme, 1 μM **^{5m}C 12 bp dsDNA**, 200 μM sodium L-ascorbate, 50 μM $(NH_4)_2Fe(SO_4)_2$, and 50 μM disodium 2-OG. Data shown as the mean of 2 experimental replicates for **B** (n = 2; mean ± SD); data shown as single replicates for **C** and **D**.

Based on the determined linear enzymatic activity range, cyclic peptides TiP1–3 and KS01–13 were screened against hTET2 (Figure 3.8 A–D). The peptide library was generally compatible with the assay with the exception of KS07: KS07 activity could not be assessed at the two highest peptide concentrations (30 and 100 μM) due to the appearance of unidentified peaks overlapping with and suppressing ^{5hm}C analyte peaks (Figure 3.8 E, F). The relative intensity of the unidentified peaks increased in a KS07 concentration-dependent manner, suggesting KS07 as the source of the contamination. The KS07-related species (1,221.485 m/z) ionised in the $[M-2H]^{2-}$ state, with deconvolution revealing a mass of 2,444.99 Da. The impurity mass could not be observed by MALDI-TOF MS analysis of KS07 (data not shown), while LC-MS analysis of

KS07 confirmed the high purity of the peptide (98.3%, Appendix 2). Thus, the SPE-MS-exclusive appearance of the 2,444.99 Da peaks may be due to the relatively higher sensitivity of the Q-TOF mass spectrometer used for analysis, and/or the differences in ionisation sources between the instruments.

Inhibition results obtained against hTET2 using SPE-MS supported the single-digit micromolar potencies obtained using AlphaScreen (Table 3.4). The majority of screened peptides (10/16), particularly AlphaScreen signal quenchers, exhibited reduced potencies under SPE-MS conditions, while AlphaScreen non-interfering peptides (TiP1, KS08, and KS09) consistently inhibited hTET2 more potently under SPE-MS conditions (ΔpIC_{50} = 0.43, 1.22, and 1.49 for TiP1, KS08, and KS09, respectively). Overall, the inhibitory potency values obtained using the two assays did not correlate well (Spearman's r_s = 0.266; p = 0.302). A comparison of Hill coefficients (n_H) for dose-response curve fits obtained using AlphaScreen and SPE-MS revealed that n_H values were on average 2-fold higher when using SPE-MS (average n_H = 1.98 and 3.96 for AlphaScreen and SPE-MS, respectively), suggesting that the majority of the tested peptides may act as aggregate-based inhibitors[219].

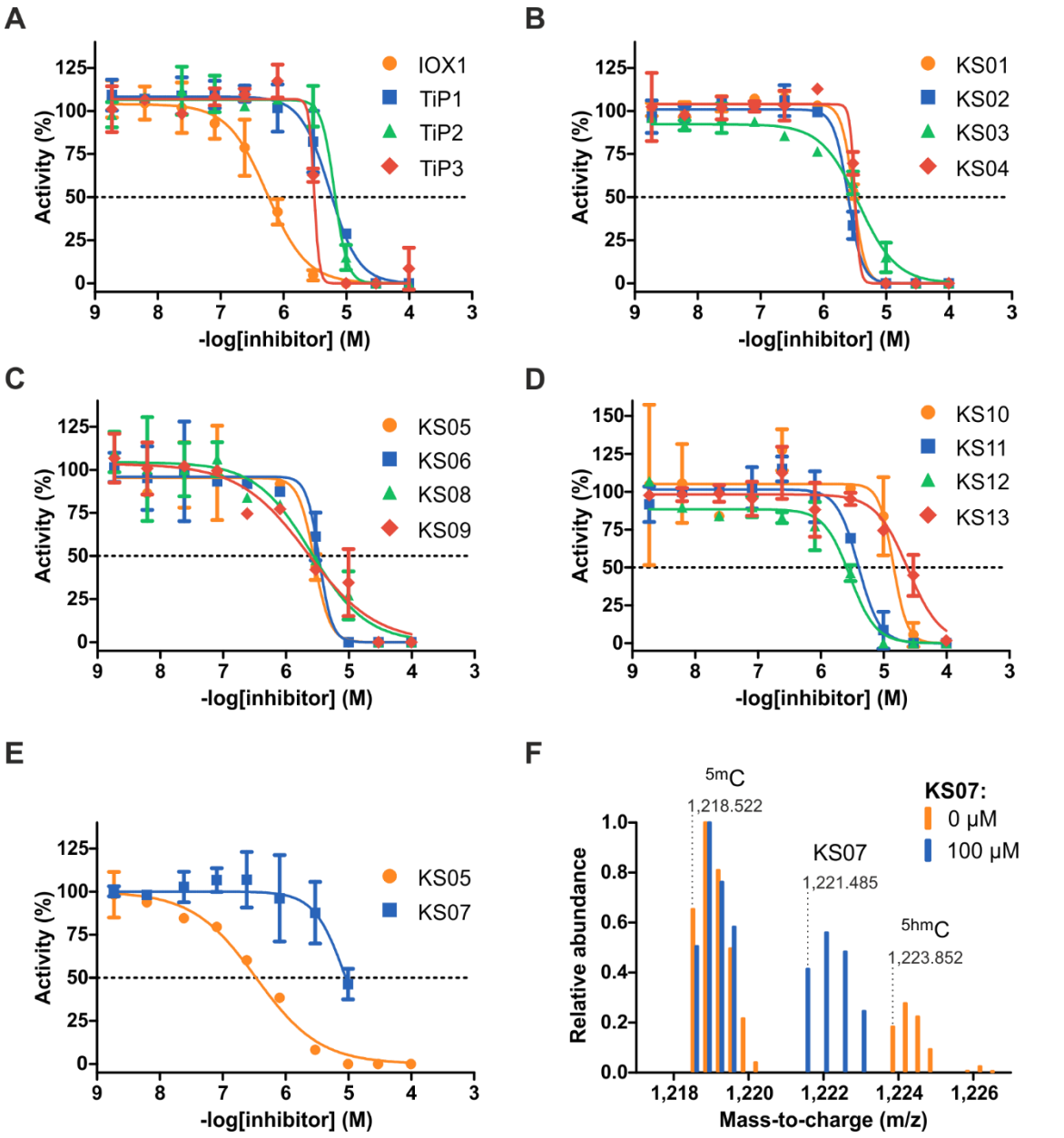


Figure 3.8 | SPE-MS dose-response curves for peptides TiP1–3 and KS01–13 against hTET2. A, **B**, **C**, **D**, dose-response curves for peptides excluding KS07; **E**, dose-response curves for IOX1 and KS07; **F**, comparison of representative centroid mass spectra for 0 and 100 µM KS07, with oligonucleotides observed in the $[M-3H]^{3-}$ charge state and the KS07-related peaks observed in the $[M-2H]^{2-}$ charge state. In **F**, 100 µM KS07 peaks (blue) are nudged by 0.1 unit to minimise peak overlap. Conditions: 300 nM hTET2, 1 µM **^{5m}C 12 bp dsDNA**, 200 µM sodium L-ascorbate, 50 µM $(NH_4)_2Fe(SO_4)_2$, and 50 µM disodium 2-OG. Data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

Table 3.4 | Tabulated AlphaScreen interference and enzymatic inhibition (AlphaScreen and SPE-MS) pIC_{50} values and Hill coefficients for cyclic peptide inhibitors TiP1–3 and KS01–13 against hTET2. Hill coefficients (n_H) for dose-response curves of hTET2 inhibition were calculated using non-linear regression fitting in GraphPad Prism 5. Δ values were calculated by subtracting AlphaScreen values from SPE-MS values. Enzyme inhibition data likely obscured by AlphaScreen interference is shaded in grey. Values reported as pIC_{50} ± SD (M). Values reported as percentage signal (%) at 100 μM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 μM). Data for anti-^{5hm}C antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for AlphaScreen hTET2 inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD); data for SPE-MS hTET2 inhibition reported as the mean of 2 experimental replicates (n = 2; mean ± SD).

Entry	Antibody interference	pIC50 ± SD (M) hTET2 inhibition			Hill coefficient (n_H)		
		AlphaScreen	SPE-MS	ΔpIC_{50}	Alpha Screen	SPE-MS	Δn_H
IOX1	n.i.	6.23 ± 0.09	6.26 ± 0.06	0.03	1.24	1.37	0.13
TiP1	48.4 ± 3.4%	4.84 ± 0.48	5.27 ± 0.05	0.43	1.24	1.88	0.64
TiP2	5.45 ± 0.13	5.49 ± 0.02	5.19 ± 0.06	−0.30	2.84	4.26	1.42
TiP3	5.48 ± 0.05	5.52 ± 0.07	5.51*	−0.01	2.12	10.6	8.48
KS01	6.08 ± 0.02	5.95 ± 0.02‡	5.52 ± 0.01	−0.43	3.88	4.65	0.77
KS02	5.72 ± 0.03	5.59 ± 0.13	5.61 ± 0.05	0.02	2.14	3.82	1.68
KS03	64.4 ± 7.4%	80.0 ± 7.0%	5.44 ± 0.04	<−1.44	n.d.	1.45	n.d.
KS04	5.29 ± 0.06	5.11 ± 0.10	5.50*	0.39	2.69	11.0	8.31
KS05	5.94 ± 0.03	5.89 ± 0.02‡	5.55 ± 0.05	−0.34	2.11	3.45	1.34
KS06	5.66 ± 0.04	5.66 ± 0.05	5.47 ± 0.09	−0.19	1.6	4.28	2.68
KS07	5.52 ± 0.03	5.52 ± 0.03	approx. 5†	−0.5	2.13	0.61†	−1.52
KS08	76.5 ± 6.1%	4.38 ± 0.39	5.60 ± 0.15	1.22	0.68	0.98	0.30
KS09	75.7 ± 7.6%	4.17 ± 0.30	5.66 ± 0.17	1.49	0.76	0.80	0.04
KS10	4.51 ± 0.04	4.65 ± 0.05	4.85 ± 0.14	0.20	1.35	3.86	2.51
KS11	6.06 ± 0.02	6.03 ± 0.03‡	5.40 ± 0.07	−0.14	3.19	2.03	−1.16
KS12	6.08 ± 0.02	6.08 ± 0.04‡	5.54 ± 0.05	−0.54	1.71	2.11	0.40
KS13	5.17 ± 0.07	5.41 ± 0.16	4.62 ± 0.08	−0.79	1.28	1.70	0.42

**Curve fitting and standard deviation are ambiguous due to unusually high Hill coefficient.*
†*Data collected up to 10 μM KS07, as higher concentrations resulted in ^{5hm}C analyte signal suppression.*
‡*Data reported as the mean of 3 experimental replicates (n = 3; mean ± SD).*

3.3.4 Evaluation of TiP1 cellular activity

Following *in vitro* TiP1 selectivity profiling using two orthogonal methods, evaluation of TiP1 cellular activity was carried out to investigate if the peptide could be used as a chemical tool. The cellular TET assay utilises doxycycline (dox)-inducible U2OS epithelial cells stably transfected with FLAG-tagged catalytic domains (CD) of hTET1–3, enabling transient expression of catalytically active hTET1–3 and enrichment of ^{5ox}C. Treatment of dox-induced cells with inhibitors of interest is followed by immunostaining using anti-FLAG and anti-^{5hm}C antibodies to quantify the relative levels of hTET enzymes and their catalytic activity (Figure 3.9 A). ^{5hm}C accumulation is not observed in cells overexpressing catalytically inactive hTET

variants, confirming the relation between hTET expression and ^{5hm}C level elevation (data not shown).

Cell treatment with IOX1 produced a dose-dependent decrease in ^{5hm}C signal (pEC_{50} = 4.65, 5.26, and 5.24 for hTET1–3, respectively), whereas TiP1 did not exhibit cellular activity (Figure 3.9 C–H). Although both inhibitors were cytotoxic at triple-digit micromolar concentrations (cell count below 50% of average, Figure 3.9 B), only TiP1-treated cells exhibited elevated hTET levels as evidenced by an increase in FLAG signals (Figure 3.9 D, F, H). Interestingly, ^{5hm}C signal was elevated for hTET2 and hTET3 at the highest concentration of TiP1 (300 µM), whereas the signal remained constant for hTET1 (Figure 3.9 C, E, G). While it could be suggested that the observed increase in hTET2/3 activity is due to TiP1-mediated allosteric hTET stabilisation, a toxicity-associated rationale is more likely. Furthermore, *in vitro* results established TiP1 as a hTET1-selective inhibitor, making TiP1-mediated selective stabilisation of hTET2/3 improbable. As such, it was concluded that TiP1 did not exhibit cellular activity at non-cytotoxic concentrations.

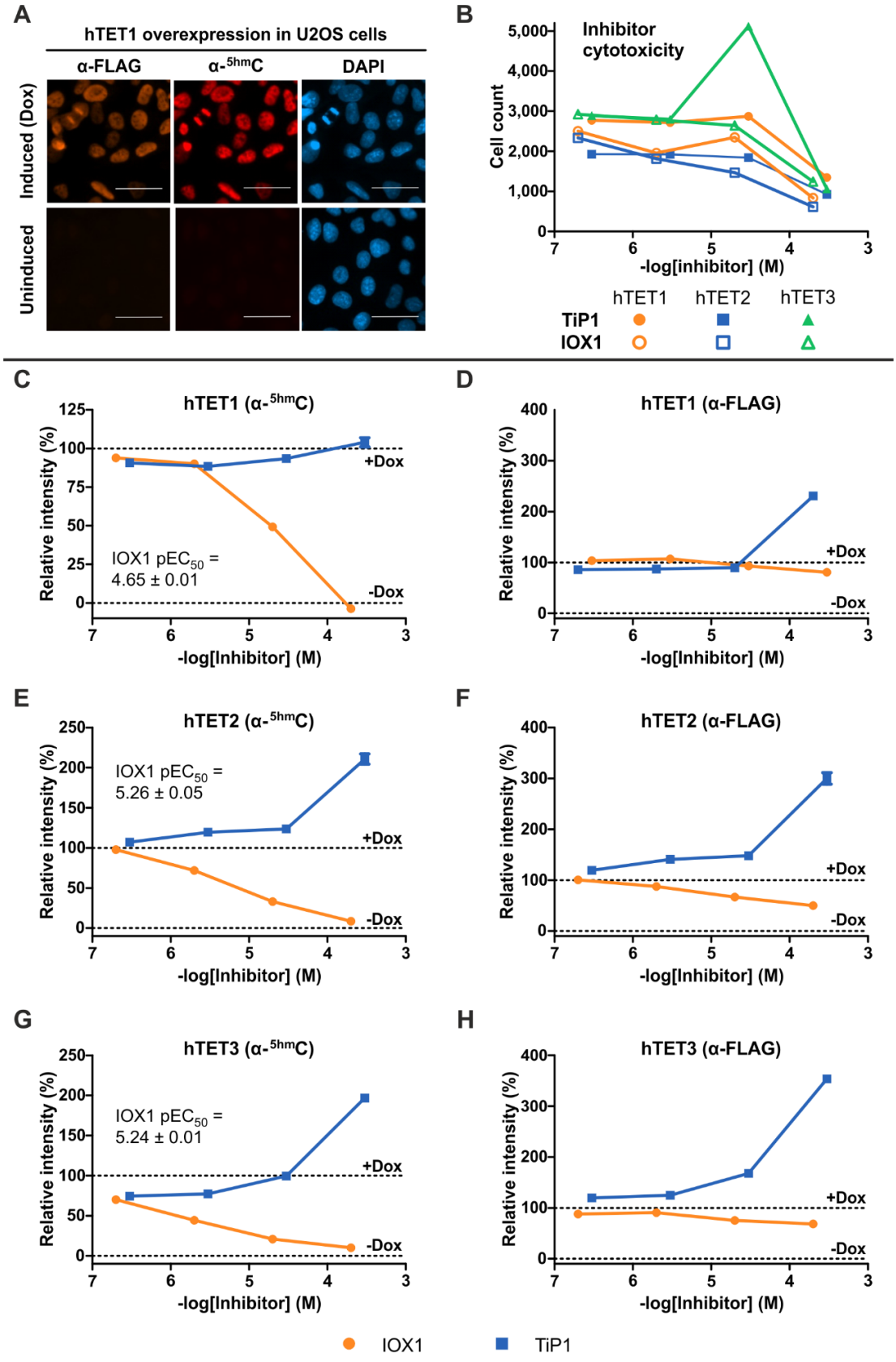


Figure 3.9 | Evaluation of TiP1 cellular activity using dox-inducible U2OS cells stably transfected with FLAG-tagged hTET1–3. A, representative immunofluorescence staining images of U2OS cells overexpressing hTET1, showing an increase in FLAG (orange, hTET1) and ^{5hm}C (red, indicative of TET activity) staining following induction with dox; DAPI nuclear stain shown in blue; scale bars represent 50 µm; **B**, FLAG-positive cell counts for every dose-response curve; **C**, **E**, **G**, anti-^{5hm}C staining dose-response curves for IOX1 and TiP1; **D**, **F**, **H**, anti-FLAG staining dose-response curves for IOX1 and TiP1. Data shown as the average ± SEM (n ≤ 5,000 cells). IOX1 pEC_{50} values (± fitting SD) reported from one biological replicate with 3 technical replicates (n = 3). Experiments were carried out by Dr Hilal Sarac.

3.3.5 Characterisation of TiP1 alanine variants

A TiP1 alanine scan was carried out to identify the key amino acid residues required for binding to hTETs; non-essential residues could subsequently be substituted with hydrophilic (e.g., Asp, Asn, Lys, Arg), α-*N*–methylated, or unnatural amino acids in order to improve the aqueous solubility and cell permeability of the peptide.

Alanine variants for every non-Ala residue within the peptide, excluding the terminal residues (^{D}Tyr and Cys) required for the formation of the thioether linkage, were synthesised by SPPS and isolated by reverse-phase HPLC (11 variants, >99% purity, Appendix 2).

The TiP1 Ala scan library was screened for anti-^{5hm}C antibody-dependent AlphaScreen interference and inhibition of hTET1–3 (Figure 3.10, Table 3.5). The variants did not interact with AlphaScreen assay components, facilitating interference-free characterisation of inhibitory potency; the sole exception to this was the evaluation of hTET2 inhibition by TiP1 Y4A (TiP1 Y4A pIC_{50} = 4.11 and 4.22 for assay interference and hTET2 inhibition, respectively). The majority of alanine variants showed at least a 3-fold ($\Delta pIC_{50} < -0.5$) loss in potency against hTET1 (Figure 3.10 C), with variants Y4A, Y7A, and W10A showing the highest impact on hTET1 potency (ΔpIC_{50} approx. −1.5). hTET1 inhibition results indicated that the sequence of TiP1 was highly optimised for hTET1 binding, concluding that sequence-based optimisation of physiochemical properties would be problematic due to the potency losses associated with the majority of amino acid substitutions. Changes in potency against hTET2/3 were less pronounced across the board compared to hTET1, indicating lower binding specificity for the enzymes (Figure 3.10 D, E), in line with the selectivity profile of TiP1 (Figure 3.10 B). W10A substitution resulted in the most substantial potency loss against hTET1/2 (TiP1 W10A ΔpIC_{50} = −1.63 and <−0.70 for hTET1 and hTET2, respectively), while only a modest decrease in potency was observed against hTET3 (ΔpIC_{50} = −0.46). Conversely, E11A, a substitution with a relatively low impact on hTET1/2 inhibition, resulted in a 20-fold loss in potency against

hTET3 (ΔpIC_{50} = −1.31), demonstrating distinct peptide binding site features within the hTETs. H12A and S13A substitutions resulted in marginal potency changes across the hTET panel (ΔpIC_{50} = −0.10 to 0.24), suggesting that the residues are not involved in protein binding. This observation is not uncommon for cyclic peptides identified using peptide display techniques, as each peptide is 'barcoded' at the C–terminus with a GSGS-linker conjugated to a specific DNA sequence *via* a puromycin linkage, hindering binding to the protein of interest *via* the C–terminal face (Figure 3.1).

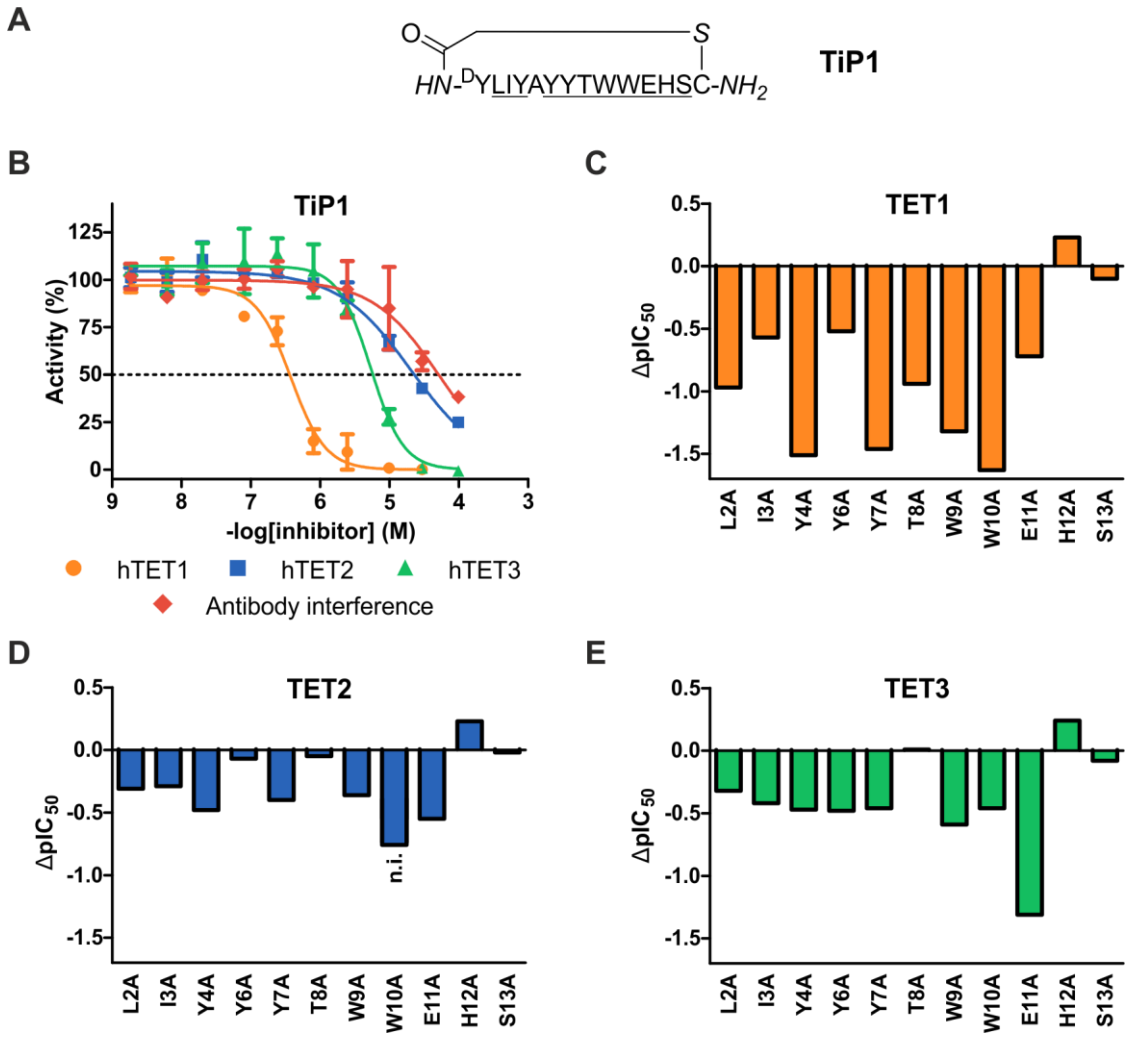


Figure 3.10 | TiP1 variable region alanine variant evaluation against hTET1–3. A, schematic structure of TiP1 (underlined residues indicate variable positions; letters representing individual atoms are italicised to distinguish from single-letter amino acid codes); **B**, representative AlphaScreen dose-response curves for TiP1 antibody interference and hTET1–3 inhibition; **C**, **D**, **E**, alanine substitution impact on potency against hTET1–3, reported as ΔpIC_{50} compared to TiP1. Conditions: TET enzyme, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. In **B**, data shown as the mean of 2–3 experimental replicates (n = 2–3; mean ± SD). N.i., no inhibition (>90% signal at 100 µM).

Table 3.5 | Tabulated AlphaScreen interference and enzymatic inhibition pIC_{50} values for TiP1 alanine mutants against hTET1–3. Changes in potency relative to parent peptide TiP1 are reported as ΔpIC_{50}. Enzyme inhibition data likely obscured by AlphaScreen interference is shaded in grey. Values reported as pIC_{50} ± SD (M). Values reported as percentage signal (%) at 100 μM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 μM). Data for AlphaScreen antibody interference reported as the mean of 2 experimental replicates (n = 2; mean ± SD); data for enzyme inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

Entry	Antibody interference	pIC_{50} ± SD (M) hTET1		hTET2		hTET3	
		pIC_{50} ± SD	ΔpIC_{50}	pIC_{50} ± SD	ΔpIC_{50}	pIC_{50} ± SD	ΔpIC_{50}
IOX1*	n.i.	6.70 ± 0.02	-	6.19 ± 0.10	-	6.00 ± 0.11	-
TiP1*	39.4 ± 3.5%	6.48 ± 0.08	0	4.70 ± 0.01	0	5.28 ± 0.04	0
TiP1 L2A	59.5 ± 1.6%	5.51 ± 0.08	−0.96	4.38 ± 0.09	−0.31	4.96 ± 0.13	−0.32
TiP1 I3A	4.19 ± 0.08	5.91 ± 0.02	−0.57	4.41 ± 0.02	−0.29	4.86 ± 0.10	−0.42
TiP1 Y4A	4.11 ± 0.12	4.97 ± 0.13	−1.51	4.22 ± 0.03	−0.48	4.80 ± 0.04	−0.47
TiP1 Y6A	58.7 ± 3.2%	5.96 ± 0.05	−0.52	4.63 ± 0.02	−0.07	4.79 ± 0.06	−0.48
TiP1 Y7A	63.1 ± 3.9%	5.02 ± 0.17	−1.46	4.30 ± 0.06	−0.40	4.81 ± 0.03	−0.46
TiP1 T8A	4.04 ± 0.14	5.54 ± 0.07	−0.94	4.65 ± 0.12	−0.05	5.28 ± 0.00	0.01
TiP1 W9A	64.5 ± 1.0%	5.16 ± 0.37	−1.32	4.34 ± 0.01	−0.36	4.69 ± 0.13	−0.59
TiP1 W10A	71.5 ± 3.7%	4.86 ± 0.14	−1.63	n.i.	<−0.70	4.82 ± 0.03	−0.46
TiP1 E11A	69.6 ± 0.2%	5.76 ± 0.17	−0.72	4.14 ± 0.03	−0.55	3.97 ± 0.12	−1.31
TiP1 H12A	4.28 ± 0.05	6.71 ± 0.12	0.23	4.93 ± 0.03	0.23	5.51 ± 0.03	0.24
TiP1 S13A	4.10 ± 0.06	6.38 ± 0.03	−0.10	4.68 ± 0.02	−0.02	5.20 ± 0.04	−0.08

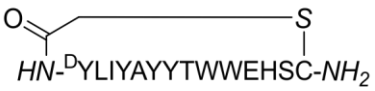
**Data for antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD).*

3.3.6 Synthesis of TiP1-cell-penetrating peptide (CPP) conjugates

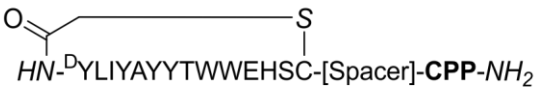
In order to improve the physiochemical properties of TiP1 without altering its sequence and potentially compromising potency and selectivity, TiP1 conjugation with cell-penetrating peptides (CPP) was investigated, as the strategy had been shown previously to enhance the cellular uptake of CPs[231]. The use of C–termini for RNA/DNA barcoding during mRNA display provided precedent for C–terminal functionalisation without loss of parent peptide inhibitory activity, so the C–terminal cysteine was selected as the site of CPP attachment. Conversely, CPP introduction *via* N–terminal elongation would result in macrocycle expansion, as the N–terminal α-amine is involved in the macrocyclisation process. Thus, CPP sequences were appended to the C–terminus of TiP1 along with a short GSGS spacer separating the cyclic peptide from the CPP handles. Three sequences (FLAG, TAT, and R8 tags) were selected for conjugation (Table 3.6). FLAG tag, which is not a CPP, was chosen to assess if improved solubility would translate to shallower Hill slopes in *in vitro* assays and mitigated cytotoxicity

in cellulo[230]; the extensively used HIV Tat protein-derived (TAT) and polyarginine (R8) cationic CPPs[232] were selected due to their relatively short sequences, permitting synthesis of TiP1-CPP conjugates using Fmoc SPPS methods without the need for sequence fragmentation and ligation.

Table 3.6 | Schematic representation of TiP1 and TiP1-CPP conjugate structures, along with selected CPP sequences. Letters representing individual atoms are italicised to distinguish from single-letter amino acid codes; letters representing the cell-penetrating peptide (CPP) are bolded.



TiP1



TiP1–CPP

TiP1-CPP conjugate	Spacer	CPP
TiP1-FLAG	GSGS	DYKDDDDK
TiP1-TAT	GSGS	YGRKKRRQRRR
TiP1-R8	GSGS	RRRRRRRR

The synthesis and purification of TiP1-CPP conjugates was challenging, likely due to the "detergent-like" nature of the peptides imposed by the addition of a hydrophilic tail to a hydrophobic cyclic peptide. Two rounds of reverse-phase HPLC purification (after SPPS synthesis and after TiP1 cyclisation) were required to isolate the desired peptides in reasonable purities (>75%), albeit in very low yields (<1% *vs* 5–10% for non-CPP conjugates).

Preparation of TiP1-FLAG proved to be particularly difficult due to the formation of aspartate-related side products. During piperidine Fmoc deprotection, a neighbouring peptide backbone amidate species can react with the ^{t}Bu ester-protected aspartate side chain, forming aspartimide (Figure 3.11 A); the cyclic imide can then be opened either by H_2O or piperidine, resulting in racemisation and further side product formation (α-/β-peptides/piperidides).

Aspartimide formation rate is controlled by the preceding residue in the peptide sequence: while Tyr or Lys residues supress side product formation, preceding Asp residues show a propensity for the undesired reactivity[233], making the DDDD motif of the FLAG tag

(DYK**DDDD**K) particularly prone to aspartimide formation. SPPS synthesis of TiP1-FLAG using standard coupling (90 °C, 2 min) and Fmoc deprotection (20% piperidine, 90 °C, 1 min) conditions afforded TiP1-FLAG in poor purity and yield (<70% purity, 0.2% yield). The purified peptide contained two types of contaminants: a −18 Da aspartimide-related impurity and a group of +27 Da adducts which could not be identified but were present for both the desired parent peptide and the aspartimide-containing side product (Figure 3.11 B).

Synthesis of an isolated FLAG sequence using standard conditions confirmed that the DDDD motif facilitated rapid aspartimide formation, with peaks corresponding to single (−18 Da) and double (−36 Da) condensation events present in the MALDI-TOF MS spectrum of the crude peptide (Figure 3.11 C). Synthesis of [spacer]-FLAG sequence (GSGS-DYKDDDDK) demonstrated that aspartimide side products accumulated further with subsequent coupling/deprotection cycles (Figure 3.11 D).

A number of approaches have been developed for the suppression of aspartimide formation. The use of sterically bulky aspartate side chain β-trialkylmethyl esters such as β-3-methylpent-3-yl (Mpe)[234] or 5-*n*-butyl-5-nonyl (Bno)[235] is a popular strategy, but the bulky nature of the monomers would likely lead to poor coupling efficiencies for consecutive Asp residues during SPPS. Alternatively, the addition of organic acids, such as hydroxybenzotriazole (HOBt), to the piperidine-based Fmoc deprotection mixture has been reported to minimise aspartimide formation[236]. Oxyma, a safer alternative to HOBt[237], was added to the deprotection mixture at a final concentration of 100 mM during synthesis of GSGS-FLAG; however, Oxyma “doping” resulted in complete loss of coupling capacity, and no product peaks were observed in the MALDI-TOF MS spectrum of the crude peptide (data not shown). Finally, reduction of reaction temperature has been shown to minimise racemisation and side product formation rates in SPPS[238]. Synthesis of GSGS-FLAG using room temperature coupling (25 °C, 60 min) and

deprotection (20% piperidine, 25 °C, 10 min) conditions completely suppressed aspartimide formation (Figure 3.11 E); however, the final Fmoc deprotection efficiency was markedly poor (approx. 80% Fmoc-protected peptide present), and the synthesis time was greatly extended (approx. 1.5 h per cycle *vs* <15 min per cycle using standard conditions). As formation of aspartimide is piperidine-mediated, a combination of microwave-assisted coupling and room temperature deprotection was explored (coupling: 90 °C, 2 min; deprotection: 20% piperidine, 25 °C, 10 min; final deprotection: 20% piperidine, 25 °C, 20 min). Furthermore, as coupling efficiency is known to correlate inversely with peptide length, double couplings were introduced from residue 17 onwards. Synthesis of TiP1-FLAG using the modified conditions furnished the desired peptide in good purity (90.3%, Figure 3.11 F, G) and slightly improved yield (up from 0.2% to 0.4%).

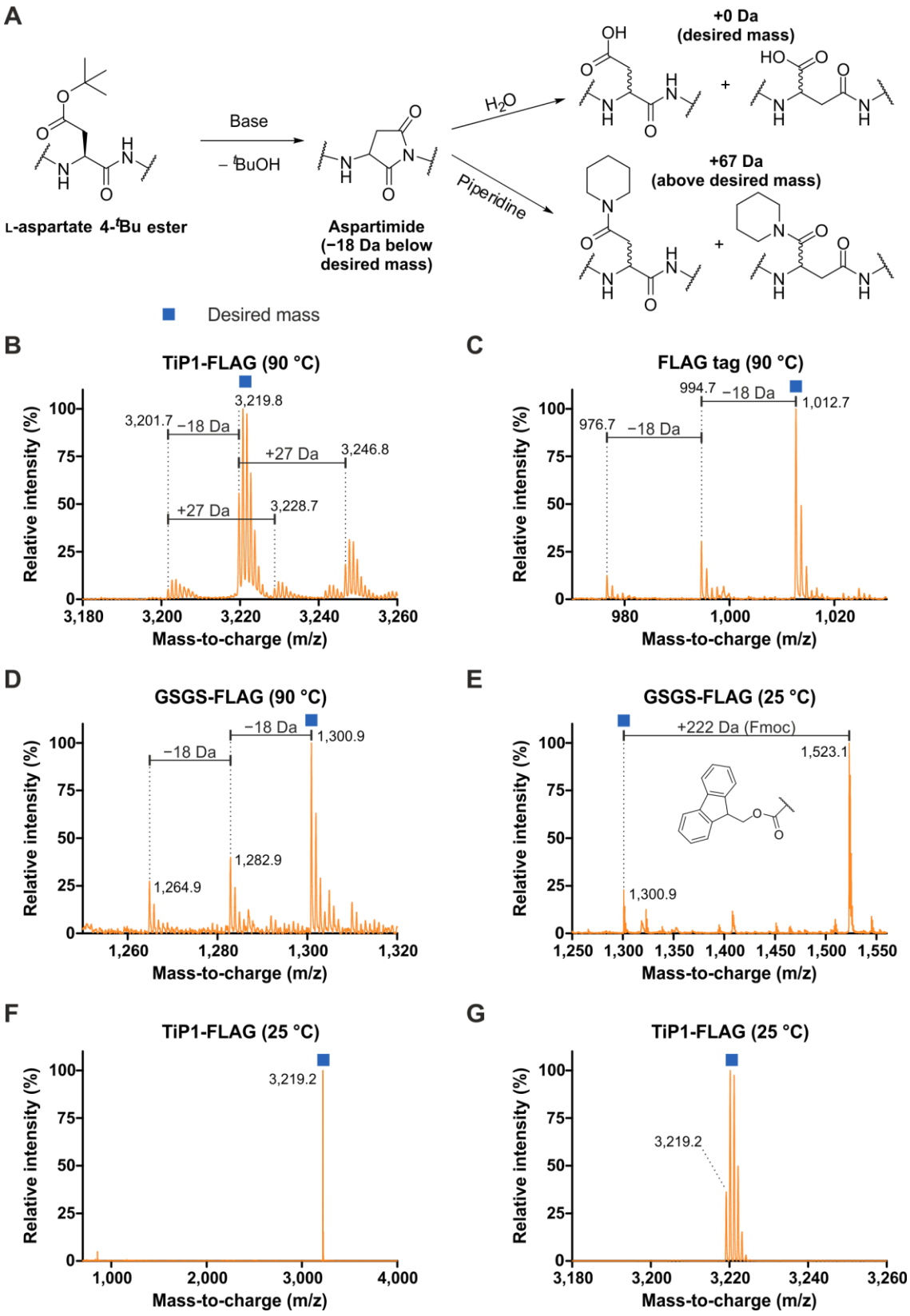


Figure 3.11 | Optimisation of SPPS conditions to suppress formation of aspartimide side products. A, aspartimide formation and succinimide ring opening scheme; **B**, MALDI-TOF spectrum of double HPLC-purified TiP1-FLAG synthesised using standard coupling (90 °C, 2 min) and deprotection conditions (90 °C, 1 min); **C**, MALDI-TOF spectrum of crude FLAG tag synthesised using standard coupling (90 °C, 2 min) and deprotection conditions (90 °C, 1 min); **D**, MALDI-TOF spectrum of crude GSGS-FLAG synthesised using standard coupling (90 °C, 2 min) and deprotection conditions (90 °C, 1 min); **E**, MALDI-TOF spectrum of crude GSGS-FLAG synthesised using room temperature coupling (25 °C, 60 min) and deprotection (25 °C, 10 min) conditions; **F**, **G**, MALDI-TOF spectrum of double HPLC-purified TiP1-FLAG synthesised using standard coupling (90 °C, 2 min) and room temperature deprotection (25 °C, 10 min) conditions, with double couplings from residue 17 onwards.

3.3.7 Evaluation of TiP1-CPP conjugate inhibitory activity using AlphaScreen

With the TiP1-CPP conjugates in hand, TiP1-FLAG and TiP1-TAT were tested against hTET1–3 using AlphaScreen; TiP1-R8 could not be tested *in vitro* due to the low amount of isolated peptide. FLAG, TAT, and R8 peptides were characterised as well to monitor for potential tag-mediated hTET inhibition.

TiP1-FLAG did not quench AlphaScreen signal and inhibited hTETs with potencies comparable to TiP1 (hTET1 pIC_{50} = 6.72 and 6.39 for TiP1-FLAG and TiP1, respectively), while the FLAG peptide did not affect enzyme activity (Figure 3.12, Table 3.7). Hill coefficients remained largely unchanged between TiP1 and TiP1-FLAG (representative hTET1 inhibition n_H = 1.04 and 1.26, respectively), although additional dose-response data points are required to accurately determine the n_H values. TiP1-TAT showed improved potency across the board (hTET1 pIC_{50} = 7.74), while the TAT peptide inhibited hTET1/2 with single-digit micromolar potencies (pIC_{50} = 5.94 and 5.39 for hTET1 and hTET2, respectively), suggesting that the observed potency enhancement was likely a consequence of increased avidity between a bivalent inhibitor and the enzyme[239,240]. Interestingly, although TiP1 and TAT CPP peptides did not interfere with AlphaScreen signal separately, the TiP1-TAT conjugate was a potent assay signal quencher (pIC_{50} = 6.08). R8 CPP was shown to inhibit hTETs more potently than TAT CPP across the board (hTET1 pIC_{50} = 6.60 and 5.94 for R8 and TAT CPPs, respectively), suggesting that TiP1-R8 may behave in a similar fashion to TiP1-TAT. Importantly, the addition of C–terminal tails to cyclic peptide TiP1 did not abrogate its inhibitory activity towards hTETs, supporting subsequent TiP1-CPP conjugate characterisation using cellular assays.

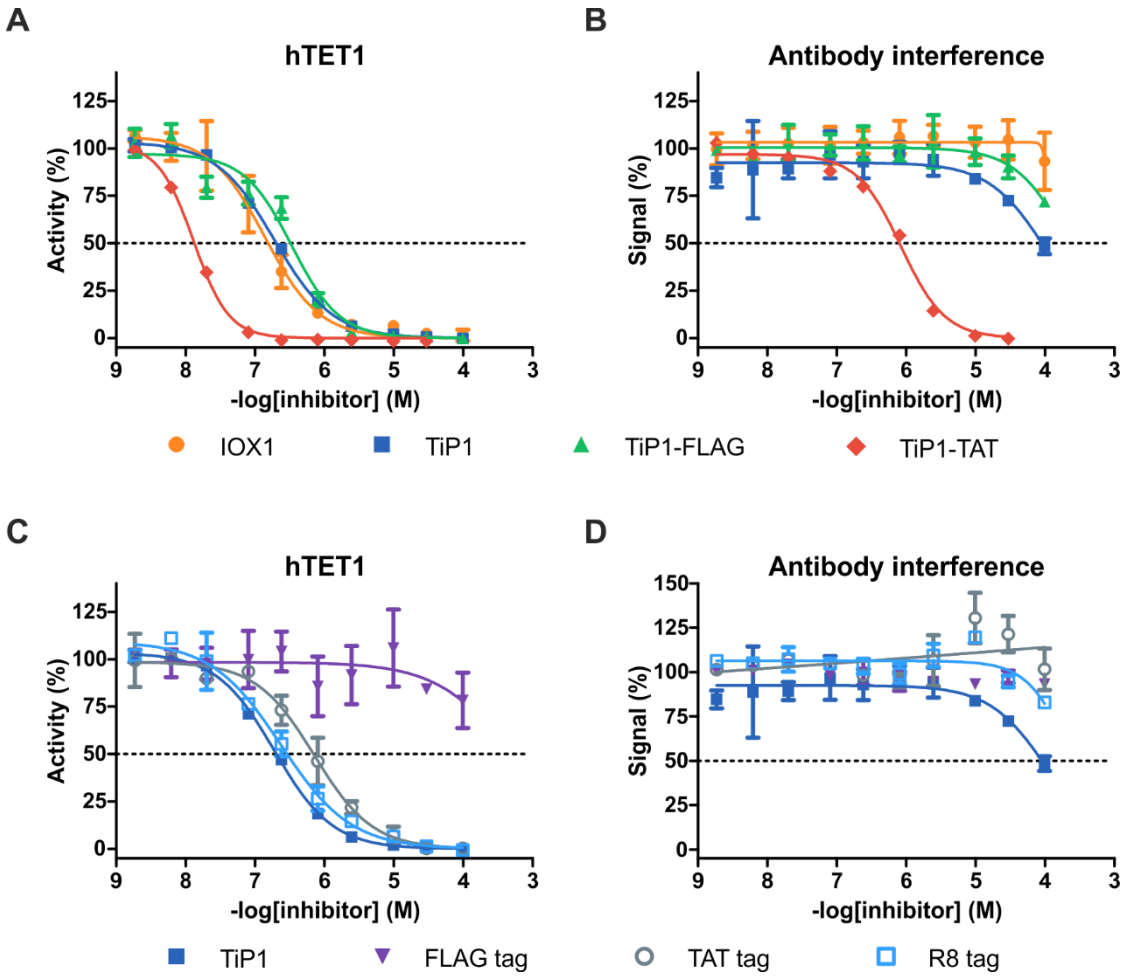


Figure 3.12 | Representative AlphaScreen hTET1 inhibition and assay interference dose-response curves for TiP1, TiP1-CPP conjugates, and CPPs. A, hTET1 inhibition by TiP1 and TiP1-CPP conjugates; **B**, antibody interference by TiP1 and TiP1-CPP conjugates; **C**, hTET1 inhibition by CPPs; **D**, antibody interference by CPPs. Conditions: 10 nM hTET1, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. Representative data shown as the mean of 2–3 experimental replicates (n = 2–3; mean ± SD).

Table 3.7 | Tabulated AlphaScreen interference and hTET1–3 inhibition pIC_{50} values for TiP1-CPP conjugates and CPPs. Enzyme inhibition data likely obscured by AlphaScreen interference is shaded in grey. Values reported as pIC_{50} ± SD (M). Values reported as percentage signal (%) at 100 µM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 µM); n.d., not determined. Data for AlphaScreen antibody interference reported as the mean of 2 experimental replicates (n = 2; mean ± SD); data for enzyme inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

Entry	Antibody interference	pIC_{50} ± SD (M) Enzyme inhibition		
		hTET1	hTET2	hTET3
IOX1*	n.i.	6.72 ± 0.66	6.23 ± 0.09	5.94 ± 0.13
TiP1*	48.4 ± 3.4%	6.70 ± 0.05	4.84 ± 0.48	5.18 ± 0.05
TiP1-FLAG	71.9 ± 1.4%	6.39 ± 0.06	4.59 ± 0.08	4.63 ± 0.05
TiP1-TAT	6.08 ± 0.03	7.74 ± 0.21	6.91 ± 0.00	6.74 ± 0.05
TiP1-R8	n.d.	n.d.	n.d.	n.d.
FLAG	n.i.	80.0 ± 2.3%	78.5 ± 2.3%	63.2 ± 11.1%
TAT	n.i.	5.94 ± 0.20	5.39 ± 0.25	81.1 ± 4.3%
R8	83.0 ± 2.0%	6.60 ± 0.02	6.45 ± 0.59	73.7 ± 5.8%

**Data for antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD).*

3.3.8 Evaluation of TiP1-CPP conjugate cellular activity

The cellular activity of TiP1-FLAG, TiP1-TAT, and TiP1-R8 conjugates was evaluated using the cell-based assay described in section 3.3.4. Although TiP1-FLAG failed to produce a dose-dependent decrease in ^{5hm}C signal, changes to hTET levels (α-FLAG signal) and cell counts were minimal at the highest peptide concentration (200 μM), unlike in the case of TiP1, suggesting that the FLAG conjugate may be less cytotoxic compared to parent peptide TiP1 (Figure 3.13). TAT and R8 conjugates of TiP1 reduced ^{5hm}C signal and increased FLAG signal in cells overexpressing hTET1 at the highest concentrations tested (100 and 30 μM, respectively, Figure 3.13 A, B). Conversely, treatment of cells overexpressing hTET2/3 with TiP1-TAT resulted in ^{5hm}C signal elevation, concomitant with an increase in FLAG signal for hTET2 (Figure 3.13 C–F). The aberrant levels of ^{5hm}C and FLAG are likely due to cytotoxicity at the highest concentrations of TAT and R8 conjugates, as evidenced by substantially reduced cell counts (<30% of average cell count, Figure 3.13 G). Overall, TiP1 conjugation with the selected peptide sequences (FLAG, TAT, and R8) did not improve cellular activity. However, conjugation of TiP1 with the anionic FLAG sequence seemed to partially ameliorate the cytotoxic effects of TiP1.

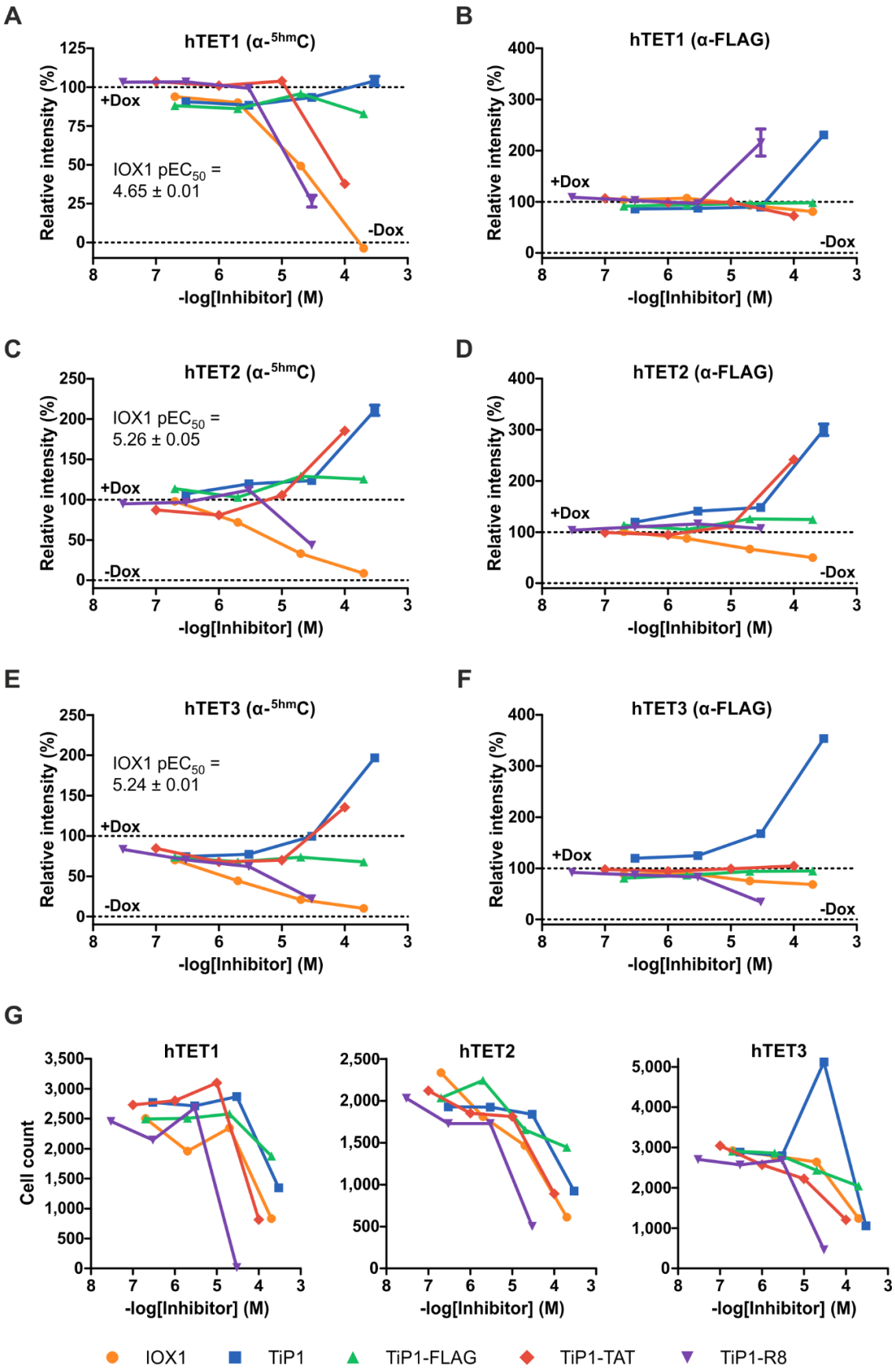


Figure 3.13 | Evaluation of TiP1-CPP conjugate cellular activity using dox-inducible U2OS cells stably transfected with FLAG-tagged hTET1–3. A, **C**, **E**, anti-^{5hm}C staining dose-response curves for IOX1 and TiP1-CPP conjugates; **B**, **D**, **F**, anti-FLAG staining dose-response curves for IOX1 and TiP1-CPP conjugates; **G**, FLAG-positive cell counts for every dose-response curve. Data shown as the average ± SEM (n ≤ 5,000 cells). IOX1 pEC_{50} values (± fitting SD) reported from one biological replicate with 3 technical replicates (n = 3). Experiments were carried out by Dr Hilal Sarac.

3.3.9 Investigation of TiP1 binding site

Although TiP1 and its CPP conjugates did not exhibit cellular activity, elucidation of TiP1 binding site for TETs was desired in order gain a better understanding of its mode of action. A deeper mechanistic insight could expedite structure-guided modifications to improve cellular activity (as shown for CP2, a cyclic peptide inhibitor of KDM4A[241]) and support investigations of TiP1 derivatives as affinity tools for TET interactome mapping studies. As attempts to crystallographically determine the interactions between TiP1 and hTET2 were unsuccessful (section 5.3.2), solution-based methods were considered.

To explore the boundaries of the binding interface between TiP1 and hTETs, a hTET1/2 tiling library was prepared by utilising the spot method on cellulose support (SPOT array). SPOT tiling libraries are constructed by splitting a protein sequence into a staggered library of short peptides (typically 10–15-mers[242]), which are immobilised at a high density on cellulose support (approx. 8 spots/cm^2 containing ~20 nmoles of peptide). Incubation of the protein tiling library with a probe of interest, followed by antibody-based signal generation, can enable high-throughput identification of the binding interface between a ligand and a protein.

A cellulose membrane was functionalised with a β-alanine handle for subsequent peptide couplings, and hTET1$_{CD}$ and hTET2$_{CD\Delta LCI}$ sequence tiling libraries consisting of 12-mer peptides staggered by two amino acids (residues 1–12, 3–14, etc.) were synthesised using reported Fmoc SPPS conditions[242] (Figure 3.14). The immobilised library was probed with TiP1-FLAG (5 μM) and the membrane was developed using primary anti-FLAG and HRP-conjugated secondary antibodies (Figure 3.14).

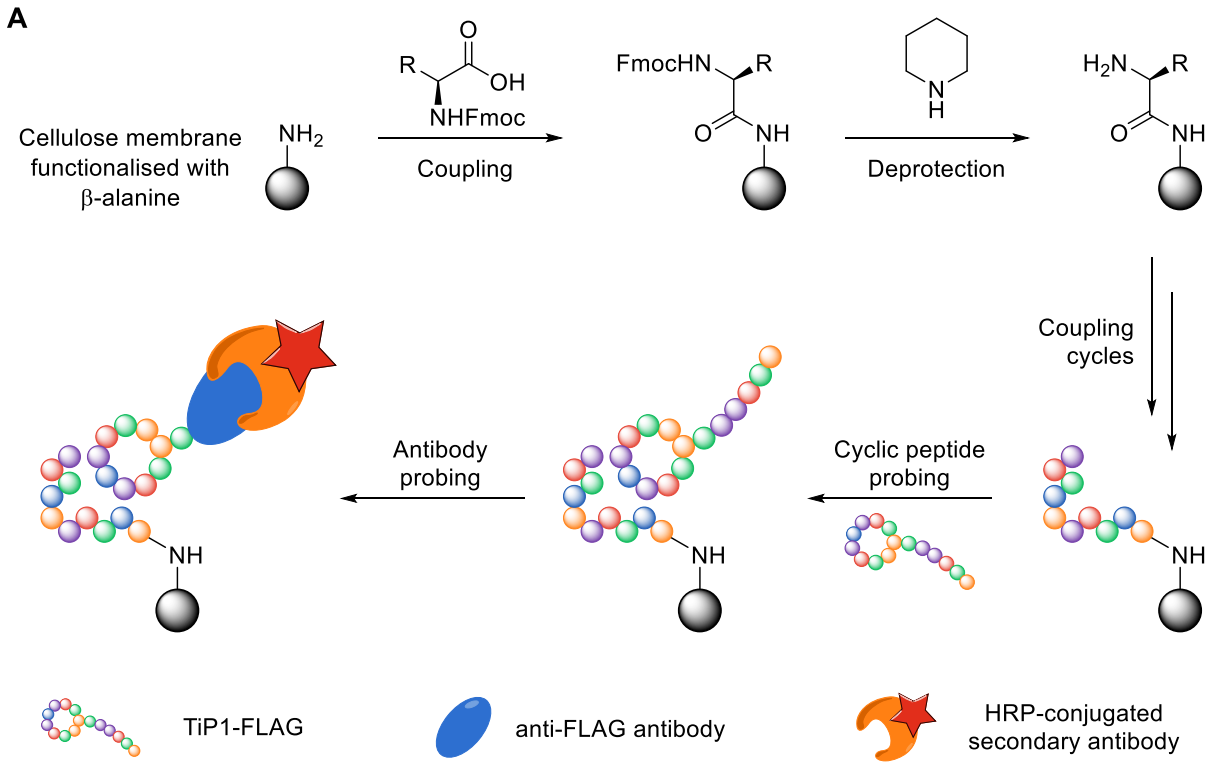


Figure 3.14 | SPOT peptide array preparation and probing. Schematic representation of SPOT peptide array synthesis on a cellulose membrane, followed by consecutive probing with a FLAG epitope-tagged TiP1, primary anti-FLAG antibody, and HRP-conjugated secondary antibody.

FLAG epitope positive controls produced a strong chemiluminescence signal, while scrambled FLAG and $(GSGS)_3$ sequences used as negative controls did not, demonstrating FLAG-dependent signal generation (Figure 3.15 A). TET tiling library signal intensities were analysed, and interaction heat maps based on rolling intensities for every two residues were created (Figure 3.15 B). Regions of interest could be identified for both hTET1 and hTET2 tiling libraries; however, the heat maps obtained from a probe-free control experiment were identical to results obtained following probing with TiP1-FLAG (Figure 3.15 B).

The observed antibody cross-reactivity was surprising, as antibody recognition is generally highly sequence- or tertiary structure-specific. Analysis of the highest intensity regions for hTET1 and hTET2 (Figure 3.15 C) revealed high sequence similarity, but no apparent similarities to the epitope sequence (DYKDDDDK and LPLYKLSDTDEF; FLAG epitope and hTET1 hot spot motifs, respectively). After excluding the hot spots arising from non-specific antibody binding, no additional hot spots could be observed after probing with TiP1-FLAG compared to

the probe-free control, suggesting that TiP1 binding events to TETs are mediated by tertiary structure interactions, and not sequence patch recognition.

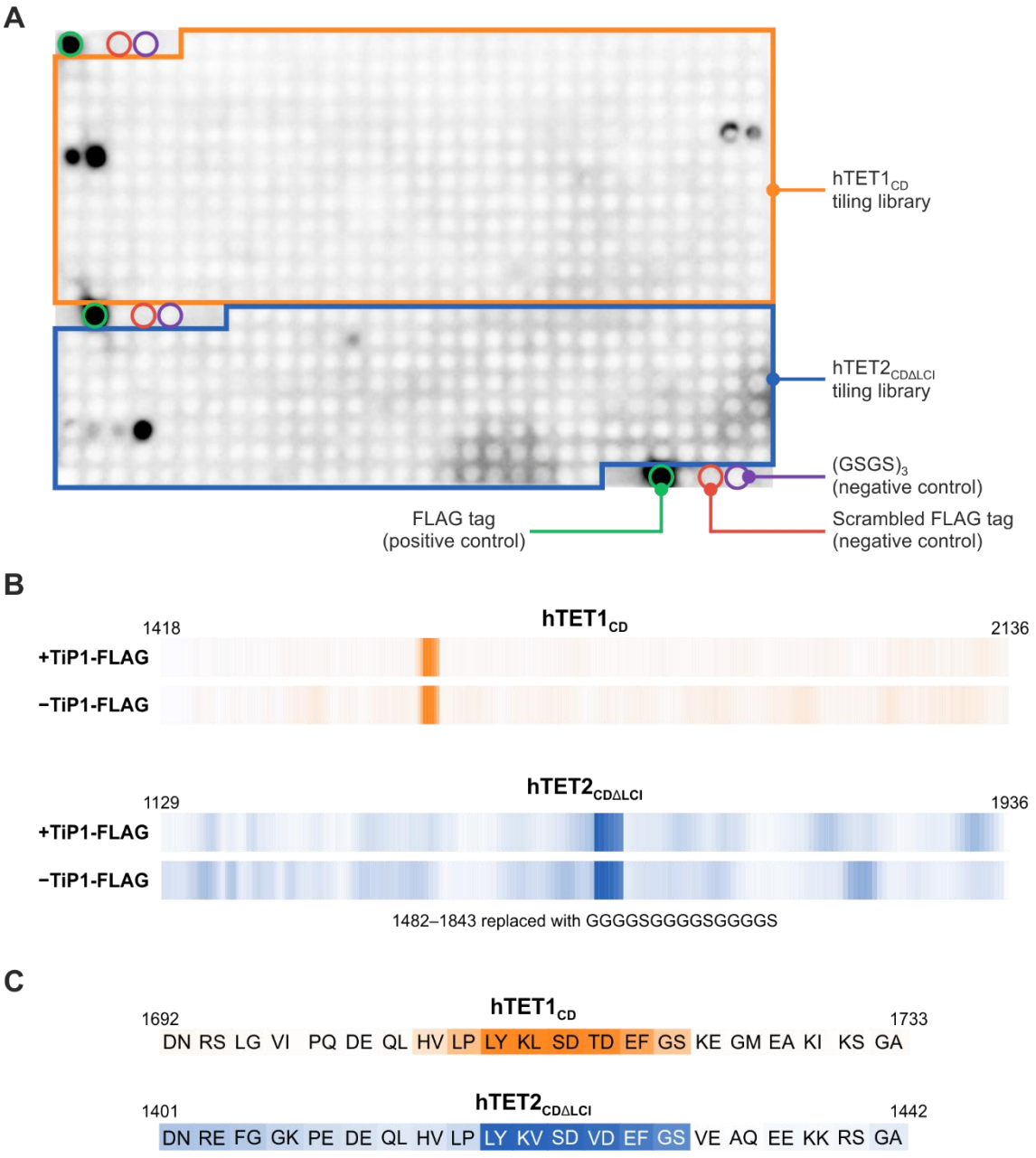


Figure 3.15 | Analysis of hTET1/2 tiling library probing using TiP1-FLAG. A, representative chemiluminescence image of a membrane containing hTET1$_{CD}$ and hTET2$_{CD\Delta LCI}$ tiling libraries probed with TiP1-FLAG and appropriate antibodies; **B**, heat maps for hTET1$_{CD}$ and hTET2$_{CD\Delta LCI}$ probing with and without TiP1-FLAG; **C**, sequence alignment for highest intensity regions of hTET1$_{CD}$ and hTET2$_{CD\Delta LCI}$ (values from probe-free control). FLAG tag was used as the positive control; scrambled FLAG tag and (GSGS)$_3$ spacer sequences were used as negative controls.

In order to investigate the interactions between TiP1 and intact catalytically active TET enzymes, a photo-crosslinking approach employing a photoreactive TiP1 derivative was explored. TiP1 alanine scanning studies (section 3.3.5) showed tolerance for substitutions at the Tyr6 position; furthermore, as Tyr6 is flanked by residues Tyr4 and Tyr7, which were shown

to be essential for hTET1 inhibition, a photoreactive group at the 6 position would likely come into close proximity to the protein surface, facilitating cross-linking. 4-Benzoyl-L-phenylalanine (Bpa), a benzophenone photophore-containing photoreactive amino acid, has been used extensively in protein-protein[243] and peptide ligand-protein[244] photocross-linking, and was thus selected as the photoreactive moiety to furnish the photoreactive variant TiP1 Y6Bpa (Figure 3.16 A). Incubation of TiP1 Y6Bpa with TET enzymes, followed by irradiation with 350–365 nm wavelength light would produce a benzophenone-derived active diradical species which reacts preferentially with unreactive C–H bonds in the presence of solvent water[245], efficiently crosslinking to the protein surface; protease digestion of the conjugated species, followed by MS/MS analysis would facilitate the identification of the crosslinking region on the protein surface (Figure 3.16 B).

Introduction of the benzophenone functionality resulted in minimal changes in the potency against hTETs as determined using AlphaScreen (ΔpIC_{50} = −0.30, 0.22, and 0.36 for hTET1–3 compared to TiP1, respectively), supporting the use of the Y6Bpa variant in subsequent photocross-linking studies (Figure 3.16 C, D).

NgTet1 was selected as a model protein for the optimisation of photocross-linking conditions, as sourcing of recombinant $hTET1_{CD}$ was challenging. Furthermore, the relatively small size of NgTet1 (38.7 kDa) enabled direct protein detection using MALDI-TOF MS, substantially reducing TiP1 Y6Bpa probe consumption. NgTet1 was shown to be stable to 365 nm irradiation (Figure 3.16 E). Incubation of NgTet1 with TiP1 Y6Bpa (50 µM) followed by UV irradiation resulted in the appearance of a new mass peak tentatively assigned as the NgTet1-TiP1 Y6Bpa cross-linking product (observed mass shift: approx. 1,900 Da; predicted mass shift: 2,023.9 Da). A peak potentially corresponding to a second cross-linking event was also present, but further experiments are required to confirm this observation. Together, these

data suggest that the photo-crosslinking conditions are in principle suitable for studies using hTETs.

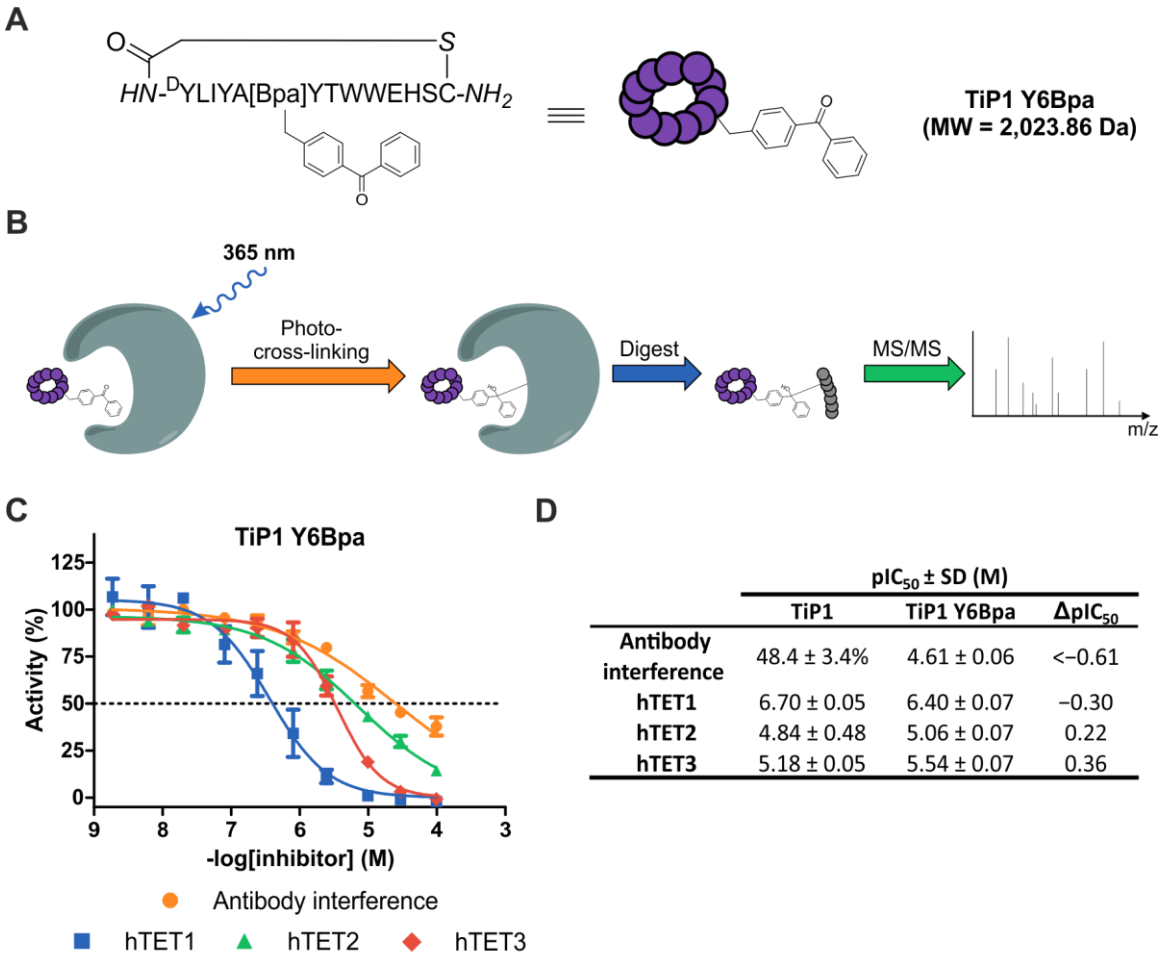


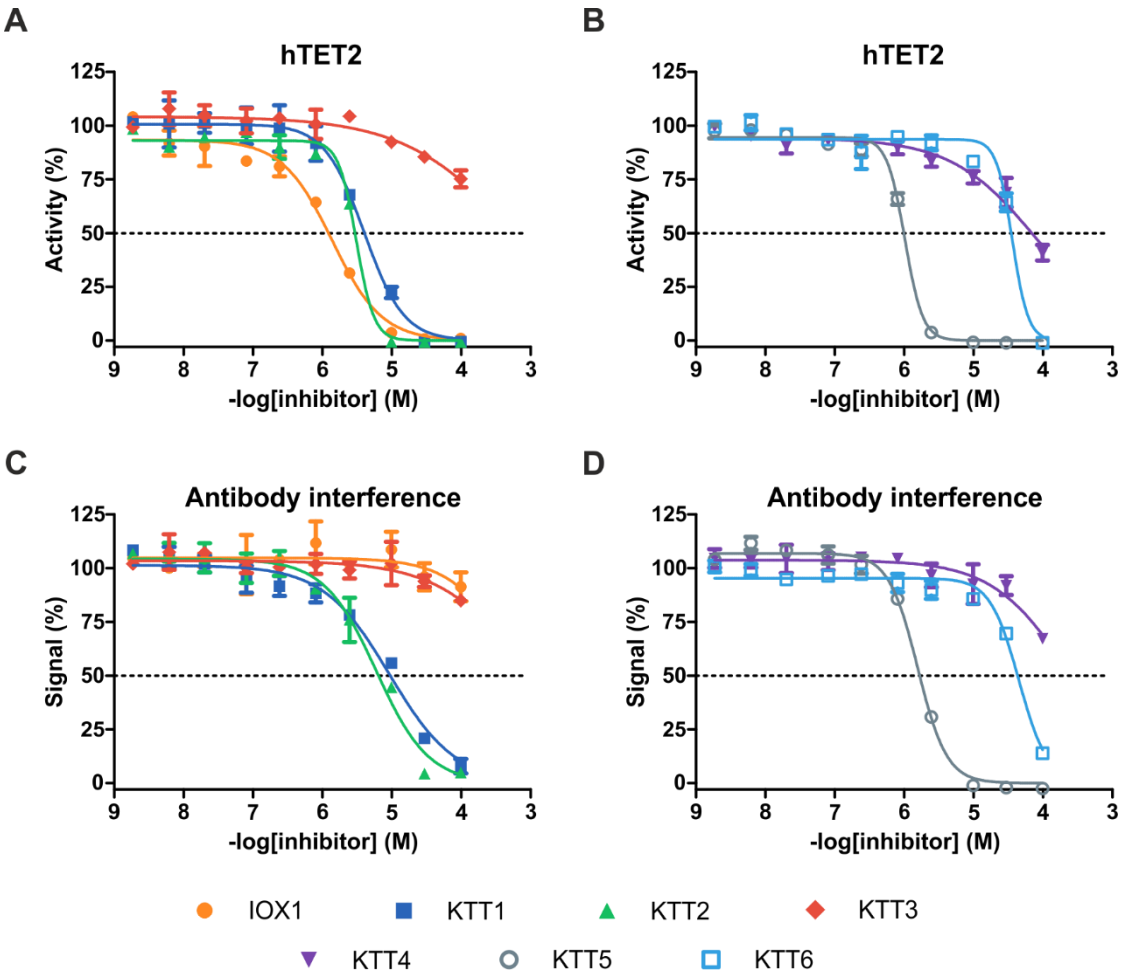
Figure 3.16 | Use of the photoreactive TiP1 derivative, TiP1 Y6Bpa, to photocross-link with model enzyme NgTet1. A, schematic structure of TiP1 Y6Bpa, highlighting the benzophenone photophore (letters representing individual atoms are italicised in the left representation of TiP1 to distinguish from single-letter amino acid codes); **B**, photocross-linking-based epitope mapping workflow; **C**, representative AlphaScreen dose-response curves for TiP1 Y6Bpa; **D,** tabulated AlphaScreen interference and hTET1–3 enzyme inhibition pIC_{50} values for TiP1 Y6Bpa; **E,** MALDI-TOF MS spectrum of NgTet1 after 365 nm irradiation for 20 min; **F**, MALDI-TOF MS spectrum of NgTet1 and TiP1 Y6Bpa after 365 nm irradiation for 20 min. AlphaScreen conditions: TET enzyme, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. In **C**, data shown as the mean of 2 experimental replicates (n = 2; mean ± SD). In **D**, data for AlphaScreen antibody interference reported as the mean of 2 experimental replicates (n = 2; mean ± SD); data for enzyme inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

3.3.10 Characterisation of cyclic peptides from RaPID selection against hTET2

A RaPID selection was carried out against hTET2 in an attempt to identify complementary hTET family CP inhibitors. The hTET2 catalytic domain with an insertion-deletion, $hTET2_{CD\Delta LCI}$, fused to an N–terminal His_6-FLAG-[thrombin cleavage site]-SUMO tag (Chapter 5) was used as the target protein in the selection (experiments were carried out by Dr Tom McAllister). Peptide sequences obtained from the selection were binned based on similarity; representative sequences from the six most enriched families (KTT1–6) were synthesised by SPPS and evaluated for hTET2 inhibition using AlphaScreen.

All six KTT peptides interfered with AlphaScreen signal generation with potencies largely comparable to the hTET2 inhibition values ($\Delta pIC_{50} < 0.35$), precluding unambiguous evaluation of inhibitory activity against hTET2 (Figure 3.17). While KTT3/4 did not exhibit strong AlphaScreen interference, the peptides were also only weakly inhibitory towards hTET2 (73.6% and 40.7% hTET2 activity at 100 µM inhibitor for KTT3 and KTT4, respectively).

KTT1–6 were orthogonally evaluated using the U2OS cell-based TET assay (Figure 3.18). Cell treatment with KTT1/5/6 did not produce changes in ^{5hm}C levels, while aberrant levels of ^{5hm}C signal were observed in cells treated with KTT3/4 at 200 µM, likely due to a cytotoxicity-related reduction in cell count (down to 10% of average cell count, Figure 3.18 G). KTT2 partially reduced ^{5hm}C content in cells overexpressing hTET2/3 at 200 µM (approx. 50% reduction) without affecting cell viability; however, additional *in vitro* characterisation of KTT2 using SPE-MS is required to determine if the effect is due to inhibition of hTETs.



E

	pIC_{50} ± SD (M)	
	Antibody interference	**hTET2 inhibition**
IOX1	80.3 ± 11.1%	6.41 ± 0.68
KTT1	5.10 ± 0.06	5.41 ± 0.04
KTT2	5.27 ± 0.04	5.53 ± 0.02
KTT3	76.9 ± 8.0%	73.6 ± 1.3%
KTT4	50.2 ± 17.1%	40.7 ± 2.3%
KTT5	5.79 ± 0.01	6.03 ± 0.09
KTT6	4.26 ± 0.08	4.45 ± 0.01

Figure 3.17 | Representative AlphaScreen dose-response curves for KTT1–6 AlphaScreen interference and hTET2 inhibition. A, **B**, hTET2 inhibition; **C**, **D**, AlphaScreen antibody interference; **E**, tabulated pIC_{50} data. Conditions: 1 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. In **E**, enzyme inhibition data likely obscured by AlphaScreen interference is shaded in grey; values reported as pIC50 ± SD (M); values reported as percentage signal (%) at 100 µM for incomplete dose-response curves. Data for antibody interference reported as the mean of 2 independent replicates (n = 2; mean ± SD); data for enzyme inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

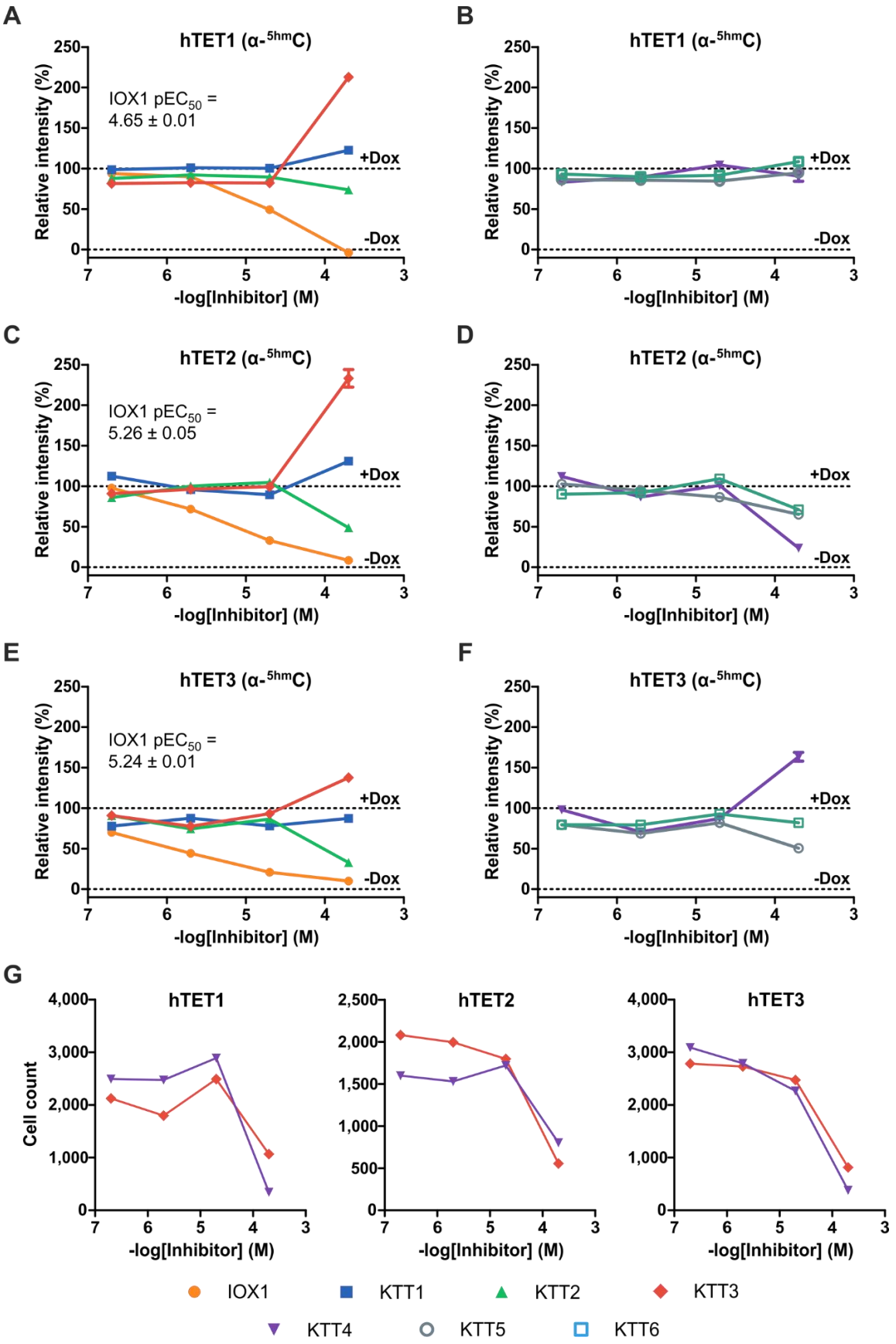


Figure 3.18 | Evaluation of KTT1–6 cellular activity using dox-inducible U2OS cells stably transfected with FLAG-tagged hTET1–3. A, **B**, anti-^{5hm}C staining dose-response curves for hTET1; **C**, **D**, anti-^{5hm}C staining dose-response curves for hTET2; **E**, **F**, anti-^{5hm}C staining dose-response curves for hTET3; **G**, FLAG-positive cell counts for KTT3 and KTT4 dose-response curves. Data shown as the average ± SEM (n ≤ 5,000 cells). IOX1 pEC_{50} values (± fitting SD) reported from one biological replicate with three technical replicates (n = 3). Experiments were carried out by Dr Hilal Sarac.

3.4 Discussion

Evaluation of reported cyclic peptide inhibitors TiP1–3 against a panel of TET enzymes reproduced the reported[191] TiP1 selectivity for hTET1. Furthermore, TiP1 exhibited excellent (>100-fold) selectivity for hTET1 over 2-OG dioxygenases KDM5B/6B. TiP1–3 interfered marginally with the antibody-independent AlphaScreen TruHits assay, but TiP2/3 were shown to be potent anti-^{5hm}C antibody-dependent AlphaScreen signal quenchers (pIC_{50} = 5.45 and 5.48), precluding selectivity profiling. The combined observations suggest that TiP2/3 may bind at the interface between protein A immobilised on acceptor beads and IgG-based antibodies. The apparent non-specific AlphaScreen interference may be partially due to peptide sequence enrichment for hydrophobic residues, as high lipophilicity is commonly associated with non-specific off-target interactions[246,247].

A number of additional CPs (KS01–13) from the RaPID selection used to identify TiP1–3 were selected for characterisation against the TET enzyme panel. Unfortunately, the majority of the novel peptides (10/13) were potent AlphaScreen signal quenchers, obscuring enzyme inhibition results. Out of the peptides not interfering with AlphaScreen signal, KS03 showed poor selectivity across the board (pIC_{50} = 4.35, < 4, and 4.50 for hTET1–3, respectively), KS09 was less potent than TiP1 (hTET1 pIC_{50} = 5.73 and 6.70 for KS09 and TiP1, respectively), and the inhibitory profile of KS08 was similar to that of TiP1.

An orthogonal SPE-MS-based assay was used to support results obtained using AlphaScreen by screening the peptide libraries against hTET2. The majority of the tested peptides inhibited hTET2 with single-digit micromolar potencies, in line with results obtained using AlphaScreen, but Hill coefficients (n_H) were uncharacteristically high (average n_H = 1.98 and 3.96 for AlphaScreen and SPE-MS, respectively). While n_H values >1 can indicate positive cooperative binding (ligand binding events increase the affinity for subsequent ligand binding), high Hill

coefficients are more often associated with colloidal aggregation of the inhibitor followed by enzyme sequestration[218], and such compounds are generally avoided in drug discovery efforts[248]. An overview of peptide library sequences highlighted a considerable enrichment in amino acids containing hydrophobic side chains (particularly aromatics Tyr and Trp, Figure 3.6), suggesting that low peptide solubility limits in aqueous environments may contribute to the apparent aggregate-based inhibition. Aggregation-prone inhibitors are often sensitive to detergents, with the addition of detergent resulting in substantial reduction of inhibitory potency[217]; in accordance with this observation, peptides TiP1, KS08, and KS09 exhibited lower potencies under AlphaScreen assay conditions containing detergent (50 mM HEPES, 150 mM NaCl, 0.1% v/v BSA, 0.01% v/v Tween® 20, pH 7.1) when compared to detergent-free SPE-MS assay conditions (50 mM HEPES, pH 7.3).

Subsequent evaluation of TiP1 cellular activity in dox-inducible U2OS epithelial cells overexpressing hTET1–3 revealed that TiP1 did not induce a dose-dependent reduction of ^{5hm}C levels and was cytotoxic at triple-digit micromolar concentrations, in line with the previously discussed relation between drug lipophilicity and off-target effects[246,247].

In an effort to identify non-essential residues amenable to substitution, an alanine scan was carried out on TiP1. Y4A, Y7A, and W10A substitutions resulted in the most substantial losses in hTET1 potency (ΔpIC_{50} approx. −1.5 compared to parent peptide TiP1), illustrating the importance of residues containing aromatic side chains, which may mimic the π−π stacking interactions between enzyme active site residues and nitrogenous bases of DNA[94]. While Trp10 was identified as the most important residue for hTET1/2 inhibition (W10A $\Delta pIC_{50} < -1.5$), loss in potency against hTET3 was relatively moderate (W10A $\Delta pIC_{50} < -0.46$). Conversely, E11A substitution exhibited the highest impact on hTET3 inhibition

(ΔpIC_{50} = −1.31), highlighting the distinctiveness of peptide binding site features within the hTET family.

The sequence of TiP1 did not prove amenable to substitutions owing to its high optimisation towards hTET1 binding, precluding physiochemical property modulation; thus, TiP1 conjugates with short property-modifying peptides were produced. Synthesis of TiP1-FLAG proved challenging due to undesired piperidine-dependent aspartimide formation, which was suppressed by carrying out the piperidine-containing Fmoc deprotection steps at room temperature. Inhibitory potencies of TiP1-CPP conjugates were evaluated against hTETs using AlphaScreen and U2OS cell-based assays. FLAG conjugation did not modulate potency *in vitro* or *in cellulo*, but resulted in reduced cytotoxicity compared to TiP1. Conversely, TiP1-TAT inhibitory potency was substantially elevated *in vitro* across the board compared to TiP1 (ΔpIC_{50} = 1.04, 2.07, and 1.56 for hTET1–3, respectively). TAT CPP was shown to be a potent inhibitor of hTET1–3, indicating a potential increase in enzyme avidity for the bivalent inhibitor TiP1-TAT[239,240]. Regardless, TiP1-TAT and TiP1-R8 did not elicit a dose-dependent reduction in ^{5hm}C levels *in cellulo*, and exhibited substantial cytotoxicity at triple-digit micromolar concentrations.

Two approaches were explored in order to investigate the interactions between TiP1 and hTETs: probing of hTET1/2 tiling libraries using TiP1-FLAG and photo-crosslinking of a photoreactive TiP1 derivative with purified TET enzymes. Probing of the TET tiling libraries with TiP1-FLAG produced false positive results due to unexpected antibody cross-reactivity; no additional hotspots could be observed, suggesting that 12-mer peptides constituting the tiling libraries may not offer sufficient surface area for TiP1 binding or that TiP1 binding affinity for TETs relies on tertiary structure recognition. Derivatisation of TiP1 with a benzophenone moiety to furnish the photoreactive probe TiP1 Y6Bpa resulted in minimal impact on potency

compared to the parent peptide (ΔpIC_{50} = −0.30, 0.22, and 0.36 for hTET1–3, respectively). Subsequent photo-crosslinking trials using 365 nm irradiation revealed a single (and a tentative double) cross-linking event between TiP1 Y6Bpa and model enzyme NgTet1 after 20 min irradiation time, encouraging extension of the approach to hTET enzymes.

Tentative hTET2-binding peptides (KTT1–6) identified using a RaPID selection against $hTET2_{CD\Delta LCI}$ were screened against hTET2 using AlphaScreen, but were found to be potent AlphaScreen signal quenchers, obscuring their potential inhibitory activities. Cellular activity evaluation of the peptides showed no inhibitory effects apart from cytotoxicity-associated reduction in cell counts and the consequently aberrant ^{5hm}C levels. KTT2 did not appear cytotoxic at 200 µM, and effected a nearly 50% reduction in ^{5hm}C levels in cells overexpressing hTET2/3, but further *in vitro* characterisation (such as inhibition studies using SPE-MS) are required to unambiguously ascribe this observation to hTET inhibition. Library enrichment for binders rather than inhibitors is a known drawback of peptide display methods[249]; the peptide hits may in fact bind allosterically to hTET2 without inhibiting its function, a property potentially desirable in TET-DNA interaction studies.

3.5 Conclusions and future work

TiP1–3 and 13 additional cyclic peptides from the same RaPID selection (KS01–13) were characterised against a panel of TET and related 2-OG dioxygenases in *vitro*, with TiP1 showing the highest degree of selectivity for hTET1. TiP1 alanine scanning confirmed high sequence optimisation for hTET1 binding and revealed divergent changes in potency against hTET1–3 for some alanine variants, highlighting structural diversity within the hTET family. The steepness of the enzyme inhibition dose-response curves obtained for the peptide library suggested potential aggregate-based inhibition, but a more detailed study on the potency dependence on detergent content should be carried out for selected peptides (such as TiP1 and KS08) using SPE-MS in order to definitively determine if the inhibitors are aggregation-prone.

Although TiP1 demonstrated excellent potency and selectivity for hTET1, the observed inhibitory activity could not be replicated *in cellulo*. Conjugation of TiP1 with FLAG, TAT, and R8 peptides did not improve cellular activity, although TiP1-FLAG exhibited an ameliorated cytotoxicity profile compared to TiP1. TiP1-TAT was markedly more potent *in vitro* compared to TiP1, but TAT and R8 CPPs alone were found to inhibit hTETs with single-digit micromolar potencies, suggesting that the increase in potency is likely due to the combined effect of the two inhibitory elements. Conjugation of two inhibitors to increase avidity is a known strategy[239,240], and could be further explored by conjugating TiP1 with solubilising motifs and less specific 2-OG dioxygenase inhibitors, such as 2-OG mimetics[250]. Furthermore, to improve peptide physiochemical properties an mRNA display selection could be carried out against hTET1 using a custom mRNA library with a partially randomised variable region: fixation of residues shown *via* alanine scanning to be essential for hTET1 binding (particularly Y4, Y7, W9, and W10) and randomisation of the remaining residues in the variable region (I3 to S13) would enable rapid variation of non-essential amino acid residues while retaining the underlying

binding mode of parent peptide TiP1. An alternative approach to improving cellular permeability is structure-guided α-*N*–methylation of the peptide backbone at positions not engaged in H–bonding interactions, as reported for the KDM4A inhibitor CP2[241]; however, the approach requires a co-crystal structure of the peptide bound to the target protein, which has proved difficult to obtain (Chapter 5).

TiP1 epitope mapping using the SPOT tiling libraries was unsuccessful, but photo-crosslinking trials between the photoreactive probe TiP1 Y6Bpa and NgTet1 have shown promising results. Follow-up work is required to unambiguously identify the cross-linking product, followed by photocross-linking with hTETs and proteomic analysis to determine the ligation site. Subsequently, development of a TiP1 Y6Bpa derivative containing a pull-down handle (i.e., biotin or FLAG tag) at the C–terminus would facilitate pull-down proteomics studies and identification of hTET1-associating proteins in cancer cell lines, analogously to a cyclic peptide affinity capture probe reported for PHD2[251].

Putative hTET2-binding peptides KTT1–6 could not be evaluated for inhibitory activity *in vitro* due to AlphaScreen interference, but KTT2 showed a cytotoxicity-independent reduction in ^{5hm}C levels at a concentration of 200 μM *in cellulo*. Evaluation of KTT peptides as hTET2 inhibitors using SPE-MS would circumvent the assay interference issues encountered using AlphaScreen and determine if KTT2 cellular activity is due to hTET2 inhibition. If no inhibition is observed, evaluation of binding affinities for hTET2 using BLI or ITC could help classify the nature of the generated library.

Chapter 4
Development of small molecule inhibitors for human TET enzymes

4.1 Introduction

Cyclic peptide (CP) inhibitors identified using affinity-driven mRNA display methods are typically exceptionally selective for their protein of interest; however, these *de novo* CPs often display poor pharmacological properties, particularly low cell permeability (as seen in Chapter 3). While a number of strategies for improving peptide cell permeability have been developed[252], optimisation of pharmacological properties remains challenging due to the structurally complex nature of CPs. Conversely, small molecule inhibitors, while typically less selective than CPs, are more amenable to rapid property optimisation.

To date, the reported small molecule TET inhibitors do not exhibit high selectivity within the TET family, and often inhibit 2-OG dioxygenases in a non-specific manner (section 1.6). Pan-specific 2-OG dioxygenase inhibitors, such as NOG, 2,4-PDCA, IOX1, and (onco)metabolites (succinate, fumarate, and 2-HG), are only moderately potent (single-digit micromolar to double-digit millimolar potencies *in vitro*) and offer virtually no selectivity. IOX1 exhibits cellular activity as the carboxylic acid (Chapter 3), while NOG, 2,4-PDCA, and (onco)metabolites are only active in cells as ester prodrugs; however, the resulting inhibition of hTETs is largely non-selective, in line with *in vitro* evaluation. Generic inhibition of 2-OG dioxygenases *in cellulo* can lead to TET-unrelated phenotypic changes, such as HIF upregulation-induced hypoxic response following FIH and PHD inhibition[253] or broader epigenetic dysregulation due to KDM inhibition[192], precluding investigation of TET-specific responses. Conversely, recently reported inhibitors TETi76[178] and C35[185] exhibit single-digit micromolar potencies against hTETs *in vitro*, with selectivity over a number of closely related

2-OG-dependent dioxygenases. However, the compounds are pan-hTET inhibitors, with little preference for a specific hTET paralogue. While selective pan-hTET inhibition is a useful tool, as exemplified by Guan *et al.*[178] in their work on induced synthetic lethality in hTET2-deficient cancer cells, the individual roles of TET paralogues cannot be interrogated in this manner. As such, chemical tools that enable selective temporal inhibition of specific TET paralogues are required to probe the distinct biological roles of TETs in a cellular context.

4.2 Aims

Identification of novel TET inhibitor scaffolds is required to facilitate the development of TET chemical tools. It was proposed that screening small molecule inhibitors of 2-OG-dependent dioxygenases against hTETs, followed by structure-activity relationship (SAR) studies on a selected series, could provide a foundation towards the development of a paralogue-selective, cell-permeable small molecule hTET inhibitor.

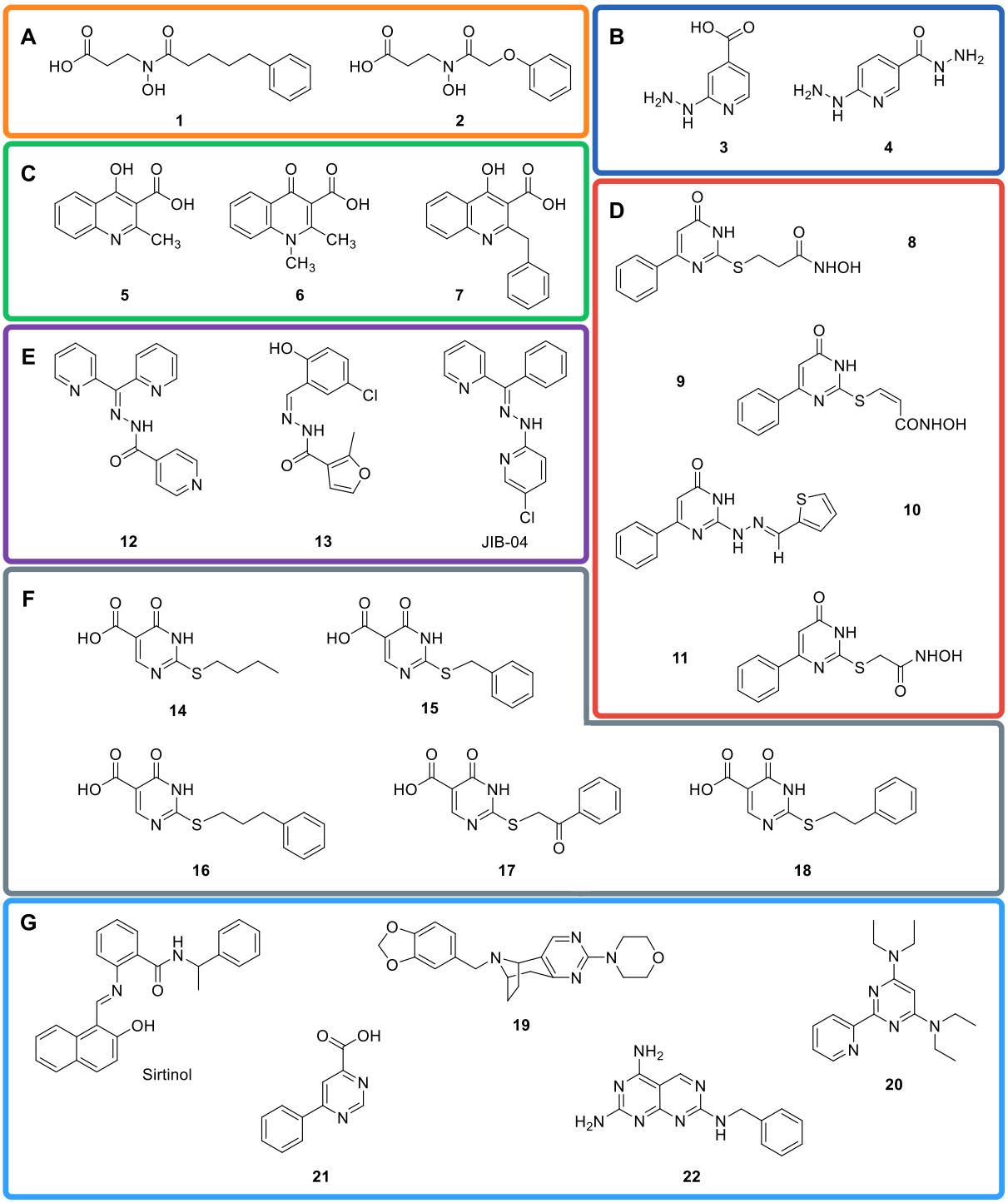
The aims of the work presented in this chapter are as follows:

- Identify a novel hTET2 small molecule inhibitor scaffold enabling expedient analogue synthesis.
- Explore the SAR between hTETs and the inhibitor series by screening analogues with various structural modifications.
- Cross-screen selected hTET inhibitors for selectivity over other Fe(II)/2-OG-dependent enzymes.
- Evaluate the cellular inhibitory activity of selected compounds.

4.3 Results

4.3.1 Identification of hTET2 inhibitor scaffolds

The development of a small molecule TET inhibitor began with the screening of a small library of compounds provided kindly by Assoc. Prof. Dante Rotili (Sapienza University of Rome) against hTET2 using AlphaScreen (1 nM $hTET2_{CD\Delta LCI}$, assay conditions optimised in Figure 3.2). Six core motifs were selected for the initial compound library (Scheme 4.1) based on prior reports on inhibitory activity against dioxygenases or potential TET substrate mimicry, including KDM4-inhibiting hydroxamic acids[254] (group **A**), pyridinedicarboxylate analogues[182] (group **B**), 4-hydroxy-3-quinolinecarboxylic acids[255] (group **C**), 6-phenyl-2-thiouracils (group **D**), lysine demethylase KDM inhibitor JIB-04[256] derivatives (group **E**), and 5-carboxy-2-thiouracil derivatives (group **F**); lysine deacetylase SIRT2 inhibitor sirtinol[257] and four potential iron-chelating scaffolds were screened as well (group **G**).



Scheme 4.1 | Structures of the hTET2 inhibitor scaffold screening library. Structures were grouped based on the core motif: **A**, KDM4-inhibiting[254] hydroxamic acid derivatives; **B**, pyridinedicarboxylate analogues[182]; **C**, 4-hydroxy-3-quinolinecarboxylic acid derivatives[255]; **D**, 6-phenyl-2-thiouracil derivatives; **E**, KDM inhibitor JIB-04[256] derivatives; **F**, 5-carboxy-2-thiouracil derivatives; **G**, SIRT2 inhibitor sirtinol[257] and four potential iron-chelating scaffolds.

Prior to enzymatic inhibition screening, the library was screened for anti-^{5hm}C antibody-dependent AlphaScreen signal interference (developed in Chapter 3). The majority of the library did not interfere with AlphaScreen signal, apart from JIB-04, which was found to be a

potent AlphaScreen signal quencher (pIC_{50} = 5.66), complicating evaluation of hTET2 inhibition (Figure 4.1).

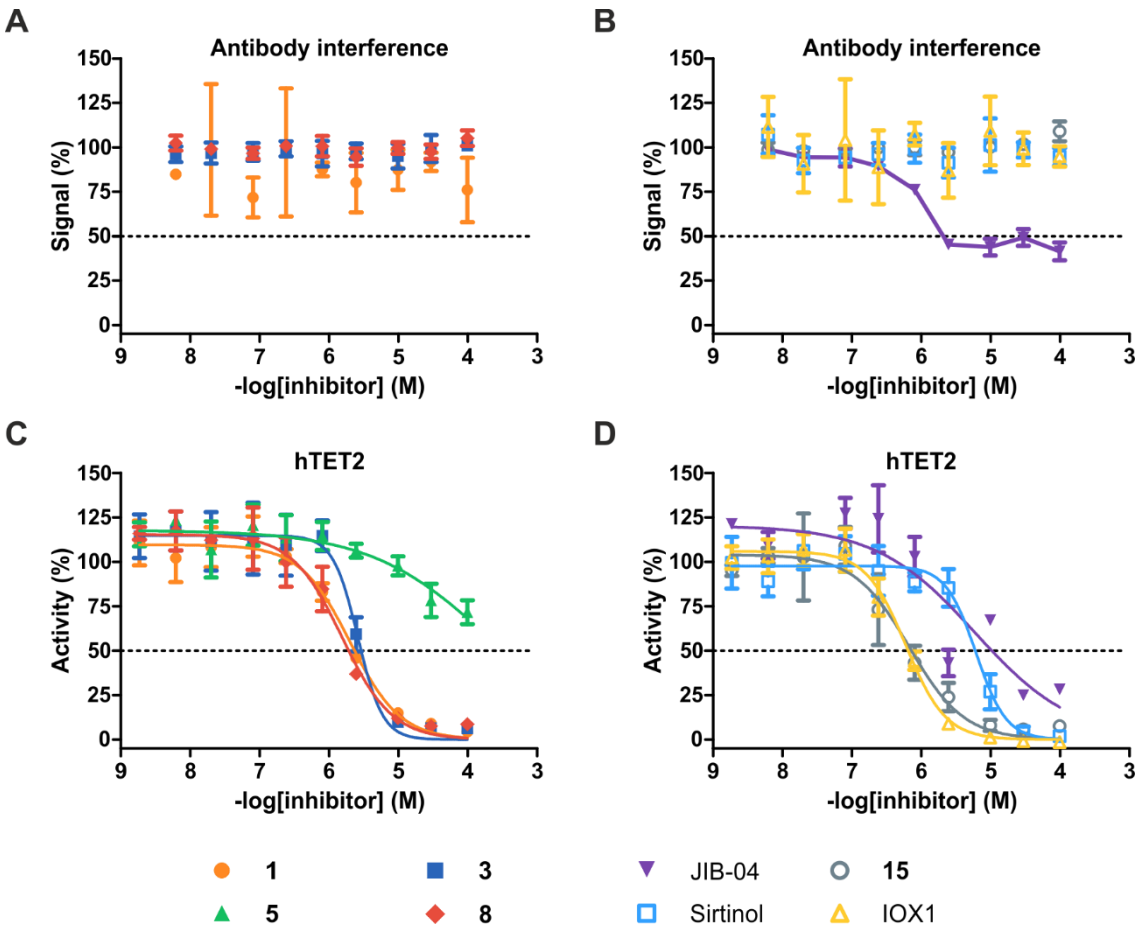


Figure 4.1 | AlphaScreen interference and hTET2 inhibition dose-response curves for representative compounds from the hTET2 inhibitor scaffold screening library. A, **B**, anti-^{5hm}C antibody interference; **C**, **D**, hTET2 inhibition. Enzymatic assay conditions: 1 nM hTET2$_{CD\Delta LCI}$, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. Data shown as the mean of 3 experimental replicates (n = 3; mean ± SD).

Single-digit micromolar inhibition of TET2 was observed for compounds from each structural motif group (Table 4.1), with the exception of the 3-quinolinecarboxylic acid derivatives (**5**–**7**), which exhibited poor potency (pIC_{50} < 4). Similarly, the majority of group **G** compounds (**19**–**22**) were shown to be poor hTET2 inhibitors (pIC_{50} ≤ 4.22). While sirtinol is reportedly selective for the NAD-dependent deacetylase sirtuin 2 (SIRT2) over histone deacetylases (HDACs)[257], the compound showed considerable potency as a hTET2 inhibitor (pIC_{50} = 5.21), suggesting that the reported *in cellulo* effects of sirtinol[257,258] may be in part be due to TET inhibition. 5-Carboxy-2-thiouracil derivatives (**14**–**18**) inhibited hTET2 with IC_{50} values around

1 µM (pIC_{50} = 5.83–6.11), and analogues of **14**–**18** were readily accessible *via* multiple functional handles, making 5-carboxy-2-thiouracil an attractive scaffold for subsequent optimisation. Furthermore, mTets, which are closely related to hTETs, have been reported to oxidise thymine to 5-hydroxymethyluracil in mESCs (section 1.2.4), suggesting that thymine derivatives may exhibit enhanced binding affinities for the TET family *via* ligand mimicry. Thus, the 5-carboxy-2-thiouracil series was selected for further SAR studies.

Table 4.1 | Tabulated AlphaScreen interference and hTET2 inhibition pIC_{50} values for the hTET2 inhibitor scaffold screening library. Enzyme inhibition data likely obscured by AlphaScreen interference is shaded in grey. Values reported as pIC_{50} ± SD (M). Values reported as percentage signal (%) at 100 µM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 µM); n.d., not determined. Data reported as the mean of 3 experimental replicates (n = 3; mean ± SD).

Group	Name	pIC_{50} ± SD (M)	
		Antibody interference	hTET2 inhibition
Control	IOX1	n.i.	6.22 ± 0.04
A	**1**	76.1 ± 14.8%	5.69 ± 0.05
	2	n.i.	5.76 ± 0.07
B	**3**	n.i.	5.58 ± 0.04
	4	n.i.	5.47 ± 0.08
C	**5**	n.d.	71.6 ± 5.5%
	6	n.d.	n.i.
	7	n.d.	61.6 ± 11.0%
D	**8**	n.i.	5.81 ± 0.06
	9	n.i.	5.65 ± 0.07
	10	n.d.	n.i.
	11	n.i.	5.57 ± 0.05
E	**12**	n.i.	4.79 ± 0.24
	13	n.i.	4.76 ± 0.04
	JIB-04	5.66 ± 0.03	5.23 ± 0.19
F	**14**	n.i.	5.83 ± 0.06
	15	n.i.	6.11 ± 0.08
	16	n.i.	5.82 ± 0.04
	17	n.i.	6.09 ± 0.04
	18	n.i.	5.92 ± 0.05
G	Sirtinol	n.i.	5.21 ± 0.06
	19	n.d.	n.i.
	20	n.d.	n.i.
	21	n.d.	16.4 ± 0.9%
	22	n.d.	86.0 ± 0.5%

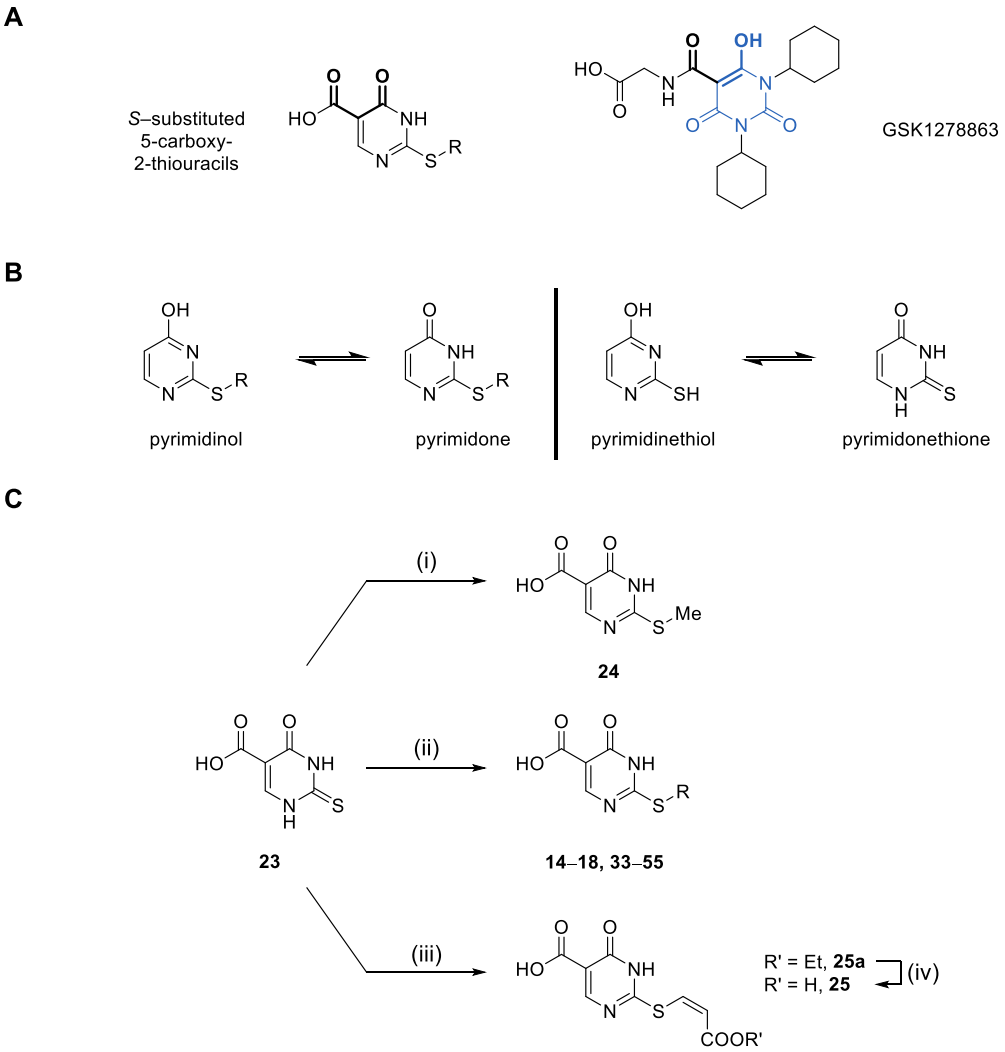
The initial motif screening revealed a number of potential compounds that could be further investigated as hTET inhibitors; however, the fact that inhibitors of other 2-OG dioxygenases also inhibit hTET2 with good potencies highlights the challenge of obtaining selectivity for a specific 2-OG hydroxylase.

4.3.2 Structural features and synthesis of *S*–substituted 5-carboxy-2-thiouracils

The 5-carboxy-2-thiouracil scaffold contains a 1,3-dicarbonyl system capable of metal chelation in a manner similar to the barbiturate-based HIF PHD inhibitor GSK1278863[259] (Scheme 4.2 A, barbiturate core highlighted in blue). Similarly to barbiturate cores[260], the 2-thiouracil scaffold exhibits a degree of keto–enol tautomerism: both *S*–substituted and *S*–unsubstituted analogues can tautomerise between enol (pyrimidinol or pyrimidinethiol) and keto (pyrimidone or pyrimidonethione) forms (Scheme 4.2 B). Core level X-ray photoemission spectroscopy (XPS) work by Giuliano *et al*.[261] identified the keto tautomers as the predominant forms in both cases, with the observation further supported by in-solution NMR studies of 2-mercaptopyrimidine[262] and isocytosine (2-amino-3*H*-pyrimidin-4-one)[263]. As such, structures reported herein are presented as the pyrimidone and pyrimidonethione tautomers; however, it should be noted that the tautomeric equilibrium is dependent on solution pH and other factors[264], and may be shifted towards the enol forms in some structurally distinct analogues.

All screened compounds were prepared by the group of Assoc. Prof. Dante Rotili (Sapienza University of Rome); thus, only a general overview of the synthetic route is presented herein. The majority of the characterised analogues were accessed from the commercially available **23**, which was alkylated using alkyl or benzyl halides under basic conditions in anhydrous DMF to afford the desired thioether analogues **14**–**18**, **33**–**55** (steps i, ii, Scheme 4.2 C). Ethyl ester **25a** was obtained by a reaction between **23** and ethyl propiolate under weakly basic

conditions (step iii); subsequent treatment of **25a** with aqueous base furnished the dicarboxylic acid **25** (step iv).

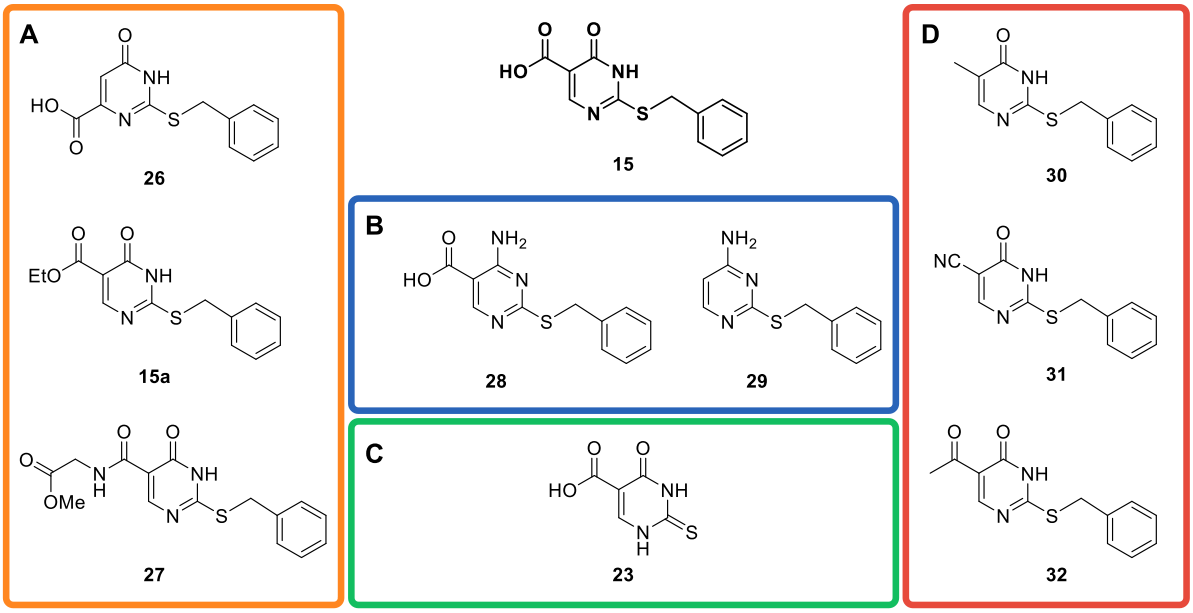


Scheme 4.2 | Overview of the structural features and synthesis of *S*–substituted 5-carboxy-2-thiouracil analogues. A, comparison of *S*–substituted 5-carboxy-2-thiouracil and PHD inhibitor GSK1278863[259] structural features (barbiturate core highlighted in blue); **B**, tautomerisation of *S*–substituted (left) and *S*–unsubstituted (right) 5-carboxy-2-thiouracil analogues; **C**, general synthetic route[265] for the majority of characterised *S*–substituted analogues. Reagents and conditions: (i) MeI, anhydrous DMF, RT; (ii) alkyl/benzyl bromide, anhydrous K_2CO_3, anhydrous DMF, RT; (iii) ethyl propiolate, TBAF, anhydrous THF, RT; (iv) aqueous KOH, EtOH, RT. All compounds were prepared by the group of Assoc. Prof. Dante Rotili (Sapienza University of Rome).

4.3.3 Scaffold modification of *S*–benzyl 5-carboxy-2-thiouracil

Having explored the effect of *S*–alkyl substitution on potency to some degree (**14**–**18**), the SAR between 5-carboxy-2-thiouracil analogues and hTET2 was further investigated *via* the synthesis of analogues with modifications at the 3-, 4-, and 5- positions of the scaffold. The analogues were based on **15**, a representative *S*–substituted compound, and a variety of scaffold modifications were explored (Scheme 4.3), including positional variation and

derivatisation of the carboxylic acid moiety (group **A**), inversion of H–bonding character at the 3-/4-positions *via* 4-amine substitution (group **B**), removal of *S*–substitution (group **C**), and introduction of alternative functional groups at the 5-position (group **D**).



Scheme 4.3 | Structures of *S*–benzyl 5-carboxy-2-thiouracil (compound 15) scaffold analogues. A, carboxylic acid analogues; **B**, 4-amino analogues; **C**, *S*–unsubstituted analogue; **D**, 5-substituted 2-thiouracil analogues.

Table 4.2 | Tabulated AlphaScreen interference and hTET2 inhibition pIC_{50} values for scaffold analogues of compound 15. Values reported as pIC_{50} ± SD (M). Changes in potency relative to parent inhibitor **15** are reported as ΔpIC_{50}. Values reported as percentage signal (%) at 100 µM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 µM). Data for AlphaScreen antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for hTET2 inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

Group	Name	pIC_{50} ± SD (M)		ΔpIC_{50}
		Antibody interference	hTET2 inhibition	
Parent	**15**	n.i.	5.93 ± 0.11	-
A	**26**	n.i.	5.26 ± 0.11	−0.67
	15a*	n.i.	75.6 ± 0.3%	<−1.93
	27	n.i.	n.i.	<−2
B	**28***	n.i.	35.9 ± 3.2%	<−1.93
	29	n.i.	n.i.	<−2
C	**23**	n.i.	5.71 ± 0.08	−0.22
D	**30**	n.i.	n.i.	<−2
	31	n.i.	n.i.	<−2
	32*	n.i.	65.1 ± 8.4%	<−1.93

**Data for hTET2 inhibition reported as the mean of 2 independent replicates (n = 2; mean ± SD).*

15 and its scaffold derivatives did not interfere with AlphaScreen signal, facilitating subsequent screening for hTET2 inhibition (Table 4.2). 6-Carboxy analogue **26** exhibited a

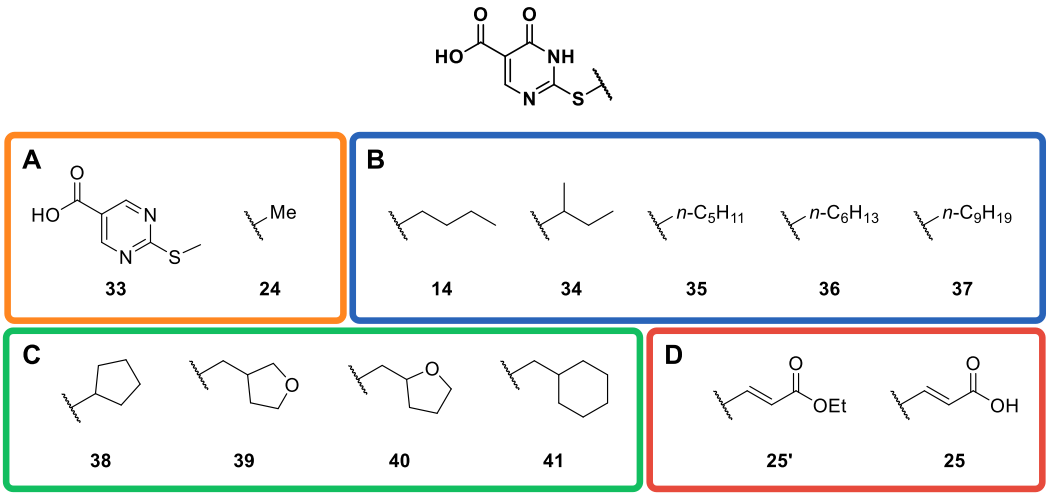
nearly 5-fold loss in potency against hTET2 ($\Delta pIC_{50} = -0.67$), while derivatisation of 5-carboxyl to ethyl (**15a**) or glycinamide methyl (**27**) esters produced a substantially negative effect on inhibition ($pIC_{50} < 4$), indicating that the free 5-carboxylic acid moiety is likely essential for inhibitory activity.

Inversion of H–bonding character by substitution of 4-oxo with a 4-amino group (**28**) reduced the potency 100-fold, with a further potency loss observed following the removal of 5-carboxylic acid (**29**). Thione **23** inhibited hTET2 with potency comparable to **15**, suggesting that *S*–substituents may not contribute substantially to the binding affinity. However, *S*–substitution may offer a means of selectivity optimisation, as *S*–substituents could disfavour binding to other 2-OG-dependent dioxygenases.

Finally, substitution of 5-carboxylic acid with methyl (**30**), nitrile (**31**), or acetyl (**32**) groups did not preserve the potency of the parent inhibitor ($pIC_{50} < 4$ for **30**–**32**); however, the ketone analogue **32** exhibited a small degree of inhibitory activity (65% hTET2 activity at 100 μM inhibitor), potentially supporting Fe(II) chelation *via* the β-ketoacid moiety as the mode of binding.

4.3.4 *S*–alkyl 5-carboxy-2-thiouracil analogues

As modification of the 5-carboxy-2-thiouracil core was shown to be poorly tolerated, *S*–alkyl analogues, including linear alkyl chains (groups **A**, **B**, Scheme 4.4), (cyclo)alkyl chains (group **C**), and acrylate derivatives (group **D**), were explored in order to determine the size tolerance limits for *S*–substituents and probe potential hydrophobic or H–bonding interactions within the binding site.



Scheme 4.4 | Structures of *S*–alkyl 5-carboxy-2-thiouracil analogues. A, *S*–methyl analogues; **B**, linear alkyl chain analogues; **C**, (cyclo)alkyl analogues; **D**, acrylate analogues.

Table 4.3 | Tabulated AlphaScreen interference and hTET2 inhibition pIC_{50} values for *S*–alkyl 5-carboxy-2-thiouracil analogues. Values reported as pIC_{50} ± SD (M). Changes in potency relative to reference inhibitor **15** are reported as ΔpIC_{50}. N.i., no inhibition (>90% signal at 100 µM). Data for AlphaScreen antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for hTET2 inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

Group	Name	pIC_{50} ± SD (M)		ΔpIC_{50}
		Antibody interference	hTET2 inhibition	
Reference	**15**	n.i.	5.93 ± 0.11	-
A	**33**	n.i.	n.i.	<−1.93
	24	n.i.	6.11 ± 0.10	0.18
B	**14**	n.i.	5.90 ± 0.10	−0.03
	34	n.i.	6.03 ± 0.05	0.10
	35	n.i.	5.82 ± 0.13	−0.11
	36	n.i.	5.58 ± 0.09	−0.35
	37	n.i.	5.22 ± 0.08	−0.71
C	**38**	n.i.	6.07 ± 0.13	0.14
	39	n.i.	6.12 ± 0.09	0.19
	40	n.i.	6.08 ± 0.07	0.15
	41	n.i.	5.39 ± 0.05	−0.54
D	**25a**	n.i.	6.12 ± 0.13	0.19
	25	n.i.	5.99 ± 0.11	0.06

While the inhibitory potency of the *S*–methylated analogue **24** was comparable to **15** (ΔpIC_{50} = 0.18, Table 4.3), removal of the 4-oxo group (**33**) completely abrogated inhibitory activity, further demonstrating the importance of the β-ketoacid functionality (*cf.* **30**, **28** in section 4.3.3). Expansion of the flexible *S*–alkyl substituents to *n*-butyl (**14**), *sec*-butyl (**34**) and

n-pentyl (**35**) chains did not affect potency, whereas some potency loss was observed for *n*-hexyl (**36**) and *n*-nonyl (**37**) analogues (ΔpIC_{50} = −0.35 and −0.71 compared to **15**, respectively).

S–alkylation with cyclopentyl (**38**) or methyltetrahydrofuran (**39**, **40**) groups did not impact inhibition, indicating that the presence of more rigid alkyl side chains and H–bond acceptors (THF analogues) did not offer additional interactions within the inhibitor binding site. Introduction of methylcyclohexane (**41**) led to an approximately 3.5-fold loss in potency compared to **15** (ΔpIC_{50} = −0.54), providing an indication of the side chain size tolerance limits. Finally, functionalisation of the thione moiety with acrylate groups (**25** and **25a**) did not have an impact on hTET2 inhibition.

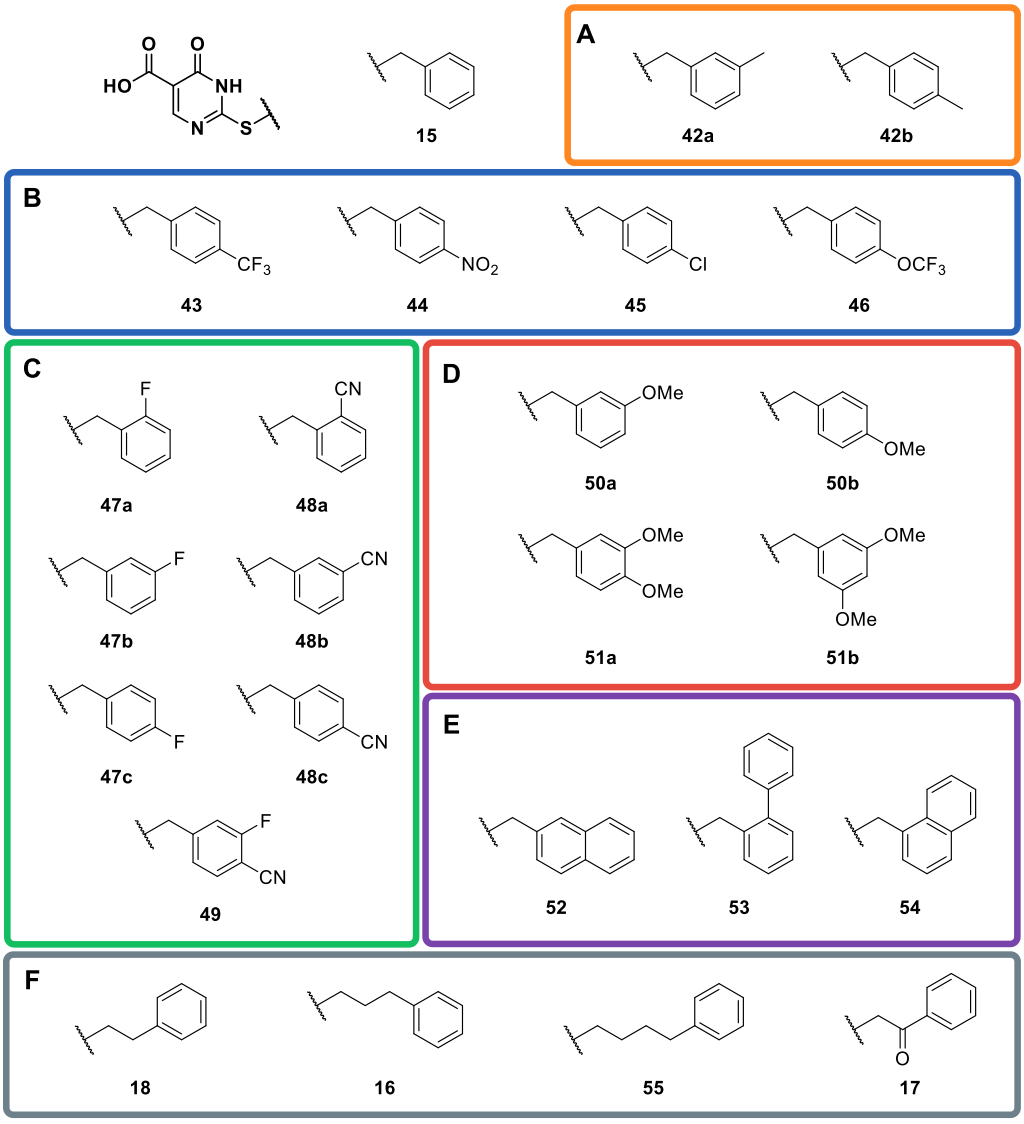
4.3.5 *S*–(aryl)alkyl 5-carboxy-2-thiouracil analogues

Conformationally flexible ligands such as the analogues containing *S*–alkyl chains (**33**–**41**) can have an entropic penalty associated with binding to the enzyme[266], potentially leading to reduced potency. In an effort to both reduce the rotational degrees of freedom within the molecules and probe for potential π-π stacking interactions within the hTET2 inhibitor binding site, side chains featuring aromatic groups were explored (Scheme 4.5).

Methylation of **15** at the *meta*- (**42a**) and *para*-positions (**42b**) marginally decreased inhibitory potency against hTET2 compared to **15** (ΔpIC_{50} = −0.13 and −0.14, respectively, Table 4.4). *Para*-nitro (**44**) and *para*-chloro (**45**) derivatisation did not impact hTET2 inhibition, whereas introduction of *para*-trifluoromethyl (**43**) and *para*-trifluoromethoxy (**46**) groups decreased inhibitory potency 3-fold (ΔpIC_{50} = −0.52 and −0.50, respectively); a comparison with *para*-methyl (**42b**) and *para*-methoxy (**50b**) analogues (ΔpIC_{50} = −0.13 and −0.16, respectively) revealed that the presence of fluorine atoms contributed to more than a 2-fold loss in potency.

Characterisation of fluoro- and cyano-substituted *S*–benzyl analogues demonstrated that *para*-substitution (**47c**, **48c**) was well-tolerated, *ortho*-substitution (**47a**, **48a**) had a minor

detrimental effect on potency, and *meta*-substitution (**47b**, **48b**) led to a 2.5-fold loss in potency. Although the effect was apparently independent of substituent size, with fluoro and cyano analogues exhibiting virtually identical potencies (e.g., ΔpIC_{50} = −0.39 and −0.40 for **47b** and **48b**, respectively), an analogue containing an electron-donating *meta*-methoxy group, **50a**, was tolerated better than its electron-withdrawing counterparts (ΔpIC_{50} = −0.24 for **50a**).



Scheme 4.5 | Structures of *S*–(aryl)alkyl 5-carboxy-2-thiouracil analogues. A, methylated **15** analogues; **B**, *para*-substituted *S*–benzyl analogues; **C**, fluoro- and cyano-substituted *S*–benzyl analogues; **D**, methoxy-substituted *S*–benzyl analogues; **E**, multiple aromatic ring analogues; **F**, *S*–benzyl homologues and acetophenone analogue.

Table 4.4 | Tabulated AlphaScreen interference and hTET2 inhibition pIC_{50} values for *S*–(aryl)alkyl 5-carboxy-2-thiouracil analogues. Values reported as pIC_{50} ± SD (M). Changes in potency relative to reference inhibitor **15** are reported as ΔpIC_{50}. N.i., no inhibition (>90% signal at 100 μM). Data for AlphaScreen antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for hTET2 inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

Group	Name	pIC_{50} ± SD (M)		ΔpIC_{50}
		Antibody interference	hTET2 inhibition	
Reference	**15**	n.i.	5.93 ± 0.11	-
A	**42a**	n.i.	5.80 ± 0.08	−0.13
	42b	n.i.	5.79 ± 0.15	−0.14
B	**43**	n.i.	5.41 ± 0.08	−0.52
	44	n.i.	5.95 ± 0.09	0.02
	45	n.i.	5.76 ± 0.14	−0.17
	46*	n.i.	5.43 ± 0.00	−0.50
C	**47a**	n.i.	5.73 ± 0.22	−0.20
	47b	n.i.	5.54 ± 0.14	−0.39
	47c	n.i.	6.02 ± 0.09	0.09
	48a	n.i.	5.72 ± 0.19	−0.21
	48b	n.i.	5.53 ± 0.18	−0.40
	48c	n.i.	6.12 ± 0.15	0.19
	49	n.i.	5.68 ± 0.17	−0.25
D	**50a**	n.i.	5.69 ± 0.04	−0.24
	50b	n.i.	5.77 ± 0.05	−0.16
	51a	n.i.	5.81 ± 0.03	−0.12
	51b	n.i.	5.50 ± 0.15	−0.43
E	**52**	n.i.	5.36 ± 0.05	−0.57
	53	n.i.	5.49 ± 0.05	−0.44
	54	n.i.	5.26 ± 0.14	−0.67
F	**18**	n.i.	5.97 ± 0.08	0.04
	16	n.i.	5.80 ± 0.16	−0.13
	55	n.i.	5.66 ± 0.09	−0.27
	17	n.i.	6.05 ± 0.03	0.12

**Data for hTET2 inhibition reported as the mean of 2 independent replicates (n = 2; mean ± SD).*

Taken together, the observations suggest that electron-withdrawing groups may affect binding *via* electronic effects, such as eliciting changes to contributions of π–π stacking or π–cation bonding. Introduction of a second *meta*-substituent to afford the 3,5-dimethoxy analogue **51b** resulted in a further reduction of potency (ΔpIC_{50} = −0.43), indicating a limit to the steric tolerance of the inhibitor binding site. Multiple ring systems were not tolerated, with both the naphthalene-based analogues (**52** and **54**) and the *ortho*-phenyl analogue **53** exhibiting a 3–4-fold loss in potency compared to **15**. A trend of decreasing potency was

observed for *S*–benzyl analogues with increasing connecting alkyl chain length (methyl **15**, ethyl **18**, propyl **16**, and butyl **55**; ΔpIC_{50} = 0.04 and −0.27 for **18** and **55**, respectively), while introduction of a ketone moiety *via* the acetophenone analogue **17** did not have an impact on hTET2 inhibition.

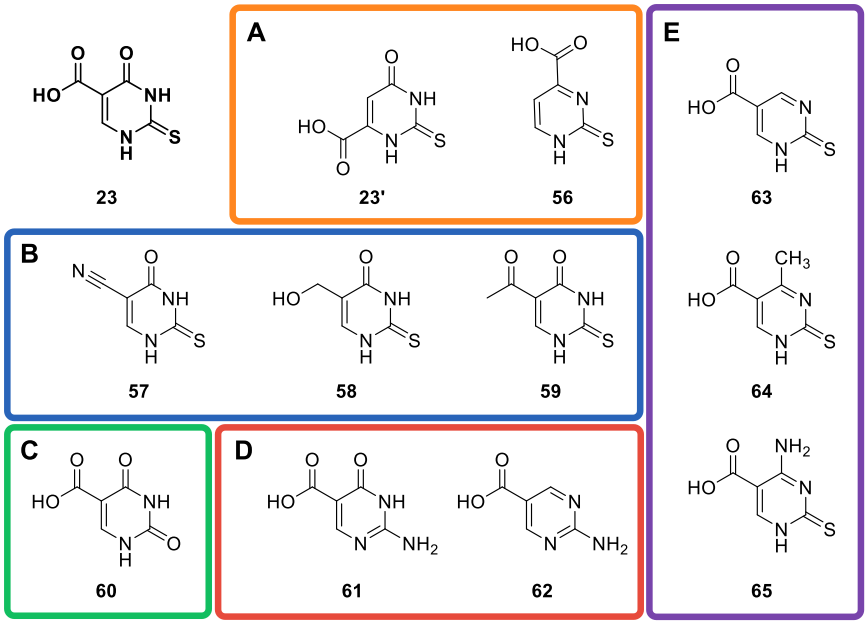
Overall, the *S*–(aryl)alkyl series demonstrated that structural variation of *S*–substituents had little impact on potency. No substantial potency gains could be actualised, whereas steric clashing of larger *S*–substituents (e.g., **46**, **54**) led to reduced potency; conversely, *meta*-substitution with electron-withdrawing groups on the *S*–benzyl benzene ring was tolerated poorly regardless of size, suggesting a potential electronic effect. Importantly, the observed toleration of a range of alkyl and (aryl)alkyl S–substituents may potentially provide a strategy for improving selectivity for TETs within the 2-OG-dependent dioxygenase superfamily by discriminating against enzymes with lower steric tolerance for substitution at the 2-position.

4.3.6 *S*–unsubstituted (thione) analogues

Although scaffold derivatisation of the *S*–benzyl **15** typically produced a negative effect on hTET2 inhibition, tolerance for *S*–unsubstituted scaffold modifications had not been explored. Thione parent 5-carboxy-2-thiouracil (**23**) was previously shown to be a potent hTET2 inhibitor (pIC_{50} = 5.71, section 4.3.3); thus, analogues of **23** were generated and characterised against hTET2 (Scheme 4.6).

6-Carboxy analogue **23'** showed similar potency to **23**, while 4-carboxy analogue **56** exhibited >10-fold higher potency (ΔpIC_{50} = 1.14, Table 4.5). Substitution of 5-carboxylic acid with other functional groups, including nitrile (**57**), hydroxymethyl (**58**), and acetyl (**59**), resulted in an approx. 10-fold loss in potency. Interestingly, *S*–unsubstituted analogues without the 5-carboxylic acid were well-tolerated compared to the direct *S*–benzyl analogues **31** and **32**,

which showed practically no inhibitory activity towards hTET2 (section 4.3.3). The discrepancy of 5-carboxylic acid contribution to potency suggests that thione compounds may inhibit hTET2 *via* a distinct mechanism from *S*–substituted analogues.



Scheme 4.6 | Structures of *S*–unsubstituted (thione) analogues. A, carboxylic acid positional analogues; **B**, 5-substituted 2-thiouracil analogues; **C**, 2–*O* analogue (uracil); **D**, isocytosine analogues; **E**, 4-oxo analogues.

Table 4.5 | Tabulated AlphaScreen interference and hTET2 inhibition pIC_{50} values for *S*–unsubstituted (thione) analogues. Values reported as pIC_{50} ± SD (M). Changes in potency relative to parent inhibitor **23** are reported as ΔpIC_{50}. N.i., no inhibition (>90% signal at 100 µM). Data for AlphaScreen antibody interference reported as the mean of 3 experimental replicates (n = 3; mean ± SD); data for hTET2 inhibition reported as the mean of 3 independent replicates (n = 3; mean ± SD).

Group	Name	pIC_{50} ± SD (M)		ΔpIC_{50}
		Antibody interference	hTET2 inhibition	
Reference	**23**	n.i.	5.71 ± 0.08	-
A	**23'**	n.i.	5.83 ± 0.02	0.12
	56	n.i.	6.85 ± 0.07	1.14
B	**57**	n.i.	4.63 ± 0.14	−1.08
	58	n.i.	4.79 ± 0.08	−0.92
	59	n.i.	4.90 ± 0.03	−0.81
C	**60**	n.i.	4.91 ± 0.34	−0.80
D	**61**	n.i.	5.44 ± 0.15	−0.27
	62	n.i.	4.93 ± 0.02	−0.78
E	**63***	n.i.	6.71 ± 0.06	1.00
	64	n.i.	5.78 ± 0.04	0.07
	65	n.i.	5.82 ± 0.11	0.11

**Data for hTET2 inhibition reported as the mean of 2 independent replicates (n = 2; mean ± SD).*

2–*O* analogue **60** (uracil) was approx. 6-fold less potent compared to reference **23**, demonstrating the importance of the thione moiety. The 2-aminouracil (isocytosine) analogue

61 was only 2-fold less potent than **23** (ΔpIC_{50} = −0.27), hinting at potential H–bonding interactions with the 2-amino group. Removal of the 4-oxo moiety (**62**) resulted in further loss of inhibitory activity (ΔpIC_{50} = −0.78), similarly to the *S*–methylated analogue **33** (section 4.3.4). Interestingly, the 4–*H* 2-thiopyrimidine derivative **63** exhibited a 10-fold enhancement in activity (ΔpIC_{50} = 1.00), while an identical modification completely abrogated the inhibitory activity of the direct *S*–methylated analogue **33** (section 4.3.4), providing further indication of divergent modes of action for *S*–substituted and thione analogues. Finally, substitution of 4-oxo with 4-methyl (**64**) or 4-amino (**65**) groups did not have an adverse effect on potency; conversely, the direct *S*–benzyl 4-amino derivative **28** (section 4.3.3) showed almost no inhibition (35.9% enzyme activity at 100 μM inhibitor). Modifications to the thione scaffold were typically much better tolerated than analogous modifications to the *S*–benzyl analogue **15**, suggesting that *S*–unsubstituted compounds bind to hTET2 through a distinct set of interactions.

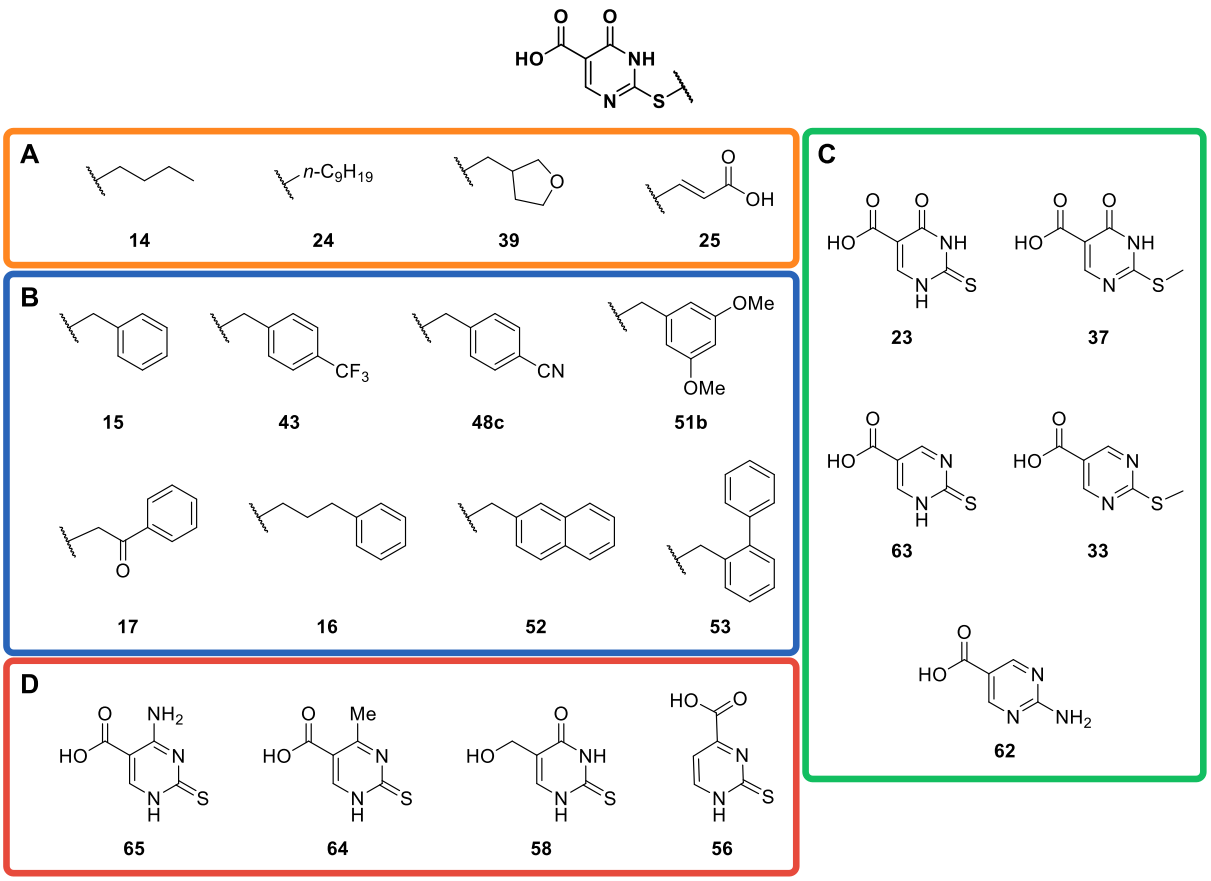
4.3.7 hTET paralogue selectivity profiling

Although the collected SAR for hTET2 inhibition did not present definite trends, analogue hTET paralogue selectivity profiles warranted investigation. To assess paralogue selectivity, representative compounds from *S*–alkyl, *S*–(aryl)alkyl, and thione series were selected for evaluation against hTET1 and hTET3 using AlphaScreen (Scheme 4.7). The conditions used for screening were identical to those optimised in Figure 3.2 (assays carried out with 10 nM $hTET1_{CD}$, 1 nM $hTET2_{CD\Delta LCI}$, and 5 nM $hTET3_{CD}$).

Table 4.6 | Tabulated AlphaScreen hTET1–3 inhibition pIC_{50} values for selected analogues. Values reported as pIC_{50} ± SD (M). Positive ΔpIC_{50} values represent selectivity for hTET2 over hTET1/3. N.i., no inhibition (>90% signal at 100 μM); n.d., not determined. Data for enzyme inhibition reported as the mean of at least 3 independent replicates (n ≥ 3; mean ± SD).

Group	Name	pIC_{50} ± SD (M)			ΔpIC_{50}	
		hTET1	hTET2	hTET3	hTET2–1	hTET2–3
IOX1		5.86 ± 0.03	6.18 ± 0.17	5.96 ± 0.08	0.32	0.22
A	**14**	5.02 ± 0.16	5.90 ± 0.10	4.23 ± 0.18	0.88	1.67
	24	4.50 ± 0.05	5.22 ± 0.08	4.84 ± 0.10	0.72	0.38
	39	5.36 ± 0.22	6.12 ± 0.09	5.37 ± 0.18	0.76	0.75
	25	4.98 ± 0.03	5.99 ± 0.11	5.31 ± 0.14	1.01	0.68
B	**15**	4.92 ± 0.06	5.93 ± 0.11	4.92 ± 0.20	1.01	1.01
	43	4.77 ± 0.04	5.41 ± 0.08	n.d.	0.64	-
	48c	5.21 ± 0.08	6.12 ± 0.15	5.14 ± 0.22	0.91	0.98
	51b	4.89 ± 0.09	5.50 ± 0.15	4.73 ± 0.12	0.61	0.77
	17	5.15 ± 0.11	6.05 ± 0.03	4.63 ± 0.23	0.90	1.42
	16	5.20 ± 0.11	5.80 ± 0.16	4.63 ± 0.13	0.60	1.17
	52	4.91 ± 0.12	5.36 ± 0.05	5.15 ± 0.07	0.45	0.21
	53	4.70 ± 0.06	5.49 ± 0.05	4.77 ± 0.01	0.79	0.72
C	**23**	4.84 ± 0.24	5.71 ± 0.08	4.38 ± 0.22	0.87	1.33
	37	5.49 ± 0.20	6.11 ± 0.10	5.56 ± 0.02	0.62	0.55
	63*	5.80 ± 0.08	6.71 ± 0.06	5.31 ± 0.06	0.91	1.40
	33	n.i.	n.i.	n.i.	-	-
	62	n.i.	4.93 ± 0.02	4.70 ± 0.01	>0.93	0.23
D	**65**	5.14 ± 0.20	5.82 ± 0.11	4.22 ± 0.09	0.68	1.60
	64	5.44 ± 0.08	5.78 ± 0.04	4.80 ± 0.11	0.34	0.98
	58	4.44 ± 0.14	4.79 ± 0.08	4.03 ± 0.02	0.35	0.76
	56	6.70 ± 0.03	6.85 ± 0.07	6.38 ± 0.02	0.15	0.47

**Data for hTET2 inhibition reported as the mean of 2 independent replicates (n = 2; mean ± SD).*



Scheme 4.7 | Structures of selected analogues for hTET paralogue selectivity profiling. A, non-aromatic *S*–substitution; **B**, aromatic *S*–substitution; **C**, 4-oxo and 4–*H* analogues; **D**, thione scaffold analogues.

Although the majority of the selected analogues preferentially inhibited hTET2 to varying degrees, Spearman's rank order correlation revealed a very strong relation between hTET1/2 inhibition (Spearman's r_s = 0.910, Figure 4.2 A; pIC_{50} values for n.i. compounds were set to 3 for ranking and visualisation purposes), while correlations between hTET1/3 and hTET2/3 inhibition were moderate (Spearman's r_s = 0.661 and 0.691, respectively, Figure 4.2 B, C). The observed relationships signified that the modest impact of *S*–substituent structural variation on potency against hTET2 extended to hTET1/3, and suggested that optimisation of hTET paralogue selectivity would be challenging within the context of the 5-carboxy-2-thiouracil series.

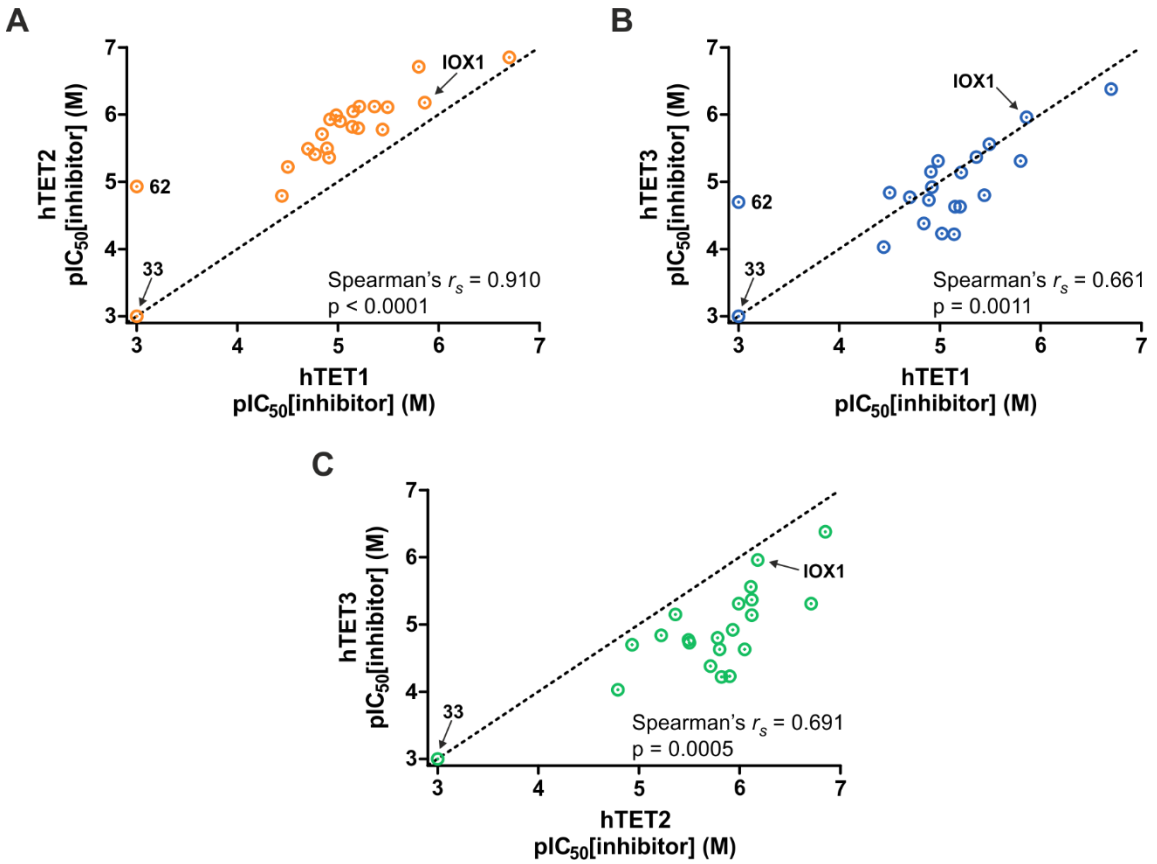


Figure 4.2 | AlphaScreen pIC_{50} correlation plots for selected inhibitors against hTET1–3. **A**, hTET1/hTET2 correlation plot; **B**, hTET1/hTET3 correlation plot; **C**, hTET2/hTET3 correlation plot. pIC_{50} values for n.i. compounds were set to 3 for ranking and visualisation purposes. Spearman's rank order correlation coefficients (r_s) and p-values for each correlation plot were calculated using GraphPad Prism 5. Dashed lines in each plot represent identity lines ($y = x$).

S–butyl **14** was the most selective *S*–alkyl analogue, exhibiting 8-fold selectivity for hTET2 over hTET1, and nearly 50-fold selectivity over hTET3. *S*–benzyl **15** was 10-fold more potent against hTET2 than hTET1/3, with the more decorated *S*–benzyl analogues and alkyl chain linkage homologues showing either comparable or poorer selectivity profiles. Thione **23** showed 7-fold selectivity for hTET2 over hTET1, and was >20-fold selective over hTET3; the *S*–methylated analogue **37** was more potent against hTET2 (pIC_{50} = 6.11 and 5.71 for **37** and **23**, respectively), but less selective than **23**.

Removal of the 4-oxo group (**63**) improved the potency 10-fold across the board compared to **23**, while concurrent *S*–methylation (**33**) resulted in a complete loss of hTET inhibition, highlighting the substantial SAR differences between *S*–substituted analogues and thiones. Conversely, substitution of the thione group with a primary amine (**62**) resulted in a >100-fold

potency loss against hTET1, but preserved low double-digit micromolar potencies against hTET2/3. While scaffold modifications on thione **23** at positions 4- and 5- (group **D**) typically did not enhance selectivity for hTET2, **65** showed better selectivity over hTET3 compared to **23** (40-fold and 21-fold, respectively). **56** was shown to be the most potent inhibitor of hTET2 (pIC_{50} = 6.85), but exhibited a pan-specific inhibitory profile within the hTET panel.

4.3.8 Cellular activity evaluation for selected 5-carboxylic acid analogues

Following hTET selectivity profiling, representative 5-carboxy-2-thiouracil analogues were selected for evaluation of inhibitory activity in U2OS cells overexpressing hTET1/2 (Figure 4.3, top panel). IOX1 treatment produced a dose-dependent reduction in ^{5hm}C signal, whereas none of the screened 5-carboxy-2-thiouracil derivatives exhibited cellular activity (conditions identical to those in Figure 3.9). Small, hydrophilic compounds, particularly those containing carboxylic acid functional groups, are known to suffer from poor cell permeability, and often require derivatisation of the carboxylic acid moiety to elicit an inhibitory response *in cellulo*[267]. Pan-specific 2-OG dioxygenase inhibitors NOG[175] and 2,4-PDCA[182] are notable examples of this issue, and are utilised as dimethyl ester prodrugs to improve cell permeability (section 1.6). However, esterification of **15** to ethyl ester **15a** did not result in enhanced cellular activity. Hydrolysis of **15a** to release the active carboxylic acid compound **15** may have been too slow to elicit a cellular response within the time frame of the experiment, as the closely related ethyl benzoate is poised to be among the most stable benzoic acid esters in human plasma[268].

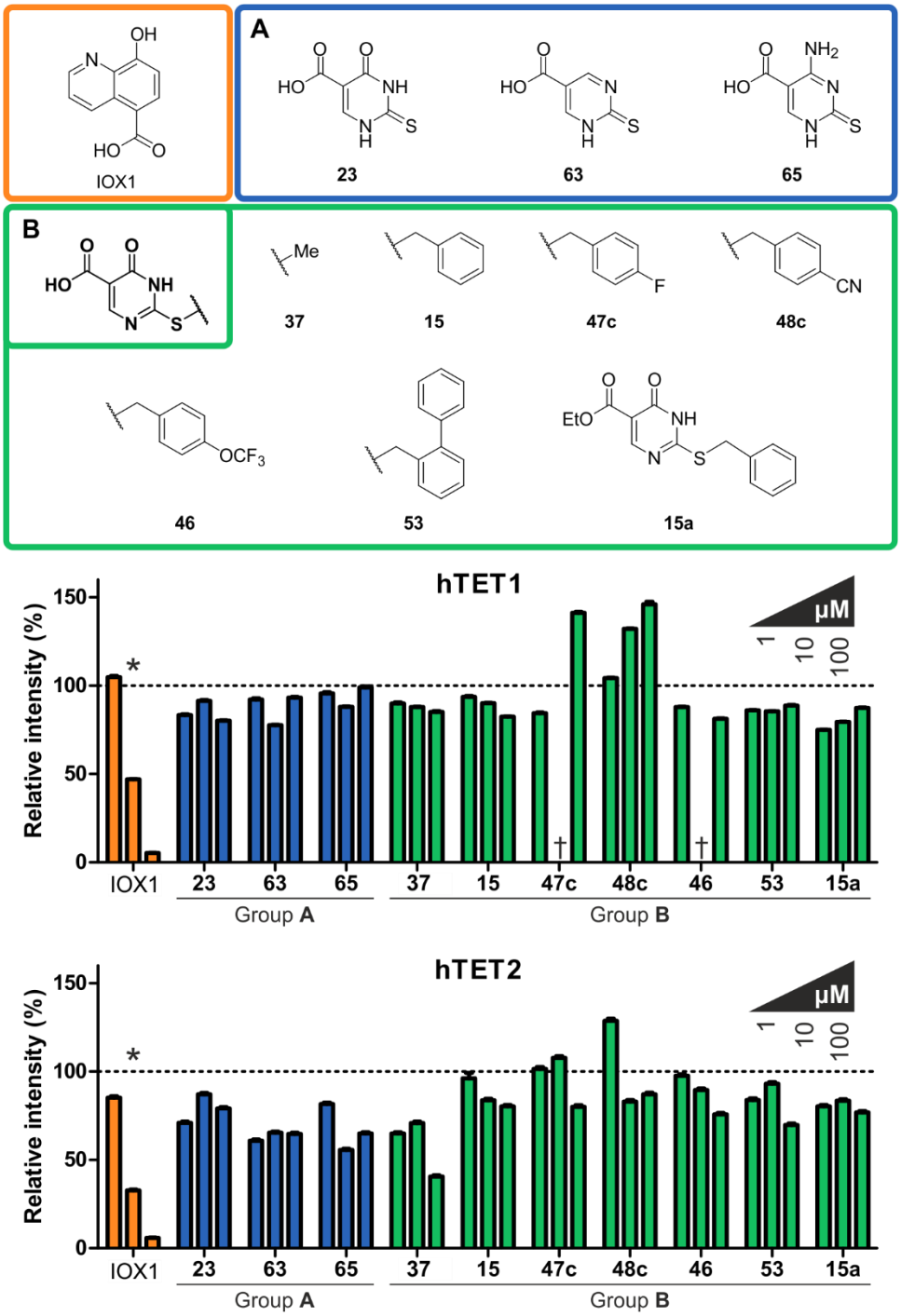
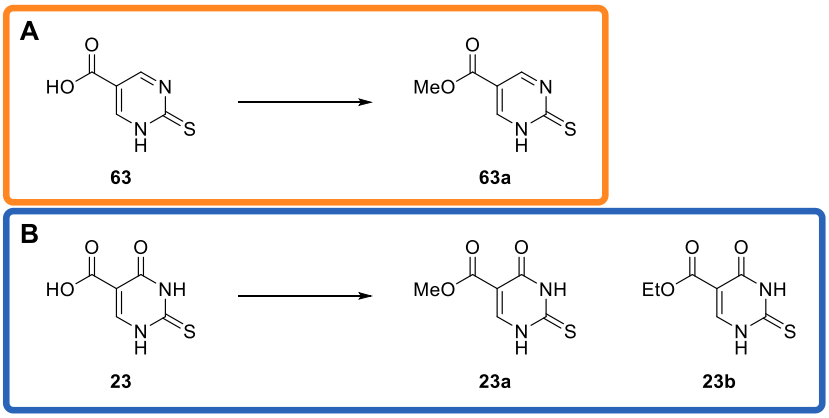


Figure 4.3 | Evaluation of cellular activity for selected 5-carboxy-2-thiouracil derivatives using dox-inducible U2OS cells stably transfected with FLAG-tagged hTET1/2. Top, structures of selected analogues; **middle**, anti-^{5hm}C staining bar charts for hTET1; **bottom**, anti-^{5hm}C staining bar charts for hTET2. Anti-^{5hm}C staining intensities were normalised to 1% DMSO vehicle control (100% relative intensity). Data shown as the average ± SEM (n ≤ 5,000 cells). *Final concentrations of IOX1 were 2, 20, and 200 μM. †Data excluded due to unusually low cell count (<20% of average cell count). Experiments were carried out by Dr Hilal Sarac.

4.3.9 Evaluation of ester prodrug cellular activity

In an effort to directly mimic the strategy employed to improve cell permeability for NOG and 2,4-PDCA, thione analogues **63** and **23** were selected for the preparation of the corresponding

methyl esters (**63a** and **23a**); an ethyl ester derivative of **23** (**23b**) was also prepared to facilitate a direct comparison of methyl and ethyl esters (Scheme 4.8).



Scheme 4.8 | Structures of parent thiones and their ester derivatives selected for cellular activity evaluation. **A**, **63** and its methyl ester, **63a**; **B**, **23** and its methyl (**23a**) and ethyl (**23b**) esters.

Table 4.7 | Tabulated AlphaScreen interference and hTET1–3 inhibition pIC_{50} values for parent thiones and their ester derivatives. Values reported as pIC_{50} ± SD (M). Values reported as percentage signal (%) at 100 µM for incomplete dose-response curves. N.i., no inhibition (>90% signal at 100 µM); n.d., not determined. Data for AlphaScreen antibody interference reported as the mean of 2–3 experimental replicates (n = 2–3; mean ± SD); data for enzyme inhibition reported as the mean of at least 3 independent replicates (n ≥ 3; mean ± SD).

Group	Name	Antibody interference	pIC_{50} ± SD (M) hTET1	hTET2	hTET3
IOX1		n.i.	5.86 ± 0.03	6.18 ± 0.17	5.96 ± 0.08
	63	n.i.	5.80 ± 0.08	6.71 ± 0.06*	5.31 ± 0.06
A	**63a**	n.i.	5.47 ± 0.14	5.69 ± 0.04	5.13 ± 0.19
	Δ_{63a-63}	-	*−0.33*	*−1.02*	*−0.18*
	23	n.i.	4.84 ± 0.24	5.71 ± 0.08	4.38 ± 0.22
	23a	n.i.	61.5 ± 9.5%*	4.54 ± 0.05	n.i.
B	Δ_{23a-23}	-	*<−0.7*	*−1.17*	*<−0.38*
	23b	n.i.	n.d.	62.1 ± 5.3%*	n.d.
	Δ_{23b-23}	-	-	*<−1.5*	-

**Data for enzyme inhibition reported as the mean of 2 independent replicates (n = 2; mean ± SD).*

Methyl esterification of **63** and **23** resulted in a >10-fold reduction in potency against hTET2 *in vitro* (Table 4.7). Inhibitory potency of methyl ester **63a** against hTET1/3 was largely preserved (ΔpIC_{50} = −0.33 and −0.18 for hTET1 and hTET3, respectively), whereas the half-maximal inhibitory concentrations of methyl ester **23a** against hTET1/3 were outside the range of the concentrations tested. Ethyl esterification (**23b**) resulted in approx. 100-fold

reduction in potency against hTET2, similarly to the *S*–benzyl ethyl ester **15a** (section 4.3.3), demonstrating that ethyl esters are not tolerated *in vitro*.

Treatment of cells with parent carboxylic acids (**63** and **23**) and ethyl ester **23b** elicited no cellular response (Figure 4.4), while methyl ester derivatives (**63a** and **23a**) reduced ^{5hm}C levels in a dose-dependent manner, with EC_{50} values around 10 μM (Table 4.8). Both **63a** and **23a** were most potent in cells overexpressing hTET3, but the compounds showed largely pan-specific activities (e.g., 2-fold and <4-fold selectivity for hTET3 over hTET2 and hTET1, respectively, for **63a**). Cellular inhibitory potencies of **63a** and **23a** were comparable to those of IOX1, in agreement with *in vitro* inhibition data for **63** and **23**, although additional biological replicates are required to accurately determine the EC_{50} values.

Table 4.8 | Tabulated cellular hTET1–3 inhibition pEC_{50} values for parent thiones and their ester prodrugs. Values reported as pEC_{50} ± SD (M). N.i., no inhibition (>90% signal at 100 μM). Data reported as the mean ± SD of n independent biological replicates (n number noted in brackets) with 2 technical replicates. Experiments were carried out by Dr Hilal Sarac.

Group	Name	pEC_{50} ± SD (M)		
		hTET1	**hTET2**	**hTET3**
IOX1		4.81 ± 0.04 (3)	5.39 ± 0.24 (2)	5.56 ± 0.32 (3)
A	**63**	n.i. (5)	n.i. (4)	n.i. (4)
	63a	4.98 ± 0.21 (5)	5.24 ± 0.17 (3)	5.56 ± 0.09 (4)
B	**23**	n.i. (2)	n.i. (1)	n.i. (2)
	23a	4.69 ± 0.02 (2)	4.76* (1)	4.99 ± 0.28 (2)
	23b	n.i. (2)	n.i. (1)	n.i. (2)

**Curve fitting and standard deviation are ambiguous due to unusually high Hill coefficient.*

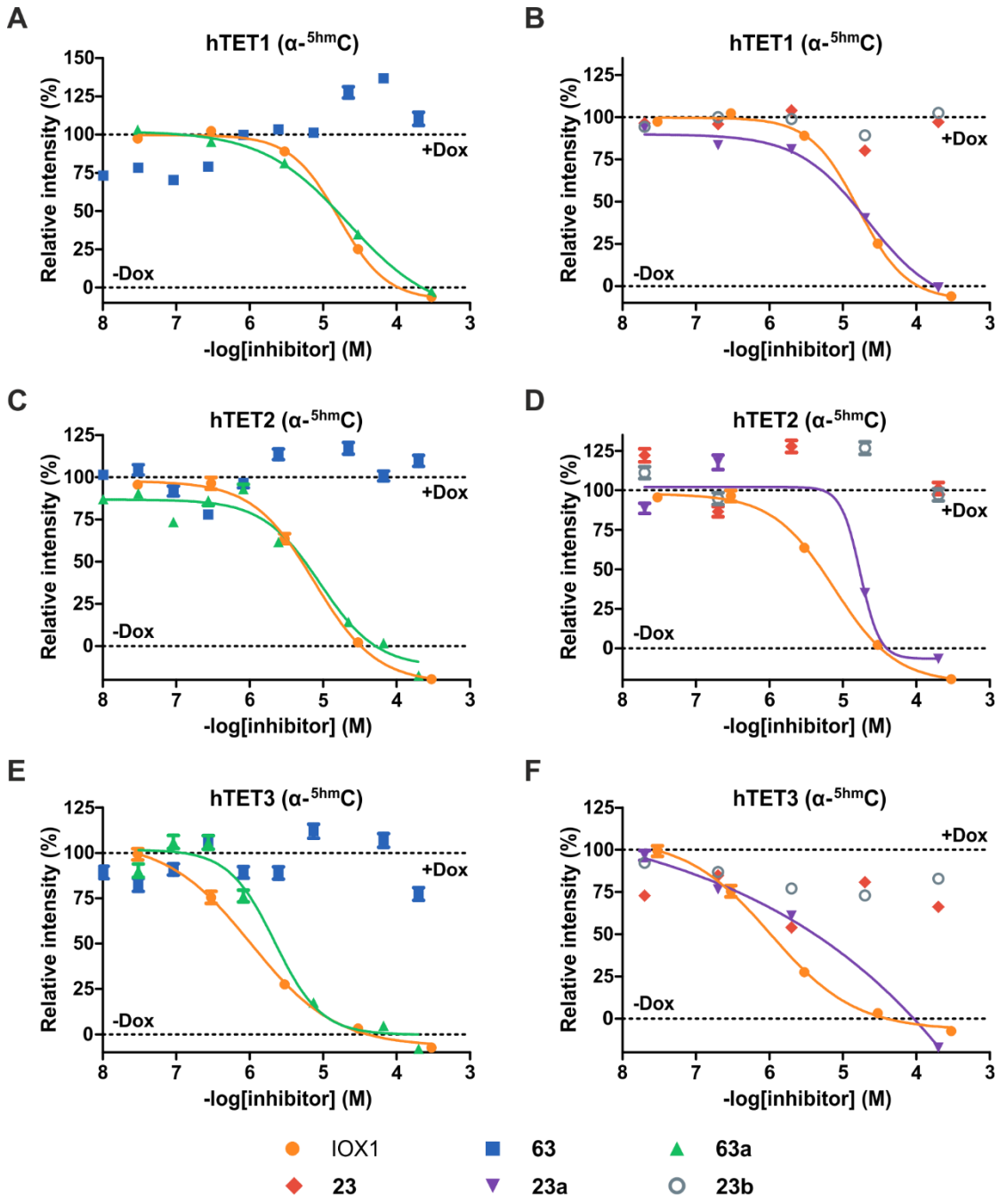


Figure 4.4 | Evaluation of the cellular activity of parent thiones and their ester prodrugs using dox-inducible U2OS cells stably transfected with FLAG-tagged hTET1–3. A, **B**, anti-^{5hm}C staining dose-response curves for hTET1; **C**, **D**, anti-^{5hm}C staining dose-response curves for hTET2; **E**, **F**, anti-^{5hm}C staining dose-response curves for hTET3. Data shown as the average ± SEM (n ≤ 5,000 cells); representative curves shown. Experiments were carried out by Dr Hilal Sarac.

4.4 Discussion

Screening of a library containing analogues of six scaffolds identified *S*–substituted 5-carboxy-2-thiouracils as single-digit micromolar inhibitors of hTET2. The thiouracil series was selected for further optimisation as the scaffold offered facile access to analogues *via* multiple functional handles and the compounds did not interfere with AlphaScreen signal, enabling rapid evaluation of inhibitory activity against hTET paralogues.

A report on mTet-catalysed oxidation of thymine to 5-hydroxymethyluracil[269] outlined the series as potential ligand mimetic-based inhibitors, although the scaffold also shared structural features with PHD inhibitor GSK1278863[259] containing a 1,3-dicarbonyl moiety that acts as an Fe(II)-chelating motif. Contrary to this, a crystal structure of **17** bound to KDM5B (PDB ID: 5FZI, IC_{50} = 45 μM)[265] proposed that the *S*–alkylated analogue occupied the 2-OG binding site without directly interacting with the active site metal ion, Mn(II). However, the observed binding mode may not necessarily extend to hTETs given the differences in active site configuration between the two families. Further work is required to determine the 2-thiouracil derivative binding site for hTETs, particularly for *S*–unsubstituted derivatives, which may be binding in a manner distinct from *S*–substituted analogues as suggested by SAR data collected herein.

Characterisation of *S*–alkyl and *S*–(aryl)alkyl analogue inhibitory activity against hTET2 revealed that *S*–substituent variation did not substantially affect potency; an analogue that was significantly more potent than **15** (pIC_{50} = 5.93) was not identified, while the observed reductions of inhibitory potency were driven primarily by the increasing steric bulk of *S*–substituents. The relatively broad tolerance of functional groups and *S*–substituent sizes suggested that the vector of *S*–substitution may be pointing away from key structural features within the inhibitor binding site or outwards into the solvent. *S*–substitution may thus offer a

means of obtaining selectivity over 2-OG dioxygenases with less sterically accommodating binding sites.

Modification of the 4-oxo and 5-carboxylic acid positions on *S*–benzyl analogue **15** resulted in a >30-fold reduction of hTET2 potency, highlighting the importance of the β-ketoacid moiety. Conversely, the 6-carboxylic acid analogue **26** exhibited only a partial loss of potency ($\Delta pIC_{50} = -0.67$), suggesting that the compound may bind to hTET2 in a manner different from the 5-carboxy-2-thiouracils. Scaffold modifications at 4-/5-positions on the *S*–unsubstituted 5-carboxy-2-thiouracil **23** (hTET2 $pIC_{50} = 5.71$) were tolerated well compared to *S*–substituted analogues, typically resulting in a less than 10-fold loss in potency. Variation of the carboxylic acid position was not only well-tolerated, but afforded the most potent hTET2 inhibitor within the series, **56** ($pIC_{50} = 6.85$). 4–*H* thione analogue **63** exhibited enhanced potency compared to **23** (hTET2 $\Delta pIC_{50} = 1.00$), in contrast with the *S*–methylated 4–*H* analogue **33** (n.i. against hTET2), providing further indication of distinct modes of binding between *S*–substituted and thione analogues. While thiones **56** and **63** were the most potent hTET2 inhibitors within the series, the fragment-like nature of the structures cautioned that the compounds may act in a pan-specific manner.

hTET paralogue selectivity profiling using AlphaScreen revealed that the 5-carboxy-2-thiouracil series inhibited hTET2 preferentially to varying extent, but strong correlations between hTET1–3 pIC_{50} values (Spearman's r_s = 0.661–0.910) signified that further improvement of selectivity profiles would be difficult. *S*–benzyl analogue **15** exhibited 10-fold selectivity over both hTET1 and hTET3, while *S*–butyl **14** was markedly selective over hTET3 (8- and 50-fold selectivity over hTET1/3, respectively). Profiling of **15** and **14** against a panel of 2-OG-dependent dioxygenases by Nowak *et al.*[265] (Table 4.9) suggested that **15** was >10-fold selective across the board with the exception of ribosomal oxygenase MINA53 (3-fold

selectivity), while **14** exhibited a similar selectivity profile with good selectivity over MINA53 but reduced selectivity over KDM4A (7-fold). 5-Carboxy-2-thiouracil **23** was shown to be a single-digit micromolar inhibitor of hTET2 with some paralogue selectivity (pIC_{50} = 4.84, 5.71, and 4.38 for hTET1–3, respectively); furthermore, the compound reportedly showed poor inhibitory potencies (pIC_{50} < 4) against the majority of the 2-OG dioxygenase panel (Table 4.9), demonstrating close to a 100-fold selectivity for hTET2. However, care should be taken while inferring inhibitor selectivity profiles from IC_{50} data, as, unlike the more reliable absolute inhibition constant K_i, IC_{50} values are relative, and can thus vary substantially with different enzyme, substrate, and cofactor concentrations[270,271]. Regardless, the collected and reported data together provide at least some evidence towards selective hTET2 inhibition.

Table 4.9 | Reported[265] 2-OG-dependent dioxygenase inhibition pIC_{50} values for compounds 15, 14, and 23. Values reported as pIC_{50} (M). N.i., no inhibition (pIC_{50} < 4). Data for enzyme inhibition reported as the mean of at least 3 independent replicates (n ≥ 3).

Name	pIC_{50} (M)								
	NO66	**MINA53**	**KDM2A**	**KDM3B**	**KDM4A**	**KDM4D**	**KDM5B**	**KDM6B**	**PHD2**
15	4.32	5.42	4.17	n.i.	4.61	4.46	4.07	4.13	4.41
14	n.i.	4.88	4.62	4.21	5.04	4.81	4.67	4.37	n.i.
23	n.i.	n.i.	n.i.	4.27	n.i.	n.i.	n.i.	n.i.	n.i.

While 4–*H* thione analogue **63** exhibited a strong preference for hTET2 (5-fold over hTET1 and 40-fold over hTET3), 4-carboxylic acid analogue **56** inhibited the hTET panel in a non-specific manner (1.4-fold selective over hTET1 and 3-fold over hTET3). Loss of the bidentate iron-chelating moiety resulted in a 10-fold enhanced potency against hTET2 in both cases, whereas variation of the carboxylic acid position caused a considerable separation in selectivity profiles; together, the results suggest that **63** and **56** interact with hTETs in distinct manners.

5-carboxy-2-thiouracil derivatives exhibited no activity against hTET1/2 in U2OS cells, although **16** and **18** were reportedly active against MINA53 in a variety of solid cancer cell lines[265]. Ethyl ester derivatives of **23** and **15** were similarly inactive in cells; conversely,

derivatisation of **63** and **23** to methyl esters (**63a** and **23a**) rescued cellular activity against hTETs. However, the selectivity profiles of the methyl esters were only comparable to the pan-specific profile of IOX1 (<5-fold paralogue selectivity). While ethyl esters were generally inactive *in vitro*, methyl ester analogues (particularly **63a**) maintained a reasonable level of inhibitory activity against hTETs by AlphaScreen, albeit with substantially deteriorated selectivity profiles. Hence, the observed pan-specific effect on hTETs *in cellulo* may be due to non-selective inhibition by the methyl esters; in contrast, **23a** showed poor to no inhibitory potency against hTET1/3 *in vitro*, but still exhibited a pan-specific profile *in cellulo*. The observed discrepancies in the inhibitory profiles could be attributed to differences between the biochemical and cell-based systems, particularly with regard to hTET2 construct use (*in vitro*, $hTET2_{CD\Delta LCI}$; *in cellulo*, $hTET2_{CD}$), but the identity of the active inhibitor species in a cellular context warrants further investigation. Characterisation of methyl ester pharmacokinetics, particularly metabolism to the carboxylic acid forms, could be carried out to determine the proportion of ester and acid inhibitor forms at different time points *in cellulo*.

4.5 Conclusions and future work

S–alkylated and *S*–unsubstituted (thione) 5-carboxy-2-thiouracils were identified as single-digit micromolar hTET2 inhibitors. SAR around *S*–substitution did not show definite trends, while scaffold derivatisation at positions 4- and 5- indicated that the β-ketoacid moiety was essential for the activity of *S*–substituted analogues. Conversely, thione derivatives tolerated a number of scaffold modifications, suggesting distinct mechanisms of inhibition for the two groups of analogues. Reported structural characterisation of *S*–alkylated **17** bound to KDM5B (ref. [265]) indicated that the inhibitor occupied the 2-OG binding site without interacting with the active site metal, but structural differences between KDM5B and hTET families preclude analogous binding site assignment for hTETs. Evaluation of potency dependence on DNA substrate and 2-OG concentration could be carried out to ascertain if the inhibitors are competitive with regards to substrate and/or 2-OG.

hTET paralogue selectivity profiling revealed varying degrees of selectivity for hTET2: *S*–benzyl analogue **15** showed 10-fold selectivity over hTET1/3, while *S*–butyl **14** exhibited excellent selectivity over hTET3 (nearly 50-fold) and 8-fold selectivity over hTET1. However, strong Spearman's rank order correlations between potencies for hTET1–3 indicated that any subsequent selectivity profile optimisation would likely be challenging. Importantly, the compounds reportedly showed an indication of broad selectivity over representative Fe(II)/2-OG-dependent dioxygenase superfamily enzymes[265]. Co-crystallisation of selected analogues with $hTET2_{CD\Delta LCI}$ could substantially expedite the design of subsequent inhibitors, as it would not only inform the binding mode, but also the tolerated vectors and steric constraints for analogue derivatisation (work towards $hTET2_{CD\Delta LCI}$ crystallisation is discussed in Chapter 5).

5-Carboxy-2-thiouracils and their ethyl ester derivatives did not exhibit cellular activity, while methyl esters of *S*–unsubstituted **63** and **23** showed pan-like inhibitory profiles *in cellulo*. It is unclear whether the active species *in cellulo* are the methyl esters or the carboxylic acids, as the observed cellular activity was largely in line with the *in vitro* selectivity profiles of the methyl ester analogues. Metabolic profiling of the methyl esters to measure the concentrations of ester and free acid species at different time points would potentially allow to identify the primary inhibitory species in the cellular environment. Furthermore, methyl esters of *S*–benzyl **15** and *S*–butyl **14** should be investigated to determine if *S*–substituted analogues exhibit a more favourable selectivity profile *in cellulo*.

Chapter 5
Production and crystallisation of human TET2 truncated catalytic domain, hTET2$_{CD\Delta LCI}$

5.1 Introduction

Structural studies of 2-OG dioxygenases in complex with natural substrates or small molecule inhibitors have produced a number of insights, particularly in the case of the well-characterised KDM[184] and PHD[272] families. Conversely, structural characterisation of the TET family has been limited, with only hTET2 and NgTet1 crystal structures in complex with DNA substrates and NOG available to date, as well as structures of the functionally distinct *C. reinhardtii* TET homologue CMD1 (section 1.4). Only three crystal structures have been reported for hTET2 in complex with Fe(II), NOG, and DNA substrates containing ^{5m}C, ^{5hm}C, or ^{5f}C nucleotides (Figure 1.4).

Limited structural information of hTET2 in complex with inhibitors (other than NOG) hampers inhibitor design. Co-crystallisation of TETs with TiP1 (Chapter 3) and selected 5-carboxy-2-thiouracil analogues (Chapter 4) would expedite structure-guided optimisation of TiP1 cellular activity (a strategy successful for KDM4A inhibitor CP2[241]) and design of small molecule inhibitors with enhanced selectivity and potency profiles, respectively.

5.2 Aims

The yield of the crystallisable hTET2 construct, $hTET2_{CD\Delta LCI}$ (Figure 1.3), recombinantly expressed in *E. coli* is approx. 0.3–0.5 mg/L (personal correspondence with authors[94,96]), highlighting the difficulty of protein isolation. Thus, optimisation of hTET2 production will be required to facilitate isolation of enzyme in quantities necessary for crystallography studies. Subsequent identification of a robust set of crystallisation conditions could allow crystal soaking with a number of small molecules of interest, including pan-2-OG dioxygenase inhibitors (IOX1, 2,4-PDCA) and 5-carboxy-2-thiouracil analogues (Chapter 4). Furthermore, co-crystallisation with cyclic peptide inhibitor TiP1[191] would support subsequent structure-guided optimisation of TiP1 properties for improved cellular activity, a strategy applied successfully to the KDM4A inhibitor CP2[241]. Structures of hTET2 with small molecule or CP inhibitors would not only improve the understanding of inhibitor mode of action, but could also enable rapid design of selective inhibitors that target active site features specific to hTETs.

The aims of the work presented in this chapter are as follows:

- Produce catalytically active $hTET2_{CD\Delta LCI}$ for biochemical characterisation.
- Optimise production of $hTET2_{CD\Delta LCI}$ to obtain high purity protein in quantities sufficient for crystallisation.
- Reproduce hTET2 crystallisation using previously reported conditions[94] / identify novel crystallisation conditions.

5.3 Results

5.3.1 Production of SUMO-tagged hTET2$_{CD\Delta LCI}$

hTET2 was expressed as the reported crystallisable catalytic domain construct[94,96] with the low complexity insert substituted for a GS-linker (CDΔLCI, Figure 1.3). Removal of the unstructured low complexity insert facilitates crystallisation by improving unit packing without having a deleterious effect on catalytic activity. Furthermore, the construct is expressible in *E. coli* expression systems owing to its lower molecular weight compared to the full-length and non-truncated catalytic domain constructs (223,811.1 Da and 90,165.3 Da, respectively, *vs* 51,260.3 Da for the CDΔLCI construct).

Initial production of catalytically active hTET2 followed conditions similar to those reported previously[94]: a modified pET-28b plasmid kindly gifted by Prof. Yanhui Xu[94] (State Key Laboratory of Genetic Engineering, Fudan University, China) containing an open reading frame coding for hTET2$_{CD\Delta LCI}$ (abbreviated from hereon as hTET2) with an N–terminal His$_6$-FLAG-[thrombin cleavage site]-SUMO (ubiquitin-like protein Smt3) tag was transformed into *E. coli* BL21(DE3) strain; transformants containing the plasmid were grown in LB or TB medium at 37 °C until appropriate cell density was reached; cells were induced with IPTG (500 μM) and incubated at 16 °C for 18 h (Figure 5.1 A). hTET2 was purified from cell lysate by immobilised metal affinity chromatography using Ni Sepharose® High Performance resin (HisTrap IMAC), followed by size exclusion chromatography using Superdex® 200 resin (SEC S200, Figure 5.1 B). Cells were lysed by sonication, and His$_6$-tagged proteins were captured using HisTrap IMAC. HisTrap afforded a single peak by UV containing the protein of interest by SDS-PAGE (Figure 5.1 C, D), although the efficiency of affinity capture (input *vs* FT lanes, SDS-PAGE gel) could not be confirmed due to low levels of hTET2 overexpression.

Subsequent SEC revealed the presence of multiple species by UV, all of which contained the protein of interest by SDS-PAGE (Figure 5.1 E, F): species eluting immediately after column void volume (orange circle) were likely protein aggregates; the peak eluting at approx. 75 mL (blue square) was identified as monomeric hTET2 based on calibration curves provided by the column manufacturer[273]; the final peaks (green triangle) corresponded to lower molecular weight impurities. Fractions containing monomeric hTET2 were pooled, affording tagged hTET2 in consistent albeit low yields and moderate purity (Figure 5.1 F, G, approx. 0.4 mg/L, stored at 9.8–10.0 mg/mL in 20 mM HEPES, 150 mM NaCl, pH 7.5).

A

Strain	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
BL21(DE3)	LB TB	37 °C	0.6–0.7 0.9–1.0	500 µM	16 °C	18 h

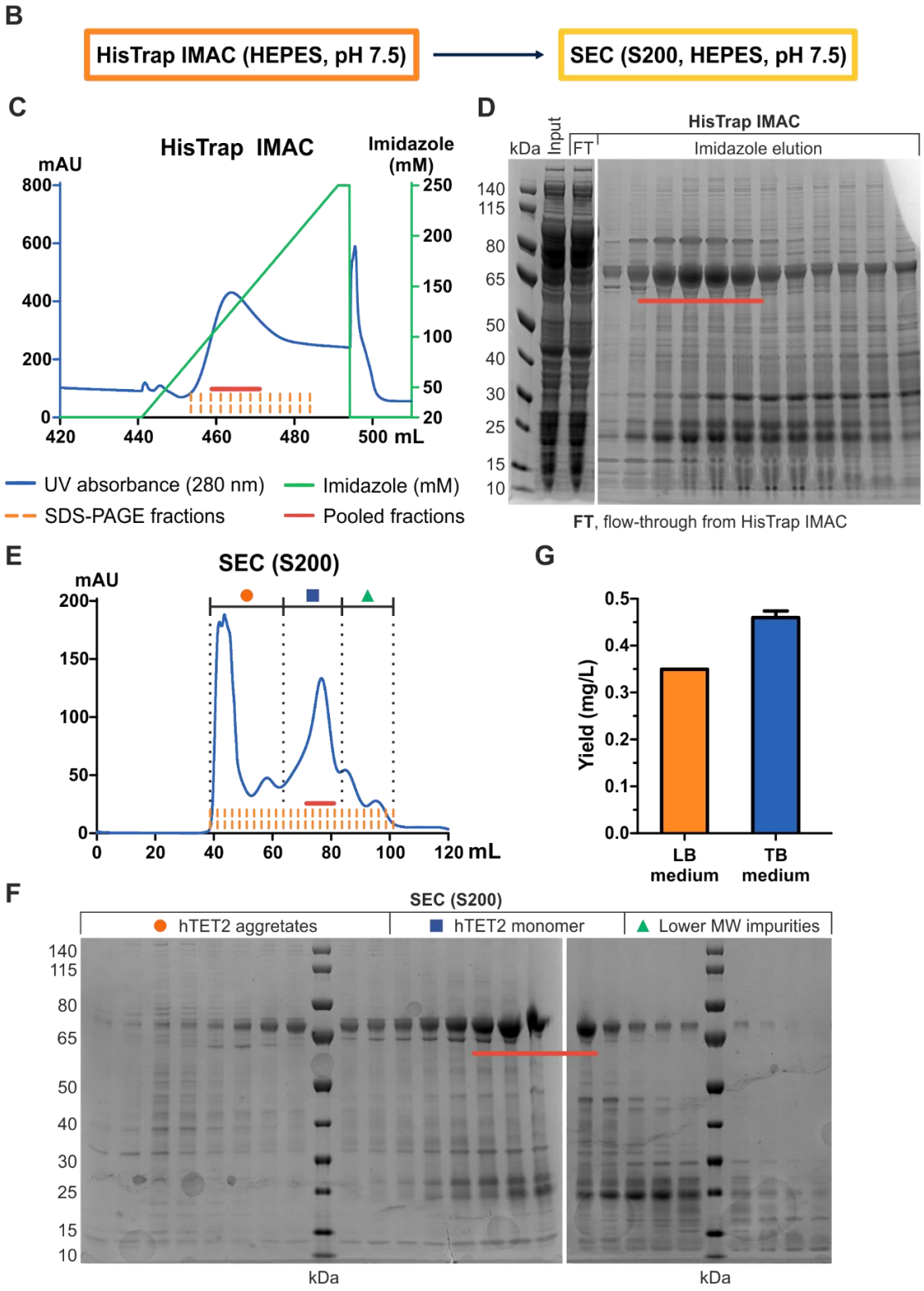


Figure 5.1 | Production of His_6-FLAG-[thrombin]-SUMO-tagged hTET2. A, cell pellet growth conditions (LB, Luria-Bertani; TB, Terrific Broth); **B**, purification workflow; **C**, representative HisTrap IMAC 280 nm elution profile; **D**, representative SDS-PAGE gel for HisTrap IMAC (input, clarified lysate; FT, HisTrap loading flow-through, predicted mass: 66.0 kDa); **E**, representative size exclusion (S200) 280 nm elution profile; **F**, representative SDS-PAGE gel for size exclusion (S200); **G**, tagged hTET2 yields in mg/L of culture (n=1 for LB, n=2 for TB, mean ± SD shown). Red lines indicating pooled fractions are placed underneath the protein of interest bands in SDS-PAGE gels.

The identity of the isolated protein could not be unambiguously confirmed using LC-MS, as a cluster of peaks was observed around the predicted mass for hTET2 (predicted mass: 65,951.6 Da, Figure 5.2 A). The main peak in the region could correspond to a loss of N–terminal methionine *via* NME[199] and α-*N*–gluconoylation[200] (predicted mass: 65,998.4 Da, observed mass: 65,997.8 Da), post-translational modifications also present in recombinantly produced NgTet1 (Chapter 2). LC-MS analysis also suggested the presence of *E. coli* Hsp70 molecular chaperone, DnaK, a common contaminant in fusion protein purification[274] (predicted mass with loss of N–terminal methionine: 68,983.8 Da, observed mass: 68,982.8 Da). Although the purity of the isolated hTET2 was low, ^{5m}C hydroxylase activity could be observed using AlphaScreen (conditions optimised in section 2.3.3). Catalytic activity could be observed with 10 nM of monomeric hTET2 (Figure 5.2 B), producing similar turnover levels as 200 nM of NgTet1 after 30 min (Figure 2.5). A comparison of monomer and aggregate hTET2 species revealed that the enzyme was only active in the monomeric state (Figure 5.2 C).

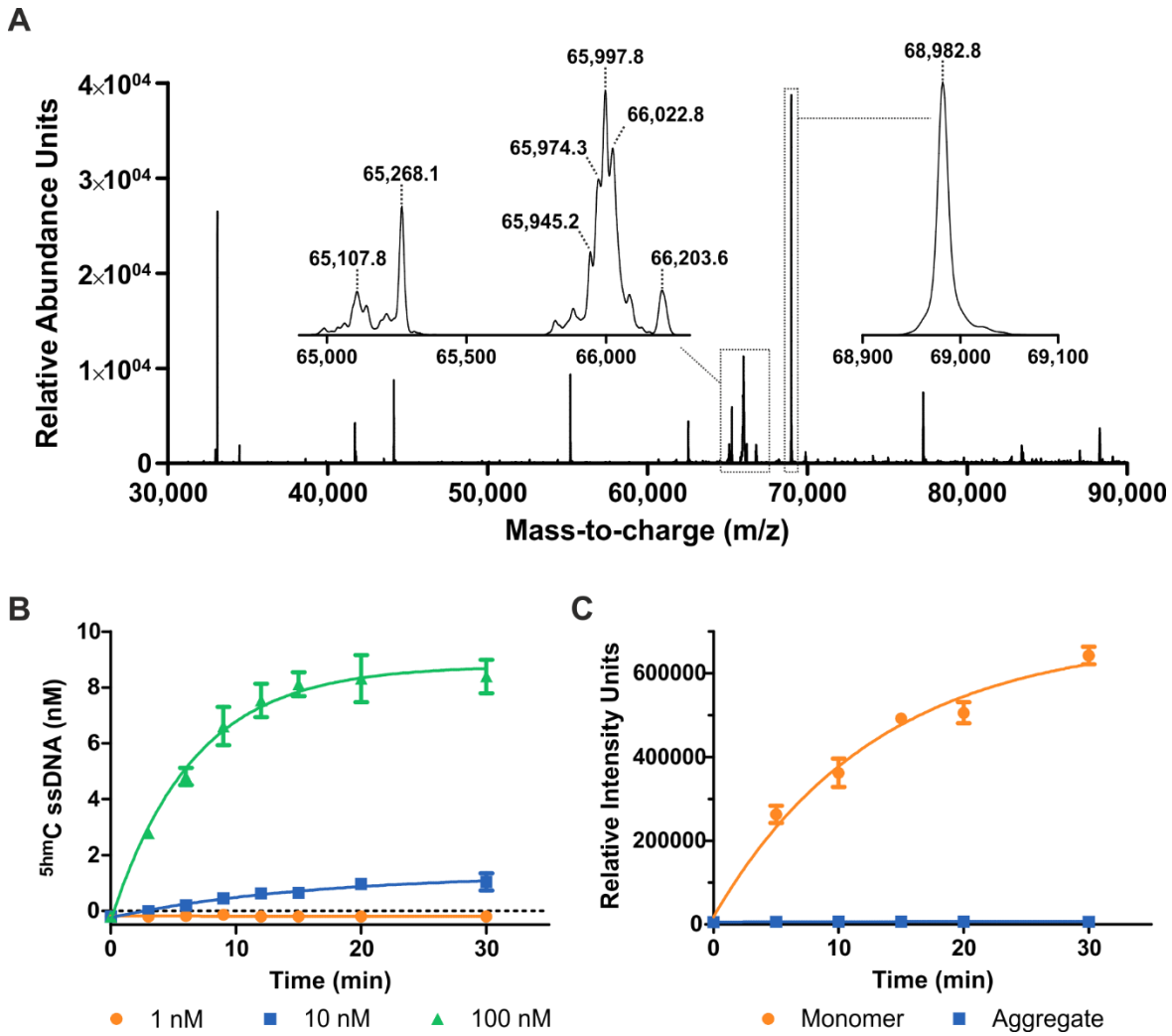


Figure 5.2 | Characterisation of purified tagged hTET2. A, LC-MS analysis (predicted mass for methionine-excised[199] α-*N*–gluconoylated[200] hTET2$_{CD\Delta LCI}$: 65,998.4 Da, observed mass: 65,997.8 Da); **B**, monomeric hTET2 catalytic activity detection using AlphaScreen; **C**, hTET2 monomer and aggregate (10 nM) catalytic activity comparison using AlphaScreen. AlphaScreen conditions: 1–100 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. Data shown as the mean of 3 experimental replicates (n = 3; mean ± SD).

5.3.2 SUMO-tagged hTET2$_{CD\Delta LCI}$ crystallisation trials

While crystallisation of proteins containing large solubility tags is generally considered challenging due to the conformational heterogeneity introduced by the presence of the fusion tag, the low yield of hTET2 discouraged the potentially problematic tag cleavage step and additional purification steps. A number of proteins have been successfully crystallised with solubility tags present, enabled by a rigid three- to five-amino acid linkage between the solubility tag and the protein of interest[275]. Although most of the reported structures contained either MBP, GST, or Trx tags, a number of crystal structures containing the Smt3 fusion tag have been deposited (e.g., PDB ID: 6P81, 4FN2, 6O4A). The presence of a relatively

short spacer between Smt3 and hTET2 in the construct used herein (…Smt3-GGS-hTET2$_{CD\Delta LCI}$), coupled with precedent of Smt3 fusion protein crystallisation encouraged attempts at crystallisation of the SUMO-tagged hTET2.

Tagged hTET2 co-crystallisation trials were carried out with cyclic peptide TiP1, which was shown to inhibit hTET2 with IC_{50} = 25 µM (Table 3.2). Tagged hTET2 (9.83 mg/mL, 149 µM) was complexed with $MnCl_2$ (2.5 mM), NOG (2 mM), and TiP1 (2.2 mM, 1% v/v DMSO final concentration), plated with PEG/Ion, Crystal Screen (1+2), SaltRx, and JCSG Plus™ broad crystallisation screens, and stored at 22 °C. Although the reported hTET2 structures contained Fe metal in the active site[94,96], Mn(II) was used to prevent spontaneous metal oxidation. After 14 days, 1/3 of the wells showed either flocculent or granular precipitation, while the remaining 2/3 of the wells remained clear, indicating that the protein concentration was likely too dilute. However, concentrating the tagged hTET2 further proved challenging, as substantial protein loss was observed when concentrating above 10 mg/mL using 30 kDa MWCO spin concentrators. Furthermore, the low yields of tagged hTET2 precluded exploration of follow-up crystallisation conditions. Evidently, production of hTET2 in higher yield and purity was a requisite for a sustained crystallography effort.

5.3.3 Optimisation of hTET2$_{CD\Delta LCI}$ expression

5.3.3.1 Fusion construct preparation

Optimisation of hTET2 expression began with the preparation of fusion constructs containing different solubility tags. The gene coding for hTET2$_{CD\Delta LCI}$ was amplified by PCR from the pSUMO plasmid used in section 5.3.1, and subcloned into five destination vectors using restriction or ligation-independent cloning (LIC, Table 5.1, construct sequences and information reported in Appendix 4). For brevity, constructs containing solubility tags will be referred to as their solubility tags from hereon (SUMO, Trx, GST, GFP, and MBP); pCOLD I will be referred to as pCOLD.

Table 5.1 | hTET2 fusion construct information. His_6, hexahistidine tag; FLAG, DYKDDDDK tag; Thr, thrombin protease cleavage site; SUMO, ubiquitin-like protein Smt3; TEV, TEV protease cleavage site; Trx, thioredoxin; GST, glutathione *S*–transferase; GFP, green fluorescent protein; MBP, maltose-binding protein; TEE, translation enhancing element; Xa, factor Xa cleavage site; Kan, kanamycin; Amp, ampicillin.

Name	Cloning method	Tag	Construct MW (kDa)	Terminus	Resistance
pSUMO	Parent vector	His_6-FLAG-Thr-SUMO	66.0	N–terminal	Kan
pNH-TrxT	LIC	His_6-Trx-TEV	65.3	N–terminal	Kan
pGTVL2	LIC	His_6-GST-TEV	79.6	N–terminal	Kan
1GFP	LIC	His_6-GFP-TEV	80.7	N–terminal	Kan
1M	LIC	His_6-MBP-TEV	94.4	N–terminal	Kan
pCOLD I	Restriction	TEE-His_6-Xa	55.5	N–terminal	Amp

Codon optimisation was considered prior to the subcloning effort, as the gene coding for hTET2 in the parent vector (SUMO) was taken directly from *Homo sapiens* chromosome 4, and contained a number of rare *E.coli* codons, including AGG, AGA, and CGA, which are known to lead to decreased translation rates or premature termination[276,277]. However, elimination of rare codons that act as translational pauses and retard the translation rate has been shown to lead to protein misfolding and aggregation[278]. Furthermore, codon optimisation has been linked to changes in translational kinetics leading to altered protein conformation in *E. coli*[279,280] as well as higher expression systems[281]. In order to reproduce the reported crystallisable protein form[94] as closely as possible, the gene coding for hTET2 was left unaltered.

Three *E. coli* strains were selected for the expression trials (Table 5.2): BL21(DE3)-R3-pRARE2 was selected as a rare codon tRNA-enriched strain[282]; Lemo21(DE3) was selected to enable translation rate control *via* L-rhamnose-dependent control of T7 lysozyme LysY levels[283]; and SHuffle® T7 strain was selected for the chaperone activity of disulfide bond isomerase DsbC[284,285]. For brevity, BL21(DE3)-R3-pRARE2 will be referred to as R3-pRARE2; Lemo21(DE3) as Lemo21; SHuffle® T7 as SHuffle from hereon.

Table 5.2 | *E. coli* strains selected for expression optimisation. Camp, chloramphenicol.

Name	Parent strain	Properties	Resistance
BL21(DE3)-R3-pRARE2	BL21(DE3)	Contains pRARE2 plasmid, which carries seven rare-codon tRNA genes (isolated from Rosetta 2 strain)[282]	Camp
Lemo21(DE3)	BL21(DE3)	Tuneable T7 expression by L-rhamnose titration[283]	Camp
SHuffle® T7	K-12	Expresses disulfide bond isomerase with chaperone activity, DsbC[284,285]	-

5.3.3.2 Fusion construct and *E. coli* strain optimisation using detergent-based lysis

All six constructs were expressed in each of the three *E. coli* strains under conditions similar to those used in section 5.3.1. Transformants were grown at 30–37 °C in TB media until sufficient optical density (OD_{600}) was reached; cells were induced with IPTG (500 µM) and incubated at 16 °C for 18 h (Table 5.3). All transformants carrying the pCOLD vector were cold-shocked (4 °C, 30 min) prior to IPTG addition.

Table 5.3 | Cell growth conditions for fusion construct and *E. coli* strain optimisation. TB medium was used for all growths. All cells were induced with IPTG (500 µM). Transformants carrying the pCOLD vector were cold-shocked (4 °C, 30 min) prior to IPTG addition.

Strain	Cultivation temperature	OD_{600} prior to induction	Post-induction temperature	Post-induction time
R3-pRARE2	37 °C	0.9–1.0	16 °C	18 h
Lemo21	37 °C	1.2–1.4	16 °C	18 h
SHuffle	30 °C	0.6–1.0	16 °C	18 h

Cell pellets were lysed using detergent (BugBuster® protein extraction reagent), and His_6-tagged proteins were purified using free-flowing Ni Sepharose® High Performance IMAC resin. SDS-PAGE analysis of total (unclarified lysate), soluble (clarified lysate), and Ni Sepharose-purified fractions revealed that hTET2 expression was poor across all tested conditions (Figure 5.3 A), with hTET2 overexpression levels too low to compare hTET2 quantities in the soluble protein fractions (S). Affinity purification did not effect a substantial enrichment of His_6-tagged hTET2 fusion proteins, hindering quantitative comparison of the fusion construct expression levels. Furthermore, SUMO and Trx construct bands in SDS-PAGE

gels (66.0 and 65.3 kDa, respectively) were obscured by overlap with co-running bands (approx. 70 kDa), further complicating analysis. The combination of low overexpression levels and poor purity after affinity purification necessitated a more selective target protein visualisation method.

As all 6 fusion constructs contained a His_6 tag, expression levels could be assessed by Western blotting for His_6 (Figure 5.3 B). The tagged hTET2 (SUMO) produced in section 5.3.1 was used as a positive control (2 μg loading), validating the anti-His_6 antibody (lane C). Total protein fraction analysis by Western blot revealed the presence of His_6 tags in the majority of conditions explored, although a large number of bands not corresponding to the fully translated protein of interest was observed, particularly prevalent in GFP and MBP constructs. While the multiple bands observed by Western blotting could be the result of premature translation termination, Dyson *et al*. demonstrated that GFP and MBP fusions may be particularly prone to proteolysis during cell lysis[286]. hTET2 expression levels in *E. coli* Shuffle strain were consistently lower than in R3-pRARE2 and Lemo21 strains; thus, Shuffle strain was excluded from further optimisation trials.

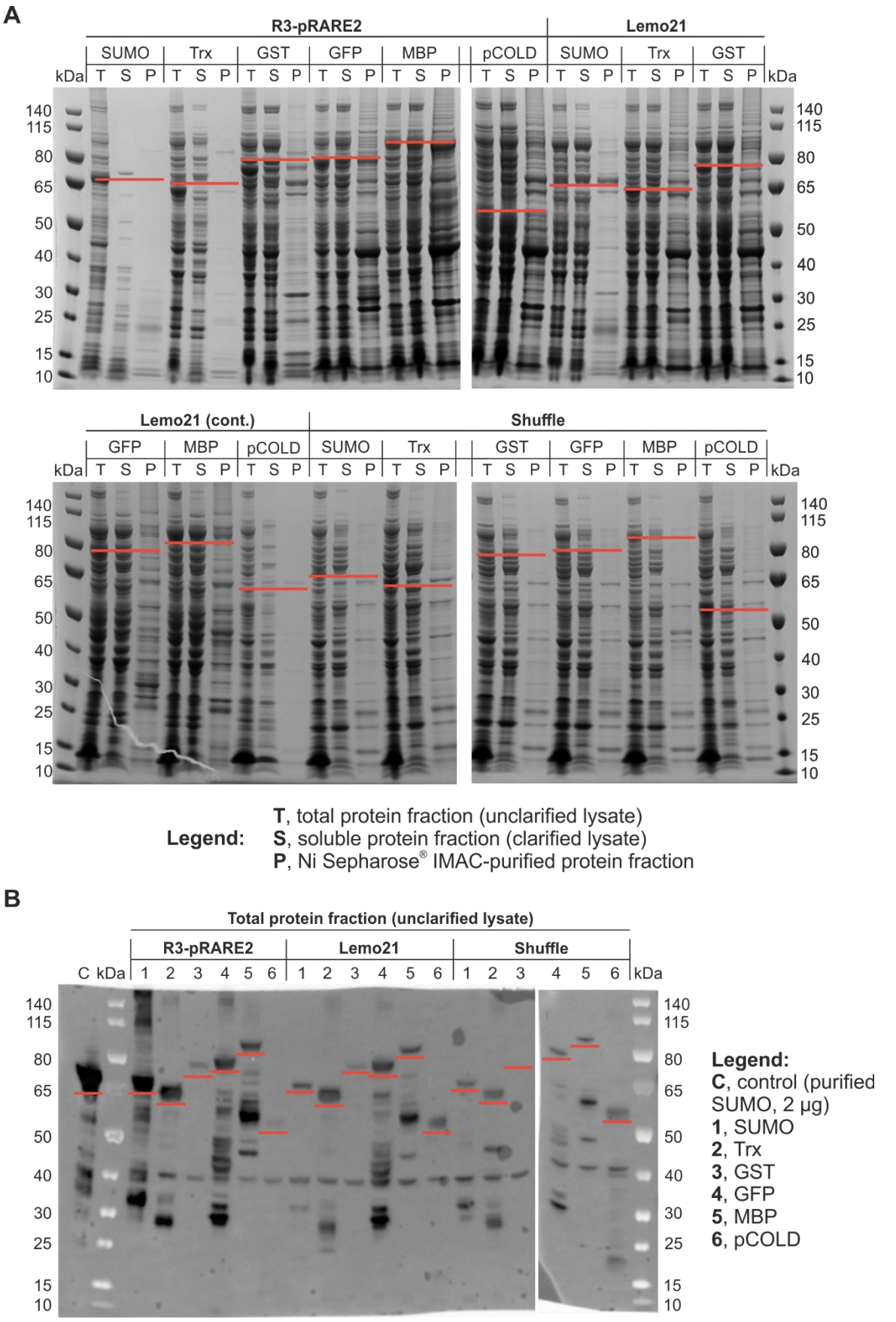


Figure 5.3 | Fusion construct and *E. coli* strain optimisation using detergent-based lysis. A, SDS-PAGE analysis of total (unclarified lysate), soluble (clarified lysate), and Ni Sepharose® IMAC-purified protein fractions; **B**, Western blot analysis of total (unclarified lysate) protein fractions. Anti-His_6 antibody was used for Western blot visualisation. Tagged hTET2 (SUMO) produced in section 5.3.1 was used as a positive control in Western blotting (2 µg loading). Red lines are placed underneath the approximate position of protein of interest bands.

5.3.3.3 Fusion construct and *E. coli* strain optimisation using physical disruption lysis

Detergent-based cell lysis (e.g., BugBuster® protein extraction reagent) provides a solubilising effect which is not present when using large-scale physical disruption methods such as sonication or high-pressure homogenisation. To ensure that results from small-scale expression trials translate well to larger-scale protein production, sonication was used as the lysis method in all subsequent expression trials.

Expression of all six constructs was repeated in R3-pRARE2 and Lemo21 strains using conditions from Table 5.3, but detergent-based cell lysis was replaced with sonication. Western blot analysis of total (unclarified lysate) and Ni Sepharose-purified protein fractions (Figure 5.4) revealed trends comparable to results from the initial optimisation round (Figure 5.3). Expression levels in Lemo21 strain were on average lower than in R3-pRARE2 strain. GST and pCOLD constructs showed very low expression levels, while SUMO, Trx, GFP, and MBP constructs expressed relatively well. SUMO and Trx construct expression levels were comparable in the total protein fraction, but affinity purification resulted in a substantial decrease in Trx construct levels, suggesting that the majority of the enzyme was produced in insoluble form.

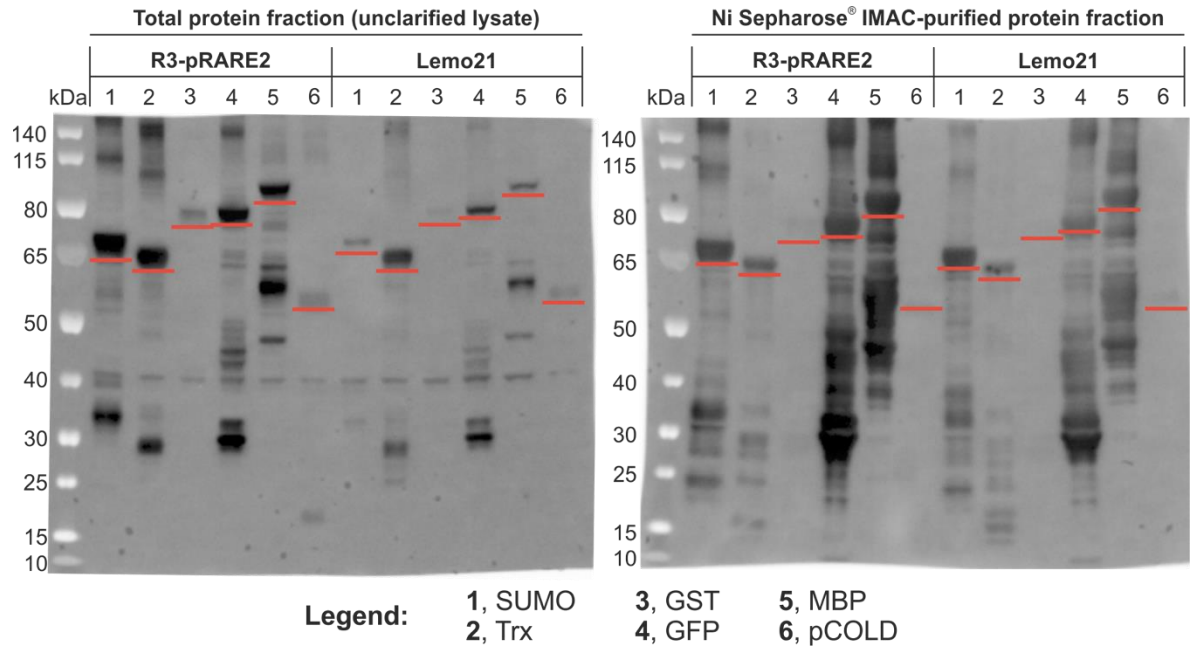


Figure 5.4 | Western blot analysis of fusion construct and *E. coli* strain optimisation using physical disruption (sonication) lysis. Anti-His_6 antibody was used for Western blot visualisation. Red lines are placed underneath the approximate position of protein of interest bands.

5.3.3.4 Expression level tuning in Lemo21 strain using L-rhamnose

Lemo21 strain contains the plasmid for T7 lysozyme LysY, an inhibitor of the T7 RNA polymerase responsible for target gene transcription. L-Rhamnose acts as a promoter for LysY expression, effectively controlling expression levels of the target protein[283]. In an attempt to improve the folding (and consequently the solubility) of recombinant hTET2 by retarding the transcription rate, L-rhamnose was titrated onto Lemo21 strain cells overexpressing SUMO and Trx constructs. Cells were grown according to conditions in Figure 5.5 A; the media was supplemented with L-rhamnose during cell growth (0, 0.5, 1.0, or 2.0 mM). Western blot analysis of total (unclarified lysate) and soluble (clarified lysate) protein fractions revealed that titration of L-rhamnose effected a dose-dependent decrease of hTET2 levels in both total and soluble protein fractions (Figure 5.5 B). As the levels of soluble hTET2 could not be improved by L-rhamnose titration, use of Lemo21 strain was not explored further.

A

Strain	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
Lemo21	TB	37 °C	1.4–1.7	500 µM	16 °C	18 h

B

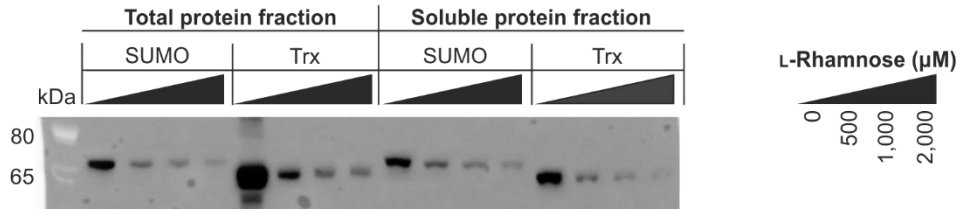


Figure 5.5 | Western blot analysis of expression level control in Lemo21 strain by L-rhamnose titration. A, cell growth conditions; **B**, Western blot analysis of total (unclarified lysate) and soluble (clarified lysate) protein fractions. Media was supplemented with L-rhamnose during cell growth (0, 500, 1,000, or 2,000 µM). Anti-His_6 antibody was used for Western blot visualisation.

5.3.3.5 Medium-scale preparation of SUMO, Trx, and GFP fusion proteins

While the R3-pRARE2 strain was identified as the optimal expression system, selection of the optimal fusion construct was complicated due to similar expression levels across the board. In order to more accurately determine the yield (mg/L of culture) and catalytic activity of the different fusions, medium-scale preparations were carried out for SUMO, Trx, and GFP fusion proteins.

SUMO, Trx, and GFP fusion constructs were overexpressed in R3-pRARE2 strain according to conditions in Figure 5.6 A (1.2 L culture volume) and purified using HisTrap IMAC. Solubility tags were cleaved using appropriate proteases (Ulp3/SenP2 for SUMO; TEV for Trx and GFP), and cleaved hTET2 was isolated by reverse HisTrap IMAC (Figure 5.6 B). Protein purity was low after the initial HisTrap, but was substantially improved following tag cleavage and reverse HisTrap, with complete removal of the highly prevalent 25–30 kDa impurity bands (Figure 5.6 C). Western blot analysis recapitulated the previously observed poor solubility of Trx fusion protein, and revealed that a considerable amount of His_6-tagged protein was present in the HisTrap loading flow-through (FT lanes, Figure 5.6 D), indicating that a substantial portion of the target protein was lost in the initial affinity capture stage.

Although GFP fusion afforded the highest yield after initial affinity capture (7, 9, and 22 mg/L of culture for SUMO, Trx, and GFP, respectively, Figure 5.6 E), the observed value was likely skewed by the presence of large 30 kDa impurity bands (Figure 5.6 C), with SUMO fusion affording the highest amount of hTET2 after tag cleavage (3.4, 1.9, and 2.6 mg/L for SUMO, Trx, and GFP, respectively). Construct catalytic activity was evaluated using AlphaScreen (Figure 5.6 F), revealing that cleaved SUMO and Trx constructs showed activity comparable to SUMO-tagged hTET2 (produced in section 5.3.1), while cleaved hTET2 from GFP fusion was substantially less active.

The SUMO fusion construct gave a higher yield than Trx, and was substantially more catalytically active than GFP. Having confirmed the optimal construct (SUMO) and *E. coli* strain (R3-pRARE2) out of the conditions screened, focus shifted towards the optimisation of cell growth conditions.

A

Strain	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
R3-pRARE2	TB	37 °C	1.6–1.7	500 µM	16 °C	18 h

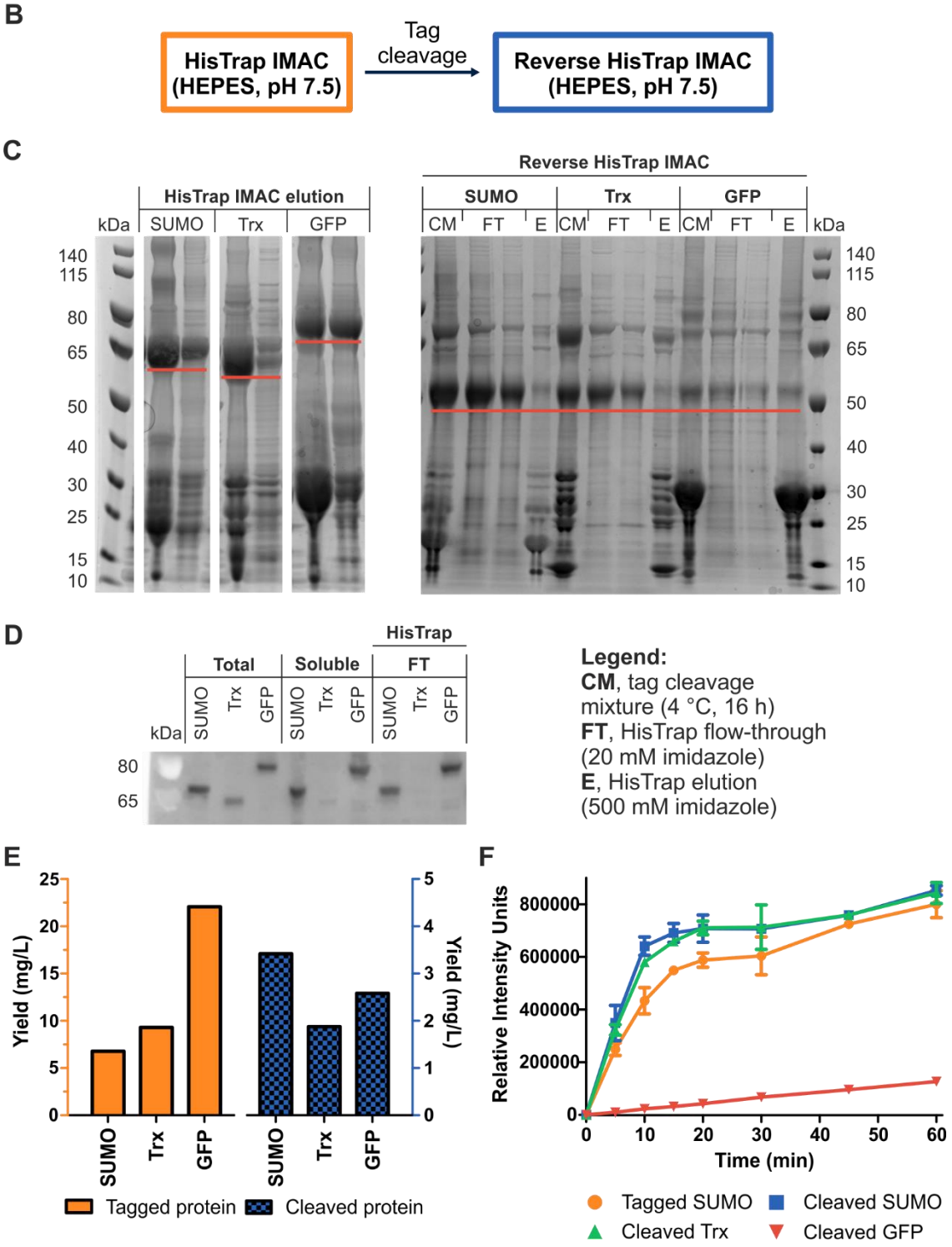


Figure 5.6 | Medium-scale preparation of SUMO, Trx, and GFP fusion proteins. A, cell growth conditions; **B**, purification workflow; **C**, SDS-PAGE analysis of HisTrap and reverse HisTrap IMAC; **D**, Western blot analysis of total (unclarified lysate), soluble (clarified lysate), and HisTrap loading flow-through protein fractions; **E**, tagged and cleaved hTET2 yields in mg/L of culture; **F**, catalytic activity comparison using AlphaScreen (SUMO-tagged hTET2 produced in section 5.3.1 used as a positive control). Red lines are placed underneath the protein of interest bands in SDS-PAGE gels. Anti-His_6 antibody was used for Western blot visualisation. AlphaScreen conditions: 10 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. AlphaScreen data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

5.3.3.6 Cell growth and induction optimisation

Optimisation of cell density prior to induction (OD_{600}) showed that OD_{600} values of up to 2.0 corresponded with an increase in the amount of soluble protein, whereas cell densities above 2.0 were detrimental to soluble protein yield (Figure 5.7 B). Induction with 100 or 500 µM IPTG produced slightly more soluble protein than induction with 300 or 1,000 µM IPTG (Figure 5.7 B), demonstrating that variation of IPTG concentration did not result in a dose-dependent change in expression levels.

Standard growth conditions were supplemented with growth medium additives at the start of cultivation, including sorbitol, sucrose, $ZnCl_2$, and EtOH (Figure 5.7 C). Sugar-based osmolytes sorbitol[287–289] and sucrose[290] were included for their known role in the solubilisation and stabilisation of proteins. The reported hTET2 crystal structure[94] contained 3 zinc-binding regions, so 50 µM $ZnCl_2$ was added to enrich Zn(II) levels approx. 3-fold, as the concentration of Zn(II) in TB medium is estimated to be under 30 µM[291]. Finally, addition of up to 3% v/v EtOH to the growth medium has been reported to enhance recombinant protein expression in *E. coli*[292–294]. Addition of sucrose, $ZnCl_2$, and 1% v/v EtOH did not alter expression levels, while addition of sorbitol and 3% v/v EtOH produced a deleterious effect.

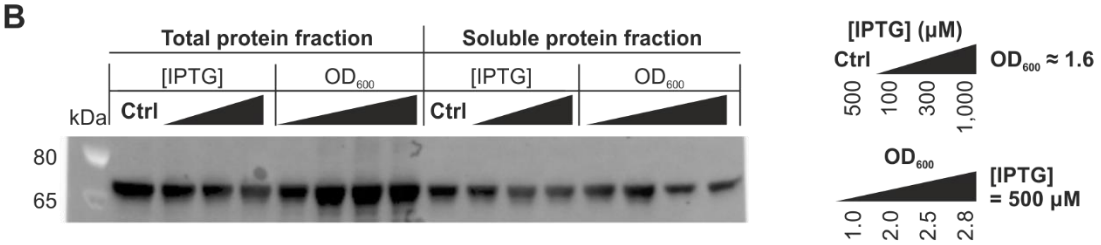
Finally, post-induction temperature and time were optimised for auto-inducing ZYM-5052 and non-inducing TB media (Figure 5.7 D). ZYM-5052 cultures were grown at 37 °C to OD_{600} = 0.8 before reducing the temperature for protein expression. Glycerol content was doubled (from 0.5% to 1% v/v) in condition B (16 °C, 48 h) to ensure sustained carbon source levels throughout the expression period[295]. A range of induction temperatures (15–30 °C) and times (5–48 h) were explored for TB media; similarly to condition B, glycerol content was doubled (to 0.8% v/v) in condition 4. Soluble protein levels were largely unchanged across the conditions screened, with a slight increase in the amount of soluble hTET2 observed in

conditions B (ZYM-5052, 16 °C, 48 h, 2× glycerol) and 2 (TB, 25 °C, 5 h) compared to control (lane C, TB, 16 °C, 18 h).

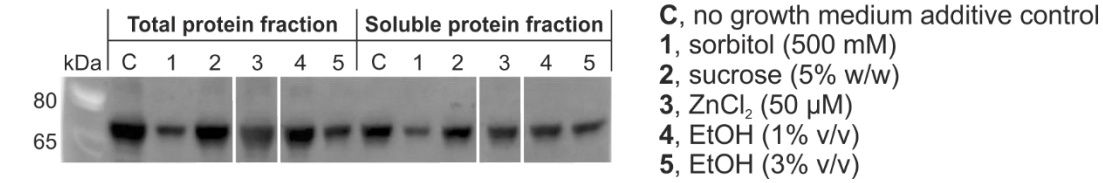
A

Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
TB	37 °C	1.5–1.7	500 µM	16 °C	18 h

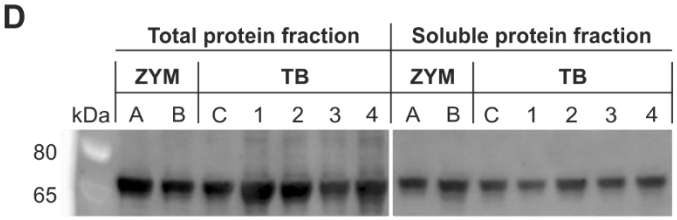
B



C



D



Condition	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time	Additional change
A	ZYM-5052	37 °C	0.8	–	20 °C	24 h	–
B				–	16 °C	48 h	2× glycerol
C	TB	37 °C	1.6–1.8	500 µM	16 °C	18 h	–
1					30 °C	5 h	–
2					25 °C	5 h	–
3					15 °C	24 h	–
4					15 °C	48 h	2× glycerol

Figure 5.7 | Cell growth and induction optimisation. A, general cell growth conditions; **B**, Western blot analysis of cell density (OD_{600}) and IPTG concentration optimisation; **C**, Western blot analysis of growth medium additive optimisation (additives added at the start of cultivation); **D**, Western blot analysis of medium and post-induction time and temperature optimisation. SUMO hTET2 fusion protein was overexpressed in R3-pRARE2 *E. coli* strain. Anti-His_6 antibody was used for Western blot visualisation.

5.3.3.7 Medium-scale evaluation of selected expression conditions

Condition 2 from Figure 5.7 D (TB, 25 °C, 5 h) was compared to control (TB, 16 °C, 18 h) in a medium-scale preparation experiment (1.2 L culture volume) to determine if hTET2 catalytic activity could be improved by increasing post-induction temperature and reducing post-

induction time (Figure 5.8). R3-pRARE2 cells overexpressing SUMO hTET2 were grown according to conditions in Figure 5.8 A, and tagged hTET2 was isolated using HisTrap IMAC. However, virtually no difference could be observed in tagged hTET2 purity, yield, or catalytic activity, indicating that changes in post-induction temperature and time did not affect hTET2 production.

A

Condition	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
2 C	TB	37 °C	1.8–1.9	500 µM	25 °C 16 °C	5 h 18 h

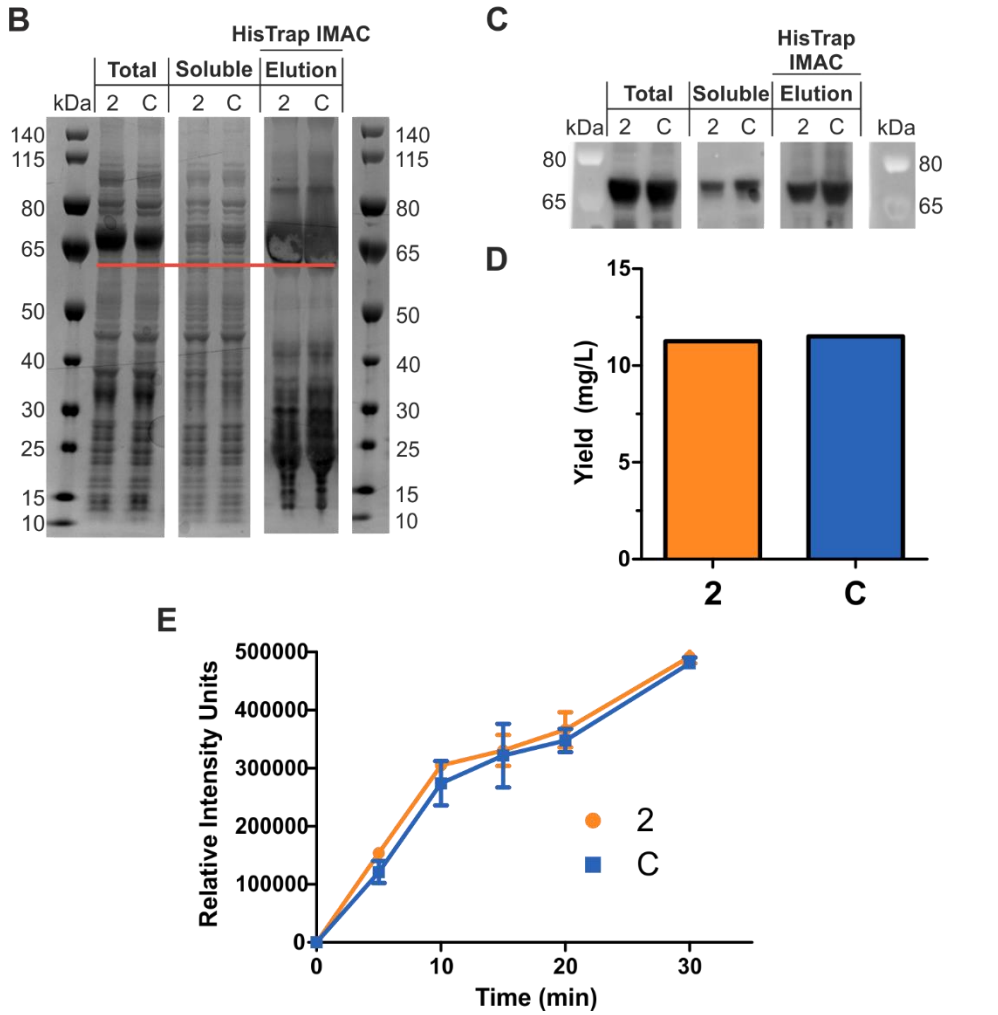


Figure 5.8 | Evaluation of post-induction time and temperature effect on catalytic activity of SUMO-tagged hTET2. A, cell growth conditions; **B**, SDS-PAGE analysis of total (unclarified lysate), soluble (clarified lysate), and HisTrap elution protein fractions; **C**, Western blot analysis of total, soluble, and HisTrap elution protein fractions; **D**, tagged hTET2 yields in mg/L of culture; **E**, catalytic activity comparison using AlphaScreen. SUMO hTET2 fusion protein was overexpressed in *E. coli* R3-pRARE2 strain. Red lines are placed underneath the protein of interest bands in SDS-PAGE gel. Anti-His_6 antibody was used for Western blot visualisation. AlphaScreen conditions: 10 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. AlphaScreen data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

5.3.4 Optimisation of hTET2 purification

Variation of solubility tags, *E. coli* strains, and expression conditions offered little improvement to the purity and catalytic activity of hTET2, while improvements in the expression levels were not significant enough for unambiguous identification of optimal conditions. Consequently, focus shifted towards optimisation of hTET2 purification.

5.3.4.1 Optimisation of cell lysate preparation

Two key changes were implemented for all subsequent hTET2 purifications based on observations from previous medium-scale preparations: (i) periplasmic material was removed from *E. coli* cell pellets by osmotic shock (cells washed with 20% w/v sucrose) to eliminate Ni(II)-chelating metallophores, which were shown to affect affinity capture of low abundance recombinant proteins from cell lysates by reducing the amount of nickel ions immobilised on the resin[296]; and (ii) sonication was replaced with high pressure homogenisation to increase the degree of cell disruption and reduce energy transfer into the sample in the form of heat[297].

5.3.4.2 HisTrap IMAC optimisation

Considerable amounts of tagged hTET2 were observed in the HisTrap lysate loading flow-through of previous medium-scale preparations, suggesting poor nickel chelation efficiency by the His_6 affinity tag (FT lane, Figure 5.6 D). In order to increase the efficiency of affinity capture, imidazole concentration in HisTrap binding buffer was reduced from 20 mM to 10 mM, and a purification of SUMO hTET2 fusion overexpressed in TB and ZYM-5052 media was carried out (Figure 5.9). Cells were grown according to conditions in Figure 5.9 A, underwent osmotic shock, and were lysed using high pressure homogenisation. Affinity capture of tagged hTET2 using HisTrap was followed by tag cleavage and reverse HisTrap (Figure 5.9 B).

A

Condition	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
1	TB	37 °C	1.8–2.0	500 µM	16 °C	18 h
2	ZYM-5052		0.8–0.9	–	20 °C	24 h

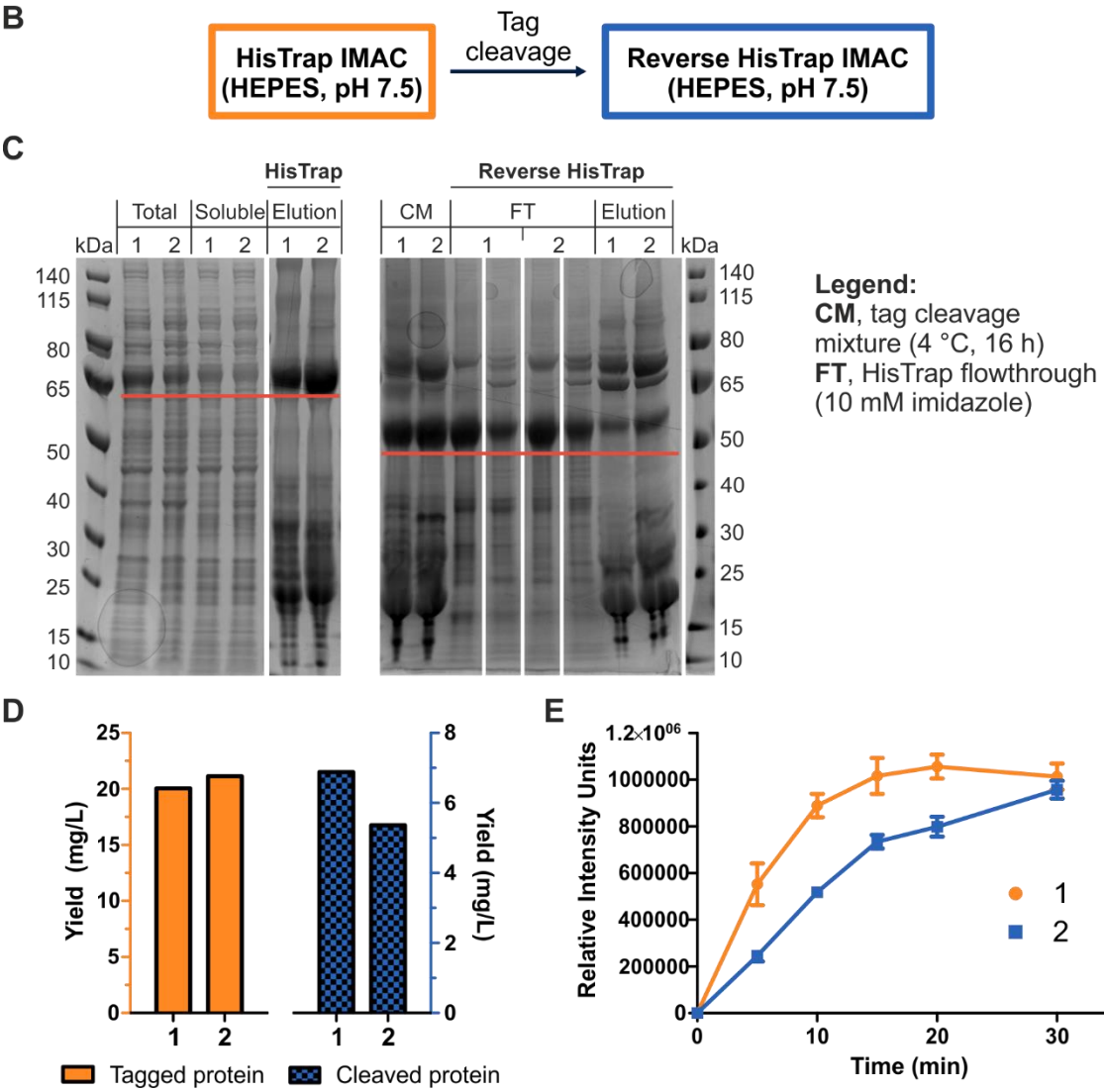


Figure 5.9 | Medium-scale preparation of SUMO hTET2 overexpressed in TB and ZYM-5052 media using optimised HisTrap IMAC conditions. A, cell growth conditions; **B**, purification workflow; **C**, SDS-PAGE analysis of HisTrap and reverse HisTrap IMAC; **D**, tagged and cleaved hTET2 yields in mg/L of culture; **E**, catalytic activity comparison using AlphaScreen. SUMO hTET2 fusion protein was overexpressed in *E. coli* R3-pRARE2 strain. Red lines are placed underneath the protein of interest bands in SDS-PAGE gels. AlphaScreen conditions: 10 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. AlphaScreen data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

While protein purity by SDS-PAGE was comparable to previous purifications, the yields of both tagged and cleaved protein were improved over 2-fold, affording 20 and 6.9 mg/L of tagged and cleaved hTET2, respectively, from condition 1 after reverse HisTrap (*vs* 6.8 and 3.4 mg/L for tagged and cleaved SUMO hTET2, respectively, in Figure 5.6). Evaluation of catalytic

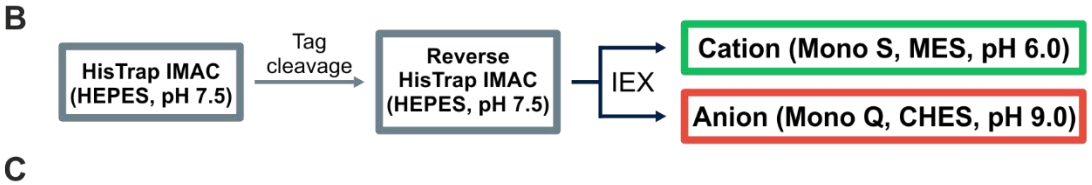
activity by AlphaScreen revealed that overexpression in TB medium produced nearly 2-fold more active protein compared to ZYM-5052 medium (Figure 5.9 E).

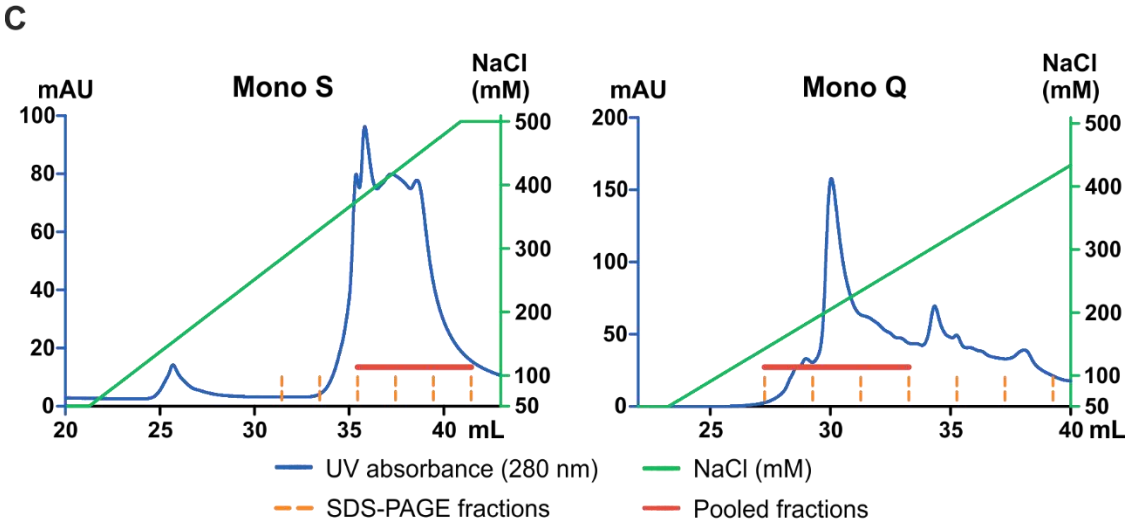
5.3.4.3 Cation and anion IEX chromatography optimisation

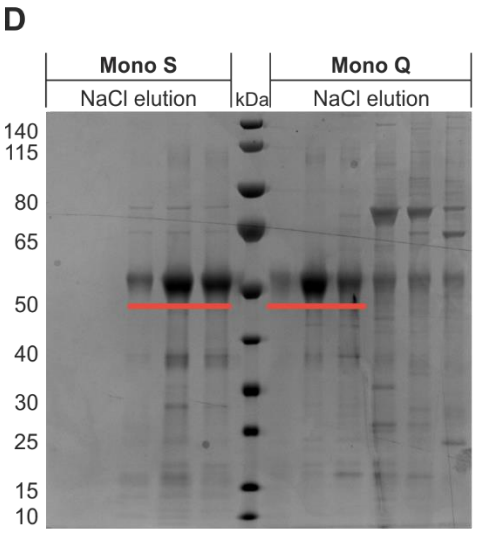
Directly after reverse HisTrap, cleaved hTET2 from condition 1 was further purified using cation-exchange chromatography (cation IEX, Mono S), while protein from condition 2 was purified using anion-exchange chromatography (anion IEX, Mono Q, Figure 5.10 A, B). As the predicted pI of cleaved hTET2 is 7.48 (Appendix 4)[298], the protein was buffer exchanged into MES, pH 6.0 for cation IEX, and into CHES, pH 9.0 for anion IEX to effect sufficient sample ionisation for IEX capture. Substantial precipitation was observed upon buffer exchange into MES, pH 6.0, while no protein was lost during buffer exchange into CHES, pH 9.0, indicating that the enzyme is more stable in basic solutions. Purification using both Mono S and Mono Q columns separated hTET2 from the main impurity band at approx. 70 kDa (Figure 5.10 C, D). However, both approaches suffered from relatively low protein recovery; only approx. 25% of cleaved hTET2 loaded onto the columns was recovered post-chromatography (Figure 5.10 E). Interestingly, specific enzyme activity increased after anion IEX, but remained unchanged after cation IEX (Figure 5.10 F*, cf*. Figure 5.9 E), further indicating that buffer conditions used in cation IEX may not be optimal for maintaining protein folding.

A

Condition	Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
Mono S	TB	37 °C	1.8–2.0	500 µM	16 °C	18 h
Mono Q	ZYM-5052		0.8–0.9	–	20 °C	24 h







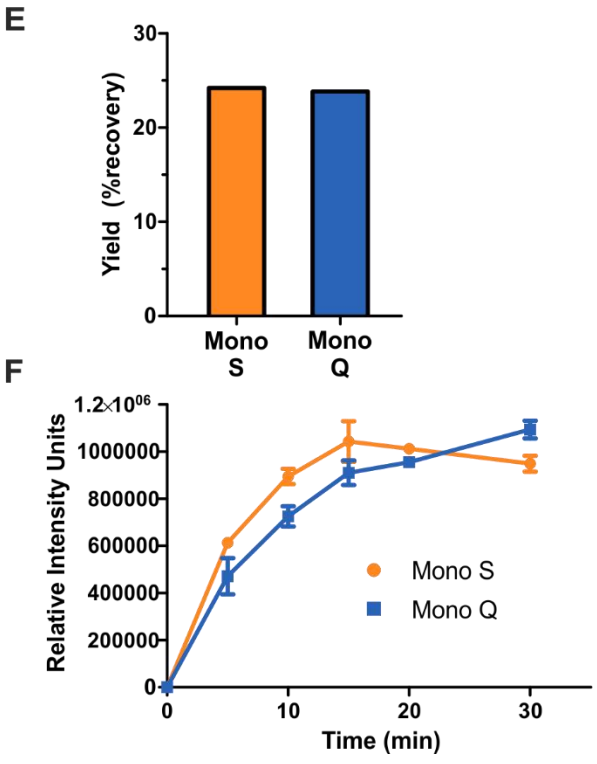


Figure 5.10 | Ion-exchange (IEX) chromatography of cleaved hTET2 overexpressed in TB and ZYM-5052 media. **A**, cell growth conditions; **B**, purification workflow (steps in grey described in Figure 5.9); **C**, IEX chromatography 280 nm elution profiles; **D**, SDS-PAGE analysis of IEX elution fractions; **E**, hTET2 yields after IEX, reported as percentage recovery of cleaved hTET2 input; **F**, catalytic activity comparison using AlphaScreen. SUMO hTET2 fusion protein was overexpressed in *E. coli* R3-pRARE2 strain. Red lines indicating pooled fractions are placed underneath the protein of interest bands in SDS-PAGE gels. AlphaScreen conditions: 10 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. AlphaScreen data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

5.3.4.4 Comparison of cation IEX and heparin affinity chromatography

Higher protein recovery was desirable for the secondary purification step in order to maximise the overall yield of hTET2. The DNA-binding nature of hTET2 enables ligand-specific affinity purification using heparin, a highly sulfated glycosaminoglycan. Heparin interaction with proteins is weakened by increasing ionic strength, facilitating affinity capture and release in a NaCl concentration-dependent manner. As such, purification of hTET2 using heparin-functionalised Sepharose® (Heparin HP) was compared with the previously validated cation IEX (Mono S) method. Cells overexpressing SUMO hTET2 were grown in TB, and hTET2 was purified uniformly up to the IEX step (Figure 5.11 A, B). The amount of Ni Sepharose® High Performance resin was increased 5-fold (from 1 to 5 mL) for initial and reverse HisTrap steps, resulting in a nearly 2-fold increase in the yield of cleaved hTET2 after reverse HisTrap (Figure 5.11 E, compared to results from HisTrap IMAC optimisation in Figure 5.9 D). After reverse HisTrap, cleaved hTET2 was split into two equal portions, buffer exchanged into suitable binding buffers, and further purified (Figure 5.11 C). Protein of interest bound to both resins with high efficiency, while the majority of impurity bands were eluted in the loading flow-through (FT lanes, Figure 5.11 D). Strikingly, heparin affinity purification afforded a nearly 2-fold improvement in protein recovery over the cation IEX method (4.2 and 2.3 mg/L for Heparin HP and Mono S, respectively). In line with what was observed previously for anion IEX, heparin purification also furnished slightly more catalytically active protein compared to cation IEX (Figure 5.11 F). The significant improvement in yield coupled with retention of purity and catalytic activity established heparin affinity as the optimal secondary purification method.

A

Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
TB	37 °C	1.9–2.0	500 µM	16 °C	18 h

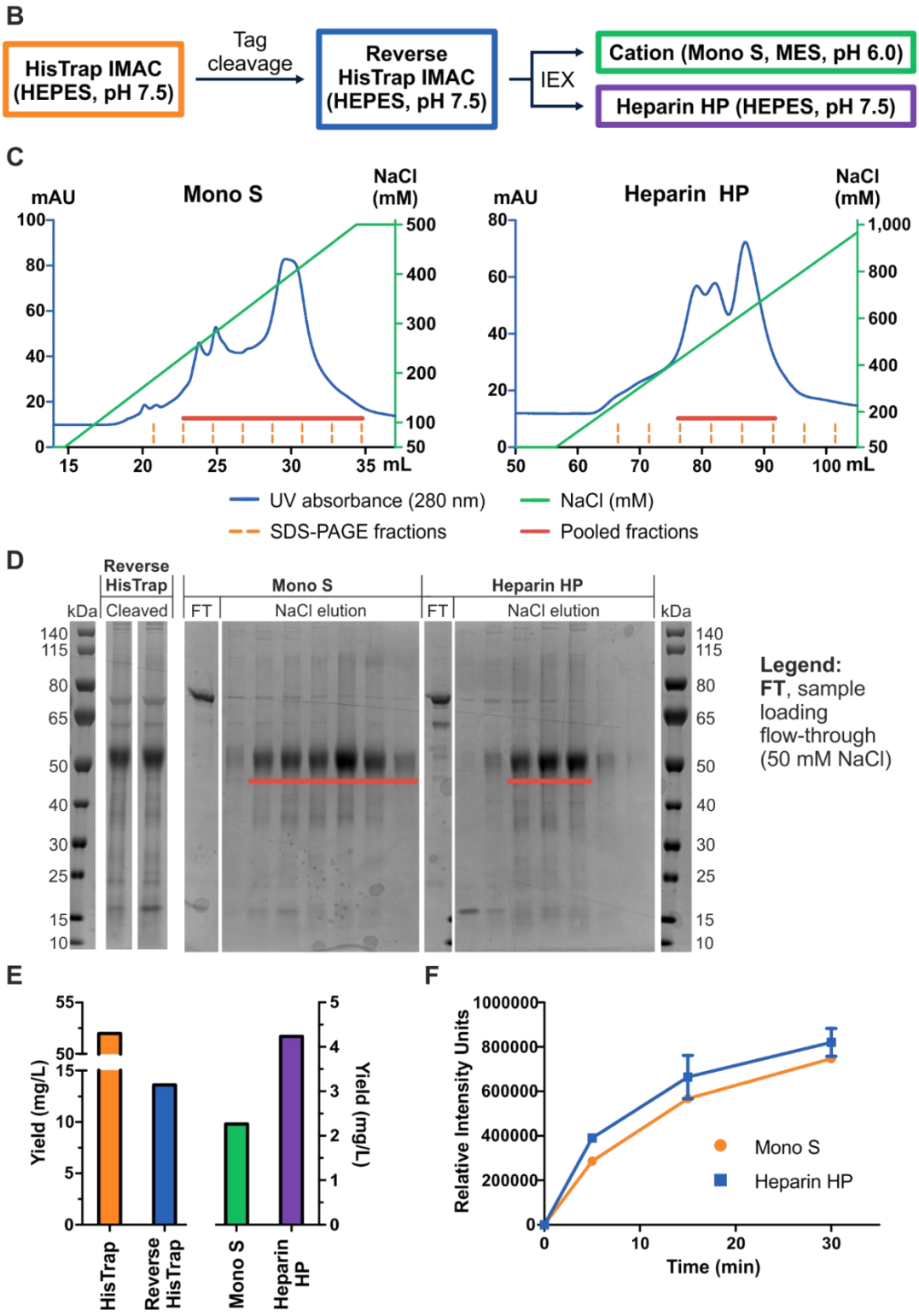


Figure 5.11 | Comparison of cation IEX and heparin affinity as secondary purification methods. A, cell growth conditions; **B**, purification workflow; **C**, chromatography 280 nm elution profiles; **D**, SDS-PAGE analysis of cleaved hTET2 after reverse HisTrap, Mono S, and Heparin HP; **E**, hTET2 yields after each step in mg/L of culture; **F**, catalytic activity comparison using AlphaScreen. SUMO hTET2 fusion protein was overexpressed in *E. coli* R3-pRARE2 strain. Red lines indicating pooled fractions are placed underneath the protein of interest bands in SDS-PAGE gels. AlphaScreen conditions: 5 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. AlphaScreen data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

5.3.4.5 Buffer pH optimisation

The concentration of hTET2 used in crystallographic studies was reported to be 20–25 mg/mL[94,96]; however, concentrating purified hTET2 above 15 mg/mL had proven challenging thus far, with further concentration resulting in rapid protein loss. One explanation for the poor hTET2 solubility at higher concentrations may be the similarity between protein storage buffer pH and the predicted isoelectric point (pI) of cleaved hTET2 (pH 7.50 *vs* 7.48, respectively); as a protein approaches zero net charge, concentrated solutions are more likely to aggregate or precipitate due to reduced repulsive electrostatic forces between molecules[299].

In order to assess the impact of pH on highly concentrated sample solubility, two hTET2 purifications were carried out in Tris buffer at pH 8.0 and 8.5 (Figure 5.12 B). The two conditions produced largely identical profiles throughout the purification process, delivering hTET2 in yields and purities comparable to purification using HEPES, pH 7.5 (Figure 5.12 D, E). Size exclusion chromatography using Superdex® 200 resin (SEC S200) was included in the purification workflow, with monomeric cleaved hTET2 eluting as a symmetrical peak (Figure 5.12 C). However, a peak was observed immediately after column void volume (40 mL), suggesting that protein aggregation was still present to some degree. While the effect of pH on protein solubility above 15 mg/mL could not be unambiguously determined, purification in Tris, pH 8.0 afforded markedly more active protein compared to Tris, pH 8.5 or HEPES, pH 7.5 (Figure 5.12 F).

A

Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
TB	37 °C	1.9–2.0	500 µM	16 °C	18 h

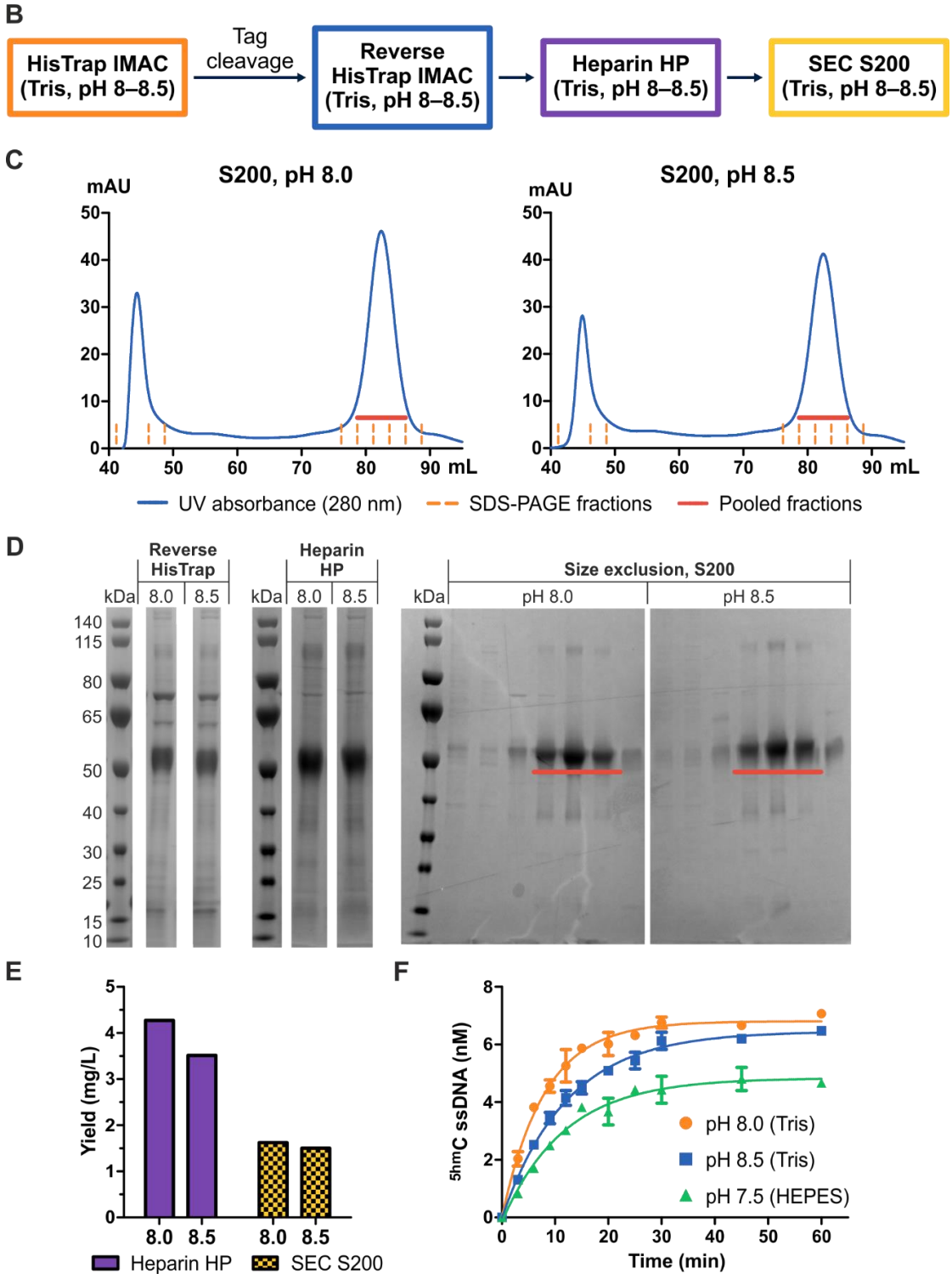


Figure 5.12 | Optimisation of buffer pH. A, cell growth conditions; **B**, purification workflow; **C**, size exclusion chromatography (SEC S200) 280 nm elution profiles; **D**, SDS-PAGE analysis of cleaved hTET2 after reverse HisTrap, Heparin HP, and SEC; **E**, hTET2 yields after heparin and size exclusion steps in mg/L of culture; **F**, catalytic activity comparison using AlphaScreen. SUMO hTET2 fusion protein was overexpressed in *E. coli* R3-pRARE2 strain. Red lines indicating pooled fractions are placed underneath the protein of interest bands in SDS-PAGE gels. AlphaScreen conditions: 1 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. AlphaScreen data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

5.3.4.6 Large-scale preparation of hTET2$_{CD\Delta LCI}$

With the optimised conditions in hand, a large-scale (14.4 L culture volume, 220 g cell pellet) preparation of hTET2 was carried out (Figure 5.13). SUMO hTET2 fusion protein was overexpressed in *E. coli* BL21(DE3)-R3-pRARE2 strain in TB medium and isolated as the pure cleaved hTET2 over 4 chromatography steps (20 mL of HisTrap IMAC resin and 25 mL of Heparin HP resin were used). The results from medium-scale experiments translated well into the large-scale preparation, affording 2.76 mg/L of highly pure hTET2 (>95% by SDS-PAGE) in a final buffer containing 20 mM Tris, 300 mM NaCl, pH 8.0 (Figure 5.13 C, D, E). Catalytic activity evaluation using AlphaScreen revealed that specific activity was improved nearly 100-fold compared to initial production attempts (Figure 5.13 F*, cf.* Figure 5.2 B). Analysis of the A_{260}/A_{280} UV light absorbance ratio confirmed that the protein sample was essentially DNA-free (purified hTET2 A_{260}/A_{280} = 0.55; reported A_{260}/A_{280} for DNA-free protein samples = 0.57[300]). Protein storage at higher pH (8.0 *vs* 7.5) and higher NaCl (300 *vs* 150 mM) enabled concentration of the final hTET2 sample to 19.7 mg/mL with minimal protein loss. LC-MS analysis confirmed cleaved hTET2 as the most prevalent species (predicted mass: 51,260.3 Da, observed mass: 51,262.5 Da), with only minute levels of undesired species present. A number of low molecular weight adducts (+32 to +175 Da, Figure 5.13 G inlet) was observed, but the mass shifts could not be unambiguously identified; the atypical MS profile could not be compared to literature profiles as no mass spectra have been reported for hTET2.

Overall, the protein yield, purity, and catalytic activity were improved significantly compared to the initial hTET2 production attempts, where only 0.4 mg/L of poorly active tagged hTET2 could be isolated (Figure 5.1, Figure 5.2).

A

Medium	Cultivation temperature	OD_{600} prior to induction	[IPTG]	Post-induction temperature	Post-induction time
TB	37 °C	1.9–2.0	500 µM	16 °C	18 h

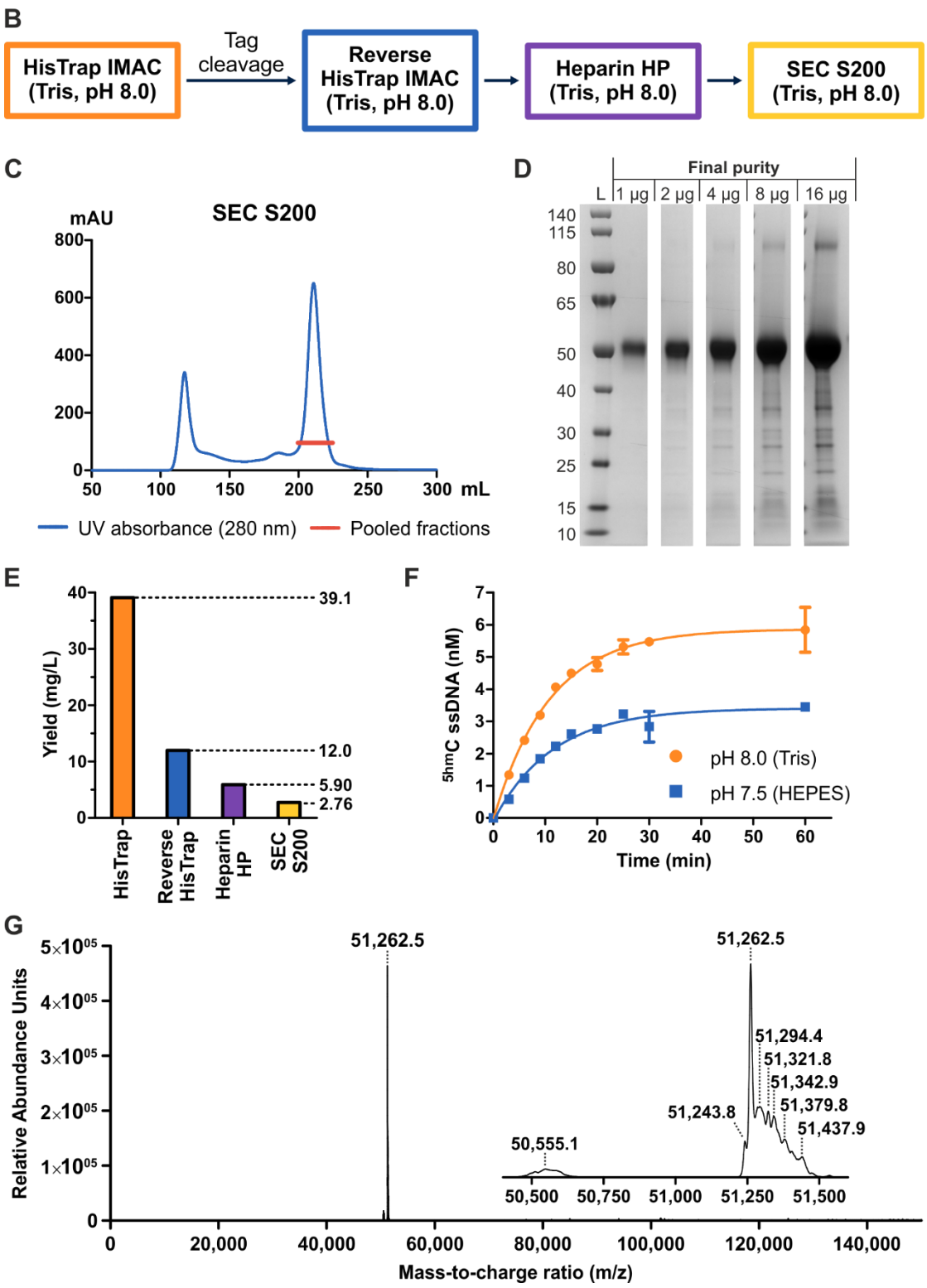


Figure 5.13 | Large-scale preparation of cleaved hTET2. A, cell growth conditions; **B**, purification workflow; **C**, SEC S200 280 nm elution profile; **D**, titration of fully purified hTET2 on an SDS-PAGE gel (1–16 µg loading); **E**, hTET2 yields after each purification step in mg/L of culture; **F**, catalytic activity comparison using AlphaScreen; **G**, LC-MS analysis of fully purified cleaved hTET2 (predicted mass: 51,260.3 Da, observed mass: 51,262.5 Da). AlphaScreen conditions: 1 nM hTET2, 10 nM **^{5m}C 32 nt bt-ssDNA**, 100 µM sodium L-ascorbate, 10 µM $(NH_4)_2Fe(SO_4)_2$, and 10 µM disodium 2-OG. SUMO hTET2 fusion protein was overexpressed in *E. coli* R3-pRARE2 strain. AlphaScreen data shown as the mean of 2 experimental replicates (n = 2; mean ± SD).

5.3.5 Crystallisation of cleaved hTET2

5.3.5.1 Reproduction of reported conditions

With the purified cleaved hTET2 in hand, focus shifted to reproducing the TET2-DNA complex crystals using conditions described by Hu *et al*.[94,96] The reported well solution conditions contained 100 mM MES, pH 6.3–6.4, and 21–26% w/v PEG 2000 monomethyl ether (MME) as the precipitant, with crystal growth carried out at 4 °C. A 96-condition optimisation screen was prepared, covering a range of pH (5.8–6.8, 8 points) and PEG 2000 MME concentrations (14–30% w/v, 10×1.5% increments, 1×1% increment). A solution containing hTET2 (approx. 19.3 mg/mL, 376 µM), $(NH_4)_2Fe(SO_4)_2$ (2 mM), NOG (2 mM), and **^{5m}C 12 bp dsDNA†** (1.5 eq.) was plated and stored at 4 °C or 22 °C. Droplets were observed from day 2 onwards, revealing clear drops in the lowest PEG conditions, slight granular precipitation in the medium PEG conditions (Figure 5.14, A), and flocculent precipitation in the highest PEG content conditions at 22 °C. After 21 days, microcrystalline precipitate was observed across a broad range (14–26%) of PEG conditions at both 4 °C and 22 °C (Figure 5.14, B). No crystals of mountable size were detected; thus, further optimisation of PEG concentrations and multiple rounds of crystal seeding would likely be required before high-quality crystals are produced.

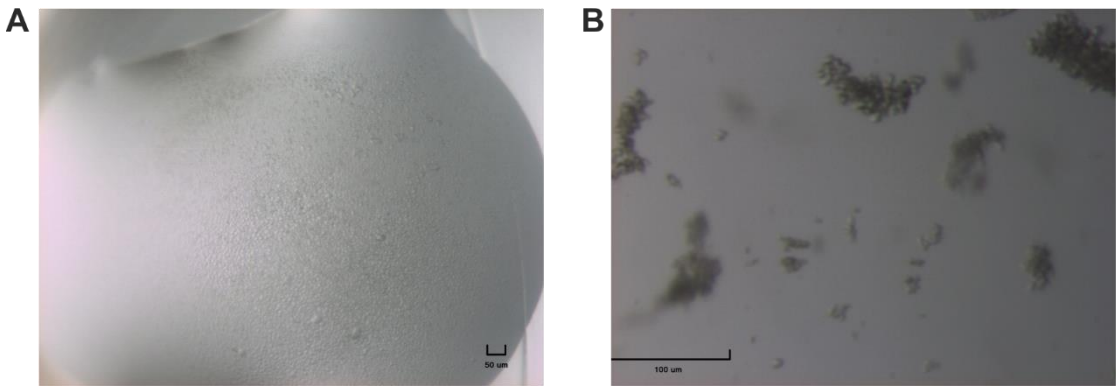


Figure 5.14 | Representative droplets from reported condition optimisation plates. A, day 2, pH 6.0, 23% PEG 2000 MME, 22 °C (bar represents 50 µm); **B**, day 22, pH 6.5, 14% PEG 2000 MME, 4 °C (bar represents 100 µm).

† ^{5m}C refers to a 5-methyldeoxycytidine monophosphate deoxynucleotide in a DNA oligomer.

5.3.5.2 Broad screen crystallisation trials

The reported[94,96] crystallisation conditions lacked a number of experimental details, making them difficult to reproduce. Thus, crystallisation trials using broad screens were carried out concurrently in an effort to identify novel crystallisation conditions. Broad screens, including JCSG Plus™, Structure Screen (1+2), PACT Premier™, Morpheus®, Morpheus® Fusion, Hampton Index, NeXtal AmSO4, and SaltRx, were set up using the same procedure as in section 5.3.5.1 and stored at 20 °C. Promising crystalline precipitate or microcrystals were observed in three conditions after 14 days (Figure 5.15). Interestingly, one of the three promising conditions contained calcium acetate (200 mM), sodium cacodylate (100 mM), and PEG 300 (40% w/v), pH 6.5 (Figure 5.15 A), bearing a high similarity to the reported crystallisation conditions in terms of pH (pH 6.5 *vs* 6.3–6.4) and precipitant (PEG in both cases).

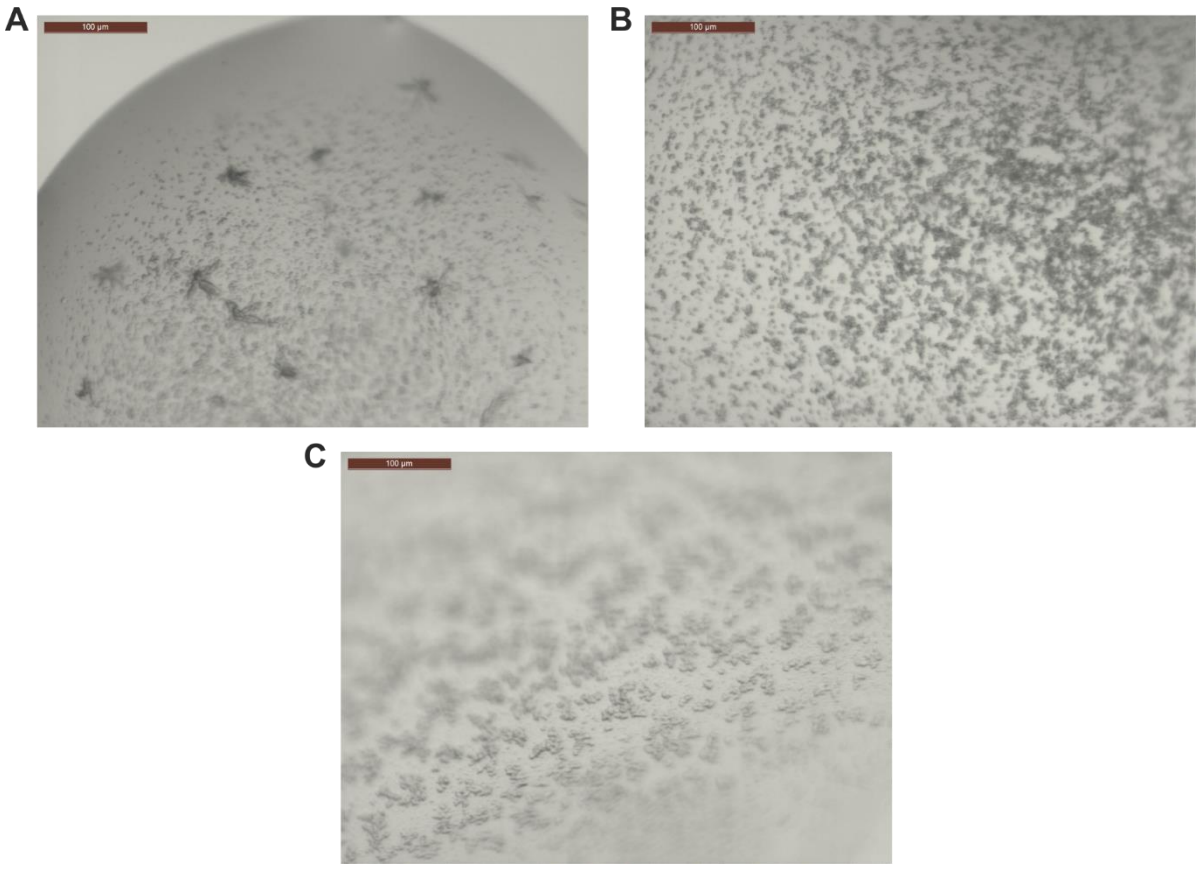


Figure 5.15 | Representative droplets from broad screen crystal plates. A, 0.2 M calcium acetate, 0.1 M sodium cacodylate, 40% PEG 300, pH 6.5; **B**, 0.5 M sodium chloride, 0.1 M magnesium chloride hexahydrate, 0.01 M CTAB; **C**, 10 % w/v PEG 1000, 10 % w/v PEG 8000. Plates stored at 20 °C. Bars represent 100 µm. Plates prepared and images provided by Dr Arnaud Basle.

5.4 Discussion

Overexpression of N–terminally His_6-FLAG-[thrombin cleavage site]-SUMO-tagged hTET2 truncated catalytic domain, $hTET2_{CD\Delta LCI}$, in *E. coli* BL21(DE3) strain afforded the enzyme in low purity (by SDS-PAGE and LC-MS) and yield (approx. 0.4 mg/L). While identity confirmation of the isolated protein by LC-MS proved challenging due to the low purity, a peak likely corresponding to methionine-excised α-*N*–gluconoylated tagged hTET2 was observed. Subsequent preparation of cleaved hTET2 confirmed that the tentative post-translational modifications were located within the N–terminal solubility tag. The activity of monomeric tagged hTET2 was found to be 20-fold higher than NgTet1 by AlphaScreen, while aggregated hTET2 was found inactive.

The SUMO-tagged hTET2 was used for initial crystallisation trials without removal of the solubility tag due to the low yield of the isolated protein. Crystallisation trials were carried out in an attempt to co-crystallise the recombinant tagged hTET2 with TiP1 (IC_{50} = 25 μM). Four protein crystallisation screens (PEG/Ion, Crystal Screen, SaltRx, and JCSG Plus™) were used, with the majority of droplets remaining clear after 2 weeks, indicating insufficiently concentrated protein sample. However, concentration of tagged hTET2 above 10 mg/mL resulted in rapid loss of protein. Furthermore, the low isolated yield of tagged hTET2 prohibited extensive screening of crystallisation conditions, necessitating optimisation of hTET2 production.

Optimisation of hTET2 expression began with the exploration of fusion vectors and *E. coli* expression strains. Six vectors (five containing solubility tags SUMO, Trx, GST, GFP, and MBP, and cold-shock expression vector pCOLD) were overexpressed in three *E. coli* strains (BL21(DE3)-R3-pRARE2, Lemo21(DE3), and SHuffle® T7). Evaluation of expression levels by SDS-PAGE was hindered by poor overexpression levels, so analysis was carried out by Western

blotting with an anti-His_6 tag antibody. Initial trials revealed that GST and pCOLD fusions expressed substantially worse than SUMO, Trx, GFP, and MBP fusions, while overexpression in SHuffle strain produced lower levels of protein across the board. Titration of L-rhamnose onto Lemo21 cells did not increase soluble SUMO and Trx fusion protein levels; instead, L-rhamnose reduced the levels of target protein in a dose-dependent manner. Medium-scale preparation of SUMO, Trx, and GFP fusion proteins enabled accurate comparison of protein yield and activity: cleaved hTET2 was isolated in 3.4, 1.9, and 2.6 mg/L yields for SUMO, Trx, and GFP fusions, respectively. Activity of cleaved hTET2 from SUMO and Trx fusions was comparable to tagged SUMO hTET2 fusion protein, while cleaved hTET2 from GFP fusion was markedly less active. As such, the original SUMO fusion and the rare codon tRNA-enriched *E. coli* BL21(DE3)-R3-pRARE2 strain were identified as the optimal construct and expression strain.

Next, optimisation of cell growth and induction conditions was carried out. Variation of cell density prior to induction (OD_{600}) revealed that OD_{600} = 2 produced the highest amount of soluble protein per L of culture, while variation of IPTG concentration (100–1,000 μM) demonstrated no dose-dependent change in expression levels. Addition of growth medium additives (sorbitol, sucrose, $ZnCl_2$, and EtOH) during cultivation did not improve soluble protein levels. Finally, overexpression in non-inducing (TB) and auto-inducing (ZYM-5052) media along with optimisation of post-induction temperature (15–30 °C) and time (5–48 h) resulted in largely comparable levels of soluble hTET2. Comparison of catalytic activity of tagged hTET2 isolated from cells grown in TB medium under short (25 °C, 5 h) and long (16 °C, 18 h) post-induction conditions revealed that induction times and temperatures did not affect enzyme catalytic activity.

Overall, optimisation of hTET2 expression was mostly unsuccessful: a construct superior to the original SUMO fusion could not be identified; variation of *E. coli* strains did not increase production of target protein; protein expression levels remained virtually unchanged across the majority of tested conditions, with a small number of conditions further reducing expression levels. Higher average expression levels in rare codon tRNA-enriched strain compared to non-Rosetta strains (R3-pRARE2 *vs* Lemo21) suggest that codon optimisation may indeed be essential for enhanced expression of soluble hTET2. Furthermore, insoluble target protein amounted to a substantial portion of total hTET2 expression, which may be due to chaperone capacity limitations in the host cells; co-overexpression of chaperone systems with target protein in *E. coli* has been shown to result in an up to 40-fold increase in soluble protein yield[301]. Finally, an analysis of the overexpression of a large set (>100) of mammalian proteins[286] in *E. coli* revealed an inverse correlation between protein molecular weight and soluble protein yield, indicating that the size of the SUMO construct (66.0 kDa) may make it an inherently challenging recombinant expression target. Baculovirus-insect cell and mammalian cell expression systems are much better suited for the production of active, soluble proteins from higher order organisms; however, production of large amounts of protein required for crystallography is similarly challenging, and the costs associated with protein production are substantially higher compared to *E. coli*. This observation may partially explain the lack of structural work carried out on human TETs to date, and why murine and *Naegleria* Tets are often used as model systems.

Expression optimisation was followed by optimisation of protein purification. Introduction of osmotic shock to remove metallophores, substitution of sonication with high pressure homogenisation, and a 2-fold reduction in imidazole concentration in HisTrap binding buffer improved the initial affinity capture step nearly 3-fold (6.8 to 20 mg/L of tagged SUMO hTET2). A subsequent 5-fold increase in Ni Sepharose resin volume further improved affinity capture

efficiency to 52 mg/L of tagged hTET2. Yields of cleaved hTET2 followed a similar trend, improving from 3.4 to 13.6 mg/L. Cation IEX, anion IEX, and heparin affinity were explored as secondary purification methods. While protein recovery after cation/anion IEX chromatography was very similar (approx. 25%), heparin affinity chromatography afforded cleaved hTET2 in nearly 2-fold higher yield (2.3 *vs* 4.2 mg/L). Furthermore, purification under acidic conditions (cation IEX) conferred no change to catalytic activity, while utilisation of basic conditions (anion IEX and heparin affinity) slightly improved the catalytic activity. Variation of buffer pH identified Tris, pH 8.0 as the optimal buffer for enzyme activity. Finally, a large-scale preparation of hTET2 using the optimised conditions afforded highly pure and highly active hTET2 in a tractable yield of 2.76 mg/L of culture, representing a more than 5-fold improvement over previous yields (0.3–0.5 mg/L, correspondence with publication authors[94,96]).

One of the main hurdles encountered throughout the optimisation was the need for three separate chromatographic steps (HisTrap IMAC, Heparin HP, and SEC S200), each reducing the protein yield nearly 2-fold. However, sample purity and homogeneity requirements for crystallographic applications necessitated thorough protein purification; as such, omission of any chromatographic step was unfeasible. The second major issue was the poor efficiency of affinity capture. Intra- or intermolecular interactions between the hexahistidine tag and other parts of the fusion protein may result in reduced tag accessibility to solvent, leading to compromised Ni(II) ion chelating capacity. In this work, the issue was addressed by removing metallophores *via* osmotic shock and increasing the volume of Ni Sepharose resin; however, alternative solutions, including expansion of the polyhistidine tag to His_{10} or more, migration of the affinity tag to the C–terminus, or introduction of a different affinity tag (e.g., Strep-tag II[302]) could prove to be more successful. Finally, the fuzzy appearance of purified hTET2 bands on SDS-PAGE gels combined with the multiple low molecular weight adducts observed by

LC-MS challenged the homogeneity of the isolated enzyme. Use of old SDS-PAGE gels has been reported to result in poor quality bands[303]; however, SDS-PAGE on freshly purchased precast polyacrylamide gels did not improve band appearance (data not shown), dismissing the possibility of SDS-PAGE artefacts. While some of the mass shifts observed by LC-MS could correspond to common PTMs, such as methionine oxidation (+32 Da) and phosphorylation (+80 Da)[304,305], the majority of adducts could not be identified. Identification of the exact nature and position of the PTMs would require a thorough MS/MS analysis, especially for mass adducts corresponding to unknown PTMs. While elucidation of the adduct sources may facilitate PTM suppression in consequent protein preparations, the diversity of the mass shifts suggests that complete protein homogenisation would be challenging.

Crystallisation of $hTET2_{CD\Delta LCI}$ at 4 °C and 22 °C in complex with Fe(II), NOG, and **^{5m}C 12 bp dsDNA** in an optimisation block based on the reported[94,96] conditions did not afford high-quality crystals, but microcrystalline precipitate was observed in a number of wells, necessitating further optimisation. Concurrently, eight broad crystallisation screens were set up at 20 °C, but no macrocrystals could be observed. Crystalline precipitate was detected in three conditions, providing conditions for further optimisation. One of the broad screen conditions yielding microcrystals was highly analogous to the reported crystallisation conditions, suggesting that slightly acidic pH (6.3–6.5) and PEG precipitants encourage hTET2 crystallisation.

5.5 Conclusions and future work

Catalytically active hTET2$_{CD\Delta LCI}$ was overexpressed in *E. coli* and isolated following a 2-stage purification (HisTrap and size exclusion). Small-scale hTET2-TiP1 complex crystallisation trials were unsuccessful, with poor protein yield and purity precluding detailed crystallographic work. A thorough optimisation of expression conditions did not result in an appreciable increase in soluble protein levels. Conversely, optimisation of the purification workflow, including the addition of tag cleavage and secondary heparin affinity-based purification steps, granted a simultaneous improvement to yield (0.4 to 2.76 mg/L) and purity (>95% by SDS-PAGE). The marked improvement in the yield of isolable hTET2 expedited subsequent crystallisation work by substantially reducing the time required to produce the desired amount of protein and enabling production of larger batches of homogeneous sample. A number of low molecular weight mass adducts, potentially corresponding to post-translational modifications, was observed for the purified hTET2 by LC-MS. Confirmation of the protein sequence by Edman sequencing in conjunction with protein digest and MS/MS analysis could be used to identify the origin of the observed mass adducts.

Crystallisation trials of hTET2$_{CD\Delta LCI}$ in complex with ^{5m}C-containing DNA using broad screens and an optimisation block based on the reported[94,96] crystallisation conditions yielded microcrystalline precipitate across a broad range of conditions, necessitating further condition optimisation and crystal seeding. The metal ion used for crystallisation, Fe(II), is relatively stable at slightly acidic pH, but oxidises rapidly at basic pH[306], potentially contributing to an underestimation of condition hits. T_M shift assays could be utilised to identify an alternative metal ion to Fe(II); concurrently, the explored broad screens should be repeated at 4 °C to more closely mimic the reported crystallisation conditions. Finally, if macrocrystallisation conditions cannot be identified, microcrystal electron diffraction (MicroED)[307], which is an electron cryomicroscopy (cryo-EM) technique, could be explored as an alternative.

Fragmentation of larger, imperfect crystals using sonication or glass bead milling methods has been shown to afford nanocrystals required for high-resolution structure determination using MicroED[308,309]. Additionally, nanocrystal soaking in crystallisation solutions containing inhibitors has also been reported[310], demonstrating the utility of this technique for co-crystallisation with small molecule inhibitors and TiP1.

Chapter 6
Summary and conclusions

Epigenetic modification of chromatin plays a key role in the regulation of gene expression and, consequently, phenotype presentation[8]. Deviation from the tightly controlled deposition and removal of epigenetic marks thus almost invariably results in cell death or disease[7]. The most abundant epigenetic DNA modification, ^{5m}C, is typically associated with transcriptional silencing[20], although context-dependent gene activation is possible[31,32]. ^{5m}C is iteratively oxidised to ^{5hm}C, ^{5f}C, and ^{5ca}C by the Fe(II)/2-OG-dependent TET enzyme family[39–41]. Identification of TETs prompted the discovery of thymine DNA glycosylase (TDG)-mediated active DNA demethylation pathways[45,46], as well as recognition of ^{5ox}C products as active epigenetic markers with distinct enrichment patterns and biological functions[117–149]. Both ^{5m}C hydroxylase and catalytic activity-independent functions of TETs are crucial for key developmental processes, including epigenetic reprogramming[150], pluripotency and differentiation control[152–155], and histone PTM reorganisation[156–158]. Accordingly, aberrant TET activity has been implicated in a number of diseases, including solid tumours[162] and haematopoietic malignancies[161], most notably AML[37]. Although TET dioxygenase significance in disease models has been well-characterised *in vivo* and in cells, biochemical and structural characterisation has been lacking. A number of TET inhibitors have been identified, but achieving sufficient paralogue and inter-family selectivity has proven difficult due to structural similarities within the 2-OG dioxygenase superfamily; three functionally non-redundant paralogues of TETs are present in humans[108], but no paralogue-specific cell-permeable inhibitors are available. The lack of selective chemical tools thus limits further studies on TET function in cancer progression and therapeutic target validation.

Chapter 2 outlines the biochemical characterisation of NgTet1 and inhibitor screening using ^{5m}C and 2-OG turnover-based detection methods. Following recombinant production of NgTet1, AlphaScreen, MALDI-TOF MS, Succinate-Glo™, and ^{1}H NMR assays were developed to facilitate bimodal TET activity detection. Investigation of NgTet1 kinetic parameters for 2-OG using ^{1}H NMR revealed substantial uncoupled 2-OG turnover levels in the absence of DNA, which increased 3.5-fold in the presence of non-substrate DNA. Addition of ^{5m}C-containing substrate DNA enhanced the total 2-OG decarboxylation rate 5.6-fold, but simultaneous analysis of ^{5m}C oxidation using MALDI-TOF MS suggested the proportion of uncoupled turnover remained substantial (up to 60%) even within the linear window of ^{5m}C substrate oxidation, demonstrating poor regulation of 2-OG decarboxylation. To determine if dioxygenase inhibitors affected both ^{5m}C substrate oxidation and 2-OG turnover, Fe(II)-chelating 2-OG mimetics (NOG[207], IOX1[182], and 2,4-PDCA[208]) and cyclic peptide inhibitor TiP1[191] were screened against NgTet1 using AlphaScreen and ^{1}H NMR. 2-OG mimetics suppressed DNA oxidation and uncoupled 2-OG turnover to a similar degree, in line with their proposed mechanism of action. Analogously, TiP1 inhibitory potency was virtually unchanged across the two assays, but a significant agonistic effect was observed for uncoupled 2-OG turnover at TiP1/NgTet1 ratios approaching 1:1 stoichiometry; a 2-OG saturation curve in the presence of 1.2 equivalents of TiP1 revealed that the peptide affected K_M^{app} and V_{max}^{app} parameters of NgTet1 without affecting the catalytic efficiency (k_{cat}/K_M), suggesting a potential reconfiguration of the enzyme active site following TiP1 binding. Although TiP1 inhibitory potency was unaffected by the addition of substrate, no catalytic activity enhancement could be observed in the presence of ^{5m}C-containing DNA, supporting a substrate-competitive mode of inhibition for TiP1, in line with analogous studies on other 2-OG-dependent dioxygenases[77,206].

Kinetic characterisation of TETs has been limited due to challenging protein production and dearth of high-throughput assays. Here, kinetic analysis of NgTet1 using ^{1}H NMR and MALDI-TOF MS identified uncoupled 2-OG decarboxylation as an underappreciated but highly prevalent facet of TET function. TETs may be implicated in the indirect regulation of other 2-OG dioxygenase activity *via* depletion of the 2-OG cofactor, although excessive uncoupled 2-OG turnover may be detrimental due to the release of oxidative potential within the cell; modulation of 2-OG turnover by small molecules and oligonucleotides observed *in vitro* suggests that 2-OG decarboxylation *in cellulo* could be regulated by cellular factors, such as metabolites or binding partner proteins. As NgTet1 originates from a lower eukaryote species, comparison with higher eukaryote TET homologues could offer insight into the evolution of 2-OG decarboxylase activity regulation within the TET family; to that end, analogous kinetic analysis could be carried out on the readily accessible hTET2$_{CD\Delta LCI}$ (Chapter 5). Importantly, the degree of uncoupled turnover depends on the assay conditions and nature of the substrate, as shown for other 2-OG dioxygenases[81,82]; thus, subsequent studies should initially focus on adjusting assay conditions (pH, temperature) and substrate (variation of flanking sequence on short oligomer substrates or use of genomic DNA isolates) to more closely emulate physiologically relevant conditions before any potential biological implications are asserted.

In **Chapter 3**, TiP1[191] and other TET-binding CPs identified using RaPID selection were evaluated for *in vitro* and cellular activity against hTETs and other 2-OG dioxygenases. All hTET1-binding CPs showed single-digit micromolar potencies against hTET2 using SPE-MS, but the majority of CPs could not be profiled for selectivity due to interference with AlphaScreen signal, with steep Hill slopes suggesting potential non-specific aggregation-based binding[219] to assay components. Conversely, TiP1 inhibited hTET1 with 200 nM potency and exhibited >30-fold selectivity over other TETs and Fe(II)/2-OG-dependent demethylases KDM5B/6B. Characterisation of TIP1 alanine variants identified Trp10 as a key residue for inhibition of

hTET1/2, while Glu11 was crucial for potency against hTET3, highlighting distinctions in peptide binding site features within hTETs. TiP1 did not elicit an inhibitory response in U2OS cells overexpressing hTETs, but was cytotoxic at concentrations above 100 μM; C–terminal conjugation of TiP1 with cell-penetrating peptides TAT and R8 failed to produce cellular activity, but the *in vitro* potency of TiP1-TAT was improved 10-fold compared to parent TiP1, likely due to conjugation with the TAT tag, which was identified as a single-digit micromolar hTET inhibitor. Investigation of the interaction patches between TiP1 and hTETs by probing hTET1 and hTET2 sequence tiling libraries with TiP1-FLAG was unsuccessful, but the absence of binding signals suggested that TiP1 binding to TETs may be controlled by tertiary structure interactions, rather than sequence patch recognition. Photo-crosslinking-mediated identification of TiP1 binding site was explored using a photoreactive TiP1 variant, TiP1 Y6Bpa, which largely retained the potency profile of parent TiP1; ligation of TiP1 Y6Bpa and NgTet1 could be observed using intact protein MALDI-TOF MS following UV irradiation, demonstrating the viability of Tyr6 variants as photo-affinity probes. hTET2-binding peptides KTT1–6 (identified using RaPID selection) could not be unambiguously validated as hTET2 inhibitors due to strong interference with AlphaScreen signal; however, treatment of hTET-overexpressing cells with KTT2 produced a cytotoxicity-unrelated decrease in ^{5hm}C levels at concentrations above 100 μM, indicating that KTT2 may inhibit TET enzymes.

Following extensive characterisation against TETs and other 2-OG dioxygenases, TiP1 has been established as the first ever selective inhibitor of a single hTET paralogue, supporting utility of TiP1 derivatives as probes for hTET1. The excellent selectivity profile of the peptide presents an alternative to hTET1 antibodies without the drawbacks associated with antibody-based approaches, such as batch-to-batch variability and host species-specific effects[311]. Importantly, TiP1-based affinity and photo-affinity probes could be used to characterise native hTET1-protein binding partner complexes in lysates of clinically relevant cancer cell lines.

Although TiP1 and CPP conjugates were inactive in cells, the SAR data generated for TiP1 may guide the future design of small molecule peptidomimetic chemical probes, with similar approaches successfully employed to develop severe acute respiratory syndrome-coronavirus (SARS-CoV) protease inhibitors[312] and von Hippel−Lindau (VHL) ligands[313].

Chapter 4 describes the identification of a novel small molecule hTET2 inhibitor scaffold, SAR studies, and cellular activity evaluation of selected analogues. Screening of a small compound library with reported or predicted 2-OG dioxygenase inhibitory activity against hTET2 using AlphaScreen identified *S*–substituted 5-carboxy-2-thiouracils as single-digit micromolar hTET2 inhibitors. Structural variation of the *S*–substituent was typically tolerated well with minimal effect on hTET2 inhibition, although analogues containing sterically large side chains, such as a nonyl chain (**36**) or multiple ring systems (**52**–**54**), exhibited up to 5-fold reductions in potency. Conversely, scaffold alterations at positions 3-, 4-, and 5- resulted in a substantial decrease in potency (up to 100-fold), demonstrating the importance of the iron-chelating β-ketoacid moiety for the activity of *S*–substituted analogues. Interestingly, thione derivatives tolerated a large degree of scaffold derivatisation, with a 10-fold increase in potency observed for 4-carboxyl **56** and 4–*H* **63** compared to **23** (hTET2 IC_{50} = 141, 195, and 1,950 nM, respectively), suggesting distinct binding modes for thioether and thione compounds. A selection of analogues was profiled for hTET paralogue selectivity; *S*–benzyl analogue **15** exhibited 10-fold selectivity for hTET2 over hTET1/3, whereas *S*–butyl analogue **14** exhibited 50-fold selectivity over hTET3. However, strong correlations for potency changes within the hTET family predicted difficulties in further selectivity profile optimisation. Importantly, screening of **15**, **14**, and **23** against a panel of 2-OG dioxygenases indicated a high degree of selectivity for hTETs[265]. A selection of 5-carboxy-2-thiouracil analogues was tested for cellular activity, but no response could be observed. Derivatisation of the carboxylic acid moiety to ethyl esters (**15a** and **23b**) did not improve activity *in cellulo*, whereas methyl esters **63a** and

23a reduced cellular ^{5hm}C levels in a dose-dependent manner with EC_{50} values of approx. 10 µM. The cellular selectivity profiles of **63a** and **23a** were largely pan-specific, and the methyl ester derivatives also showed considerably reduced paralogue selectivity *in vitro* compared to parent carboxylic acids **63** and **23**, suggesting that the esters may not be readily hydrolysed by endogenous esterases, instead acting as the primary inhibitory species *in cellulo*.

Methyl ester derivatives of the 5-carboxy-2-thiouracil series expand the TET molecular toolkit and constitute a promising starting point for cell-active selective inhibitors of TETs. While esterification approaches complicate *in vivo* work, they are useful for cell-based studies focusing on temporal inhibition and target validation. Cell treatment with thione esters decreased cellular ^{5hm}C levels without affecting hTET overexpression, suggesting a specific on-target effect, although chemoproteomic validation will be required to confirm target engagement in a cellular environment. Future work could focus on replicating the *in vitro* paralogue selectivity in a cellular context: exploration of ester derivatives, both in terms of parent inhibitor structure and ester moiety, could lead to improved cellular selectivity, furnishing the first ever hTET2-specific inhibitors.

Chapter 5 presents the optimisation of catalytically active, crystallisable[94] hTET2 construct ($hTET2_{CD\Delta LCI}$) production for crystallographic applications, and subsequent work towards crystallisation of hTET2 in complex with DNA and TET inhibitors. Overexpression of a gene encoding for $hTET2_{CD\Delta LCI}$ fused to a SUMO solubility tag and a His_6 affinity tag in *E. coli*, followed by Ni(II) affinity and size exclusion purification furnished poorly active tagged hTET2 in low yield (approx. 0.4 mg/L) and poor purity, leading to unsuccessful co-crystallisation trials with TiP1. Western blot-guided expression optimisation exploring different solubility tags, *E. coli* strains, cultivation and post-induction temperatures, induction methods, and growth

medium additives did not afford an appreciable increase in soluble protein levels. Subsequent optimisation of buffer pH and affinity purification along with the addition of solubility tag cleavage and secondary heparin affinity purification steps substantially improved specific activity, yield (2.8 mg/L), and purity of the isolated enzyme, enabling detailed crystallisation efforts. Crystals of the hTET2:dsDNA:NOG complex could not be readily reproduced using reported[94] conditions at 4 °C and 22 °C, although granular and flocculent precipitation was observed across a range of pH and PEG 2000 MME precipitant concentrations, suggesting that further optimisation of selected conditions could yield crystals suitable for X-ray data collection. Sparse matrix and broad crystal screens were set up at 20 °C for the hTET2:dsDNA:NOG complex; after 14 days, crystalline precipitate was observed in three conditions, one of which bore a striking similarity to the reported[94] conditions, offering additional options for further condition optimisation.

Significant improvements in the yield and purity of hTET2 in this work have made structural characterisation efforts substantially more tractable. Co-crystallisation of hTET2 in complex with small molecule and cyclic peptide inhibitors would have a tremendous impact on structure-guided design of TET inhibitors. The work also lays the foundation for resource-intensive biochemical studies on the hTET family, such as in-depth kinetics on substrate and 2-OG turnover, by removing one of the major barriers – accessibility to recombinant enzyme. Furthermore, robust production of hTET2 could support high-throughput screening efforts for the identification of novel inhibitor scaffolds.

In conclusion, this body of work provides novel biochemical and biophysical insights into the important but understudied epigenetic enzyme family of TETs. Uncoupled 2-OG decarboxylation was identified as a tuneable aspect of TET function, and factors influencing activity were explored. The work provides novel tools for the investigation of paralogue- and

pan-specific TET roles in development and cancer contexts, delivers promising starting points for the development of TET chemical probes, and presents critical steps towards a platform for structure-guided design of TET inhibitors. The results invite further studies on this DNA-modifying family of enzymes, particularly in regard to the potential TET role in regulation of 2-OG levels and disease context-specific variation of TET interactions with protein binding partners.

Chapter 7
Materials and methods

7.1 General information

Commonly used chemicals were purchased and used as received from Merck, Thermo Fisher, and Fluorochem. Organic solvents were purchased from Merck unless mentioned otherwise. Water used refers to water purified using Milli-Q® Reference Water Purification System equipped with a Quantum® TEX Polishing Cartridge and a Millipak® Express 40 Filter unless stated otherwise. Catalogue numbers (Cat#) are provided for reagents where a particular formulation may be essential for experiment reproducibility.

7.2 Peptide synthesis

7.2.1 General information

Materials: Fmoc-protected natural amino acids, Rink amide 4-methylbenzhydrylamine (MBHA) resin, *N*,*N*'–diisopropylcarbodiimide (DIC), and Oxyma Pure were purchased from Merck, CS Bio, Novabiochem®, and Fluorochem. Fmoc-protected D-amino acids (Fmoc-D-Tyr(tBu)-OH and Fmoc-D-Trp(Boc)-OH) and α-cyano-4-hydroxycinnamic acid (α-CHCA, matrix substance for MALDI-MS, ≥99.0%) were purchased from Sigma-Aldrich. Fmoc-L-Bpa-OH was purchased from Iris Biotech. Peptide synthesis grade DMF and piperidine (20% v/v in DMF) were purchased from Cambridge Bioscience or Merck. Puradisc 4 mm non-sterile PTFE syringe filters (0.45 µm) were purchased from Cytiva.

Synthetic peptide purchasing: FLAG, R8, and TAT peptides were purchased as N–terminally acetylated C–terminal amides from GL Biochem (>95% purity by LC-MS).

7.2.2 Solid Phase Peptide Synthesis (SPPS)

Linear peptides were synthesised on 50 or 100 µmole scale using standard Fmoc solid phase peptide synthesis methods on a CEM Liberty Blue Microwave Synthesiser according to the

manufacturer's instructions[314]. All peptides were synthesised as C–terminal amides. Briefly, Rink Amide MBHA resin (0.5 mmol/g loading capacity) was soaked in *N*,*N*'–dimethylformamide (DMF, 10 mL) and transferred to the reaction vessel. Microwave-assisted Fmoc deprotection was carried out using piperidine in DMF (20% v/v, 4 mL), and the resin was washed with DMF. Fmoc-protected amino acid (0.2 M, 2.5 mL), DIC (0.5 M, 1.0 mL), and Oxyma Pure (1.0 M, +1.9% v/v *N*,*N*–diisopropylethylamine [DIPEA], 0.5 mL) were added to the reaction vessel, and the mixture was irradiated at 90 °C for 120 sec unless otherwise mentioned. His and Cys couplings were carried out at 50 °C for 600 sec; Arg couplings were carried out twice at 75 °C for 420 sec. The resin was then washed with DMF (3×) before the subsequent Fmoc deprotection was carried out.

7.2.3 Resin-bound peptide capping

The finished resin-bound, side chain-protected linear peptides were modified at their N–termini by capping with chloroacetic anhydride or acetic anhydride for cyclic or linear peptides, respectively. Capping reagent (10 eq.) in DMF (5 mL) was added onto the resin, and the mixture was incubated at room temperature (RT) for 2 h with occasional agitation. A further portion of capping reagent (5 eq.) in DMF (2.5 mL) was added, and the mixture was incubated at RT for an additional 1 h. Following N–terminal capping, the suspensions were washed with DMF and filtered.

7.2.4 Acidic deprotection of peptides

The resins were washed with CH_2Cl_2 (3×15 mL) and dried under reduced pressure (30 min). The dry resins were suspended in a 92.5:2.5:2.5:2.5 mixture of trifluoroacetic acid (TFA)/1,3-dimethoxybenzene/triisopropyl silane (TIPS)/H_2O (5 mL per peptide), and the suspensions were incubated at RT for 3 h with occasional agitation. The resin suspensions were then filtered, and crude peptides were precipitated from the filtrates in Et_2O (45 mL) at −20 °C for 16 h. The settled suspensions were pelleted (3,082 ×g, 4 °C, 10 min), resuspended

in fresh Et_2O (50 mL), and precipitated at −20 °C for 16 h, repeating for a total of 3 times. The resulting pellets were dissolved in a minimal amount of H_2O and MeCN. The volume and composition of the solution depended on peptide solubility. The solutions were snap frozen in liquid N_2 and lyophilised to afford the crude linear peptides as colourless or pale-yellow solids.

7.2.5 Peptide cyclisation in DMSO

Crude *N*–chloroacetylated linear peptides were suspended in dimethyl sulfoxide (DMSO, 30 mg/mL) and incubated at 35 °C for 10 min with sonication to ensure complete dissolution. The pH was adjusted to approx. 9 by the dropwise addition of DIPEA. The resulting solutions were incubated at 40 °C for 1.5 h or until further cyclisation could not be observed by MALDI-TOF MS (assessed by the ratio of linear to cyclic peptide peaks). The reactions were quenched by adjusting the pH to 4-5 *via* the dropwise addition of neat TFA.

7.2.6 Peptide cyclisation in 1:1 40 mM NH_4HCO_3/MeCN

Crude *N*–chloroacetylated linear peptides were suspended in a 1:1 mixture of 40 mM NH_4HCO_3/MeCN (10 mg/mL) and incubated at 35 °C for 10 min with sonication to maximise dissolution. Complete dissolution could not be achieved in the majority of cases. The pH was adjusted to approx. 9 by the dropwise addition of NaOH (1 M). The resulting solutions were incubated at 35 °C for 30 min or until further cyclisation could not be observed by MALDI-TOF MS (assessed by the ratio of linear to cyclic peptide peaks). The reactions were quenched by adjusting the pH to 4-5 *via* the dropwise addition of neat TFA. The cyclised peptides were snap frozen in liquid N_2, lyophilised, and dissolved in DMSO.

7.2.7 HPLC purification

Peptide solutions in DMSO were filtered (0.45 µm Whatman filters) and purified on a Shimadzu Prominence HPLC system equipped with Phenomenex Gemini® 5 µm NX-C18 110 Å LC columns (00G-4454-U0-AX or 00G-4454-P0-AX, depending on purification scale). A

H_2O/MeCN solvent system supplemented with 0.1% v/v TFA was used. General gradient: 10–25% (5 min), 25–60% (30 min), 60–98% (5 min), 98% (10 min) MeCN in H_2O (+0.1% v/v TFA). Fractions were analysed by MALDI-TOF MS; selected fractions were combined and lyophilised to afford the pure products.

7.2.8 Peptide analysis using MALDI-TOF MS

Spectra recording: All MALDI-TOF MS spectra were recorded on a Bruker MALDI-TOF Autoflex maX mass spectrometer (Bruker) on either MTP 384 polished steel or MSP 96 Ground Steel target plates. Data was collected using positive ion reflectron mode, 700–3,500 Da collection range.

Matrix preparation: A saturated matrix solution was prepared fresh by suspending α-CHCA (10 mg/mL) in a 1:1 mixture of H_2O/MeCN supplemented with TFA (0.1% v/v) and sonicating the mixture for 10 min. The matrix was stored at RT.

Sample deposition: 1 µL of the matrix was deposited on the target plate, followed by 1 µL of sample, and the solutions were mixed. The mixture was allowed to air-dry before analysis. Peptide concentrations were generally within 1–10 µM. Up to 1% v/v DMSO in the peptide sample was practicable, whereas higher DMSO content negatively impacted sample analysis by affecting matrix crystallisation.

7.2.9 Purity and high-resolution mass analysis

Peptide purity was determined using LC-UV analysis at 220 nm. High-resolution mass spectra (HRMS) were recorded using an ACQUITY Ultra Performance LC system coupled to a Waters Xevo G2-XS Q-TOF mass spectrometer equipped with an ESI LockSpray™ source as described in section 7.19. Purity and HRMS information for all peptides used in this work is reported in Appendix 2.

7.2.10 Accurate peptide concentration determination

Pure cyclic peptides were dissolved in H_2O or DMSO to produce concentrated peptide stock solutions (10–100 mM). Water-soluble peptides were diluted to approximately 1 mM in H_2O (+10% v/v D_2O), and accurate concentration was established on a Bruker Avance III 700 MHz equipped with an inverse TCI cryoprobe (acquisition parameters: TD = 128000, DS = 0, D1 = 60, P12 = 1800) using TSP-d_4 as the internal standard at a 0.01 mg/mL concentration. Accurate concentrations for water-insoluble peptides in DMSO were determined on a NanoDrop™ One Microvolume UV-Vis Spectrophotometer (absorbance measurement at 280 nm, normalised against DMSO) based on the extinction coefficient, ε, calculated using the Expasy ProtParam tool[298].

7.3 General molecular biology methods

7.3.1 Materials

Pipette tips, Eppendorf tubes, and H_2O were autoclaved at 121 ˚C for 20 min before use. Q5 polymerase was purchased from NEB. BioTaq polymerase was purchased from Bioline. SYBR Safe DNA gel stain, 6× DNA Gel Loading Dye, GeneJET PCR Purification Kits, and GeneJET Gel Extraction Kits were purchased from Thermo Fisher. GeneRuler 1 kb DNA ladders and 100 bp DNA ladders were purchased from Invitrogen and NEB, respectively. Custom oligonucleotide primers were purchased from Merck or Thermo Fisher (Life Technologies). XL10-Gold ultracompetent *E. coli* cells (Cat# 200317) were purchased from Agilent.

7.3.2 Polymerase Chain Reaction (PCR)

PCR reactions were carried out in standard thin wall 100 µL PCR tubes in a ^{3}Prime thermocycler (TECHNE). Reaction concentrations and parameters were tailored specifically to each experiment following the polymerase manufacturer's recommendations. A representative PCR mixture is shown below (Table 7.1).

Table 7.1 | Representative composition of a PCR mixture for gene insert amplification. Primer sequences for PCR are reported in section 7.4.2 (restriction cloning of pCOLD I) and Table 7.6 (ligation-independent cloning).

Component	Stock concentration	Volume (µL)	Final concentration
$hTET2_{CD\Delta LCI}$ in pET-28b	5 ng/µL	20	2 ng/µL
Forward primer	10 µM	2.5	0.5 µM
Reverse primer	10 µM	2.5	0.5 µM
Q5 buffer	5×	10	1×
dNTP mix	10 mM	1	200 µM
Sterile H_2O	-	13.5	-
Q5 polymerase	-	0.5	-
Final volume	-	50	-

Q5 polymerase was used for gene insert amplification; BioTaq polymerase was used for high-throughput identification of clone colonies containing the gene of interest. Annealing temperature was calculated using the NEB T_m calculator[315]. The prepared PCR mixture was incubated in a thermocycler for 25–40 cycles. A representative example of a thermocycler programme used for gene insert amplification is shown in Table 7.2.

Table 7.2 | Representative thermocycler programme for gene insert amplification.

Step		Temperature (°C)	Time (sec)
Initial denaturation		98	120
30 cycles	Denaturation	98	10
	Annealing	68	20
	Extension	72	60
Final extension		72	120

7.3.3 Agarose gel electrophoresis

Buffer: TAE buffer was prepared as a 50× stock solution and stored at RT; the 50× stock was diluted to 1× prior to use (1× composition: 40 mM Tris, 20 mM acetate, 1 mM EDTA, pH approx. 8.6).

PCR and restriction enzyme digestion products were separated by agarose gel electrophoresis prior to downstream applications. Agarose gels were prepared by suspending agarose (0.5–2% w/v) in 1× TAE buffer. Microwave irradiation was used to fully dissolve the agarose. After the solution had partially cooled, SYBR Safe was added (10,000× stock, 1× final) and the

resulting mixture was allowed to set in a gel casting tray. GeneRuler 1 kb DNA ladder and 100 bp DNA ladder were used as gel ladders (5 µL loading). Samples were mixed with 6× DNA Gel Loading Dye in a 5:1 ratio and loaded onto the gels (30 µL maximum loading). The gels were run at 100 V for 30–60 min in 1× TAE buffer and imaged under UV light on a BIO-RAD ChemiDoc MP Universal Hood III ("SYBR Safe" method). In cases where band excision was desired, gels were imaged under blue light ("SYBR Safe Excision" method) to avoid DNA degradation.

7.4 Restriction cloning of pCOLD I

7.4.1 Materials

Modified empty pCOLD I plasmid was kindly gifted by Dr Patrick Rabe (University of Oxford, Table 7.3). Q5 polymerase, T4 DNA ligase, and restriction enzymes (SalI-HF and NotI-HF) were purchased from NEB. Other materials used were as described in section 7.3.1.

7.4.2 $hTET2_{CD\Delta LCI}$ insert preparation

Gene insert coding for $hTET2_{CD\Delta LCI}$ (D1129–S1480_GGGGSGGGGSGGGGS_N1844–G1936) was amplified with appropriate overhangs from a modified pET-28b donor plasmid kindly gifted by Prof. Yanhui Xu[94] (State Key Laboratory of Genetic Engineering, Fudan University, China). A PCR mixture was prepared and incubated in a thermocycler using appropriate conditions for 30 cycles (see section 7.3.2) with the following primers: 5'-CTA GCG CAC GTC GAC GAT TTC CCA TCT TGC AGA TGT G-3' (forward primer) and 5'-CTA GCG CAT GCG GCC GCT TAG CCA TAC TTT TCA CAC TCT TCC TC-3' (reverse primer). The primers were designed to have at least 15 bp overlap with the target gene. A random 6–9 nt extension sequence was added at the 5' of the primers to facilitate restriction enzyme binding. The PCR mixture was separated on a 2% agarose gel and the amplified gene insert was extracted using a GeneJET Gel Extraction Kit.

7.4.3 pCOLD I linearisation and hTET2$_{CD\Delta LCI}$ insert digestion

General information for the modified pCOLD I vector used herein can be found in Table 7.3.

Table 7.3 | General information of modified pCOLD I vector. TEE, translation enhancing element; His_6, hexahistidine tag; Xa, factor Xa cleavage site; Amp, ampicillin.

Name	Source	Addgene #	Tag	Terminus	Resistance
pCOLD I (modified)	Dr Patrick Rabe/Takara Bio	N/A	TEE-His_6-Xa	N–terminal	Amp

Recipient plasmid pCOLD I (modified to contain a NotI restriction site) and the hTET2$_{CD\Delta LCI}$ gene insert were digested at 37 °C for 16 h using SalI-HF and NotI-HF restriction enzymes (Table 7.4). The linearised pCOLD I was separated on a 1% agarose gel and the desired band was extracted using a GeneJET Gel Extraction Kit. The digested hTET2$_{CD\Delta LCI}$ gene insert was purified using a GeneJET PCR Purification Kit.

Table 7.4 | Reaction mixture composition for the digestion of pCOLD I vector and hTET2$_{CD\Delta LCI}$ gene insert.

Component	Stock concentration	Volume (µL)	Final amount
Modified pCOLD I plasmid ***or*** hTET2$_{CD\Delta LCI}$ gene insert	varied	varied	2 µg
CutSmart buffer	10×	5	1×
SalI-HF	20 U/µL	2	40 U
NotI-HF	20 U/µL	2	40 U
Sterile H_2O	-	varied	-
Final volume	-	50	-

7.4.4 Ligation and transformation

The digested gene insert and the linearised pCOLD I were mixed in a 3:1 molar ratio along with T4 DNA ligase (Table 7.5) and incubated at RT for 48 h. Plasmid propagation was achieved by transforming the ligation mixture (5 µL) into XL10-Gold ultracompetent cells (35 µL) as described in section 7.6.4. The transformed bacterial cells were plated onto LB agar plates containing the appropriate antibiotics. Colony PCR was performed to confirm gene insert presence in selected colonies. Plasmids were purified from 10 mL scale bacterial growths using

a GeneJET Plasmid Miniprep Kit and sequence-verified by Sanger sequencing (TubeSeq, Eurofins Genomics).

Table 7.5 | Ligation mixture composition for pCOLD I (hTET2$_{CD\Delta LCI}$) ligation.

Component	Stock concentration	Volume (µL)	Final amount
Linearised pCOLD I plasmid	varied	varied	0.02 pmol
Digested hTET2$_{CD\Delta LCI}$ gene insert	varied	varied	0.06 pmol
T4 DNA ligase buffer	10×	2	1×
T4 DNA ligase	400 U/µL	1	400 U
Sterile H_2O	-	varied	-
Final volume	-	20	-

7.5 Ligation-independent cloning (LIC): general protocol

7.5.1 Materials

Restriction enzymes BsaI-HF and SspI-HF were purchased from NEB. LIC-qualified T4 DNA polymerase (Cat# 70099-M) was purchased from Merck Novagen. Other materials used were as described in section 7.3.1.

7.5.2 LIC insert preparation

The procedure was adapted from reported protocols[316]. Primers for LIC were designed according to Addgene or SGC vector information with at least a 15 bp overlap with the target gene. Gene insert coding for hTET2$_{CD\Delta LCI}$ (D1129–S1480_GGGGSGGGGSGGGGS_N1844–G1936) was amplified with appropriate overhangs from a modified pET-28b donor plasmid kindly gifted by Prof. Yanhui Xu[94] (State Key Laboratory of Genetic Engineering, Fudan University, China). PCR mixtures were prepared and incubated in a thermocycler using appropriate conditions for 30 cycles (see section 7.3.2) with the appropriate primers (Table 7.6). The PCR mixtures were separated on a 2% agarose gel and the desired bands were extracted using a GeneJET Gel Extraction Kit.

Table 7.6 | Forward and reverse primer sequences for LIC cloning. LIC tag sequences are shown in bold.

Vectors	Primer	Sequence
pNH-TrxT and pGTVL2	Forward	5′-**TAC TTC CAA TCC ATG** GAT TTC CCA TCT TGC AGA TGT G-3′
	Reverse	5′-**TAT CCA CCT TTA CTG** TTA GCC ATA CTT TTC ACA CTC TTC CTC-3′
1GFP and 1M	Forward	5′-**TAC TTC CAA TCC AAT GCA** GAT TTC CCA TCT TGC AGA TGT G-3′
	Reverse	5′-**TTA TCC ACT TCC AAT GTT ATT A** TTA GCC ATA CTT TTC ACA CTC TTC CTC-3′

Insert cohesive ends were generated using LIC-qualified T4 DNA polymerase. Reaction mixtures were prepared (Table 7.7) and incubated at 25 °C for 30 min, followed by enzyme inactivation at 75 °C for 20 min (thermocycler lid pre-heated to 90 °C).

Table 7.7 | General composition of insert cohesive end generation mixture.

Component	Stock concentration	Volume (µL)	Final amount
Gene insert	varied	8	varied
LIC-qualified T4 DNA polymerase buffer	10×	1	1×
dCTP	100 mM	0.25	2.5 mM
DTT (0.20 µm filtered)	100 mM	0.5	5 mM
LIC-qualified T4 DNA polymerase	N/A	0.25	N/A
Final volume	-	10	-

7.5.3 LIC vector preparation

General information for vectors used in LIC can be found in Table 7.8.

Table 7.8 | General information of LIC vectors used for cloning. His_6, hexahistidine tag; TEV, TEV protease cleavage site; Trx, thioredoxin; GST, glutathione *S*–transferase; GFP, green fluorescent protein; MBP, maltose binding protein; Kan, kanamycin.

Name	Source	Addgene #	Tag	Terminus	Resistance
pNH-TrxT	Opher Gileadi (SGC)	26106	His_6-Trx-TEV	N–terminal	Kan
pGTVL2	Pavel Savitsky (SGC)	N/A	His_6-GST-TEV	N–terminal	Kan
1GFP	Scott Gradia	29663	His_6-GFP-TEV	N–terminal	Kan
1M	Scott Gradia	29656	His_6-MBP-TEV	N–terminal	Kan

LIC vectors containing a negative sucrose selection gene, SacB, were digested at 37 °C for 2–3 h using appropriate restriction enzymes: BsaI-HF was used for the digestion of pNH-TrxT and

pGTVL2 vectors; SspI-HF was used for the digestion of 1GFP and 1M vectors (Table 7.9). 0.5–5 µg of intact vector was used for each digestion. The digested vectors were purified using a GeneJET PCR Purification Kit. Small portions of the purified digested vectors (50–150 ng) were analysed on a 0.8% agarose gel to verify the linearisation.

Table 7.9 | General composition of a LIC vector digestion mixture.

Component	Stock concentration	Volume (µL)	Final amount
Intact LIC vector	varied	varied	2 µg
CutSmart buffer	10×	5	1×
BsaI-HF / SspI-HF	20 U/µL	1.5	30 U
Sterile H_2O	-	varied	-
Final volume	-	50	-

Linearised vector cohesive ends were generated using LIC-qualified T4 DNA polymerase. Reaction mixtures were prepared (Table 7.10) and incubated at 25 °C for 30 min, followed by enzyme inactivation at 75 °C for 20 min (thermocycler lid pre-heated to 90 °C).

Table 7.10 | General composition of linearised vector cohesive end generation mixture.

Component	Stock concentration	Volume (µL)	Final amount
Linearised vector	varied	18	varied
LIC-qualified T4 DNA polymerase buffer	10×	4	1×
dGTP	100 mM	1	2.5 mM
DTT (filtered)	100 mM	2	5 mM
LIC-qualified T4 DNA polymerase	N/A	1	N/A
Final volume	-	40	-

7.5.4 Annealing and transformation

Mixtures of LIC-qualified T4 DNA polymerase-treated gene insert and vector were prepared in a 1:2 and 2:1 ratio (3 µL final volume) and transformed into XL10-Gold ultracompetent *E. coli* cells (25 µL) as described in section 7.6.4. LB agar plates contained sucrose (5% w/v), lethal to cells transformed with empty LIC vectors containing the SacB gene. Colony PCR was performed to confirm gene insert presence in selected colonies. Plasmids were purified from 10 mL scale

bacterial growths using a GeneJET Plasmid Miniprep Kit and sequence-verified by Sanger sequencing (TubeSeq, Eurofins Genomics).

7.6 General microbiology methods

7.6.1 Materials

All microbiology experiments were performed under standard sterile conditions. Pipette tips, Eppendorf tubes, and solutions were autoclaved at 121 ˚C for 20 min before use. Transformations and culture inoculations were performed in a laminar flow hood. Water purified using the Elix® Essential 10 UV Water Purification System equipped with a Progard® TS2 Pretreatment Pack was used for making growth media and agar plates. All components for growth media were purchased from Sigma-Aldrich. SOC outgrowth medium was purchased from NEB. Ampicillin, kanamycin (sulphate, 1:1), and chloramphenicol were purchased from Apollo Scientific. XL10-Gold ultracompetent *E. coli* cells (Cat# 200317) were purchased from Agilent. Competent *E. coli* BL21(DE3)-R3-pRARE2 Rosetta-type cells were kindly gifted by Eidarus Salah (produced in-house by SGC)[317]. Competent BL21(DE3) (Cat# C2527), Lemo21(DE3) (Cat# C2528), and SHuffle® T7 (Cat# C3026) *E. coli* cells were purchased from NEB. NuPAGE™ 4-12% Bis-Tris, 1.0 mm, 17-well pre-cast gels and NuPAGE™ MES SDS Running Buffer were purchased from Thermo Fisher. InstantBlue™ Ultrafast Protein Stain (discontinued) was purchased from Sigma-Aldrich. QuickBlue Protein Stain was purchased from LubioScience.

7.6.2 Media preparation

Growth medium was prepared fresh for every cell pellet growth and autoclaved at 121 ˚C for 20 min before use. 3 types of non-inducing media were used (Table 7.11).

Table 7.11 | Compositions of non-inducing media.

Medium	Composition
LB (Luria-Bertani)	1% w/v tryptone, 0.5% w/v yeast extract, 1% w/v NaCl
2×YT	1.6% w/v tryptone, 1% w/v yeast extract, 0.5% w/v NaCl
TB (Terrific Broth)	2% w/v tryptone, 2.4% w/v yeast extract, 0.4% v/v glycerol, phosphate buffer* [24.6 mM KH_2PO_4, 75.4 mM K_2HPO_4, pH 7.4]

**Phosphate buffer was autoclaved separately from the other components as a 10× stock and re-combined afterwards to avoid phosphate salt precipitation.*

ZYM-5052 autoinduction medium was used in combination with MDG or MDAG-135 non-inducing minimal media for starter culture growths. MDAG-135 minimal medium was used in cases where overnight growth in MDG could not be observed. The media were prepared following the reported procedures[318] with the exception of substituting the metal trace mixture with $FeCl_3$ (100 µM final concentration). The final compositions of the media are presented in Table 7.12.

Table 7.12 | Compositions of autoinduction medium ZYM-5052 and non-inducing minimal media MDG and MDAG-135.

Medium	Composition
ZYM-5052	1% w/v N-Z-Amine® AS, 0.5% w/v yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 100 µM $FeCl_3$, 0.5% w/w glycerol, 0.05% w/v glucose, 0.2% w/v α-lactose
MDG	25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 100 µM $FeCl_3$, 0.5% w/v glucose, 0.25% w/v Asp
MDAG-135	25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 100 µM $FeCl_3$, 0.35% w/v glucose, 0.1% w/v Asp, 200 µg/mL of all amino acids except Cys and Tyr

7.6.3 Antibiotic stock preparation

Antibiotics were prepared as 1000× stock solutions, filter sterilised (0.20 µm), and stored at −20 °C prior to use. Stock preparation information is shown in Table 7.13. 2× kanamycin (100 µg/mL) was required when using ZYM-5052 autoinduction media to prevent growth of untransformed bacteria.

Table 7.13 | Antibiotic stock preparation information.

Antibiotic	1000× stock concentration	1× stock concentration	Solvent
Ampicillin (Amp)	100 mg/mL	100 µg/mL	H_2O
Kanamycin (Kan)	50 mg/mL	50 µg/mL	H_2O
Chloramphenicol (Camp)	34 mg/mL	34 µg/mL	EtOH

7.6.4 Bacterial transformation

Agar plate preparation: Agar stock for agar plates was prepared by adding agar (1.5% w/v) to LB medium (section 7.6.2), and autoclaving the mixture at 121 ˚C for 20 min before use. In the case of LB agar supplemented with sucrose (5% w/v), LB agar was made up to 90% volume, autoclaved, and sucrose in sterile water (50% w/v, 0.20 µm filter sterilised) was added as the final 10% of volume.

Transformation: Plasmid containing desired vector (1–3 µL) was incubated with (ultra)competent *E. coli* cells (15–30 µL) on ice for 20 min, followed by heat shock (42 °C, 13–30 sec, as per the manufacturer's instructions). The mixture was returned to ice for 20 sec, SOC outgrowth media (300 µL) was added, and the cells were incubated at 37 °C for 1 h. The transformants were plated on an LB agar plate (supplemented with appropriate antibiotics) and incubated at 37 °C for 16 h to produce single colonies. XL10-Gold ultracompetent *E. coli* cells were used for plasmid propagation. BL21(DE3), Lemo21(DE3), SHuffle® T7, and BL21(DE3)-R3-pRARE2 competent *E. coli* cells were used for protein expression.

Antibiotic selection: When using XL10-Gold, BL21(DE3)-R3-pRARE2, or Lemo21(DE3) cell lines, chloramphenicol was added to the LB agar as a secondary antibiotic selection.

7.6.5 SDS-PAGE

NuPAGE™ 4-12% Bis-Tris, 1.0 mm, 17-well precast protein gels were used in most cases. For hand-cast gels, separating and stacking gel mixtures (Table 7.14) were initially prepared without APS. APS was added to the separating gel mixture (12.5% acrylamide), and the mixture

was poured into the cast. A small amount of EtOH/^{i}PrOH was added on top to remove bubbles and reduce moisture loss, and the separating gel was allowed to set at RT for 30 min, after which time the alcohol was removed. APS was added to the stacking gel mixture, and the mixture was poured into the cast. The gels were allowed to solidify at RT for 30 min before storing at 4 °C.

Table 7.14 | SDS-PAGE gel preparation stock solutions for 2 gels.

Type	Tris, 1.5 M, pH 8.8	Tris, 0.1 M, pH 6.8	H_2O	30% w/v acrylamide	10% w/v SDS	TEMED	10% w/v APS
Separating	5.88 mL	-	890 µL	5.00 mL	120 µL	10 µL	100 µL
Stacking	-	2.45 mL	1.89 mL	500 µL	50 µL	15 µL	100 µL

Sample Preparation: Protein samples were mixed in a 1:1 ratio with 2× loading buffer (108 mM Tris-HCl, pH 6.8, 20% v/v glycerol, 3.6% w/v SDS, 200 mM DTT, 0.004% w/v Bromophenol Blue) or a 5:1 ratio with 6× loading buffer (375 mM Tris-HCl, pH 6.8, 50% v/v glycerol, 9% w/v SDS, 200 mM DTT, 0.006% w/v Bromophenol Blue), and denatured at 95 °C for 5 min. DTT stock was made fresh and added to the loading buffer for each SDS-PAGE experiment.

Sample loading: up to 20 µL of sample was loaded per well.

Electrophoresis: the gels were run at 180 V for 50 min or until the dye front had reached the bottom of the gel.

Hand-cast gel running buffer: 25 mM Tris, 192 mM glycine, 0.1% w/w SDS, pH 8.3.

Precast gel running buffer: 50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3 (NuPAGE™ MES SDS Running Buffer).

Gel staining: InstantBlue™ Ultrafast Protein Stain or QuickBlue Protein Stain staining at RT for 30 min on a 3D orbital shaker.

Gel imaging: Gels were imaged on a BIO-RAD ChemiDoc MP Universal Hood III ("Protein gels – Coomassie" method).

7.6.6 Transformed cell glycerol stock preparation and use

Preparation: Cells from an appropriate cell line were transformed as described in section 7.6.4. A culture was prepared by inoculating 100 mL of LB, 2×YT, or TB media (supplemented with appropriate antibiotics) with 1–5 single colonies and incubating at 37 °C, 180 rpm for 4–5 h. In sterile screw cap vials, cell culture (250 µL) was mixed to uniformity with sterile glycerol (250 µL, 50% v/v), and stored at –80 °C.

Revival: Glycerol stocks were placed on dry ice. To prepare a starter culture, a sterile 200 µL pipette tip was used to scrape the surface of the frozen glycerol stock, and was placed into 100 mL of 2×YT, TB, MDG, or MDAG-135 media (supplemented with appropriate antibiotics). The inoculated media was incubated at 37 °C, 180 rpm for 16 h.

7.7 Recombinant protein expression

7.7.1 Materials and general information

Materials: Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was purchased from Apollo Scientific.

General information: Media preparation, antibiotic stock preparation, bacterial transformation, and glycerol stock revival protocols are described in section 7.6. Vector antibiotic resistance information is reported in sections 7.4.3 and 7.5.3.

7.7.2 Recombinant protein expression for expression trials using non-inducing media

Cells from an appropriate cell line were transformed fresh or a glycerol stock was used for starter culture inoculation. A starter culture was prepared by inoculating 10 mL of 2×YT or TB media (supplemented with appropriate antibiotics) with 1–5 single colonies or glycerol stock

scrapings, and incubating at 37 °C, 180 rpm for 16 h. The starter culture was expanded into 100 mL of TB medium (supplemented with appropriate antibiotics) by inoculating with the saturated starter culture (1% v/v). The cultures were incubated at 37 °C, 180 rpm for 6 h, at which point the cultures were moved to 4 °C for 30 min before induction. Cells were induced by the addition of IPTG (500 mM stock concentration in H_2O, 500 µM final concentration, unless stated otherwise) and incubated at 18 °C, 160 rpm for a further 16–18 h. The cells were harvested by centrifugation (4,255 ×g, 4 °C, 20 min) and stored at −80 °C.

7.7.3 Recombinant protein expression using non-inducing media

Cells from an appropriate cell line were transformed fresh or a glycerol stock was used. A starter culture was prepared by inoculating 100 mL of LB, 2×YT, or TB media (supplemented with appropriate antibiotics) with 1–5 single colonies or glycerol stock scrapings, and incubating at 37 °C, 180 rpm for 16 h. The starter culture was expanded into 2 L flasks (preferably baffled) containing 600/1,000 mL of LB, 2×YT, or TB media (supplemented with appropriate antibiotics) by inoculating with the saturated starter culture (1% v/v). The cultures were incubated at 37 °C, 180 rpm until cell density reached OD_{600} = 0.6–2.0, at which point the cultures were moved to 4 °C for 30 min before induction. Protein expression was induced by the addition of IPTG solution in water (500 mM stock concentration, 500 µM final concentration). The cells were incubated at 18 °C, 160 rpm for a further 16–18 h. The cells were harvested by centrifugation (7,741 g, 4 °C, 20 min) and stored at −80 °C.

7.7.4 Recombinant protein expression for expression trials using autoinduction media

Cells from an appropriate cell line were transformed fresh or a glycerol stock was used. A starter culture was prepared by inoculating 10 mL of MDG or MDAG-135 media (supplemented with appropriate antibiotics) with 1–5 single colonies or glycerol stock scrapings, and incubating at 37 °C, 200 rpm for 16 h. The starter culture was expanded into

100 mL of pre-warmed (37 °C) autoinduction media (supplemented with appropriate antibiotics; 2× kanamycin was required to prevent untransformed growth in autoinduction media) by inoculating with the saturated started culture (0.1% v/v). The cultures were incubated at 37 °C, 200 rpm for 5 h. Afterwards, the cultures were further shaken at 20 °C, 200 rpm for 24 h or at 16 °C, 200 rpm for 48 h. The cells were harvested by centrifugation (4,255 ×g, 4 °C, 20 min) and stored at −80 °C.

7.7.5 Recombinant protein expression using autoinduction media

Cells from an appropriate cell line were transformed fresh or a glycerol stock was used. A starter culture was prepared by inoculating 100 mL of MDG or MDAG-135 media (supplemented with appropriate antibiotics) with 1–5 single colonies or glycerol stock scrapings, and incubating at 37 °C, 200 rpm for 16 h. The starter culture was expanded into baffled 2 L flasks containing 600 mL of autoinduction media (supplemented with appropriate antibiotics; 2× kanamycin was required to prevent untransformed growth in autoinduction media) by inoculating with the saturated starter culture (0.1% v/v). The cultures were incubated at 37 °C, 200 rpm for 3–4 hours or until the cell density reached OD_{600} = 0.6. Afterwards, the cultures were further shaken at 20 °C, 200 rpm for 24 h. The cells were harvested by centrifugation (7,741 ×g, 4 °C, 20 min) and stored at −80 °C.

7.8 Recombinant protein purification

7.8.1 Materials and general information

Benzonase® Nuclease (Cat# 70746), DNase I (grade II, from bovine pancreas), and Calbiochem® protease inhibitor cocktail set III (EDTA-Free) were purchased from Merck. cOmplete protease inhibitor cocktail (EDTA-free) was purchased from Sigma-Aldrich. TCEP hydrochloride was purchased from Fluorochem. HisTrap HP Ni High Performance IMAC resin charged with Ni^{2+}, Mono Q™/Mono S™ ion exchange resins, Heparin HP affinity resin, and Superdex 75/200 resins were purchased as either loose resins or pre-packed columns from

Cytiva (previously GE Healthcare Life Sciences). 10 kDa and 30 kDa MWCO centrifugal filters were purchased from Amicon. PD-10 desalting columns packed with Sephadex G-25 resin were purchased from Cytiva. Filtropur S PES membrane syringe filters (0.20/0.45 µm) were purchased from Sarstedt. SnakeSkin™ Dialysis Tubing, 7 kDa MWCO, 22 mm was purchased from Fisher Scientific. TEV protease was kindly gifted by Dr Patrick Rabe. SUMO protease was kindly gifted by Eidarus Salah.

All protein purifications were carried out on ÄKTA Pure FPLC systems. Where practically feasible, all protein manipulations were carried out at 4 °C. All buffers were filtered through 0.20 µm filters prior to use. Buffer used for IEX or SEC were thoroughly degassed prior to use. All buffers contain 500 µM TCEP unless otherwise stated. Buffers used for column maintenance generally do not contain TCEP.

7.8.2 Cell lysate preparation

Sonication method: Cell sonication was carried out using a Sonics Vibra-Cell™ VC 505 Ultrasonic Liquid Processor. Cell pellet was re-suspended at 4 °C in binding buffer (4× v/w) supplemented with Benzonase® (0.1% v/v, 25.7 U/µL), Calbiochem® protease inhibitor cocktail set III (EDTA-free, 0.1% v/v, 1,000×), and cOmplete protease inhibitor cocktail (EDTA-free, 5 tablets/L buffer). Cells were lysed by sonication (60% amplitude; 4 sec on, 6 sec off; 4–6 min sonication time depending on sample volume). The insoluble cell debris was pelleted (63,988 ×g, 4 °C, 25–60 min), and the cell lysate supernatant was further filtered by passing through a 0.45 µm filter.

High pressure cell disruption method: High pressure cell disruption was carried out using a CF21 model from Constant Cell Disruption Systems equipped with a Julaba F1000 Compact Recirculating Cooler. Cell pellet was re-suspended at 4 °C in binding buffer (5× v/w) supplemented with $MgCl_2$ (2 mM), $CaCl_2$ (120 µM), Benzonase® (0.1% v/v, 25.7 U/µL), DNase I

(10 mg/L), Calbiochem® protease inhibitor cocktail set III (EDTA-Free, 0.1% v/v, 1,000×), and cOmplete protease inhibitor cocktail (EDTA-free, 5 tablets/L buffer). The cell disruptor was washed with H_2O (2×100 mL), Decon 90 (5% v/v, 1×100 mL), H_2O (2×100 mL), and HisTrap binding buffer (2×100 mL) prior to cell lysis *via* single pass disruption at 20,000 psi. An additional portion of Benzonase® (0.05% v/v, 25.7 U/µL) and DNase I (0.1% v/v, 10 mg/mL) was added, and the lysate was incubated at 4 °C for 30 min. The insoluble cell debris was pelleted (63,988 ×g, 4 °C, 40–60 min), and the cell lysate supernatant was further filtered by passing through a 0.45 µm filter. The cell disruptor was cleaned with H_2O (2×100 mL) and EtOH (2×100 mL).

7.8.3 HisTrap immobilised metal affinity chromatography (IMAC)

Buffers: Imidazole concentration in binding and wash buffers was reduced 2-fold (indicated in brackets in Table 7.15) in cases where substantial amounts of unbound protein of interest were observed in the HisTrap binding flow-through. Buffering agents were selected based on their buffering range: MES for pH 6.0; HEPES for pH 7.5; Tris for pH 8.0 and 8.5 (measured at 4 °C).

Table 7.15 | Compositions of HisTrap IMAC buffers.

Buffer type	Buffering agent (mM)	NaCl (mM)	Imidazole (mM)	Glycerol (% v/v)
Binding	20	500	20 (10)	5
Wash	20	500	40 (20)	5
Elution	20	500	500	5

Maintenance: Buffers used for HisTrap maintenance did not contain TCEP. 1 mL or 5 mL HisTrap HP columns containing Ni Sepharose® High Performance IMAC resin were stripped and re-charged with Ni^{2+} prior to every purification. Ni^{2+} was stripped by washing with strip buffer (20 mM Na_2HPO_4, 500 mM NaCl, 50 mM EDTA, pH 7.4; 5 CV, 1 CV/min). When cleaning-in-place was required, the column was washed with H_2O (10 CV, 1 CV/min), NaCl (1.5 M, 10 CV, 0.6 CV/min), H_2O (10 CV, 1 CV/min), NaOH (1 M, 2 h contact time, 0.5 mL/min), and H_2O

(15 CV, 1 CV/min). The column was re-charged by equilibrating with $NiSO_4$ (100 mM, 3 CV, 0.4 CV/min). The column was further washed with H_2O (10 CV, 1 CV/min), imidazole to clear residual unbound Ni^{2+} (500 mM, 10 CV, 1 CV/min), and HisTrap binding buffer to re-equilibrate the column (10 CV, 1 CV/min). The columns were stored in 20% EtOH for long-term storage.

Purification: The filtered cell lysate was loaded onto 1×1 mL, 1×5 mL, or 2×5 mL HisTrap HP columns connected in series (0.7–2 mL/min), washed with HisTrap binding buffer (3 CV, 0.2–1 CV/min) followed by HisTrap wash buffer (20 CV, 0.2–1 CV/min), and eluted with HisTrap elution buffer (100%, 5 CV, 0.2–1 CV/min). Protein fractions were analysed by SDS-PAGE and processed depending on the subsequent step.

7.8.4 Protease-assisted tag cleavage and reverse HisTrap IMAC

TEV protease or SUMO protease was used to cleave the combined solubility-affinity tags after the initial HisTrap purification. The protein samples were buffer exchanged into HisTrap binding buffer prior to tag cleavage.

Small scale preparations: For small scale preparations (up to 2 L of cell culture growth), the protein was concentrated to under 2.5 mL (or above 10 mg/mL) using 30 kDa MWCO centrifugal filters (3,082 ×g, 4 °C, 5 min runs) and buffer exchanged into HisTrap binding buffer using PD-10 desalting columns as per the manufacturer's instructions. TEV protease (1:50 w/w) or SUMO protease (1:100 w/w) was added, and the mixture was incubated on a roller mixer at 4 °C for 16 h.

Large scale preparations: For large scale preparations (more than 12 L of cell culture growth), the protein was transferred into dialysis tubing (7 kDa MWCO) and submerged in HisTrap binding buffer (3.5 L) with minimal agitation at 4 °C for 5 h. SUMO protease (1:350 w/w) was added to the dialysis tubing. The tubing was submerged in fresh HisTrap binding buffer (3.5 L) with minimal agitation at 4 °C for a further 16 h.

After tag cleavage, a small portion of the cleavage mixture was set aside for SDS-PAGE and MS analysis. The remaining mixture was centrifuged (4,255 xg, 4 °C, 20 min) to remove any precipitated protein. The clarified cleavage mixture was loaded onto 1×1 mL, 1×5 mL, or 2×5 mL HisTrap HP columns connected in series (0.2–0.5 CV/min) and the flow-through was collected. The columns were further washed with HisTrap binding buffer (2 CV, 0.2–0.5 CV/min), followed by HisTrap wash buffer (3 CV, 0.5–2 mL/min) and HisTrap elution buffer (5 CV, 0.4–1 CV/min). Protein fractions were analysed by SDS-PAGE and processed depending on the subsequent step.

7.8.5 Ion-exchange chromatography (IEX)

Buffers: All buffers used for ion-exchange chromatography (IEX) were thoroughly degassed prior to use. Different buffering agents were used for cation and anion exchange chromatography: MES at pH 6.0 (at 4 °C) was used for cation exchange; CHES at pH 9.0 (at 4 °C) was used for anion exchange.

Table 7.16 | Compositions of ion-exchange chromatography buffers.

Buffer type	Buffering agent (mM)	NaCl (mM)	Glycerol (% v/v)
Binding	50	50	5
Elution	50	1,000	5

Maintenance: A 1 mL Mono S™ cation exchange or a 1 mL Mono Q™ anion exchange column was washed with H_2O (5 CV, 0.25 CV/min) and cleaned with NaCl (2 M, 2 CV, 0.25 CV/min), NaOH (1 M, 4 CV, 0.25 CV/min), and NaCl (2 M, 10 CV, 0.25 CV/min). The column was re-equilibrated by washing with IEX binding buffer (10 CV, 0.25 CV/min). The columns were stored in 20% EtOH in H_2O for long-term storage.

Purification: Protein sample obtained after HisTrap IMAC, tag cleavage, and reverse HisTrap was buffer exchanged into IEX binding buffer using PD-10 desalting columns as described in section 7.8.4. The buffer exchanged sample was centrifuged (10,000 xg, 4 °C, 10 min) to

remove any precipitated protein. The sample was loaded onto an IEX column (0.15–0.5 CV/min), and the column was washed with IEX binding buffer (10 CV, 0.2–0.5 CV/min). The protein sample was eluted using gradient elution (0–50% IEX elution buffer in IEX binding buffer, 20 CV, 0.2–0.5 CV/min), followed by flushing with IEX elution buffer (10 CV, 0.2–0.5 CV/min). Protein fractions were analysed by SDS-PAGE and processed depending on the subsequent step.

7.8.6 Heparin affinity chromatography

Buffers: Buffering agents were selected based on their buffering range: HEPES for pH 7.5; Tris for pH 8.0 and 8.5 (measured at 4 °C).

Table 7.17 | Compositions of heparin affinity chromatography buffers.

Buffer type	Buffering agent (mM)	NaCl (mM)
Binding	50	50
Elution	50	2,000

Maintenance: 5 mL HiTrap™ Heparin HP or a 25 mL Heparin HP column (packed in-house by Dr Adam Hardy) was washed with H_2O (5 CV, 0.12–0.2 CV/min), NaCl (2 M, 2 CV, 0.12–0.2 CV/min), H_2O (5 CV, 0.12–0.2 CV/min), NaOH (0.1 M, 4 CV, 0.12–0.2 CV/min), and H_2O (5 CV, 0.12–0.2 CV/min). The column was re-equilibrated by washing with heparin binding buffer (15 CV, 0.12–0.2 CV/min). The columns were stored in 20% EtOH for long-term storage.

Purification: Protein sample obtained after HisTrap IMAC, tag cleavage, and reverse HisTrap was buffer exchanged into heparin binding buffer as described in section 7.8.4. The buffer exchanged sample was centrifuged (4,255 xg, 4 °C, 20 min) to remove any precipitated protein. The sample was loaded onto the column (0.016–0.04 CV/min) and the column was washed with heparin binding buffer (8 CV, 0.12–0.2 CV/min). The protein sample was eluted using gradient elution (0–50% heparin elution buffer in heparin binding buffer, 8 CV, 0.12–0.2 CV/min), followed by flushing with heparin elution buffer (8 CV, 0.12–0.2 CV/min).

Protein fractions were analysed by SDS-PAGE and processed depending on the subsequent step.

7.8.7 Size exclusion chromatography (SEC)

Buffers: Buffering agents were selected based on their buffering range: HEPES for pH 7.5; Tris for pH 8.0 and 8.5 (measured at 4 °C). High NaCl concentration was used to improve the stability of highly concentrated protein samples.

Table 7.18 | Compositions of size exclusion chromatography buffers.

Gel filtration buffer	Buffering agent (mM)	NaCl (mM)
Low NaCl	20	150
High NaCl	20	300

Maintenance: HiLoad® Superdex® 75 pg or 200 pg size exclusion column (16/600 or 26/600) was washed with H_2O (1.1 CV, 1 mL/min), NaOH (500 mM, 1.1 CV, 0.5 mL/min), and H_2O (1.1 CV, 1 mL/min), and equilibrated with appropriate gel filtration buffer (2.1 CV, 1 mL/min) prior to use. The columns were stored in 20% EtOH for long-term storage.

Purification: Protein sample obtained after affinity and/or ion exchange chromatography was concentrated to a volume below 2% of total column volume using 30 kDa MWCO centrifugal filters (3,082 xg, 4 °C, 5 min runs). The concentrate was loaded onto the appropriate size exclusion column *via* a 2 mL capillary loop or a 10 mL Superloop™ (0.5 mL/min) and eluted with gel filtration buffer (1.1 CV, 1–2 mL/min). Protein fractions were analysed by SDS-PAGE and processed depending on purity.

7.8.8 Concentration and storage

Pure protein fractions (as assessed by SDS-PAGE) were pooled and concentrated using 30 kDa MWCO centrifugal filters (3,082 xg, 4 °C, 5 min runs) to the desired concentration. Protein concentration was evaluated using a NanoDrop™ One Microvolume UV-Vis Spectrophotometer. In some cases, protein loss was observed before the desired

concentration was reached, necessitating storage at a lower concentration. Protein was aliquoted, snap frozen in liquid N_2, and stored at –80 °C.

NgTet1 was stored in low NaCl gel filtration buffer at a final concentration of 15.9–18.6 mg/mL (411–481 µM).

Tagged hTET2 (pSUMO) was stored in low NaCl gel filtration buffer at a final concentration of 9.8–10.0 mg/mL (147–150 µM).

Cleaved hTET2 (pSUMO) was stored in high NaCl gel filtration buffer at a final concentration of 19.7 mg/mL (384 µM).

7.9 Ni Sepharose resin IMAC for expression trials

7.9.1 Materials and buffers

Buffers used in this section are identical to those described in section 7.8.3. Materials used in this section are identical to those described in section 7.8.1. Lysozyme from chicken egg white and SIGMAFAST™ protease inhibitor cocktail tablets (EDTA-free) were purchased from Sigma-Aldrich. BugBuster® Protein Extraction Reagent (Cat# 70923) was purchased from Merck Millipore.

All buffers were filtered through 0.20 µm filters prior to use. All buffers contain 500 µM TCEP unless otherwise stated.

7.9.2 Cell lysate preparation

Cell pellet preparation: Cell pellets were prepared from 100 mL cell cultures as described in sections 7.7.2 and 7.7.4.

Sonication method: Cell sonication was carried out using Sonics Vibra-Cell™ VCX 130 Ultrasonic Liquid Processor. Cell sonication was carried out as described in section 7.8.2 with slight modifications. Cell pellets from 100 mL cell cultures were re-suspended in identical

volumes (10 mL) of binding buffer supplemented with SIGMAFAST™ protease inhibitor cocktail (EDTA-free, 10 tablets/1 L buffer) and DNase I (0.1% v/v, 10 mg/mL). Cells were lysed by sonication (60% amplitude; 5 sec on, 5 sec off; 3.5 min sonication time). Additional DNase I (approx. 300 µg per sample) was added to each sample after sonication, and the lysates were incubated at 4 °C for 30 min. The insoluble cell debris was pelleted (4,255 xg, 4 °C, 2 h), and the clarified cell lysates were used without further filtration.

BugBuster® method: BugBuster® Protein Extraction Reagent was used in cases where the cell pellets were obtained from 10 mL cell cultures. The thawed cells were re-suspended in 1.5 mL BugBuster® buffer supplemented with SIGMAFAST™ protease inhibitor cocktail (10 tablets/1 L buffer), DNase I (0.1% v/v, 10 mg/mL), and lysozyme (0.1% v/v, 1 mg/mL), and incubated on a tube rotator at 4 °C, 25 rpm for 30–90 min. The samples were then centrifuged (13,000 xg, 30 min, 4 °C), and the cell lysate supernatants were used for downstream applications.

7.9.3 Loose resin IMAC purification

Resin equilibration: HisTrap resin stored in EtOH (20% v/v) was re-suspended and transferred to a 50 mL Falcon tube (250 µL of resin per sample). The resin was pelleted (1,000 xg, 4 °C, 10 min) and the supernatant was removed. The resin was washed with binding buffer (4 times, 5 mL each), pelleting by centrifugation (500 xg, 4 °C, 10 min) following each wash step. The washed resin was re-suspended in an equal amount of binding buffer (250 µL of binding buffer per 250 µL of resin).

Sample purification: Equilibrated HisTrap resin (500 µL) was mixed with clarified lysate (10 mL for sonication lysis, 1.5 mL for BugBuster® lysis) and incubated on a rolling mixer at 4 °C for 2 h. The mixture was centrifuged (500 xg, 4 °C, 10 min), and the lysate supernatant was carefully removed. The resin was washed (2×5 mL) with HisTrap binding buffer supplemented

with SIGMAFAST™ protease inhibitor cocktail (10 tablets/1 L buffer), followed by washing with HisTrap wash buffer (2×5 mL). The resin was pelleted by centrifugation (500 xg, 4 °C, 10 min) following each wash step. The resin was re-suspended in a minimum amount of HisTrap wash buffer and transferred to 1.5 mL Eppendorf tubes. The resin was pelleted by centrifugation (500 xg, 4 °C, 10 min), the supernatant was removed, and HisTrap elution buffer was added (200 µL). The sample was incubated on a tube rotator at 4 °C, 30 rpm for 30 min. The resin was pelleted by centrifugation (1,000 xg, 10 min, 4 °C), and the supernatant was used for SDS-PAGE and Western blot analysis.

7.10 Western Blot

7.10.1 Materials and buffers

Materials: Amersham™ Protran® 0.2 µm nitrocellulose membrane (Cat# 10600001) and Amersham™ ECL™ Prime Western Blotting Detection Reagents (Cat# RPN2232) were purchased from Cytiva. Trizma® base and Tween® 20 were purchased from Merck. Tesco Instant Dried Skimmed Milk was purchased from Tesco. Ponceau S staining solution was purchased from VWR. Grade 3MM Chr Cellulose Chromatography Paper was purchased from Cytiva. Anti-His_6 (hexahistidine) mouse monoclonal antibody was purchased from ProteinTech (RRID: AB_11232599). Anti-mouse IgG, HRP-linked horse antibody was purchased from Cell Signaling Technology (RRID: AB_330924).

Transfer buffer: Tris base (25 mM), glycine (200 mM), MeOH (20% v/v).

PBS-T buffer: Na_2HPO_4 (10 mM), KH_2PO_4 (1.8 mM), NaCl (137 mM), KCl (2.7 mM), Tween® 20 (0.1% v/v).

Blocking solution: Dried skimmed milk (5% w/v) dissolved in PBS-T.

7.10.2 Western blot method

Membrane preparation: Protein samples were separated on an SDS-PAGE gel as described in section 7.6.5, and transferred under wet conditions onto a nitrocellulose membrane in ice-cold transfer buffer at 100 V for 60 min. The transfer sandwich for wet transfer consisted of a sponge, three layers of pre-wet Whatman paper, polyacrylamide gel, nitrocellulose membrane, three layers of pre-wet Whatman paper, and a sponge. Protein transfer was confirmed by staining with Ponceau S stain at RT for 5 min; the membrane was then de-stained in PBS-T.

Membrane development: All incubation and washing steps were carried out on a 3D orbital shaker unless otherwise noted. The membrane was blocked with blocking solution at RT for 1 h. The blocking solution was replaced with fresh blocking solution containing anti-His_6 primary antibody (1:1,000 dilution), and the membrane was incubated at 4 °C for 16 h. The primary antibody solution was decanted, and the membrane was washed with PBS-T (3×10 min). Fresh blocking solution containing anti-mouse secondary antibody (1:10,000 dilution) was added, and the membrane was incubated at RT for 1 h. The membrane was washed with PBS-T (3×10 min) and developed using ECL reagents as per the manufacturer's instructions. The membrane was incubated in the dark for 5 min before imaging using a BIO-RAD ChemiDoc MP Universal Hood III ("Blots – Chemi" method). "Protein gels – Stain Free Gel" method was used to image the protein ladder, and the two images were superimposed using ImageLab (5.2.1) software.

7.11 SPOT peptide array

7.11.1 Materials and buffers

Materials: Whatman® grade 1 Chr cellulose chromatography paper was purchased from Merck. Fmoc-protected amino acids, biotin, DIC, Oxyma Pure, and *N*–methylimidazole (NMI) were purchased from Merck, Novabiochem®, or CEM. Peptide synthesis grade *N*–methyl-2-

pyrrolidone (NMP), DMF, and piperidine (20% v/v in DMF) were purchased from Cambridge Bioscience or Merck. Anti-FLAG mouse monoclonal antibody was purchased from Merck (RRID: AB_262044). Anti-mouse IgG, HRP-linked horse antibody was purchased from Cell Signaling Technology (RRID: AB_330924). TiP1-FLAG was synthesised in-house as described in sections 3.3.6 and 7.2.

PBS buffer: Na_2HPO_4 (10 mM), KH_2PO_4 (1.8 mM), NaCl (137 mM), KCl (2.7 mM).

PBS-T buffer: Na_2HPO_4 (10 mM), KH_2PO_4 (1.8 mM), NaCl (137 mM), KCl (2.7 mM), Tween® 20 (0.1% v/v).

Blocking solution: 5% w/v of dried skimmed milk dissolved in PBS-T.

7.11.2 SPOT peptide array synthesis

SPOT peptide arrays were synthesised as described previously[242] on a MultiPep (CEM) automated peptide synthesiser. Briefly, a Whatman® grade 1 Chr cellulose chromatography paper membrane was incubated in a DMF solution containing Fmoc-β-alanine-OH (0.64 mg/mL), DIC (3.74% v/v), and NMI (3.17% v/v) for 16 h. The membrane was washed with DMF (3×30 sec), Fmoc deprotected by incubating in piperidine (20% v/v in DMF, 2×5 min), and washed again with DMF (4×30 sec).

The functionalised membrane was placed on the MultiPep automated peptide synthesiser and synthesis was carried out as per the manufacturer's instructions (coupling, deprotection, washing, and drying step sequences were provided by the ResPep SLi Spotter software). Fmoc-protected amino acids (500 mM solutions in DMF; Phe, His, and Pro were dissolved in NMP at 500 mM; biotin was dissolved in NMP at 300 mM) were pre-activated by mixing with DIC, Oxyma Pure, and NMP, and incubating at RT for 15 min. The pre-activated mix was dispensed to the programmed positions (100 nL) and allowed to react at RT (2×5 min).

An initial coupling of β-alanine was used to define the peptide array pattern on the membrane. The membrane was covered with capping solution containing acetic anhydride (2% v/v) and DIPEA (2% v/v) in DMF and left to react at RT for 15 min. The capped membrane was vacuum-dried, washed thoroughly with DMF, washed with EtOH, and vacuum-dried again. The membrane was covered in piperidine (20% v/v in DMF) and allowed to react at RT for 15 min. The membrane was vacuum-dried, washed thoroughly with DMF, washed with EtOH, and vacuum-dried again. All subsequent coupling steps followed the same procedure.

Following the final Fmoc deprotection, the membrane was covered with capping solution containing acetic anhydride (2% v/v) and DIPEA (2% v/v) in DMF and left to react at RT for 15 min. The capped membrane was vacuum-dried, washed thoroughly with DMF, washed with EtOH, and vacuum-dried again.

The dry membrane was treated with a 92.5:2.5:2.5:2.5 deprotection mixture of TFA/H_2O/TIPS/1,3-dimethoxybenzene and incubated at RT for 2 h. The membrane was washed with CH_2Cl_2 (5×) and EtOH (5×), and finally air-dried. The deprotected membrane was stored at −80 °C until probing.

7.11.3 SPOT peptide array probing

Membrane development: All incubation and washing steps were carried out on a 3D orbital shaker unless otherwise noted. The membrane was thawed and imaged on a BIO-RAD ChemiDoc MP Universal Hood III (“Blots – Fluorescein” and “Nucleic Acid Gels – EtBr” methods). The membrane was pre-wet with PBS-T (3×10 min), blocked with blocking solution at RT for 2 h, and washed with PBS-T (3×10 min). Fresh blocking solution containing TiP1-FLAG (5 μM, 0.05% v/v DMSO) was added, and the membrane was incubated at RT for 5 h. The TiP1-FLAG solution was recovered, and the membrane was washed with PBS-T (3×10 min). Fresh blocking solution containing anti-FLAG primary antibody (1:1,250 dilution) was added,

and the membrane was incubated at 4 °C for 16 h. The primary antibody solution was decanted, and the membrane was washed with PBS-T (3×10 min). Fresh blocking solution containing anti-mouse secondary antibody (1:5,000 dilution) was added, and the membrane was incubated at RT for 1.5 h. The membrane was washed with PBS-T (4×10 min), dried using a paper towel, and developed using ECL reagents as per the manufacturer's instructions. The membrane was incubated in the dark for 5 min before imaging using a BIO-RAD ChemiDoc MP Universal Hood III ("Blots – Chemi" method).

Membrane regeneration: Membrane regeneration was carried out at RT. The membrane was rinsed with PBS-T immediately after probing (4×), washed with H_2O (2×10 min), and incubated in regeneration solution A (8 M urea, 1% w/v SDS in PBS) for 16 h. The membrane was washed with H_2O (2×10 min) and incubated in regeneration solution B (6:3:1 mixture of TFA/EtOH/H_2O) for 2h. The membrane was then washed with H_2O (2×10 min), PBS (3×10 min), and MeOH (2×1 min), air-dried, and stored at –80 °C.

7.12 2-Oxoglutarate (2-OG) decarboxylation monitoring by ^{1}H NMR

7.12.1 General information

Materials: (+)-Sodium L-ascorbate, $(NH_4)_2Fe(SO_4)_2$, and disodium 2-OG were purchased from Sigma-Aldrich. 3 mm NMR tubes (Cat# 2001724) were purchased from Hilgenberg. Tris-d_{11} and 2,4-pyridinedicarboxylic acid (2,4-PDCA) were purchased from Sigma-Aldrich. *N*–Oxalylglycine (NOG) and 8-hydroxy-5-quinolinecarboxylic acid (IOX1) were purchased from Cayman Chemicals. Micro Bio-Spin™ 6 columns were purchased from BIO-RAD. DNA oligomers used in NMR assays were purchased from ATDBio or Life Technologies (Thermo Fisher).

Spectra recording: All spectra were recorded at 25 °C using D_2O as a reference for internal deuterium lock on a Bruker Avance III 700 MHz equipped with an inverse TCI cryoprobe in

3 mm diameter Bruker MATCH NMR tubes with a total sample volume of 160 µL. Water suppression was achieved using Perfect Echo excitation sculpting suppression.

Buffers: Compositions of buffers used in NMR experiments are reported in Table 7.19. Tris-d_{11} (200 mM, pH 7.5, measured at RT) stock was stored at 4 °C.

Table 7.19 | Compositions of buffers used in ^{1}H NMR 2-OG turnover assays.

Buffer	Composition
Deuterated storage buffer	50 mM Tris-d_{11}, 150 mM NaCl, pH 7.5
Assay buffer	50 mM Tris-d_{11}, pH 7.5
Quenching solution	9.08 mM NOG, 9.18 mM $ZnCl_2$, 91 mM Tris-d_{11}, pH 7.5

Cofactor preparation: (+)-Sodium L-ascorbate and disodium 2-OG stock solutions (200 mM) were prepared fresh in H_2O. $(NH_4)_2Fe(SO_4)_2$ (100 mM) and $ZnCl_2$ (2 M) stock solutions were prepared fresh in 20 mM HCl. NOG stock solutions (10/100 mM) were prepared in H_2O or buffered with Tris-d_{11} (100 mM) and stored at –20 °C.

7.12.2 NgTet1 buffer exchange for ^{1}H NMR

NgTet1 stock solutions in non-deuterated storage buffer were buffer exchanged into deuterated storage buffer using Micro Bio-Spin™ 6 columns as per the manufacturer's instructions. The columns were pre-equilibrated with the deuterated storage buffer (4×500 µL) prior to sample application. The buffer exchanged enzyme stock was aliquoted, snap frozen in liquid N_2, and stored at –80 °C. Protein recovery was typically above 85%.

7.12.3 Continuous enzymatic 2-OG decarboxylation monitoring by ^{1}H NMR

Each 2-OG decarboxylation reaction was prepared according to Table 7.20 with the enzyme added last to initiate the enzymatic reaction (160 µL final volume). The solution was briefly mixed, transferred to a 3 mm NMR tube, and subjected to ^{1}H NMR measurements (32 scans were used for 2-OG concentrations ≤100 µM; 16 scans were used for 2-OG concentrations

>100 µM). A control sample without enzyme was used to lock, tune, match, and shim the sample.

Table 7.20 | General composition of continuous 2-oxoglutarate (2-OG) turnover monitoring reactions. Final concentration refers to concentration in 160 µL volume.

Component	Stock concentration	Final concentration
Tris-d_{11}, pH 7.5	200 mM	50 mM
D_2O	100% v/v	10% v/v
L-Ascorbate	200 mM	2 mM
$(NH_4)_2Fe(SO_4)_2$	10 mM	100 µM
2-OG	1 or 10 mM	50–4,000 µM
NgTet1	varied	5 µM

In cases where oligonucleotides or peptides were used in the reactions, mixtures containing the additional component (Table 7.21), but no 2-OG were prepared. Enzyme was added, and the resulting solution was incubated at RT for 5 min. The reaction was then initiated by the addition of 2-OG.

Table 7.21 | Additional components added to continuous 2-oxoglutarate (2-OG) turnover monitoring reactions. Final concentration refers to concentration in 160 µL volume.

Component	Stock concentration	Final concentration
^{5m}C 12 bp dsDNA*	200 µM	20 µM
^{5}C 12 bp dsDNA†	200 µM	20 µM
TiP1‡	61.7 µM (in 10% v/v DMSO)	6.17 µM

Self-complementary 12 base-pair double-stranded DNA (sequence: 5'-ACC AC^{5m}C** GGT GGT-3').*
†*Self-complementary 12 base-pair double-stranded DNA (sequence: 5'-ACC AC**C** GGT GGT-3').*
‡*Thioether-linked macrocyclic peptide (sequence: ^{D}YLIYAYYTWWEHSC).*

7.12.4 Enzyme inhibition assay

End-point readout was used for NgTet1 enzyme inhibition assays. Inhibitor dilution series were prepared manually. 2.5× enzyme mix (64 µL, Table 7.22) was plated into wells containing inhibitors (16 µL, 10% v/v DMSO) and incubated at RT for 10 min. The reaction was initiated by the addition of 2× substrate mix (80 µL) and quenched by the addition of quenching solution (19.38 µL, Table 7.19) within the enzyme linear activity window (determined separately). In cases where inhibition of coupled turnover was evaluated, **^{5m}C 12-bp dsDNA** (40 µM) was added to the 2× substrate mix.

Table 7.22 | General compositions of end-point 2-OG turnover monitoring reactions. Final concentration refers to concentration in 160 µL volume.

	Component	Stock concentration	Mix concentration	Final concentration
2.5× enzyme mix	Tris-d_{11}, pH 7.5	200 mM	62.5 mM	50 mM
	D_2O	100% v/v	12.5% v/v	10% v/v
	NgTet1	varied	12.5 µM	5 µM
2× substrate mix	Tris-d_{11}, pH 7.5	200 mM	50 mM	50 mM
	D_2O	100% v/v	10% v/v	10%
	L-Ascorbate	200 mM	4 mM	2 mM
	$(NH_4)_2Fe(SO_4)_2$	10 mM	200 µM	100 µM
	2-OG	1 mM	100 µM	50 µM
	^{5m}C 12 bp dsDNA*	200 µM	0/40 µM	0/20 µM

Self-complementary 12 base-pair double-stranded DNA (sequence: 5'-ACC AC^{5m}C** GGT GGT-3').*

The quenched reaction mixtures were transferred to 3 mm NMR tubes (160 µL transfer volume) and subjected to ^{1}H NMR analysis (256 scans). In cases where precipitation was observed upon quenching, the samples were transferred to 1.5 mL Eppendorf tubes and centrifuged (13,000 rpm, Fisher Scientific AccuSpin Micro centrifuge, RT, 3 min) prior to transfer to 3 mm NMR tubes.

All enzyme activity data was normalised against enzyme activity in reactions containing vehicle only (1% v/v DMSO). Data was analysed using GraphPad Prism 5 software. Mean and standard deviation were plotted, and data was fitted using "log(inhibitor) *vs*. response - Variable slope (four parameters)" or "log(inhibitor) *vs*. normalised response - Variable slope" non-linear regression models.

7.13 ^{5m}C oxidation monitoring by Amplified Luminescent Proximity Homogenous Assay (AlphaScreen)

7.13.1 General information

Materials: DTPA-purified bovine serum albumin (BSA), 7.5%, was purchased from PerkinElmer. (+)-Sodium L-ascorbate, $(NH_4)_2Fe(SO_4)_2$, disodium 2-OG, and Tween® 20 were purchased from Sigma-Aldrich. Anti-5-hydroxymethylcytosine (^{5hm}C) polyclonal rabbit

antibody was purchased from Active Motif (RRID: AB_10013602). All AlphaScreen kits and plates were purchased from PerkinElmer. AlphaScreen General IgG (Protein A) Detection Kit (Cat# 6760617) was used for enzyme activity, inhibition, and antibody-dependent compound interference assays. AlphaScreen TruHits Kit (Cat# 6760627) was used for antibody-independent compound interference assays. All manipulations of AlphaScreen beads were carried out under subdued lighting. All AlphaScreen assays were performed in shallow-well 384-well light gray ProxiPlates (Cat# 6008350). DNA oligomers used in AlphaScreen assays were prepared by Dr Roman Belle in collaboration with Prof. Tom Brown (University of Oxford). Active $hTET1_{CD}$ enzyme was purchased from Epigentek (1418–2136 a.a., Cat# E12002-1). Active $hTET1_{CD}$ and $hTET3_{CD}$ enzymes were kindly produced by Eidarus Salah. 8-Hydroxy-5-quinolinecarboxylic acid (IOX1) was purchased from Cayman Chemicals.

Liquid handling: An E1-ClipTip™ Bluetooth™ 16-channel 125 µL Electronic Multichannel Pipette (Thermo Fisher, Cat# 4671060BT) or a Thermo Multidrop™ Combi equipped with a small tube plastic tip dispensing cassette (Thermo Fisher, Cat# 24073290) were used for liquid handling in 384-well plates.

Chemiluminescence measurement: Chemiluminescence was measured on an Envision Multimode Plate Reader (PerkinElmer) or a PHERAstar FS (BMG Labtech) equipped with an AlphaScreen 680 570 module.

Buffers: Compositions of buffers used in AlphaScreen assays are reported in Table 7.23. Assay buffer was prepared fresh for each assay from component stock solutions. HEPES (1 M, pH 7.3) stock was stored at 4 °C in dark conditions. Quenching solution was stored at RT.

Table 7.23 | Compositions of buffers used in AlphaScreen assays.

Buffer	Composition
Assay buffer	50 mM HEPES, 150 mM NaCl, 0.1% v/v BSA, 0.01% v/v Tween® 20, pH 7.1 (final)
Quenching solution	30 mM EDTA, 358 mM NaCl, pH 4.2

Cofactor preparation: (+)-Sodium L-ascorbate (500 mM) and disodium 2-OG (50 mM) stock solutions were prepared fresh in assay buffer. $(NH_4)_2Fe(SO_4)_2$ (50 mM) stock solution was prepared fresh in 20 mM HCl.

7.13.2 Acoustic dispensing of compounds

Prior to compound characterisation experiments, DMSO-soluble compounds were pre-dispensed in a logarithmically linear concentration gradient (9–11 concentration points) onto 384-well ProxiPlates using an Echo® 550 acoustic liquid handler (Labcyte). The final volume of DMSO in all wells was adjusted to 100 nL. The final reaction volume was 10.1 µL (10 µL reaction mixture + 0.1 µL DMSO), resulting in a 0.99% v/v final DMSO concentration.

7.13.3 Enzyme activity assay (General IgG (Protein A) Detection Kit)

Typical enzyme activity assay conditions for TET proteins were adapted from literature[319]. The protocol for enzyme activity assays was optimised by Dr Roman Belle. All other protocols using General IgG (Protein A) Detection Kit are adapted from the enzyme activity assay protocol.

AlphaScreen bead mixture preparation: Streptavidin-coated donor and protein A-coated acceptor beads (5 mg/mL stock concentration) were combined and diluted in assay buffer (1:62.5, 80 µg/mL), and pre-incubated in the dark with anti-^{5hm}C antibody (1:2,000) for at least 30 min.

Assay protocol: 2× enzyme mix (5 µL) was mixed with 2× substrate mix (5 µL) and incubated at RT (Table 7.24) until the reaction was quenched by the addition of quenching solution (5 µL, Table 7.23). Pre-incubated AlphaScreen bead mixture (5 µL) was added to the quenched assay, and the resulting mixture was incubated in the dark for 45–60 min before luminescence was measured. In a 20 µL final volume, the AlphaScreen bead concentration was 20 µg/mL, and the anti-^{5hm}C antibody dilution factor was 1:8,000.

Table 7.24 | General compositions of AlphaScreen end-point assay solutions in assay buffer used for enzyme activity assays. Final concentration refers to concentration in 10 µL volume.

	Component	Stock concentration	2× mix concentration	Final concentration
2× enzyme mix	TET enzyme	varied	2–800 nM	1–400 nM
2× substrate mix	L-Ascorbate	2 mM	200 µM	100 µM
	$(NH_4)_2Fe(SO_4)_2$	200 µM	20 µM	10 µM
	2-OG	200 µM	20 µM	10 µM
	^{5m}C 32 nt bt-ssDNA*	1 µM	20 nM	10 nM

32 nucleotide biotinylated single-stranded DNA (sequence: 5'-[Biotin]-TCG GAT GTT GTG GGT CAG ^{5m}C**GC ATG ATA GTG TA-3').*

7.13.4 Antibody content optimisation (General IgG (Protein A) Detection Kit)

Assay buffer (5 µL) was mixed with modified 2× substrate mix (5 µL, Table 7.25). Quenching solution (5 µL) and pre-incubated AlphaScreen bead mixture (5 µL) were added. The resulting mixture was incubated in the dark for 45–60 min before luminescence was measured. The final antibody dilution factor was varied from 1:4,000 to 1:24,000 to identify conditions producing the best signal-to-noise (S/N) ratio.

Table 7.25 | Composition of modified 2× substrate mix in assay buffer used for AlphaScreen antibody content optimisation. Final concentration refers to concentration in 10 µL volume.

	Component	Stock concentration	2× mix concentration	Final concentration
Modified 2× substrate mix	L-Ascorbate	2 mM	200 µM	100 µM
	$(NH_4)_2Fe(SO_4)_2$	200 µM	20 µM	10 µM
	2-OG	200 µM	20 µM	10 µM
	^{5m}C 32 nt bt-ssDNA*	1 µM	17 nM	8.5 nM
	^{5hm}C 32 nt bt-ssDNA†	1 µM	3 nM	1.5 nM

32 nucleotide biotinylated single-stranded DNA (sequence: 5'-[Biotin]-TCG GAT GTT GTG GGT CAG ^{5m}C**GC ATG ATA GTG TA-3'). †32 nucleotide biotinylated single-stranded DNA (sequence: 5'-[Biotin]-TCG GAT GTT GTG GGT CAG ^{5hm}**C**GC ATG ATA GTG TA-3').*

7.13.5 Standard curve generation (General IgG (Protein A) Detection Kit)

Assay buffer (5 µL) was mixed with 2× substrate mix (5 µL, as in Table 7.24), quenching solution (5 µL), and pre-incubated AlphaScreen bead mixture (5 µL). The resulting mixture was incubated in the dark for 45–60 min before luminescence was measured.

3 dilution series of 2× substrate mix solutions containing different DNA substrates were prepared:

(i) 0–20 nM of **^{5m}C 32 nt bt-ssDNA** (5'-[Biotin]-TCG GAT GTT GTG GGT CAG **^{5m}C**GC ATG ATA GTG TA-3');

(ii) 0–20 nM of **^{5hm}C 32 nt bt-ssDNA** (5'-[Biotin]-TCG GAT GTT GTG GGT CAG **^{5hm}C**GC ATG ATA GTG TA-3');

(iii) 0–100% of **^{5hm}C 32 nt bt-ssDNA** in **^{5m}C 32 nt bt-ssDNA** at a 20 nM total DNA concentration.

7.13.6 Enzyme inhibition assay (General IgG (Protein A) Detection Kit)

2× enzyme mix (5 µL) was plated into wells containing inhibitors (100 nL, DMSO) and incubated at RT for 10 min (Table 7.24). The reaction was initiated by the addition of 2× substrate mix (5 µL) and quenched by the addition of quenching solution (5 µL) within the enzyme linear activity window (determined separately for each enzyme). Pre-incubated AlphaScreen bead mixture (5 µL) was added, and the resulting mixture was incubated in the dark for 45–60 min before luminescence was measured.

Pre-quenched reactions containing vehicle (0.99% v/v DMSO) were used for background determination. IOX1 was used as a positive inhibition control. All enzyme activity data was normalised against enzyme activity in reactions containing vehicle only (0.99% v/v DMSO). Data was analysed using GraphPad Prism 5 software. Mean and standard deviation were plotted, and data was fitted using "log(inhibitor) *vs*. response - Variable slope (four parameters)" or "log(inhibitor) *vs*. normalised response - Variable slope" non-linear regression models.

7.13.7 Antibody-independent compound interference assay (TruHits Kit)

No antibody is required when using the AlphaScreen TruHits Kit. Donor and acceptor beads were diluted in assay buffer (1:62.5) and used without pre-incubation.

Assay buffer (10 µL) was plated into wells containing compounds (100 nL, DMSO). Quenching solution (5 µL) and the AlphaScreen bead mixture (5 µL) were added, and the resulting mixture was incubated in the dark for 45–60 min before luminescence was measured.

7.13.8 Antibody-dependent compound interference assay (General IgG (Protein A) Detection Kit)

Assay buffer (5 µL) was plated into wells containing compounds (100 nL, DMSO). Modified 2× substrate mix (5 µL, Table 7.25), quenching solution (5 µL), and the pre-incubated AlphaScreen bead mixture (5 µL) were added. The resulting mixture was incubated in the dark for 45–60 min before luminescence was measured.

7.14 Succinate-Glo™ bioluminescent succinate detection assay

7.14.1 General information

Materials: Succinate-Glo™ assay kit (Cat# V7991) was purchased from Promega. (+)-Sodium L-ascorbate, $(NH_4)_2Fe(SO_4)_2$, and disodium 2-OG were purchased from Sigma-Aldrich. Disodium succinate hexahydrate was purchased from BDH Chemicals. DNA oligomers used for Succinate-Glo™ assays were purchased from ATDBio or Sigma-Aldrich. All Succinate-Glo™ assays were performed in 384-well light gray AlphaPlates (Cat# 6005350, PerkinElmer).

Luminescence measurement: Luminescence was measured on a CLARIOstar Plus multi-mode microplate reader (BMG Labtech).

Buffers: Composition of assay buffer used in Succinate-Glo™ assays is reported in Table 7.26. Assay buffer was prepared fresh for each assay from component stock solutions. HEPES (1 M, pH 7.3) stock was stored at 4 °C in dark conditions.

Table 7.26 | Composition of Succinate-Glo™ assay buffer.

Buffer	Composition
Assay buffer	50 mM HEPES, 100 mM NaCl, pH 7.1

Cofactor preparation: (+)-Sodium L-ascorbate (500 mM), disodium 2-OG (50 mM) stock solutions were prepared fresh in assay buffer. $(NH_4)_2Fe(SO_4)_2$ (50 mM) stock solution was prepared fresh in 20 mM HCl. Disodium succinate stock (500 mM) was prepared fresh in H_2O.

7.14.2 Standard curve generation

Succinate Detection Reagent (SDR) Preparation:

SDR I: Acetoacetyl-CoA (100×) and Succinate-Glo™ (100×) solutions were thawed on ice and diluted to 1× in assay buffer shortly before use.

SDR II: Succinate Detection Reagent II was thawed on ice shortly before use.

A standard curve was generated for every Succinate-Glo™ experiment to quantitate succinate production. A 2-fold dilution series of standard curve solution was prepared fresh (Table 7.27). Standard curve solution (10 µL) was mixed with SDR I (10 µL), and the resulting mixture was incubated at RT for 1 h. SDR II (20 µL) was added, and the final mixture was incubated at RT for a further 10 min before luminescence was measured.

Table 7.27 | Composition of Succinate-Glo™ standard curve solution in assay buffer. Final concentration refers to concentration in 10 µL reaction volume.

	Component	Stock concentration	Final concentration
Standard curve solution	L-Ascorbate	10 mM	1 mM
	$(NH_4)_2Fe(SO_4)_2$	1 mM	100 µM
	2-OG	1 mM	100 µM
	Succinate	250 µM	0–15 µM

7.14.3 Enzyme activity assay

Typical enzyme activity assay conditions for TET proteins were optimised in-house based on literature protocols[319].

Assay protocol: 2× enzyme mix (5 µL) was mixed with 2× substrate mix (5 µL) and incubated at RT for the desired time (Table 7.28). The reaction was quenched by the addition of SDR I

(10 µL), and the resulting mixture was incubated at RT for 1 h. SDR II (20 µL) was added, and the final mixture was incubated at RT for a further 10 min before luminescence was measured.

For uncoupled 2-OG turnover experiments, **^{5m}C 12 bp dsDNA** was substituted with **^{5}C 12 bp dsDNA** or no DNA in the 2× substrate mix. Reactions containing no enzyme were used for background determination.

Table 7.28 | General compositions of Succinate-Glo™ assay solutions in assay buffer. Final concentration refers to concentration in 10 µL reaction volume.

	Component	Stock concentration	2× mix concentration	Final concentration
2× TET enzyme mix	NgTet1	varied	2 µM	1 µM
2× substrate mix	L-Ascorbate	10 mM	2 mM	1 mM
	$(NH_4)_2Fe(SO_4)_2$	10 mM	200 µM	100 µM
	2-OG	1 mM	200 µM	100 µM
	^{5m}C 12 bp dsDNA* / **^{5}C 12 bp dsDNA†** / no DNA	1 mM	10 µM	5 µM

Self-complementary 12 base-pair double-stranded DNA (sequence: 5''-ACC AC^{5m}C** GGT GGT-3').*
†*Self-complementary 12 base-pair double-stranded DNA (sequence: 5'-ACC AC**C** GGT GGT-3').*

7.15 ^{5m}C oxidation monitoring using solid-phase extraction-mass spectrometry (SPE-MS)

7.15.1 Materials and buffers

Materials: (+)-Sodium L-ascorbate, $(NH_4)_2Fe(SO_4)_2$, and disodium 2-OG were purchased from Sigma-Aldrich. *N*–Oxalylglycine (NOG) and 8-hydroxy-5-quinolinecarboxylic acid (IOX1) were purchased from Cayman Chemicals. DNA oligomers used for SPE-MS experiments were purchased from ATDBio. 96-well, 2 mL, PP, V-bottom plates (Cat# 780271) used for multiple live reaction sampling assays and 384-well, PP, V-bottom microplates (Cat# 781280) used for end-point enzyme inhibition assays were purchased from Greiner Bio-One. SPE-MS-grade H_2O (LiChrosolv®, Cat# 115333) and MeCN (LiChrosolv®, Cat# 100030) were purchased from Merck. Octyl ammonium acetate (OAA) was prepared by Dr Roman Belle following literature

procedure[320]. hTET2$_{CD\Delta LCI}$ was produced in-house (Chapter 5). hTET1$_{CD}$ was kindly produced by Eidarus Salah.

Spectra recording: All SPE-MS spectra were recorded on an Agilent RapidFire 365 High-throughput Mass Spectrometry System coupled to an Agilent 6550 iFunnel Q-TOF mass spectrometer. An Agilent RapidFire C4 (type A, Cat# G9203A) or C18 (type C, Cat# G9205A) solid-phase extraction (SPE) cartridge was used for analyte separation.

Buffers: Compositions of buffers used in SPE-MS experiments are reported in Table 7.29. HEPES (1 M, pH 7.5) stock was stored at 4 °C in dark conditions. Quenching solution was stored at RT.

Table 7.29 | Composition of buffers used in SPE-MS TET enzyme activity assays.

Buffer	Composition
Assay buffer	50 mM HEPES, pH 7.3
Quenching solution	2 mM NOG

Cofactor preparation: (+)-Sodium L-ascorbate (500 mM) and disodium 2-OG (50 mM) stock solutions were prepared fresh in assay buffer. $(NH_4)_2Fe(SO_4)_2$ stock solution (50 mM) was prepared fresh in 20 mM HCl.

7.15.2 SPE-MS method parameters

RapidFire: Pump flow rates were set at 1.5 mL/min (pump 1), 1.25 mL/min (pump 2), and 0.8 mL/min (pump 3). Samples were sequentially aspirated with vacuum into the sample loop (40 µL) from a 96/384-well plate and loaded directly onto the SPE cartridge. The cartridge was washed with solvent A (4,000 msec) and the samples were eluted with solvent B (4,000 msec) (Table 7.30).

Table 7.30 | SPE-MS system solvents.

Solvent	Composition
Solvent A	6 mM octylammonium acetate (OAA) in H_2O
Solvent B	80% v/v MeCN in H_2O

Concurrent with sample elution, the sample loop was washed with solvent B. The system was re-equilibrated using solvent A, and the needle was washed prior to the next sample injection. The SPE cartridge was washed with 3–5 cycles of 'inorganic' wash (H_2O) followed by 'organic' wash (MeCN), with a final 'inorganic' wash (H_2O) prior to the next sample injection.

Q-TOF MS: The Q-TOF settings used for MS analysis of DNA oligomers are reported in Table 7.31.

Table 7.31 | Q-TOF settings used for DNA oligomer MS analysis.

Parameter	Value
Ion polarity	Negative
Drying gas temperature	280 °C
Drying gas flow	13 L/min
Nebuliser gas pressure	40 psig
Sheath gas temperature	350 °C
Sheath gas flow	12 L/min
Capillary voltage (Vcap)	4,000 V
Nozzle voltage	0 V
Fragmentor voltage	250 V
Spectrum acquisition rate	5 Hz

7.15.3 Continuous reaction sampling assay

6× enzyme mix (100 µL) was mixed with 1.2× substrate mix (500 µL) in a 96-well, 2 mL plate (Table 7.32) and incubated at RT. The reaction mixture was sampled directly from the plate at the desired time intervals (typically between 60 and 180 sec). The spectrum for the initial time point (*t*=0 sec) was acquired by sampling the 1.2× substrate mix prior to enzyme addition. Agilent MassHunter Qualitative Analysis™ (B.07.00) and RapidFire Integrator software was used for data processing.

Table 7.32 | General compositions of SPE-MS continuous reaction sampling assay solutions in assay buffer. Final concentration refers to concentration in 600 µL reaction volume.

	Component	Stock concentration	Mix concentration	Final concentration
6× TET enzyme mix	$hTET1_{CD}$/$hTET2_{CD\Delta LCI}$	varied	1.2–3.0 µM	200–500 nM
1.2× substrate mix	L-Ascorbate	10 mM	240 µM	200 µM
	$(NH_4)_2Fe(SO_4)_2$	1 mM	60 µM	50 µM
	2-OG	1 mM	60 µM	50 µM
	^{5m}C 12 bp dsDNA*	100 µM	1.2–2.4 µM	1–2 µM

Self-complementary 12 base-pair double-stranded DNA (sequence: 5'-ACC AC^{5m}C** GGT GGT-3').*

7.15.4 Enzyme inhibition end-point assay

Compound dispensing: Prior to enzyme inhibition experiments, DMSO-soluble compounds were pre-dispensed in a logarithmically linear concentration gradient (9–11 concentration points) onto 384-well, PP, V-bottom microplates using an Echo® 550 acoustic liquid handler (Labcyte). The final volume of DMSO in all wells was adjusted to 250 nL. The final reaction volume for enzyme inhibition assays was 25.25 µL (25 µL reaction mixture + 0.25 µL DMSO), resulting in a 0.99% v/v final DMSO concentration.

Assay protocol: 2× enzyme mix (12.5 µL) was plated into wells containing inhibitors (250 nL, DMSO) and incubated for 10 min at RT (Table 7.33). The reaction was initiated by the addition of 2× substrate mix (12.5 µL) and quenched by the addition of quenching solution (25 µL) within the enzyme linear activity window (determined separately). Agilent MassHunter Qualitative Analysis™ (B.07.00) and RapidFire Integrator software was used for data processing.

Table 7.33 | Compositions of SPE-MS enzyme inhibition end-point assay solutions in assay buffer. Final concentration refers to concentration in 25 µL reaction volume.

	Component	Stock concentration	Mix concentration	Final concentration
2× TET enzyme mix	$hTET2_{CD\Delta LCI}$	17.6 µM	600 nM	300 nM
2× substrate mix	L-Ascorbate	10 mM	400 µM	200 µM
	$(NH_4)_2Fe(SO_4)_2$	1 mM	100 µM	50 µM
	2-OG	1 mM	100 µM	50 µM
	^{5m}C 12 bp dsDNA*	100 µM	2 µM	1 µM

***Self-complementary 12 base-pair double-stranded DNA (sequence: 5'-ACC AC^{5m}**C** GGT GGT-3').*

Pre-quenched reactions containing vehicle (0.99% v/v DMSO) were used for background determination. IOX1 was used as a positive inhibition control. All enzyme activity data was normalised against enzyme activity in reactions containing vehicle only (0.99% v/v DMSO). Data was analysed using GraphPad Prism 5 software. Mean and standard deviation were plotted, and data was fitted using "log(inhibitor) *vs*. response - Variable slope (four parameters)" or "log(inhibitor) *vs*. normalised response - Variable slope" non-linear regression models.

7.16 ^{5m}C oxidation monitoring using Matrix-Assisted Laser Desorption/Ionisation-Time of Flight Mass Spectrometry (MALDI-TOF MS)

7.16.1 General information

Materials: (+)-Sodium L-ascorbate, $(NH_4)_2Fe(SO_4)_2$, and disodium 2-OG were purchased from Sigma-Aldrich. AG® 50W-X8 cation exchange resin (biotechnology grade, 100–200 mesh, hydrogen form) was purchased from BIO-RAD. 3-Hydroxypicolinic acid (3-HPA, matrix substance for MALDI-MS, ≥99.0%) was purchased from Sigma-Aldrich. DNA oligomers used in MALDI-TOF MS experiments were purchased from ATDBio. Ziptip C_{18} pipette tips were purchased from Millipore. T4 Phage β-glucosyltransferase (T4 β-GT) was purchased from NEB. UDP-α-D-glucose was purchased from Cayman Chemicals.

Spectra recording: All MALDI-TOF MS spectra were recorded on a Bruker MALDI-TOF Autoflex maX mass spectrometer (Bruker) on MSP 96 Ground Steel target plates. Data was collected using positive ion linear mode, 1–4 kDa collection range.

Buffers: Composition of assay buffer used in MALDI-TOF MS TET enzyme activity assays is reported in Table 7.34. Assay buffer was prepared fresh for each assay from component stock solutions. HEPES (1 M, pH 7.3) stock was stored at 4 °C in dark conditions.

Table 7.34 | Composition of MALDI-TOF MS TET enzyme activity assay buffer.

Buffer	Composition
Assay buffer	50 mM HEPES, 100 mM NaCl, pH 7.1

7.16.2 TET enzyme activity assay

Assay protocol: 2× enzyme mix (5 µL) was mixed with 2× substrate mix (5 µL) and incubated at RT (Table 7.35). The reaction was quenched by the addition of AG® 50W-X8 cation exchange resin suspension in H_2O (10 µL). The suspension was mixed and allowed to settle before decanting the supernatant for analysis.

Table 7.35 | Compositions of MALDI-TOF MS TET enzyme activity assay solutions in assay buffer. Final concentration refers to concentration in 10 µL reaction volume.

	Component	Stock concentration	2× mix concentration	Final concentration
2× TET enzyme mix	NgTet1	varied	4 µM	2 µM
2× substrate mix	L-Ascorbate	10 mM	1 mM	500 µM
	$(NH_4)_2Fe(SO_4)_2$	10 mM	100 µM	50 µM
	2-OG	1 mM	100 µM	50 µM
	^{5m}C 12 bp dsDNA*	100 µM	4 µM	2 µM

Self-complementary 12 base-pair double-stranded DNA (sequence: 5'-ACC AC^{5m}C** GGT GGT-3').*

7.16.3 ^{5hm}C glucosylation using T4 phage β-glucosyltransferase (T4 β-GT)

2× enzyme mix (60 µL) was mixed with 2× substrate mix (60 µL) and incubated at RT (Table 7.35). The reaction was quenched by heat inactivation (65 °C, 10 min). A portion of the quenched reaction mixture (5 µL) was mixed with a mixture containing T4 β-GT (0.2 U/µL) and UDP-α-D-glucose (0–500 µM) in NEB4 buffer (5 µL). The resulting mixture was incubated at 37 °C for 2 h. The samples were purified using reverse-phase C_{18} purification (ZipTip) and further desalted using AG® 50W-X8 cation exchange resin suspension in H_2O (10 µL) for 10 min before sample deposition.

7.16.4 DNA oligomer analysis using MALDI-TOF MS

ZipTip purification: reverse-phase C_{18} pipette tips were sequentially equilibrated with MeCN, 30% MeCN in H_2O, and 3% MeCN in H_2O. The nucleotides were bound to the C_{18} resin by repeat aspiration (10 aspiration cycles), washed with 3% MeCN in H_2O (3 aspiration cycles), and eluted with 30% MeCN in H_2O (10 µL elution volume, 5 aspiration cycles).

Matrix preparation: 3-HPA was dissolved to saturation (50 mg/mL) in a 1:1 mixture of H_2O/MeCN (+0.1% v/v TFA) containing diammonium hydrogen citrate (10 mg/mL). The matrix was prepared fresh for each experiment.

Sample deposition: Matrix solution (1 µL) was spotted on the target plate and allowed to dry. Quenched reaction sample (1 µL) was spotted onto the crystalline matrix spot and allowed to dry before analysis.

7.17 Photoreactive probe cross-linking

7.17.1 General information

Materials: Photoreactive probe TiP1 Y6Bpa was synthesised in-house using the methods outlined in section 7.2. Manganese (II) chloride tetrahydrate and diammonium hydrogen citrate (DAC) were purchased from Sigma-Aldrich. NgTet1 enzyme used for the experiments was prepared in-house (Chapter 2). UV-STAR® 384-well, F-bottom microplates (Cat# 781801) were purchased from Greiner Bio-One. 2,5-Dihydroxyacetophenone (2,5-DHAP, Cat# 8231829) was purchased from Bruker.

Light source: Irradiation was carried out using a 3UV Lamp (365/302/254 nm, 230 V, 50 Hz, Cat# 95035) from Thermo Fisher.

Spectra recording: All MALDI-TOF MS spectra were recorded on a Bruker MALDI-TOF Autoflex maX mass spectrometer (Bruker) on MSP 96 Ground Steel target plates. Data was collected using positive ion linear mode, 30–210 kDa collection range.

Buffers: Composition of buffer used in photocross-linking experiments is reported in Table 7.36. Photocross-linking buffer was prepared fresh for each experiment from component stock solutions. HEPES (1 M, pH 7.3) stock was stored at 4 °C in dark conditions. $MnCl_2$ (100 mM) stock was prepared fresh in H_2O.

Table 7.36 | Composition of buffer used in photocross-linking experiments.

Buffer	Composition
Photocross-linking buffer	20 mM HEPES, 150 mM NaCl, 20 µM $MnCl_2$, pH 7.1

7.17.2 Photocross-linking assay

Sample preparation: A mixture of NgTet1 (2 µM) and TiP1 Y6Bpa (25–100 µM) was prepared in assay buffer (Table 7.37) and incubated at 4 °C for 45 min. The samples were clarified (10,000 xg, 5 min, 4 °C) prior to irradiation.

Table 7.37 | General composition of photocross-linking reactions in assay buffer.

Component	Stock concentration	Final concentration
NgTet1	106.6 µM	2 µM
TiP1 Y6Bpa	500 µM (10% v/v DMSO)	25–100 µM (0.5–2% v/v DMSO)

Sample irradiation: Irradiation was carried out at 365/302 nm in UV-transparent 384-well microplates at a 5 cm distance from the light source. Irradiation time was 5–40 min. The samples were stored on ice during irradiation to prevent thermal decomposition.

7.17.3 Protein analysis using MALDI-TOF MS

Matrix preparation: 2,5-DHAP (7.6 mg) was dissolved in EtOH (375 µL). DAC in H_2O (125 µL, 18 mg/mL) was added. The matrix was stored at –20 °C.

Sample deposition: Protein sample (3 µL) was mixed with H_2O (1 µL, 4% v/v TFA). Matrix solution (2 µL) was added, and the resulting ternary mixture was mixed until crystallisation was observed. The crystal slurry (1 µL) was spotted onto the target plate and allowed to dry before analysis.

7.18 Protein LC-MS

7.18.1 Materials

Materials: H_2O (LiChrosolv®, Cat# 115333) and MeCN (LiChrosolv®, Cat# 100030) were purchased from Merck. Formic acid (99.0+%, Optima™ LC/MS Grade, Cat# A117-50) was purchased from Fisher Scientific.

LC-MS system: Protein LC-MS was carried out on an Agilent 1290 Infinity II LC system coupled to an Agilent 6550 iFunnel Q-TOF mass spectrometer. A ProSwift™ RP-4H HPLC Column (50 mm, Cat# 069477, Thermo Fisher) was used for analyte separation.

7.18.2 Method

Protein samples (1 µM) were stored at 10 °C and were injected (2–3 µL) using an Agilent 1290 series autosampler. The samples were eluted at 0.2 mL/min using a H_2O/MeCN solvent system supplemented with formic acid (0.1% v/v). The following gradient was used: 5–20% (0–1 min), 20–55% (1–9 min), 55–100% (9–10 min), 100% (10–11 min), 100–5% (11–12 min), 5% (12–15 min) MeCN in H_2O (+0.1% v/v FA).

Q-TOF MS: The Q-TOF settings used for intact protein MS analysis are reported in Table 7.38.

Table 7.38 | Q-TOF settings used for intact protein MS analysis.

Parameter	Value
Ion polarity	Positive
Drying gas temperature	225 °C
Drying gas flow	13 L/min
Nebuliser gas pressure	20 psig
Sheath gas temperature	350 °C
Sheath gas flow	12 L/min
Capillary voltage (Vcap)	3,500 V
Nozzle voltage	1,500 V
Fragmentor voltage	365 V
Spectrum acquisition rate	5 Hz

Spectra were processed on Agilent MassHunter Qualitative Analysis™ software (B.07.00) using the Maximum Entropy deconvolution algorithm.

7.19 Peptide LC-MS

7.19.1 Materials

Materials: H_2O was purified on a Milli-Q® Direct 8 Water Purification System equipped with a Progard® T3 Pretreatment Pack and a Millipak® Express 40 Filter. MeCN (HPLC grade, Cat# 34851) was purchased from Sigma-Aldrich. Formic acid (99%, Cat# 270480010) was purchased from Acros Organics.

LC-MS system: Peptide LC-MS was carried out on an ACQUITY Ultra Performance LC system coupled to a Waters Xevo G2-XS Q-TOF mass spectrometer equipped with an ESI LockSpray™ source. An ACQUITY UPLC HSS T3 Column (100 Å, 1.8 µm, 2.1×100 mm, Cat# 186003539, Waters) was used for analyte separation.

7.19.2 Method

Peptide samples (20 µM, up to 1% v/v DMSO) were stored at 10 °C and were injected (10 µL) using an ACQUITY Ultra Performance LC Sample Manager. The samples were eluted at 0.3 mL/min using a H_2O/MeCN solvent system supplemented with formic acid (0.1% v/v).

The following gradients were used depending on peptide lipophilicity:

High lipophilicity: 5% (0–1 min), 5–95% (1–7 min), 95% (7–8 min), 95–5% (8–8.1 min), 5% (8.1–10 min) MeCN in H_2O (+0.1% v/v FA).

Medium lipophilicity: 5% (0–1 min), 5–60% (1–6 min), 60–95% (6–7 min), 95% (7–8 min), 95–5% (8–8.1 min), 5% (8.1–10 min) MeCN in H_2O (+0.1% v/v FA).

Low lipophilicity: 1% (0–1 min), 1–15% (1–6 min), 15–95% (6–7 min), 95% (7–8 min), 95–5% (8–8.1 min), 5% (8.1–10 min) MeCN in H_2O (+0.1% v/v FA).

The scan range was 100–2,500 Da with a scan time of 1 sec in positive mode. The following MS parameters were used: capillary voltage limit – 2,000 V; sampling cone – 30 V, source

offset – 80 V, source temperature – 120 °C, desolvation temperature – 200 °C; cone gas flow – 50 L/h; desolvation gas flow – 700 L/h.

Spectra were processed on Waters MassLynx (4.1) software using isotopic modelling and peak centring (centroid top 80% by area, no mass correction).

7.20 Protein crystallography

7.20.1 Materials

30 kDa MWCO centrifugal filters (0.5 mL) were purchased from Amicon. Intelli-Plate 96-3 LVR sitting drop crystallisation plates were purchased from Hampton Research. Xtra-Clear Advanced Polyolefin StarSeal (qPCR) was purchased from STARLAB. 8-well 5 µL micro-reservoir strips for use with Mosquito® Xtal3 were purchased from SPT Labtech. Broad crystallisation screens were purchased from Hampton Research (PEG/Ion, Crystal screen HT, and SaltRx) or Molecular Dimensions (JCSG Plus and PACT premier™ HT-96).

7.20.2 Tagged hTET2$_{CD\Delta LCI}$ plate setup protocol (with TiP1)

Tagged hTET2$_{CD\Delta LCI}$ (9.83 mg/mL, 149 µM) in gel filtration buffer (20 mM HEPES, 150 mM NaCl, 0.5 mM TCEP, pH 7.5) was thawed on ice. A 2× crystallisation solution (20 mM HEPES, 150 mM NaCl, pH 7.5) containing $MnCl_2$ (5 mM) and NOG (4 mM) was combined with hTET2 in a 1:1 ratio, and the resulting mixture was incubated on ice for 20 min before centrifugation (13,800 xg, 10 min, 4 °C). The supernatant was concentrated in a 0.5 mL 30 kDa MWCO centrifugal filter to approx. 50% of the initial enzyme sample volume. Peptide TiP1 (220 mM in DMSO) was added to the resulting mixture (1:100 dilution, 2.2 mM final conc., 1% v/v DMSO final conc.). Precipitation was observed upon addition of TiP1. The suspension was mixed carefully and incubated on ice for 2 h with occasional mixing before clarification (13,800 xg, 10 min, 4 °C). The concentrated supernatant was transferred to a pre-cooled 0.5 mL Eppendorf tube and plated on pre-cooled Intelli-Plate 96-3 LVR sitting drop plates in 2:1, 1:1, 1:2 ratios

of protein/well solution using a PHOENIX liquid handler (Art Robbins). 55 µL of the concentrated mixture was dispensed per plate. After plating, the plates were sealed and stored at room temperature (approx. 22 °C). Plates were observed from day 2 onwards.

7.20.3 Tag-free hTET2$_{CD\Delta LCI}$ plate setup protocol (with DNA)

Tag-free hTET2$_{CD\Delta LCI}$ (19.68 mg/mL, 384 µM) in gel filtration buffer (20 mM Tris, 300 mM NaCl, pH 8.0) was thawed on ice. A $(NH_4)_2Fe(SO_4)_2$ stock was prepared in water (200 mM); an NOG stock was prepared either in water or gel filtration buffer (200 mM, buffered with additional NaOH to maintain pH 8.0). Fe(II) and NOG were sequentially added to the protein sample in two portions (1:100 dilution, 2 mM final concentration). To the resulting mixture was added portion-wise **^{5m}C 12 bp dsDNA** as a solid (7,320 g/mol, 1.5 equivalents, 600 µM final concentration) with careful mixing until all solid was dissolved. The final mixture was incubated on ice for 20 min before clarification (21,100 xg, 10 min, 4 °C). The supernatant was transferred to 8-well strips on a chilled block (6.5 µL in each well) and plated on pre-cooled Intelli-Plate 96-3 LVR sitting drop plates in 2:1, 1:1, 1:2 ratios of protein/well solution using a Mosquito® Xtal3 liquid handler (SPT Labtech). 52 µL of the final mixture was dispensed per plate. After plating, the plates were sealed and stored at room temperature (approx. 22 °C) or 4 °C. Plates were observed from day 2 onwards.

7.20.4 Reported hTET2 crystallisation condition reproduction

A 96-condition optimisation screen in a 12×8 matrix was prepared based on the reported hTET2 crystallisation conditions[94]. The 12-point axis contained 14–30% PEG2000 monomethyl ether (MME, 10×1.5% increments, 1×1% increment); the 8-point axis contained 100 mM MES at various pH (5.8, 6.0, 6.2, 6.3, 6.4, 6.5, 6.6, 6.8).

Chapter 8
References

1. Rédei, G. P. Central Dogma. *Encycl. Genet. Genomics, Proteomics Informatics* **227**, 309–309 (2008).
2. Handy, D. E., Castro, R. & Loscalzo, J. Epigenetic modifications: Basic mechanisms and role in cardiovascular disease. *Circulation* **123**, 2145–2156 (2011).
3. Waddington, C. H. The Epigenotype. *Endeavour* **1**, 18–20 (1942).
4. D'Urso, A. & Brickner, J. H. Mechanisms of epigenetic memory. *Trends Genet.* **30**, 230–236 (2014).
5. Butler, M. G. Genomic imprinting disorders in humans: A mini-review. *J. Assist. Reprod. Genet.* **26**, 477–486 (2009).
6. Delcuve, G. P., Rastegar, M. & Davie, J. R. Epigenetic control. *J. Cell. Physiol.* **219**, 243–250 (2009).
7. Tzika, E., Dreker, T. & Imhof, A. Epigenetics and metabolism in health and disease. *Front. Genet.* **9**, 361 (2018).
8. Rivera, R. M. & Bennett, L. B. Epigenetics in humans: An overview. *Curr. Opin. Endocrinol. Diabetes Obes.* **17**, 493–499 (2010).
9. Maeshima, K. & Eltsov, M. Packaging the genome: The structure of mitotic chromosomes. *J. Biochem.* **143**, 145–153 (2008).
10. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589**, 2914–2922 (2015).
11. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
12. Gery Felsenfeld, M. G. Controlling the double helix. *Nature* **421**, 448–453 (2003).
13. Bednar, J. *et al.* Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14173–14178 (1998).
14. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
15. Zhao, Y. & Garcia, B. A. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol.* **7**, a025064 (2015).
16. Zhao, Z. & Shilatifard, A. Epigenetic modifications of histones in cancer. *Genome Biol.* **20**, 245 (2019).
17. Mirkovitch, J. The role of chromatin in HIV-1 transcriptional regulation. *Immunobiology* **198**, 279–290 (1997).
18. Strålfors, A. & Ekwall, K. *Heterochromatin and Euchromatin – Organization, Boundaries, and Gene Regulation*. *Reviews in Cell Biology and Molecular Medicine* (John Wiley & Sons, Ltd, 2011). doi:10.1002/3527600906.mcb.200400018.pub2.

19. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367 (2010).

20. Jones, P. L. & Wolffe, A. P. Relationships between chromatin organization and DNA methylation in determining gene expression. *Semin. Cancer Biol.* **9**, 339–347 (1999).

21. Newell-Price, J., Clark, A. J. L. & King, P. DNA methylation and silencing of gene expression. *Trends Endocrinol. Metab.* **11**, 142–148 (2000).

22. Free, A. *et al.* DNA Recognition by the Methyl-CpG Binding Domain of MeCP2. *J. Biol. Chem.* **276**, 3353–3360 (2001).

23. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19**, 187–191 (1998).

24. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).

25. Feng, Q. & Zhang, Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev.* **15**, 827–832 (2001).

26. Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A. & Cedar, H. Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 61–65 (1982).

27. Raiber, E.-A., Hardisty, R., van Delft, P. & Balasubramanian, S. Mapping and elucidating the function of modified bases in DNA. *Nat. Rev. Chem.* **1**, 0069 (2017).

28. Timothy H Bestor, G. L. V. DNA methyltransferases. *Curr. Opin. Cell Biol.* **6**, 380–389 (1994).

29. Lyko, F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**, 81–92 (2018).

30. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219–220 (1998).

31. Williams, B. P., Pignatta, D., Henikoff, S. & Gehring, M. Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat. *PLoS Genet.* **11**, e1005142 (2015).

32. Bahar Halpern, K., Vana, T. & Walker, M. D. Paradoxical role of DNA methylation in activation of FoxA2 gene expression during endoderm development. *J. Biol. Chem.* **289**, 23882–23892 (2014).

33. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).

34. Zhu, H., Wang, G. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, 551–565 (2016).

35. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).

36. Warnecke, P. M. & Bestor, T. H. Cytosine methylation and human cancer. *Curr. Opin. Oncol.* **12**, 68–73 (2000).

37. Ono, R. *et al.* LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in

acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res.* **62**, 4075–4080 (2002).

38. Lorsback, R. B. *et al.* TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;23). *Leukemia* **17**, 637–641 (2003).

39. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).

40. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).

41. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chemie - Int. Ed.* **50**, 7008–7012 (2011).

42. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming. *J. Biol. Chem.* **279**, 52353–52360 (2004).

43. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).

44. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).

45. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333**, 1303–1307 (2011).

46. Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).

47. Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).

48. Boorstein, R. J. *et al.* Definitive Identification of Mammalian 5-Hydroxymethyluracil DNA N-Glycosylase Activity as SMUG1. *J. Biol. Chem.* **276**, 41991–41997 (2001).

49. Nabel, C. S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).

50. Iwan, K. *et al.* 5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo. *Nat. Chem. Biol.* **14**, 72–78 (2018).

51. Kamińska, E., Korytiaková, E., Reichl, A., Müller, M. & Carell, T. Intragenomic Decarboxylation of 5-Carboxy-2'-deoxycytidine. *Angew. Chemie - Int. Ed.* **60**, 23207–23211 (2021).

52. Liutkevičiutė, Z. *et al.* Direct decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).

53. Simonelli, V., Narciso, L., Dogliotti, E. & Fortini, P. Base excision repair intermediates are mutagenic in mammalian cells. *Nucleic Acids Res.* **33**, 4404–4411 (2005).

54. Mouret, J. F., Polverelli, M., Sarrazini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chem. Biol. Interact.* **77**, 187–201 (1991).

55. Bjelland, S. *et al.* Oxidation of Thymine to 5-Formyluracil in DNA: Mechanisms of Formation, Structural Implications, and Base Excision by Human Cell Free Extracts.

Biochemistry **34**, 14758–14764 (1995).

56. Pais, J. E. *et al.* Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4316–4321 (2015).

57. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).

58. Miyabe, I. Identification of 5-formyluracil DNA glycosylase activity of human hNTH1 protein. *Nucleic Acids Res.* **30**, 3443–3448 (2002).

59. Djuric, Z. *et al.* Levels of 5-hydroxymethyl-2'-deoxyuridine in DNA from blood of women scheduled for breast biopsy. *Cancer Epidemiol. Biomarkers Prev.* **10**, 147–149 (2001).

60. Rogstad, D. K. *et al.* 5-Formyluracil-Induced Perturbations of DNA Function. *Biochemistry* **43**, 5688–5697 (2004).

61. Zhang, Q. M. *et al.* Replication of DNA templates containing 5-formyluracil, a major oxidative lesion of thymine in DNA. *Nucleic Acids Res.* **25**, 3969–3973 (1997).

62. Klungland, A. *et al.* 5-Formyluracil and its nucleoside derivatives confer toxicity and mutagenicity to mammalian cells by interfering with normal RNA and DNA metabolism. *Toxicol. Lett.* **119**, 71–78 (2001).

63. Wu, T. P. *et al.* DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).

64. Xiao, C. Le *et al.* N6-Methyladenine DNA Modification in the Human Genome. *Mol. Cell* **71**, 306–318 (2018).

65. Li, Z. *et al.* N6-Methyladenine in DNA antagonizes SATB1 in early development. *Nature* **583**, 625–630 (2020).

66. Gommers-Ampt, J. H. *et al.* β-d-glucosyl-hydroxymethyluracil: A novel modified base present in the DNA of the parasitic protozoan T. brucei. *Cell* **75**, 1129–1136 (1993).

67. Xue, J. H. *et al.* A vitamin-C-derived DNA modification catalysed by an algal TET homologue. *Nature* **569**, 581–585 (2019).

68. Gaultney, R. A. *et al.* 4-Methylcytosine DNA modification is critical for global epigenetic regulation and virulence in the human pathogen Leptospira interrogans. *Nucleic Acids Res.* **48**, 12102–12115 (2021).

69. Rodriguez, F., Yushenova, I. A., DiCorpo, D. & Arkhipova, I. R. Bacterial N4-methylcytosine as an epigenetic mark in eukaryotic DNA. *Nat. Commun.* **13**, 1072 (2022).

70. Meng, S. *et al.* Epigenetics in neurodevelopment: Emerging role of circular RNA. *Front. Cell. Neurosci.* **13**, 327 (2019).

71. Martinez, S. & Hausinger, R. P. Catalytic mechanisms of Fe(II)- and 2-Oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **290**, 20702–20711 (2015).

72. Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A. & Schofield, C. J. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40**, 4364–4397 (2011).

73. Sim, J. *et al.* The Factor Inhibiting HIF Asparaginyl Hydroxylase Regulates Oxidative Metabolism and Accelerates Metabolic Adaptation to Hypoxia. *Cell Metab.* **27**, 898–913

(2018).

74. Gorres, K. L. & Raines, R. T. Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **45**, 106–124 (2010).

75. Fedeles, B. I., Singh, V., Delaney, J. C., Li, D. & Essigmann, J. M. The AlkB family of Fe(II)/α-ketoglutarate-dependent dioxygenases: Repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.* **290**, 20734–20742 (2015).

76. Lee, J. S., Smith, E. & Shilatifard, A. The Language of Histone Crosstalk. *Cell* **142**, 682–685 (2010).

77. Counts, D. F., Cardinale, G. J. & Udenfriend, S. Prolyl hydroxylase half reaction: peptidyl prolyl-independent decarboxylation of alpha-ketoglutarate. *Proc. Natl. Acad. Sci.* **75**, 2145–2149 (1978).

78. Puistola, U., Turpeenniemi-Hujanen, T. M., Myllylä, R. & Kivirikko, K. I. Studies on the lysyl hydroxylase reaction. I. Initial velocity kinetics and related aspects. *Biochim. Biophys. Acta - Enzymol.* **611**, 40–50 (1980).

79. TUDERMAN, L., MYLLYLÄ, R. & KIVIRIKKO, K. I. Mechanism of the Prolyl Hydroxylase Reaction: 1. Role of Co-substrate. *Eur. J. Biochem.* **80**, 341–348 (1977).

80. Alves, J., Vidugiris, G., Goueli, S. A. & Zegzouti, H. Bioluminescent High-Throughput Succinate Detection Method for Monitoring the Activity of JMJC Histone Demethylases and Fe(II)/2-Oxoglutarate-Dependent Dioxygenases. *SLAS Discov.* **23**, 242–254 (2018).

81. Hopkinson, R. J., Hamed, R. B., Rose, N. R., Claridge, T. D. W. & Schofield, C. J. Monitoring the activity of 2-oxoglutarate dependent histone demethylases by NMR spectroscopy: Direct observation of formaldehyde. *ChemBioChem* **11**, 506–510 (2010).

82. Ergel, B. *et al.* Protein dynamics control the progression and efficiency of the catalytic reaction cycle of the Escherichia coli DNA-repair enzyme AlkB. *J. Biol. Chem.* **289**, 29584–29601 (2014).

83. De Jong, L., Albracht, S. P. J. & Kemp, A. Prolyl 4-hydroxylase activity in relation to the oxidation state of enzyme-bound iron. The role of ascorbate in peptidyl proline hydroxylation. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **704**, 326–332 (1982).

84. MYLLYLÄ, R., TUDERMAN, L. & KIVIRIKKO, K. I. Mechanism of the Prolyl Hydroxylase Reaction: 2. Kinetic Analysis of the Reaction Sequence. *Eur. J. Biochem.* **80**, 349–357 (1977).

85. Myllyla, R., Majamaa, K. & Gunzler, V. Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. *J. Biol. Chem.* **259**, 5403–5405 (1984).

86. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).

87. Chen, J. *et al.* Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat. Genet.* **45**, 1504–1509 (2013).

88. Liu, A., Ho, R. Y. N. & Que L., J. Alternative reactivity of an α-ketoglutarate-dependent iron(II) oxygenase: Enzyme self-hydroxylation. *J. Am. Chem. Soc.* **123**, 5126–5127 (2001).

89. Ryle, M. J. *et al.* O2- and α-ketoglutarate-dependent tyrosyl radical formation in TauD,

an α-keto acid-dependent non-heme iron dioxygenase. *Biochemistry* **42**, 1854–1862 (2003).

90. Henshaw, T. F., Feig, M. & Hausinger, R. P. Aberrant activity of the DNA repair enzyme AlkB. *J. Inorg. Biochem.* **98**, 856–861 (2004).

91. Sundheim, O. *et al.* Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO J.* **25**, 3389–3397 (2006).

92. Tamanaha, E., Guan, S., Marks, K. & Saleh, L. Distributive Processing by the Iron(II)/α-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases. *J. Am. Chem. Soc.* **138**, 9345–9348 (2016).

93. Crawford, D. J. *et al.* Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism. *J. Am. Chem. Soc.* **138**, 730–733 (2016).

94. Hu, L. *et al.* Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155**, 1545–1555 (2013).

95. Hashimoto, H. *et al.* Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. *Nucleic Acids Res.* **43**, 10713–10721 (2015).

96. Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118–122 (2015).

97. Price, J. C., Barr, E. W., Glass, T. E., Krebs, C. & Bollinger, J. M. Evidence for Hydrogen Abstraction from C1 of Taurine by the High-Spin Fe(IV) Intermediate Detected during Oxygen Activation by Taurine:α-Ketoglutarate Dioxygenase (TauD). *J. Am. Chem. Soc.* **125**, 13008–13009 (2003).

98. Pan, J. *et al.* Evidence for Modulation of Oxygen Rebound Rate in Control of Outcome by Iron(II)- And 2-Oxoglutarate-Dependent Oxygenases. *J. Am. Chem. Soc.* **141**, 15153–15165 (2019).

99. Hashimoto, H. *et al.* Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391–395 (2014).

100. Kavoosi, S. *et al.* Site- And degree-specific C-H oxidation on 5-methylcytosine homologues for probing active DNA demethylation. *Chem. Sci.* **10**, 10550–10555 (2019).

101. Ghanty, U., Denizio, J. E., Liu, M. Y. & Kohli, R. M. Exploiting Substrate Promiscuity to Develop Activity-Based Probes for Ten-Eleven Translocation Family Enzymes. *J. Am. Chem. Soc.* **140**, 17329–17332 (2018).

102. Fu, L. *et al.* Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136**, 11582–11585 (2014).

103. Denizio, J. E., Liu, M. Y., Leddin, E. M., Cisneros, G. A. & Kohli, R. M. Selectivity and Promiscuity in TET-Mediated Oxidation of 5-Methylcytosine in DNA and RNA. *Biochemistry* **58**, 411–421 (2019).

104. Schröder, A. S. *et al.* 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Commun.* **52**, 14361–14364 (2016).

105. Adam, S. *et al.* Flanking sequences influence the activity of TET1 and TET2 methylcytosine dioxygenases and affect genomic 5hmC patterns. *Commun. Biol.* **5**, 92 (2022).

106. Aik, W. S. *et al.* CHAPTER 2 Introduction to Structural Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes. in *2-Oxoglutarate-Dependent Oxygenases* (eds. Hausinger, R. P. & Schofield, C. J.) 59–94 (The Royal Society of Chemistry, 2015). doi:10.1039/9781782621959-00059.

107. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710 (2009).

108. Iyer, L. M., Abhiman, S. & Aravind, L. Chapter 2 - Natural History of Eukaryotic DNA Methylation Systems. in *Modifications of Nuclear DNA and its Regulatory Proteins* (eds. Cheng, X. & Blumenthal, R. M.) vol. 101 25–104 (Academic Press, 2011).

109. Liu, D., Li, G. & Zuo, Y. Function determinants of TET proteins: The arrangements of sequence motifs with specific codes. *Brief. Bioinform.* **20**, 1826–1835 (2019).

110. Zhang, W. *et al.* Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Mol. Cell* **64**, 1062–1073 (2016).

111. Jin, S. G. *et al.* Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep.* **14**, 493–505 (2016).

112. Xu, Y. *et al.* Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Mol. Cell* **42**, 451–464 (2011).

113. Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122–126 (2013).

114. Upadhyay, A. K., Horton, J. R., Zhang, X. & Cheng, X. Coordinated methyl-lysine erasure: Structural and functional linkage of a Jumonji demethylase domain and a reader domain. *Curr. Opin. Struct. Biol.* **21**, 750–760 (2011).

115. Parker, M. J., Weigele, P. R. & Saleh, L. Insights into the Biochemistry, Evolution, and Biotechnological Applications of the Ten-Eleven Translocation (TET) Enzymes. *Biochemistry* **58**, 450–467 (2019).

116. Li, W. *et al.* Molecular mechanism for vitamin C-derived C5-glyceryl-methylcytosine DNA modification catalyzed by algal TET homologue CMD1. *Nat. Commun.* **12**, 744 (2021).

117. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science* **324**, 929–930 (2009).

118. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chemie - Int. Ed.* **49**, 5375–5377 (2010).

119. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).

120. Lian, C. G. *et al.* Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of Melanoma. *Cell* **150**, 1135–1146 (2012).

121. Christoph, H. M. *et al.* Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* **2**, 627–637 (2011).

122. Valinluck, V. & Sowers, L. C. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* **67**, 946–950 (2007).

123. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40**, 4841–4849 (2012).

124. Valinluck, V. *et al.* Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **32**, 4100–4108 (2004).

125. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).

126. Thalhammer, A., Hansen, A. S., El-Sagheer, A. H., Brown, T. & Schofield, C. J. Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chem. Commun.* **47**, 5325–5327 (2011).

127. Lercher, L. *et al.* Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **50**, 1794–1796 (2014).

128. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152**, 1146–1159 (2013).

129. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).

130. Su, M. *et al.* 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chemie - Int. Ed.* **55**, 11797–11800 (2016).

131. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).

132. Münzel, M. *et al.* Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chem. - A Eur. J.* **17**, 13782–13788 (2011).

133. Booth, M. J., Marsico, G., Bachman, M., Beraldi, D. & Balasubramanian, S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat. Chem.* **6**, 435–440 (2014).

134. Raiber, E. A. *et al.* 5-formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).

135. Hardwick, J. S. *et al.* 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.* **24**, 544–552 (2017).

136. Dai, Q. *et al.* Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability. *ACS Chem. Biol.* **11**, 470–477 (2016).

137. Ngo, T. T. M. *et al.* Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **7**, 10813 (2016).

138. Li, F. *et al.* 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *J. Am. Chem. Soc.* **139**, 10617–10620 (2017).

139. Raiber, E. A. *et al.* 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat. Chem.* **10**, 1258–1266 (2018).

140. Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).

141. Raiber, E. A. *et al.* Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* **13**, R69 (2012).

142. Yuan, F. *et al.* 5-Carboxylcytosine is resistant towards phosphodiesterase I digestion: Implications for epigenetic modification quantification by mass spectrometry. *RSC Adv.* **9**, 29010–29014 (2019).

143. Kellinger, M. W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).

144. Wang, L. *et al.* Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* **523**, 621–625 (2015).

145. Pinto, P. A. B. *et al.* RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J.* **30**, 2431–2444 (2011).

146. De La Mata, M. *et al.* A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* **12**, 525–532 (2003).

147. Lu, X. *et al.* Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Res.* **25**, 386–389 (2015).

148. Neri, F. *et al.* Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. *Cell Rep.* **10**, 674–683 (2015).

149. Eleftheriou, M. *et al.* 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clin. Epigenetics* **7**, 88 (2015).

150. Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–612 (2011).

151. Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460–464 (2013).

152. Huang, Y. *et al.* Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1361–1366 (2014).

153. Colquitt, B. M., Allen, W. E., Barnea, G. & Lomvardas, S. Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14682–14687 (2013).

154. Li, X. *et al.* Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7120–7125 (2014).

155. Lan, J. *et al.* Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. *Nat. Commun.* **11**, 4956 (2020).

156. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–349 (2011).

157. Zhang, Q. *et al.* Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* **525**, 389–393 (2015).

158. Chrysanthou, S. *et al.* The DNA dioxygenase Tet1 regulates H3K27 modification and embryonic stem cell biology independent of its catalytic activity. *Nucleic Acids Res.* **50**, 3169–3189 (2022).

159. Xu, Y. *et al.* Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for xenopus eye and neural development. *Cell* **151**, 1200–1213 (2012).

160. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome-biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).

161. Feng, Y., Li, X., Cassady, K., Zou, Z. & Zhang, X. TET2 function in hematopoietic malignancies, immune regulation, and DNA repair. *Front. Oncol.* **9**, 210 (2019).

162. Yang, H. *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**, 663–669 (2013).

163. Langemeijer, S. M. C. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).

164. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **18**, 553–567 (2010).

165. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).

166. Wang, Y. *et al.* WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Mol. Cell* **57**, 662–673 (2015).

167. Rivas, M. P. *et al.* TET upregulation leads to 5-hydroxymethylation enrichment in hepatoblastoma. *Front. Genet.* **10**, 553 (2019).

168. Kaelin, W. G. SDH5 Mutations and Familial Paraganglioma: Somewhere Warburg is Smiling. *Cancer Cell* **16**, 180–182 (2009).

169. Selak, M. A. *et al.* Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-α prolyl hydroxylase. *Cancer Cell* **7**, 77–85 (2005).

170. Isaacs, J. S. *et al.* HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: Novel role of fumarate in regulation of HIF stability. *Cancer Cell* **8**, 143–153 (2005).

171. Laukka, T. *et al.* Fumarate and succinate regulate expression of hypoxia-inducible genes via TET enzymes. *J. Biol. Chem.* **291**, 4256–4265 (2016).

172. Belle, R. Histone Lysine and DNA Methylation: Dynamic Marks in the Chromatin. (University of Oxford, 2019).

173. Koivunen, P. *et al.* Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484–488 (2012).

174. Sudhamalla, B., Dey, D., Breski, M. & Islam, K. A rapid mass spectrometric method for the measurement of catalytic activity of ten-eleven translocation enzymes. *Anal. Biochem.* **534**, 28–35 (2017).

175. Baader, E., Tschank, G., Baringhaus, K. H., Burghard, H. & Gunzler, V. Inhibition of prolyl

4-hydroxylase by oxalyl amino acid derivatives in vitro, in isolated microsomes and in embryonic chicken tissues. *Biochem. J.* **300**, 525–530 (1994).

176. McDonough, M. A. *et al.* Selective inhibition of factor inhibiting hypoxia-inducible factor. *J. Am. Chem. Soc.* **127**, 7680–7681 (2005).

177. Rose, N. R. *et al.* Selective Inhibitors of the JMJD2 Histone Demethylases: Combined Nondenaturing Mass Spectrometric Screening and Crystallographic Approaches. *J. Med. Chem.* **53**, 1810–1818 (2010).

178. Guan, Y. *et al.* A Therapeutic Strategy for Preferential Targeting of TET2-Mutant and TET Dioxygenase-Deficient Cells in Myeloid Neoplasms. *Cancer Discov.* **2**, 146–161 (2021).

179. Tiwari, A. D. *et al.* SAR insights into TET2 catalytic domain inhibition: Synthesis of 2-Hydroxy-4-Methylene-pentanedicarboxylates. *Bioorganic Med. Chem.* **39**, 116141 (2021).

180. Sudhamalla, B. *et al.* Complementary Steric Engineering at the Protein-Ligand Interface for Analogue-Sensitive TET Oxygenases. *J. Am. Chem. Soc.* **140**, 10263–10269 (2018).

181. Islam, K. The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem. Biol.* **25**, 1171–1184 (2018).

182. King, O. N. F. *et al.* Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5**, e15535 (2010).

183. Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.* **4**, 3110–3117 (2013).

184. Thinnes, C. C. *et al.* Targeting histone lysine demethylases - Progress, challenges, and the future. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 1416–1432 (2014).

185. Singh, A. K. *et al.* Selective targeting of TET catalytic domain promotes somatic cell reprogramming. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3621–3626 (2020).

186. Guan, Y. *et al.* Eltrombopag inhibits TET dioxygenase to contribute to hematopoietic stem cell expansion in aplastic anemia. *J. Clin. Invest.* **132**, e149856 (2022).

187. Vlachodimitropoulou, E. *et al.* Eltrombopag: A powerful chelator of cellular or extracellular iron(III) alone or combined with a second chelator. *Blood* **130**, 1923–1933 (2017).

188. Spitz, A. Z., Zacharioudakis, E., Reyna, D. E., Garner, T. P. & Gavathiotis, E. Eltrombopag directly inhibits BAX and prevents cell death. *Nat. Commun.* **12**, 1134 (2021).

189. Chua, G. N. L. *et al.* Cytosine-Based TET Enzyme Inhibitors. *ACS Med. Chem. Lett.* **10**, 180–185 (2019).

190. Weirath, N. A. *et al.* Small Molecule Inhibitors of TET Dioxygenases: Bobcat339 Activity Is Mediated by Contaminating Copper(II). *ACS Med. Chem. Lett.* (2022) doi:10.1021/acsmedchemlett.1c00677.

191. Nishio, K. *et al.* Thioether Macrocyclic Peptides Selected against TET1 Compact Catalytic Domain Inhibit TET1 Catalytic Activity. *ChemBioChem* **19**, 979–985 (2018).

192. Ackloo, S., Brown, P. J. & Müller, S. Chemical probes targeting epigenetic proteins: Applications beyond oncology. *Epigenetics* **12**, 378–400 (2017).

193. Pais, J. E. *et al.* Biochemical characterization of a Naegleria TET-like oxygenase and its

application in single molecule sequencing of 5-methylcytosine. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4316–4321 (2015).

194. Ito, S. *et al.* Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).

195. Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607–1616 (2011).

196. Kim, J., Kim, K., Mo, J. S. & Lee, Y. Atm deficiency in the DNA polymerase β null cerebellum results in cerebellar ataxia and Itpr1 reduction associated with alteration of cytosine methylation. *Nucleic Acids Res.* **48**, 3678–3691 (2020).

197. Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci.* **86**, 8247–8251 (1989).

198. Giglione, C., Vallon, O. & Meinnel, T. Control of protein life-span by N-terminal methionine excision. *EMBO J.* **22**, 13–23 (2003).

199. Wingfield, P. T. N-terminal methionine processing. *Curr. Protoc. Protein Sci.* **2017**, 14–16 (2017).

200. Geoghegan, K. F. *et al.* Spontaneous α-N-6-phosphogluconoylation of a 'His tag' in Escherichia coli: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267**, 169–184 (1999).

201. Meier, S., Jensen, P. R. & Duus, J. Direct Observation of Metabolic Differences in Living Escherichia Coli Strains K-12 and BL21. *ChemBioChem* **13**, 308–310 (2012).

202. Kim, K. M., Yi, E. C., Baker, D. & Zhang, K. Y. J. Post-translational modification of the N-terminal His tag interferes with the crystallization of the wild-type and mutant SH3 domains from chicken src tyrosine kinase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 759–762 (2001).

203. Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).

204. Khan, A., Schofield, C. J. & Claridge, T. D. W. Reducing Agent-Mediated Nonenzymatic Conversion of 2-Oxoglutarate to Succinate: Implications for Oxygenase Assays. *ChemBioChem* **21**, 2898–2902 (2020).

205. Kelly, A. E., Ou, H. D., Withers, R. & Dötsch, V. Low-conductivity buffers for high-sensitivity NMR measurements. *J. Am. Chem. Soc.* **124**, 12013–12019 (2002).

206. Rao, N. V. & Adams, E. Partial reaction of prolyl hydroxylase. (Gly-Pro-Ala)(n) stimulates α-ketoglutarate decarboxylation without prolyl hydroxylation. *J. Biol. Chem.* **253**, 6327–6330 (1978).

207. Hamada, S. *et al.* Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. *Bioorganic Med. Chem. Lett.* **19**, 2852–2855 (2009).

208. Rose, N. R. *et al.* Inhibitor scaffolds for 2-oxoglutarate-dependent histone lysine demethylases. *J. Med. Chem.* **51**, 7053–7056 (2008).

209. Pirrone, G. F. *et al.* Use of MALDI-MS Combined with Differential Hydrogen-Deuterium

Exchange for Semiautomated Protein Global Conformational Screening. *Anal. Chem.* **89**, 8351–8357 (2017).

210. Welford, R. W. D., Schlemminger, I., McNeill, L. A., Hewitson, K. S. & Schofield, C. J. The selectivity and inhibition of AlkB. *J. Biol. Chem.* **278**, 10157–10161 (2003).

211. Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J. Med. Chem.* **55**, 6639–6643 (2012).

212. Schiller, R. S. Investigating the inhibitor and substrate diversity of the JmjC histone demethylases. (University of Oxford, 2016).

213. Cascella, B. & Mirica, L. M. Kinetic analysis of iron-dependent histone demethylases: α-Ketoglutarate substrate inhibition and potential relevance to the regulation of histone demethylation in cancer cells. *Biochemistry* **51**, 8699–8701 (2012).

214. Tarhonskaya, H. *et al.* Studies on the Interaction of the Histone Demethylase KDM5B with Tricarboxylic Acid Cycle Intermediates. *J. Mol. Biol.* **429**, 2895–2906 (2017).

215. Tarhonskaya, H. *et al.* Kinetic investigations of the role of factor inhibiting hypoxia-inducible factor (FIH) as an oxygen sensor. *J. Biol. Chem.* **290**, 19726–19742 (2015).

216. Dao, J. H. *et al.* Kinetic characterization and identification of a novel inhibitor of hypoxia-inducible factor prolyl hydroxylase 2 using a time-resolved fluorescence resonance energy transfer-based assay technology. *Anal. Biochem.* **384**, 213–223 (2009).

217. Feng, B. Y., Shelat, A., Doman, T. N., Guy, R. K. & Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* **1**, 146–148 (2005).

218. Shoichet, B. K. Interpreting steep dose-response curves in early inhibitor discovery. *J. Med. Chem.* **49**, 7274–7277 (2006).

219. Feng, B. Y. *et al.* A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* **50**, 2385–2390 (2007).

220. Mantri, M., Zhang, Z., McDonough, M. A. & Schofield, C. J. Autocatalysed oxidative modifications to 2-oxoglutarate dependent oxygenases. *FEBS J.* **279**, 1563–1575 (2012).

221. Hutton, J. J., Tappel, A. L. & Udenfriend, S. Cofactor and substrate requirements of collagen proline hydroxylase. *Arch. Biochem. Biophys.* **118**, 231–240 (1967).

222. Zhang, Z. *et al.* Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4667–4672 (2017).

223. Zhang, Y., Dai, M. & Yuan, Z. Methods for the detection of reactive oxygen species. *Anal. Methods* **10**, 4625–4638 (2018).

224. Frei, B. Reactive oxygen species and antioxidant vitamins: Mechanisms of action. *Am. J. Med.* **97**, 5–13 (1994).

225. Fukumura, H. *et al.* Effect of ascorbic acid on reactive oxygen species production in chemotherapy and hyperthermia in prostate cancer cells. *J. Physiol. Sci.* **62**, 251–257 (2012).

226. Dixon, S. J. & Stockwell, B. R. The role of iron and reactive oxygen species in cell death. *Nat. Chem. Biol.* **10**, 9–17 (2014).

227. Serrano, L., Roper, G. & Kawamura, A. CHAPTER 5 Cyclic Peptides as Chemical Probes. in *The Discovery and Utility of Chemical Probes in Target Discovery* 100–123 (The Royal

Society of Chemistry, 2021). doi:10.1039/9781839160745-00100.

228. Huang, Y., Wiedmann, M. M. & Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* **119**, 10360–10391 (2019).

229. Passioura, T. & Suga, H. A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* **53**, 1931–1940 (2017).

230. Bhushan, B. *et al.* An integrated platform approach enables discovery of potent, selective and ligand-competitive cyclic peptides targeting the GIP receptor. *Chem. Sci.* **13**, 3256–3262 (2022).

231. Jiang, Y. *et al.* Concise solid-phase synthesis enables derivatisation of YEATS domain cyclopeptide inhibitors for improved cellular uptake. *Bioorganic Med. Chem.* **45**, 116342 (2021).

232. Jones, A. T. & Sayers, E. J. Cell entry of cell penetrating peptides: Tales of tails wagging dogs. *J. Control. Release* **161**, 582–591 (2012).

233. Lauer, J. L., Fields, C. G. & Fields, G. B. Sequence dependence of aspartimide formation during 9-fluorenylmethoxycarbonyl solid-phase peptide synthesis. *Lett. Pept. Sci.* **1**, 197–205 (1995).

234. Karlström, A. & Undén, A. A new protecting group for aspartic acid that minimizes piperidine-catalyzed aspartimide formation in fmoc solid phase peptide synthesis. *Tetrahedron Lett.* **37**, 4243–4246 (1996).

235. Behrendt, R., Huber, S., Martí, R. & White, P. New t-butyl based aspartate protecting groups preventing aspartimide formation in Fmoc SPPS. *J. Pept. Sci.* **21**, 680–687 (2015).

236. Michels, T., Dölling, R., Haberkorn, U. & Mier, W. Acid-mediated prevention of aspartimide formation in solid phase peptide synthesis. *Org. Lett.* **14**, 5218–5221 (2012).

237. Subirós-Funosas, R., Prohens, R., Barbas, R., El-Faham, A. & Albericio, F. Oxyma: An efficient additive for peptide synthesis to replace the benzotriazole-based HOBt and HOAt with a lower risk of explosion. *Chem. - A Eur. J.* **15**, 9394–9403 (2009).

238. Palasek, S. A., Cox, Z. J. & Collins, J. M. Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. *J. Pept. Sci.* **13**, 143–148 (2007).

239. Lacham-Hartman, S. *et al.* Avidity observed between a bivalent inhibitor and an enzyme monomer with a single active site. *PLoS One* **16**, e0249616 (2021).

240. Imaide, S. *et al.* Trivalent PROTACs enhance protein degradation via combined avidity and cooperativity. *Nat. Chem. Biol.* **17**, 1157–1167 (2021).

241. Kawamura, A. *et al.* Highly selective inhibition of histone demethylases by de novo macrocyclic peptides. *Nat. Commun.* **8**, 14773 (2017).

242. Winkler, D. F. H. & Campbell, W. D. The spot technique: Synthesis and screening of peptide macroarrays on cellulose membranes. *Methods Mol. Biol.* **494**, 47–70 (2008).

243. Murale, D. P., Hong, S. C., Haque, M. M. & Lee, J. S. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **15**, 14 (2017).

244. Becker, J. M. & Naider, F. Chapter Twenty-Four - Cross-linking Strategies to Study

Peptide Ligand–Receptor Interactions. in *Membrane Proteins—Production and Functional Characterization* (ed. Shukla, A. K.) vol. 556 527–547 (Academic Press, 2015).

245. Dorman, G. & Prestwich, G. D. Benzophenone Photophores in Biochemistry. *Biochemistry* **33**, 5661–5673 (1994).

246. Wang, Y. *et al.* Effects of Lipophilicity on the Affinity and Nonspecific Binding of Iodinated Benzothiazole Derivatives. *J. Mol. Neurosci.* **20**, 255–260 (2003).

247. Johnson, T. W., Gallego, R. A. & Edwards, M. P. Lipophilic Efficiency as an Important Metric in Drug Design. *J. Med. Chem.* **61**, 6401–6420 (2018).

248. Walters, W. P. & Namchuk, M. Designing screens: How to make your hits a hit. *Nat. Rev. Drug Discov.* **2**, 259–266 (2003).

249. Pitt, A. & Nims, Z. Peptide Display Technologies. in *Cyclic Peptide Design* (ed. Goetz, G.) 285–298 (Springer New York, 2019). doi:10.1007/978-1-4939-9504-2_13.

250. Brewitz, L., Nakashima, Y. & Schofield, C. J. Synthesis of 2-oxoglutarate derivatives and their evaluation as cosubstrates and inhibitors of human aspartate/asparagine-β-hydroxylase. *Chem. Sci.* **12**, 1327–1342 (2021).

251. McAllister, T. E. *et al.* Non-competitive cyclic peptides for targeting enzyme-substrate complexes. *Chem. Sci.* **9**, 4569–4578 (2018).

252. Walport, L. J., Obexer, R. & Suga, H. Strategies for transitioning macrocyclic peptides to cell-permeable drug leads. *Curr. Opin. Biotechnol.* **48**, 242–250 (2017).

253. Singh, A., Schofield, C. J. & Chen, R. Hypoxia-inducible factor (HIF) prolyl hydroxylase inhibitors induce autophagy and have a protective effect in an in-vitro ischaemia model. *Sci. Rep.* **10**, 1597 (2020).

254. Hamada, S. *et al.* Design, synthesis, enzyme-inhibitory activity, and effect on human cancer cells of a novel series of Jumonji domain-containing protein 2 histone demethylase inhibitors. *J. Med. Chem.* **53**, 5629–5638 (2010).

255. Hong, Y. R. *et al.* [(4-Hydroxyl-benzo[4,5]thieno[3,2-c]pyridine-3-carbonyl)-amino]-acetic acid derivatives; HIF prolyl 4-hydroxylase inhibitors as oral erythropoietin secretagogues. *Bioorganic Med. Chem. Lett.* **23**, 5953–5957 (2013).

256. Wang, L. *et al.* A small molecule modulates Jumonji histone demethylase activity and selectively inhibits cancer growth. *Nat. Commun.* **4**, 2035 (2013).

257. Grozinger, C. M., Chao, E. D., Blackwell, H. E., Moazed, D. & Schreiber, S. L. Identification of a Class of Small Molecule Inhibitors of the Sirtuin Family of NAD-dependent Deacetylases by Phenotypic Screening. *J. Biol. Chem.* **276**, 38837–38843 (2001).

258. Jung-Hynes, B., Nihal, M., Zhong, W. & Ahmad, N. Role of sirtuin histone deacetylase SIRT1 in prostate cancer: A target for prostate cancer management via its inhibition? *J. Biol. Chem.* **284**, 3823–3832 (2009).

259. Ariazi, J. L. *et al.* Discovery and preclinical characterization of GSK1278863 (daprodustat), a small molecule hypoxia inducible factor-prolyl hydroxylase inhibitor for anemias. *J. Pharmacol. Exp. Ther.* **363**, 336–347 (2017).

260. Jeong, Y. C. & Moloney, M. G. Antibacterial barbituric acid analogues inspired from natural 3-acyltetramic acids; Synthesis, tautomerism and structure and physicochemical property-antibacterial activity relationships. *Molecules* **20**, 3582–3627 (2015).

261. Giuliano, B. M. *et al.* Tautomerism in 4-Hydroxypyrimidine, S-Methyl-2-thiouracil, and 2-Thiouracil. *J. Phys. Chem. A* **114**, 12725–12730 (2010).

262. Lima, M. C. P., Coutinho, K., Canuto, S. & Rocha, W. R. Reaction mechanism and tautomeric equilibrium of 2-mercaptopyrimidine in the gas phase and in aqueous solution: A combined monte carlo and quantum mechanics study. *J. Phys. Chem. A* **110**, 7253–7261 (2006).

263. Bühlmann, P. & Simon, W. Neutral hosts for the complexation of creatinine. *Tetrahedron* **49**, 7627–7636 (1993).

264. Rauf, M. A., Hisaindee, S. & Saleh, N. Spectroscopic studies of keto-enol tautomeric equilibrium of azo dyes. *RSC Adv.* **5**, 18097–18110 (2015).

265. Nowak, R. P. *et al.* First-in-Class Inhibitors of the Ribosomal Oxygenase MINA53. *J. Med. Chem.* **64**, 17031–17050 (2021).

266. Mann, A. 15 - CONFORMATIONAL RESTRICTION AND/OR STERIC HINDRANCE IN MEDICINAL CHEMISTRY. in *The Practice of Medicinal Chemistry (Second Edition)* (ed. Wermuth, C. G.) 233–250 (Academic Press, 2003). doi:https://doi.org/10.1016/B978-012744481-9/50019-2.

267. Dave, V. S., Gupta, D., Yu, M., Nguyen, P. & Varghese Gupta, S. Current and evolving approaches for improving the oral permeability of BCS Class III or analogous molecules. *Drug Dev. Ind. Pharm.* **43**, 177–189 (2017).

268. Nielsen, N. M. & Bundgaard, H. Prodrugs as drug delivery systems. 68. Chemical and plasma-catalyzed hydrolysis of various esters of benzoic acid: a reference system for designing prodrug esters of carboxylic acid agents. *Int. J. Pharm.* **39**, 75–85 (1987).

269. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).

270. Burlingham, B. T. & Widlanski, T. S. An intuitive look at the relationship of Ki and IC50: A more general use for the Dixon plot. *J. Chem. Educ.* **80**, 214–218 (2003).

271. Holdgate, G. A., Meek, T. D. & Grimley, R. L. Mechanistic enzymology in drug discovery: A fresh perspective. *Nat. Rev. Drug Discov.* **17**, 115–132 (2018).

272. Figg, W. D. *et al.* Structural Basis of Prolyl Hydroxylase Domain Inhibition by Molidustat. *ChemMedChem* **16**, 2082–2088 (2021).

273. Cytiva. *Gel filtration calibration kit manual*. https://cdn.cytivalifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=14782 (2021).

274. Rial, D. V. & Ceccarelli, E. A. Removal of DnaK contamination during fusion protein purifications. *Protein Expr. Purif.* **25**, 503–507 (2002).

275. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* **12**, 1313–1322 (2003).

276. Correddu, D. *et al.* Effect of consecutive rare codons on the recombinant production of human proteins in Escherichia coli. *IUBMB Life* **72**, 266–274 (2020).

277. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Curr. Opin. Biotechnol.* **6**, 494–500 (1995).

278. Rosano, G. L. & Ceccarelli, E. A. Rare codon content affects the solubility of recombinant

proteins in a codon bias-adjusted Escherichia coli strain. *Microb. Cell Fact.* **8**, 41 (2009).

279. Komar, A. A., Lesnik, T. & Reiss, C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* **462**, 387–391 (1999).

280. Thanaraj, T. A. & Argos, P. Ribosome-mediated translational pause and protein domain organization. *Protein Sci.* **5**, 1594–1612 (1996).

281. Alexaki, A. *et al.* Effects of codon optimization on coagulation factor IX translation and structure: Implications for protein and gene therapies. *Sci. Rep.* **9**, 15449 (2019).

282. Guo, Y., Wallace, S. S. & Bandaru, V. A novel bicistronic vector for overexpressing Mycobacterium tuberculosis proteins in Escherichia coli. *Protein Expr. Purif.* **65**, 230–237 (2009).

283. Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14371–14376 (2008).

284. Lobstein, J. *et al.* SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11**, 753 (2012).

285. Shao, F., Bader, M. W., Jakob, U. & Bardwell, J. C. A. DsbG, a protein disulfide isomerase with chaperone activity. *J. Biol. Chem.* **275**, 13349–13352 (2000).

286. Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. Production of soluble mammalian proteins in Escherichia coli: Identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4**, 32 (2004).

287. Blackwell, J. R. & Horgan, R. A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett.* **295**, 10–12 (1991).

288. Prasad, S., Khadatare, P. B. & Roy, I. Effect of chemical chaperones in improving the solubility of recombinant proteins in Escherichia coli. *Appl. Environ. Microbiol.* **77**, 4603–4609 (2011).

289. Yang, X. & Zhang, Y. Effect of temperature and sorbitol in improving the solubility of Carboxylesterases protein CpCE-1 from Cydia pomonella and biochemical characterization. *Appl. Microbiol. Biotechnol.* **97**, 10423–10433 (2013).

290. Lee, J. C. & Timasheff, S. N. The stabilization of proteins by sucrose. *J. Biol. Chem.* **256**, 7193–7201 (1981).

291. Grant, C. L. & Pramer, D. MINOR ELEMENT COMPOSITION OF YEAST EXTRACT. *J. Bacteriol.* **84**, 869–870 (1962).

292. Chhetri, G., Kalita, P. & Tripathi, T. An efficient protocol to enhance recombinant protein expression using ethanol in Escherichia coli. *MethodsX* **2**, 385–391 (2015).

293. Zheng, H. *et al.* Ethanol effects on the overexpression of heterologous catalase in Escherichia coli BL21 (DE3). *Appl. Microbiol. Biotechnol.* **103**, 1441–1453 (2019).

294. Priyanka, P. *et al.* Ethanol as additive enhances expression of Ranibizumab in Escherichia coli: Impact on cellular physiology and transcriptome. *Process Biochem.* **112**, 167–176 (2022).

295. Paliy, O. & Gunasekera, T. S. Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl. Microbiol. Biotechnol.* **73**, 1169–1172 (2007).

296. Magnusdottir, A., Johansson, I., Dahlgren, L.-G., Nordlund, P. & Berglund, H. Enabling IMAC purification of low abundance recombinant proteins from E. coli lysates. *Nat. Methods* **6**, 477–478 (2009).

297. Ho, C. W., Tan, W. S., Yap, W. B., Ling, T. C. & Tey, B. T. Comparative evaluation of different cell disruption methods for the release of recombinant hepatitis B core antigen from Escherichia coli. *Biotechnol. Bioprocess Eng.* **13**, 577–583 (2008).

298. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571.

299. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci.* **10**, 1206–1215 (2001).

300. Joint ProteomicS Laboratory (JPSL) of the Ludwig Institute for Cancer Research and Walter and Eliza Hall Institute of Medical Research (Australia). Measuring Protein Concentration in the Presence of Nucleic Acids by A280/A260: The Method of Warburg and Christian. *Cold Spring Harb. Protoc.* (2006) doi:10.1101/pdb.prot4252.

301. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC Biotechnol.* **7**, 32 (2007).

302. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5**, 172 (2014).

303. Aksamitiene, E., Hoek, J. B., Kholodenko, B. & Klyatkin, A. Multistrip Western blotting to increase quantitative data output. *Electrophoresis* **28**, 3163–3173 (2007).

304. Larsen, M. R., Trelle, M. B., Thingholm, T. E. & Jensen, O. N. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* **40**, 790–798 (2006).

305. Macek, B. *et al.* Protein post-translational modifications in bacteria. *Nat. Rev. Microbiol.* **17**, 651–664 (2019).

306. Morgan, B. & Lahav, O. The effect of pH on the kinetics of spontaneous Fe(II) oxidation by O2 in aqueous solution – basic principles and a simple heuristic description. *Chemosphere* **68**, 2080–2084 (2007).

307. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *Elife* **2**, e01345 (2013).

308. De La Cruz, M. J. *et al.* Atomic-resolution structures from fragmented protein crystals with the cryoEMEM method MicroEDED. *Nat. Methods* **14**, 399–402 (2017).

309. de la Cruz, M. J. *et al.* Micro- and nanocrystal preparation for MicroED and XFEL serial crystallography by fragmentation of imperfect crystals. *Protoc. Exch.* (2017) doi:10.1038/protex.2017.010.

310. Clabbers, M. T. B., Fisher, S. Z., Coinçon, M., Zou, X. & Xu, H. Visualizing drug binding interactions using microcrystal electron diffraction. *Commun. Biol.* **3**, 417 (2020).

311. Cooper, H. M., East, I. J., Todd, P. E. E. & Leach, S. J. Antibody response to myoglobins: Effect of host species. *Mol. Immunol.* **21**, 479–487 (1984).

312. Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S. H. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* **59**, 6595–6628 (2016).

313. Buckley, D. L. *et al.* Targeting the von Hippel-Lindau E3 ubiquitin ligase using small molecules to disrupt the VHL/HIF-1α interaction. *J. Am. Chem. Soc.* **134**, 4465–4468 (2012).

314. CEM. Liberty Blue Operation Manual. http://cem.com/media/contenttype/media/literature/600291-cemv6.pdf.

315. NEB Tm calculator. https://tmcalculator.neb.com/#!/main.

316. Gileadi, O. *et al.* High Throughput Production of Recombinant Human Proteins for Crystallography. in *Structural Proteomics: High-Throughput Methods* (eds. Kobe, B., Guss, M. & Huber, T.) 221–246 (Humana Press, 2008). doi:10.1007/978-1-60327-058-8_14.

317. Savitsky, P. *et al.* High-throughput production of human proteins for crystallization: The SGC experience. *J. Struct. Biol.* **172**, 3–13 (2010).

318. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

319. Kawamura, A. *et al.* Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* **404**, 86–93 (2010).

320. Vas, G. *et al.* Study of transaldolase deficiency in urine samples by capillary LC-MS/MS. *J. Mass Spectrom.* **41**, 463–469 (2006).

Appendix 1 DNA oligomer information

Information for all DNA oligomers used in this work. Primer sequences used for cloning and sequencing are reported in the appropriate sections in Chapter 7.

Name	Length (nt/bp)	Hybridisation	Modification	Sequence (5'–3')
^{5}C 12 bp dsDNA	12	Double-stranded self-complementary	None	ACC ACC GGT GGT
^{5m}C 12 bp dsDNA	12	Double-stranded self-complementary	5-methylcytosine	ACC AC^{5m}**C** GGT GGT
^{5m}C 32 nt bt-ssDNA	32	Single-stranded	Biotinylation, 5-methylcytosine	[Biotin]-TCG GAT GTT GTG GGT CAG ^{5m}**C**GC ATG ATA GTG TA
^{5hm}C 32 nt bt-ssDNA	32	Single-stranded	Biotinylation, 5-hydroxymethylcytosine	[Biotin]-TCG GAT GTT GTG GGT CAG ^{5hm}**C**GC ATG ATA GTG TA

Appendix 2 Peptide purity and accurate mass information

Peptide purity was determined using LC-UV analysis at 220 nm on an ACQUITY Ultra Performance LC system using a H_2O/MeCN solvent system supplemented with formic acid (0.1% v/v); high-resolution mass spectra (HRMS) were recorded using a Waters Xevo G2-XS Q-TOF mass spectrometer equipped with an ESI LockSpray™ source as described in section 7.19. Values for observed m/z are reported from +1 charge states unless otherwise noted; values from +2 charge states $[M+2H]^{2+}$ are deconvoluted and noted with an asterisk (*); values from +3 charge states $[M+3H]^{3+}$ are deconvoluted and noted with an obelus (†).

All peptides contain a C–terminal amide and are thioether macrocycles unless otherwise noted. Residues forming the macrocycle are highlighted in bold in lasso-type macrocyles. D-Amino acids are represented as lowercase letters. [Bpa] refers to 4-benzoyl-L-phenylalanine.

Name	Sequence	Chemical formula	m/z, $[M+H]^+$		Mass error (ppm)	Purity (%)
			Predicted	Observed		
TiP1	yLIYAYYTWWEHSC	$C_{96}H_{118}N_{19}O_{23}S$	1936.8369	1936.8374	0.27	>99.0
TiP2	yWYLPSYRVPWFC	$C_{93}H_{116}N_{19}O_{18}S$	1818.8466	1818.8442	−1.35	>99.0
TiP3	yYRVKTYYQITYVYLSC	$C_{109}H_{154}N_{23}O_{28}S$	2265.1054	2265.1064*	0.42	91.9
KS01	yKFYVPWWARLYFIGC	$C_{111}H_{144}N_{23}O_{20}S$	2151.0679	2151.0708*	1.36	97.4
KS02	yRIYLFYTWWNVYC	$C_{103}H_{130}N_{21}O_{21}S$	2028.9471	2028.9474	0.15	98.8
KS03	yVYYIPWWCRAVFETC	$C_{105}H_{137}N_{22}O_{23}S_2$	2137.9668	2137.9646	−1.05	84.2
KS04	yYRVKTYYQITYVYVSC	$C_{108}H_{152}N_{23}O_{28}S$	2251.0898	2251.0867	−1.37	77.7
KS05(M15L)	yYRVKTYHQITYVYLSC	$C_{106}H_{152}N_{25}O_{27}S$	2239.1010	2239.1054*	1.95	95.4
KS06	yIVYRYYTWWRC	$C_{90}H_{116}N_{21}O_{18}S$	1810.8528	1810.8499	−1.60	92.1
KS07	yRYYFAYTWWAC	$C_{88}H_{103}N_{18}O_{18}S$	1731.7418	1731.7437	1.07	98.3
KS08	yREYALYTWWLIQC	$C_{95}H_{128}N_{21}O_{22}S$	1946.9264	1946.9236	−1.42	97.8
KS09	yAPVYWNYTWWC	$C_{86}H_{100}N_{17}O_{18}S$	1690.7153	1690.7141	−0.71	>99.0
KS10	wIPLHLRSILAYNGRSC	$C_{93}H_{145}N_{28}O_{22}S$	2038.0809	2038.0862*	2.60	95.4

Name	Sequence	Chemical formula	m/z, $[M+H]^+$		Mass error (ppm)	Purity (%)
			Predicted	Observed		
KS11	wKLYYNYTWWIC	$C_{91}H_{113}N_{18}O_{18}S$	1777.8201	1777.8220	1.07	93.2
KS12	wIIYHNYTWWRC	$C_{88}H_{111}N_{22}O_{17}S$	1779.8218	1779.8207	−0.63	92.5
KS13	wTDKYYWAYTWWSC	$C_{100}H_{117}N_{20}O_{23}S$	1997.8321	1997.8306	−0.76	>99.0
TiP1 L2A	yAIYAYYTWWEHSC	$C_{93}H_{112}N_{19}O_{23}S$	1894.7899	1894.7871	−1.49	>99.0
TiP1 I3A	yLAYAYYTWWEHSC	$C_{93}H_{112}N_{19}O_{23}S$	1894.7899	1894.7904	0.25	>99.0
TiP1 Y4A	yLIAAYYTWWEHSC	$C_{90}H_{114}N_{19}O_{22}S$	1844.8107	1844.8097	−0.52	>99.0
TiP1 Y6A	yLIYAAYTWWEHSC	$C_{90}H_{114}N_{19}O_{22}S$	1844.8107	1844.8097	−0.52	>99.0
TiP1 Y7A	yLIYAYATWWEHSC	$C_{90}H_{114}N_{19}O_{22}S$	1844.8107	1844.8096	−0.57	>99.0
TiP1 T8A	yLIYAYYAWWEHSC	$C_{95}H_{116}N_{19}O_{22}S$	1906.8263	1906.8267	0.21	>99.0
TiP1 W9A	yLIYAYYTAWEHSC	$C_{88}H_{113}N_{18}O_{23}S$	1821.7947	1821.7947	0.01	>99.0
TiP1 W10A	yLIYAYYTWAEHSC	$C_{88}H_{113}N_{18}O_{23}S$	1821.7947	1821.7922	−1.36	>99.0
TiP1 E11A	yLIYAYYTWWAHSC	$C_{94}H_{116}N_{19}O_{21}S$	1878.8314	1878.8286	−1.49	>99.0
TiP1 H12A	yLIYAYYTWWEASC	$C_{93}H_{116}N_{17}O_{23}S$	1870.8151	1870.8153	0.12	>99.0
TiP1 S13A	yLIYAYYTWWEHAC	$C_{96}H_{118}N_{19}O_{22}S$	1920.8420	1920.8445	1.32	>99.0
TiP1 Y6Bpa	yLIYA[Bpa]YTWWEHSC	$C_{103}H_{122}N_{19}O_{23}S$	2024.8682	2024.8738	2.78	>99.0
TiP1-FLAG	**y**LIYAYYTWWEHS**C**GSGSDYKDDDDK	$C_{147}H_{192}N_{33}O_{48}S$	3219.3318	3219.3320*	0.05	90.3
TiP1-R8	**y**LIYAYYTWWEHS**C**GSGSRRRRRRRR	$C_{154}H_{230}N_{55}O_{37}S$	3473.7527	3473.7580†	1.51	78.2
TiP1-TAT	**y**LIYAYYTWWEHS**C**GSGSYGRKKRRQRRR	$C_{170}H_{250}N_{55}O_{42}S$	3765.8838	3765.8917†	2.09	85.0
KTT1	yVRPIPFNYFFC	$C_{81}H_{106}N_{17}O_{16}S$	1604.7724	1604.7700	−1.51	96.0
KTT2	yKQSIPWIYWFC	$C_{85}H_{110}N_{17}O_{17}S$	1672.7986	1672.7961	−1.52	97.0
KTT3	yIAQSIYRYVAC	$C_{69}H_{102}N_{17}O_{18}S$	1488.7310	1488.7313	0.24	>99.0
KTT4	yQGYSWYWIGAC	$C_{75}H_{91}N_{16}O_{18}S$	1535.6418	1535.6399	−1.24	>99.0
KTT5	yKHPYSRQYWLWTC	$C_{95}H_{125}N_{24}O_{21}S$	1969.9172	1969.9116	−2.84	>99.0
KTT6	yFPLITIYRYLC	$C_{80}H_{115}N_{16}O_{17}S$	1603.8347	1603.8330	−1.05	>99.0
TiP1.L (Ac-TiP1-NH_2)	Ac-yLIYAYYTWWEHSC-NH_2	$C_{96}H_{120}N_{19}O_{23}S$	1938.8525	1938.8502*	−1.20	97.6
Ac-FLAG-NH_2	Ac-DYKDDDDK-NH_2	$C_{43}H_{64}N_{11}O_{20}$	1054.4329	1054.4324	−0.49	97.2
Ac-R8-NH_2	Ac-RRRRRRRR-NH_2	$C_{50}H_{102}N_{33}O_9$	1308.8538	1308.8542†	0.28	95.9
Ac-TAT-NH_2	Ac-YGRKKRRQRRR-NH_2	$C_{66}H_{122}N_{33}O_{14}$	1600.9849	1600.9819†	−1.88	96.3

Appendix 3 Small molecule inhibitor numbering

Chapter 4 entry	MC code
1	MC3096
2	MC3125
3	MC4295
4	MC4296
5	MC1827
6	MC2177
7	MC2282
8	MC1648
9	MC1711
10	MC3543
11	MC3962
12	MC2191
13	MC3947
14	MC2164
15	MC2200
15a	MC4462
16	MC2204
17	MC3095
18	MC3098
19	MC3373
20	MC3484
21	MC3965
22	MC4261
23	MC2230
23’	MC4469
23a	MC4772
23b	MC4509
24	MC2209
25	MC2223
25a	MC2213
26	MC4473
27	MC4481
28	MC4489
29	MC4488
30	MC4458
31	MC4498
32	MC4466
33	MC4593
34	MC4483
35	MC2314
36	MC2315
37	MC3359
38	MC2422
39	MC4497
40	MC4487
41	MC2320
42a	MC4500
42b	MC4476
43	MC4477
44	MC4467
45	MC4468
46	MC4486
47a	MC4545
47b	MC4542
47c	MC4465
48a	MC4541
48b	MC4544
48c	MC4485
49	MC4543
50a	MC4499
50b	MC4472
51a	MC4503
51b	MC4484
52	MC4490
53	MC4506
54	MC4471
55	MC3234
56	MC4645
57	MC4478
58	MC4474
59	MC4510
60	MC4479
61	MC4480
62	MC4587
63	MC4494
63a	MC4605
64	MC4586
65	MC4482
JIB-04	MC4300

Appendix 4 Protein construct information

General information

Protein expression in *E. coli* systems was carried out by the author. Protein expression in insect/mammalian systems was carried out by Eidarus Salah (University of Oxford). Affinity/solubility tag sequences are highlighted in bold. Protease cleavage sites are highlighted using a black down-pointing triangle (▼). Protein properties were calculated using the Expasy ProtParam tool[298]. Calculated properties are reported for constructs with tags intact; calculated properties for cleaved proteins are reported in brackets.

1 Naegleria Tet-like dioxygenase (NgTet1)

Expression system: Bacterial (*E. coli*)

MGSSHHHHHHSSGLVPR▼GSHMTTFKQQTIKEKETKRKYCIKGTTANLTQTHPNGPVCVNRGEEVANT
TTLLDSGGGINKKSLLQNLLSKCKTTFQQSFTNANITLKDEKWLKNVRTAYFVCDHDGSVELAYLPNVLPK
ELVEEFTEKFESIQTGRKKDTGYSGILDNSMPFNYVTADLSQELGQYLSEIVNPQINYYISKLLTCVSSRTINY
LVSLNDSYYALNNCLYPSTAFNSLKPSNDGHRIRKPHKDNLDITPSSLFYFGNFQNTEGYLELTDKNCKVFV
QPGDVLFFKGNEYKHVVANITSGWRIGLVYFAHKGSKTKPYYEDTQKNSLKIHKETK

Number of amino acids: 341 (324)

Enzyme residues: 1–321 (full length)

Calculated molecular weight: 38,661.66 Da (36,779.61 Da)

Ext. coefficient (assuming all Cys residues reduced): 37,820 (37,820)

Calculated pI: 9.00 (8.92)

Protease: Thrombin

2 Human TET1 (hTET1)

Expression system: Insect (Sf9 *Spodoptera frugiperda* strain), produced by Eidarus Salah

MGHHHHHHSSGVDLGTENLYFQ▼SMELPTCSCLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQ
KGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVLIMVWDGIPLPM
ADRLYTELTENLKSYNGHPTDRRCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRF
RIDPSSPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLDFCAH
PHRDIHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPRR

KKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSLPTLGSNTETVQPEV
KSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLKNDATASCGFSERSSTPHCTM
PSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPSTGVTEPLTPHQPNHQPSFLTSPQDLAS
SPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSVLIECARRE
LHATTPVEHPNRNHPTRLSLVFYQHKNLNKPQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQS
SEVNELNQIPSHKALTLTHDNVVTVSPYALTHVAGPYNHWV

Number of amino acids: 743 (721)

Enzyme residues: 1418–2136 (catalytic domain, CD)

Calculated molecular weight: 82,002.28 Da (79,479.59 Da)

Ext. coefficient (assuming all Cys residues reduced): 59,820 (58,330)

Calculated pI: 6.75

Protease: Tobacco etch virus (TEV)

3.1 Human TET2 (hTET2), His_6-TEV tag

Expression system: Insect (Sf9)

MGHHHHHHSSGVDLGTENLYFQ▼SMDFPSCRCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQK
GKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLY
SELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARSKIPRKFKLLGDDP
KEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEGRPFSGVTACLDFCAHAHRDLH
NMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSDVDEFGSVEAQEEKKRSGAIQVLSSFRRKVR
MLAEPVKTCRQRKLEAKKAAAEKLSGGGGSGGGGSGGGGSDEVWSDSEQSFLDPDIGGVAVAPTHGSI
LIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGLALWEAKMAEKAREKEEECEKYG

Number of amino acids: 484 (462)

Enzyme residues: 1129–1481_linker(GGGGSGGGGSGGGGS)_1844–1936 (catalytic domain without low-complexity insert, CDΔLCI)

Calculated molecular weight: 53,914.14 Da (51,391.44 Da)

Ext. coefficient (assuming all Cys residues reduced): 46,870 (45,380)

Calculated pI: 7.02 (7.48)

Protease: Tobacco etch virus (TEV)

TEE-His_6-HRV 3C tag (translation enhancing element)

Expression system: Bacterial (*E. coli*)

MNHKVHHHHHHIEGRHMELGTLEGSEFLEVLFQ▼GPVDDFPSCRCVEQIIEKDEGPFYTHLGAGPNVA
AIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVILILV
WEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARS
KIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEGRPFSGVTACL
DFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSDVDEFGSVEAQEEKKRSGAI
QVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLSGGGGSGGGGSGGGGSDEVWSDSEQSFLDPDIG
GVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGLALWEAKMAEKAREKEE
ECEKYG

Number of amino acids: 497 (464)

Enzyme residues: 1129–1481_linker(GGGGSGGGGSGGGGS)_1844–1936 (catalytic domain without low-complexity insert, CDΔLCI)

Calculated molecular weight: 55,491.00 Da (51,541.56 Da)

Ext. coefficient (assuming all Cys residues reduced): 45,380 (45,380)

Calculated pI: 6.73 (7.25)

Protease: Human rhinovirus (HRV) 3C or PreScission protease

3.2 Human TET2 (hTET2), His_6-FLAG-SUMO (Smt3) tag

Expression system: Bacterial (*E. coli*)

**MGSSHHHHHHDYKDDDDKSSGLVPRGSHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIF
FKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG▼S**DFPSC
RCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSS
EEKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPE
TCGASFSFGCSWSMYYNGCKFARSKIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIE
YEHRAPECRLGLKEGRPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLP
LYKVSDVDEFGSVEAQEEKKRSGAIQVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLSGGGGSGGGG
SGGGGSDEVWSDSEQSFLDPDIGGVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSM
NEPKHGLALWEAKMAEKAREKEEECEKYG

Number of amino acids: 590 (461)

Enzyme residues: 1129–1481_linker(GGGGSGGGGSGGGGS)_1844–1936 (catalytic domain without low-complexity insert, CDΔLCI)

Calculated molecular weight: 65,951.55 Da (51,260.25 Da)

Ext. coefficient (assuming all Cys residues reduced): 48,360 (45,380)

Calculated pI: 6.37 (7.48)

Protease: Ulp1 or SenP3 SUMO protease

3.3 Human TET2 (hTET2), His_6-Trx-TEV (thioredoxin) tag

Expression system: Bacterial (*E. coli*)

MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA
KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGTENLYFQ▼SMDFPSCR
CVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSE
EKLLCLVRERAGHTCEAAVIVILILVWEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPET
CGASFSFGCSWSMYYNGCKFARSKIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEY
EHRAPECRLGLKEGRPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPL
YKVSDVDEFGSVEAQEEKKRSGAIQVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLSGGGGSGGGGS
GGGGSDEVWSDSEQSFLDPDIGGVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMN
EPKHGLALWEAKMAEKAREKEEECEKYG

Number of amino acids: 589 (462)

Enzyme residues: 1129–1481_linker(GGGGSGGGGSGGGGS)_1844–1936 (catalytic domain without low-complexity insert, CDΔLCI)

Calculated molecular weight: 65,318.31 Da (51,391.44 Da)

Ext. coefficient (assuming all Cys residues reduced): 60,850 (45,380)

Calculated pI: 6.49 (7.48)

Protease: Tobacco etch virus (TEV)

3.4 Human TET2 (hTET2), His_6-GFP-TEV (green fluorescent protein) tag

Expression system: Bacterial (*E. coli*)

MGSSHHHHHHGSSVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGIEENLYFQ▼SNADFPSCRCVE
QIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLL
CLVRERAGHTCEAAVIVILILVWEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGA
SFSFGCSWSMYYNGCKFARSKIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHR
APECRLGLKEGRPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKV
SDVDEFGSVEAQEEKKRSGAIQVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLSGGGGSGGGGSGG

GGSDEVWSDSEQSFLDPDIGGVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEP
KHGLALWEAKMAEKAREKEEECEKYG

Number of amino acids: 723 (463)

Enzyme residues: 1129–1481_linker(GGGGSGGGGSGGGGS)_1844–1936 (catalytic domain without low-complexity insert, CDΔLCI)

Calculated molecular weight: 80,664.21 Da (51,445.43 Da)

Ext. coefficient (assuming all Cys residues reduced): 65,780 (45,380)

Calculated pI: 6.49 (7.48)

Protease: Tobacco etch virus (TEV)

3.5 Human TET2 (hTET2), His_6-GST-TEV (glutathione S-transferase) tag

Expression system: Bacterial (*E. coli*)

**MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID
GDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEML
KMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYI
AWPLQGWQATFGGADHPPKSSSGVDLGTENLYFQ▼SM**DFPSCRCVEQIIEKDEGPFYTHLGAGPNVA
AIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVILILV
WEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSFGCSWSMYYNGCKFARS
KIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPECRLGLKEGRPFSGVTACL
DFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSDVDEFGSVEAQEEKKRSGAI
QVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLSGGGGSGGGGSGGGGSDEVWSDSEQSFLDPDIG
GVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGLALWEAKMAEKAREKEE
ECEKYG

Number of amino acids: 704 (462)

Enzyme residues: 1129–1481_linker(GGGGSGGGGSGGGGS)_1844–1936 (catalytic domain without low-complexity insert, CDΔLCI)

Calculated molecular weight: 79,613.01 Da (51,391.44 Da)

Ext. coefficient (assuming all Cys residues reduced): 89,730 (45,380)

Calculated pI: 6.70 (7.48)

Protease: Tobacco etch virus (TEV)

3.6 Human TET2 (hTET2), His_6-MBP-TEV (maltose binding protein) tag

Expression system: Bacterial (*E. coli*)

MGSSHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAAT
GDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLP
NPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKA
GLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKP
FVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQK
GEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNGIEENLYFQ▼SNADFPSCRCVEQIIEK
DEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVR
ERAGHTCEAAVIVILILVWEGIPLSLADKLYSELTETLRKYGTLTNRRCALNEERTCACQGLDPETCGASFSF
GCSWSMYYNGCKFARSKIPRKFKLLGDDPKEEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRAPE
CRLGLKEGRPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLHVLPLYKVSDV
DEFGSVEAQEEKKRSGAIQVLSSFRRKVRMLAEPVKTCRQRKLEAKKAAAEKLSGGGGSGGGGSGGGGS
DEVWSDSEQSFLDPDIGGVAVAPTHGSILIECAKRELHATTPLKNPNRNHPTRISLVFYQHKSMNEPKHGL
ALWEAKMAEKAREKEEECEKYG

Number of amino acids: 853 (463)

Enzyme residues: 1129–1480_linker(GGGGSGGGGSGGGGS)_1844–1936 (catalytic domain without low-complexity insert, CDΔLCI)

Calculated molecular weight: 94,392.05 Da (51,445.43 Da)

Ext. coefficient (assuming all Cys residues reduced): 113,220 (45,380)

Calculated pI: 6.12 (7.48)

Protease: Tobacco etch virus (TEV)

4 Human TET3 (hTET3)

Expression system: Mammalian (Sf9 *Spodoptera frugiperda* strain for virus amplification, Expi293FTM for protein expression), produced by Eidarus Salah

MEFPTCDCVEQIVEKDEGPYYTHLGSGPTVASIRELMEERYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWV
IRRHTLEEKLLCLVRHRAGHHCQNAVIVILILAWEGIPRSLGDTLYQELTDTLRKYGNPTSRRCGLNDDRTC
ACQGKDPNTCGASFSFGCSWSMYFNGCKYARSKTPRKFRLAGDNPKEEEVLRKSFQDLATEVAPLYKRL
APQAYQNQVTNEEIAIDCRLGLKEGRPFAGVTACMDFCAHAHKDQHNLYNGCTVVCTLTKEDNRCVGK
IPEDEQLHVLPLYKMANTDEFGSEENQNAKVGSGAIQVLTAFPREVRRLPEPAKSCRQRQLEARKAAAEK
KKIQKEKLSTPEKIKQEALELAGITSDPGLSLKGGLSQQGLKPSLKVEPQNHFSSFKYSGNAVVESYSVLGNC
RPSDPYSMNSVYSYHSYYAQPSLTSVNGFHSKYALPSFSYYGFPSSNPVFPSQFLGPGAWGHSGSSGSFEK
KPDLHALHNSLSPAYGGAEFAELPSQAVPTDAHHPTPHHQQPAYPGPKEYLLPKAPLLHSVSRDPSPFAQ

SSNCYNRSIKQEPVDPLTQAEPVPRDAGKMGKTPLSEVSQNGGPSHLWGQYSGGPSMSPKRTNGVGG
SWGVFSSGESPAIVPDKLSSFGASCLAPSHFTDGQWGLFPGEGQQAASHSGGRLRGKPWSPCKFGNSTS
ALAGPSLTEKPWALGAGDFNSALKGSPGFQDKLWNPMKGEEGRIPAAGASQLDRAWQSFGLPLGSSEK
LFGALKSEEKLWDPFSLEEGPAEEPPSKGAVKEEKGGGGAEEEEEELWSDSEHNFLDENIGGVAVAPAHG
SILIECARRELHATTPLKKPNRCHPTRISLVFYQHKNLNQPNHGLALWEAKMKQLAERARARQEEAARLG
LGQQEAKLYGKKRKWGGTVVAEPQQKEKKGVVPTRQALAVPTDSAVTVSSYAYTKVTGPYSRWI**AENLY**
FQ▼SHHHHHHHHHHGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEK

Number of amino acids: 1,022 (980)

Enzyme residues: 824–1795 (catalytic domain, CD)

Calculated molecular weight: 111,848.56 Da (107,305.84 Da)

Ext. coefficient (assuming all Cys residues reduced): 148,170 (137,170)

Calculated pI: 8.51 (8.51)

Protease: Tobacco etch virus (TEV)