Expanding the Inhibitor Space of the WWP1 and WWP2 Ubiquitin Ligases

Ashley Philip Dudey

School of Biological Sciences

University of East Anglia

This thesis is submitted for the degree of *Doctor of Philosophy*, July 2024

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Acknowledgements

Having enrolled at UEA as an enthusiastic undergraduate back in 2016, the submission of this thesis is not only an end to my PhD but to a significant chapter in my life. UEA has allowed me to grow as an academic and individual and is thanks to all the wonderful people I have met, studied and worked alongside throughout my journey.

Firstly, I would like to thank Dr Gregory Hughes for introducing me to laboratory life and for kindly providing the data to support my findings. I want to thank Dr Andrew Chantry for giving me the opportunity to work on the WWP project as well as for his supervision and guidance over the first 2 years of my project. A special thanks must go to Professor Andrew Hemmings, who has gone above and beyond, as both a co-supervisor and later sole supervisor, to now someone I consider a mentor. He has inspired my passion for structural biology and as such I will be forever grateful. I must thank our collaborators at the School of Chemistry, including Dr Joshual Hall for his work on the frustrating NSC-2805 project, the project supervisors Dr Tom Storr and Professor Richard Stephenson and my PhD twin Dr Jake Rigby for his incredible speed of synthesis in the NSC-217913 and I3C projects. I want to thank UEA and the Big C charity for funding my work alongside Dr Charles Brearley for his help in funding a vital extension.

On a personal note, I would like to thank Dr Ben Miller and Professor Tom Clarke for their support during a difficult period. I want to thank the BIO 2.30 lab for making me feel at home during my move-up from floor 0, and thank the lab technicians, particularly Andy Loveday, for his quite amazing service despite the challenges. I would also like to thank the Head of BIO School, Professor Tamas Dalmay, for his constant reminder of the various Ziggurat challenges and football opportunities, with us winning the science research park tournament in my final year.

Finally, I want to thank my friends and most of all family for their support throughout my time at UEA. A special thanks must go to Dr Daniel Moye, my bio buddy from the very beginning, with all the laps of the UEA campus and much-needed mental release over a regular oat milk latte and a large hot chocolate (with some GF brownies for good measure). I must thank my parents for their unwavering support of my work and their patience with my failed attempts to explain it. I want to thank my dad in particular for the various snacks and coffee deliveries to my writing cave. Finally, I should thank myself for believing that I can do anything I put my mind to, whilst being kind to myself when things did not always go to plan.

Declaration

I hereby declare that except where specific reference is made to both mine and the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Ashley Philip Dudey
July 2024

Covid-19 Impact Statement

This statement is included to account for the impact of the COVID-19 pandemic on the research undertaken throughout this thesis, particularly in the early stages where a UK-wide lockdown and social distancing protocols had affected training on key equipment and techniques required. As such, the overall volume of research undertaken over the three-year project was reduced.

Abstract

The two members of the NEDD4 family of HECT E3 ligases, WW-domain containing protein 1 (WWP1) and 2 (WWP2), are responsible for the Ubiquitin-mediated degradation and downregulation of key tumour suppressor proteins and transcription factors. As such, WWP1 and WWP2 dysregulation has been directly linked to various oncogenic, cardiovascular, osteogenic and even infectious diseases. Despite their therapeutic potential, there is a limited number of small-molecule inhibitors known in the literature.

This thesis aimed to expand the inhibitor space of WWP1 and WWP2, utilising biochemical, biophysical and structural techniques to identify and develop small molecule inhibitors through a structure-activity relationship (SAR) by synthesis approach. Following the drug discovery pipeline, Hit-identification was achieved using high-throughput differential scanning fluorimetry (DSF) and in vitro autoubiquitination assays to identify small molecule ligands and inhibitors, respectively, through screening of the NCI Diversity Set VI compound library. NSC-217913 was selected as the sole hit compound demonstrating an IC_{50} of 158.3 µM towards WWP1. In the next Hit-to-Lead phase, a series of NSC-217913 analogues were generated, with compound S11 displaying an increased potency towards WWP1 (32.7 µM) and WWP2 (269.2 µM). This approach was also used to demonstrate the first in vitro WWP1 activity of the literature-derived NEDD4-1 inhibitor Indole-3-carbinol (I3C), however, its acid-condensation product 3,3′-diindolylmethane (DIM) shown to be considerably more potent (WWP1 IC_{50} of 111.2 µM). The acid-stabilised I3C derivative N-Tosyl-I3C was also shown to non-selectively inhibit both WWP1 (IC_{50} of 218.3) and WWP2 (IC_{50} of 223.7). Finally, X-ray crystallography was used to solve improved macromolecular models of WWP1 and WWP2 for subsequent use in molecular docking studies as a means to further understand the SAR of the NSC-217913 and I3C inhibitor scaffolds. This project provides the basis for lead compound development in the next phase of the drug discovery pipeline.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Table of Contents

Acknowledgements ..... **i**

Declaration ..... **ii**

Covid-19 Impact Statement ..... **iii**

Abstract ..... **iv**

Table of Contents ..... **v**

Abbreviations ..... **xiv**

List of Figures ..... **xxi**

List of Tables ..... **xxv**

1 Chapter 1: Introduction **1**

1.1 Ubiquitination **4**

1.1.1 Ubiquitin and Ubiquitin-Like 7

1.1.2 E1 Activating Enzymes 8

1.1.3 E2 Conjugating Enzymes 11

1.1.4 E3 Ligating Enzymes 14

1.1.4.1 RING E3 ligases *14*

1.1.4.2 HECT E3 ligases *18*

1.1.5 Deubiquitinating Enzymes 21

1.2 The NEDD4 Family of HECT E3 Ligases **22**

1.2.1 WWP1 and WWP2 E3 Ligases and Cancer 25

1.3 Targeting the NEDD4 Family of HECT E3 Ligases **27**

1.3.1 Small Molecule Inhibitors 27

1.4 Thesis Aims **31**

2 Chapter 2: Material and Methods **33**

2.1 DNA Techniques **34**

2.1.1 Plasmids 34

2.1.1.1 Plasmid regeneration *34*

2.1.2 Transformations 34

2.1.2.1 Generation of chemically competent cells *35*

2.1.2.2 Heat shock *35*

2.1.2.3 Electroporation *35*

2.2 Protein Purification Techniques **35**

2.2.1 Protein Expression 35

2.2.2 Cell Lysis and Clarification 36

2.2.3 Affinity Chromatography 37

2.2.3.1 His-tagged proteins *37*

2.2.3.2 GST-tagged proteins *37*

2.2.3.3 StrepII-tagged Proteins *37*

2.2.4 Dialysis and Cleavage 37

2.2.4.1 Protease removal *38*

2.2.5 Size Exclusion Chromatography 38

2.2.6 Concentration and Storage 38

2.2.7 Protein-Specific Protocols 38

2.2.7.1 His-tagged hUba1 and UbcH7 ..... *40*

2.2.7.2 His-tagged WWP2-H *40*

2.2.7.3 His-tagged WWP1-L34H *40*

2.2.7.4 His-tagged WWP1-2L34H *40*

2.2.7.5 GST-tagged WWP2-LH *41*

2.2.7.6 GST-tagged hUba1 ..... *41*

2.2.7.7 GST-tagged UbcH7 ..... *41*

2.2.7.8 StrepII-tagged WWP2-H *41*

2.2.8 SDS-PAGE Analysis 42

2.2.8.1 Lab-cast acrylamide gels *42*

2.2.8.2 Bolt™ pre-cast acrylamide gels *42*

2.3 Biochemical and Biophysical Techniques 43

2.3.1 In Vitro Autoubiquitination Assays 43

2.3.1.1 Glutathione coated 96-well plates *43*

2.3.1.2 GST-immobilised WWP2 autoubiquitination assay *43*

2.3.1.3 His-immobilised WWP1 autoubiquitination assay *44*

2.3.1.4 Uba1 counter-assay *44*

2.3.1.5 UbcH7 counter-assay *44*

2.3.1.6 Plate reading and data processing *45*

2.3.2 Differential Scanning Fluorimetry 45

2.3.2.1 Ligand-binding analysis *45*

2.3.2.2 Buffer stability screening *45*

2.3.3 Isothermal Titration Calorimetry 46

2.3.4 X-ray Crystallography 48

2.3.4.1 Protein crystallisation *48*

2.3.4.2 Co-crystallisation, commercial screening and seeding *49*

2.3.4.3 Crystal harvesting and ligand soaking *49*

2.3.4.4 X-ray data collection and processing *49*

2.3.5 In Silico Molecular Modelling 50

2.3.5.1 Cresset flare docking *50*

2.3.5.2 Schrödinger glide docking *50*

3 Chapter 3: Discovery of Novel WWP1 and WWP2 Inhibitors 51

3.1 Introduction 52

3.1.1 Biophysical Techniques 55

3.1.1.1 Differential scanning fluorimetry 55

3.1.1.2 Enzyme-linked immunosorbent autoubiquitination assay 56

3.1.2 Experimental Aims 58

3.2 Results 59

3.2.1 DSF screening of NCI Diversity Set VI against WWP1 and WWP2 59

3.2.1.1 DSF assay: construct selection and purification 59

3.2.1.2 High-throughput DSF compound screen discovered 24 compounds of interest for both WWP1 and WWP2 62

3.2.2 Auto-ubiquitination Screening of DSF Hits Against WWP1 and WWP2 63

3.2.2.1 Auto-ubiquitination assay: construct selection and purification 63

3.2.2.2 The single-shot inhibitor screen discovered four hit-compounds for WWP1 and three for WWP2. 65

3.2.2.3 Dose-response analysis of hit-compounds highlighted three compounds with a IC_{50} below 500 µM for WWP1 and WWP2. 66

3.2.2.4 NSC-217913 was the sole-hit compound after counter screening. 69

3.2.2.5 Synthesised NSC-217913 demonstrated reduced inhibition for both WWP1 and WWP2 69

3.2.3 PAINS Analysis 71

3.2.3.1 In silico PAINS screening was used to remove problematic compounds 71

3.2.3.2 NSC-2805 was found to be a PAINS compound 73

3.3 Discussion 76

3.3.1 Justification for Construct and Assay Choice 76

3.3.2 High-throughput Screen Analysis 78

3.3.2.1 Literature analysis of hit compounds highlighted possible aggregation properties of NSC-73735 79

3.3.3 An assay-specific PAINS 80

4 Chapter 4: The Development of WWP1 and WWP2 Inhibitors 82

4.1 Introduction 83

4.1.1 Indole-3-carbinol as a NEDD4 Family Inhibitor 84

4.1.2 Biophysical Techniques 86

4.1.2.1 Isothermal titration calorimetry 86

4.1.3 Experimental Aims 87

4.2 Results 89

4.2.1 NSC-217913 and I3C Analogue Generation 89

4.2.2 DSF Screening of NSC-217913 and I3C Analogues Against WWP1 and WWP291

4.2.2.1 DSF assay: construct selection and purification 91

4.2.2.2 DSF screen of NSC-217913 and I3C analogues discovered various stabilising and destabilising compounds 92

4.2.3 Screening of NSC-217913 and I3C Analogues Against WWP1 and WWP2 Autoubiquitination Assays 95

4.2.3.1 Single-shot auto-ubiquitination assay highlighted various inhibitory compounds from both NSC-217913 and I3C analogues 95

4.2.3.2 Dose-response screening discovered analogues from both the NSC-217913 and I3C libraries had increased potency for WWP1 and WWP2. 98

4.2.3.3 Counter-assay analysis of hit analogues showed the most potent NSC-217913 and I3C analogues were acting off-target 103

4.2.4 Efforts Towards ITC Ligand Characterisation 104

4.2.4.1 WWP2 HECT purification and optimisation required the use of a StrepII-tag 104

4.2.4.2 ITC optimisation efforts did not achieve a positive result 107

4.3 Discussion 109

4.3.1 SAR analysis of NSC-217913 109

4.3.1.1 A dichlorinated heterocycle was required for high potency 111

4.3.1.2 A longer ethyl ester chain improved potency *112*

4.3.2 SAR analysis of I3C 113

4.3.2.1 DIM is a promising candidate displaying improved potency towards WWP1 114

4.3.2.2 N-Toysl-I3C had similar potency for both WWP1 and WWP2 ..... *115*

4.3.3 ITC Optimisation and the Problems Encountered 116

5 Chapter 5: Structural Analysis of WWP1 and WWP2 Inhibitor-Bound States..... 118

5.1 Introduction..... 119

5.1.1 Biophysical Techniques..... 124

5.1.1.1 X-ray crystallisation *124*

5.1.1.2 In silico molecular docking..... *126*

5.1.2 Experimental Aims 127

5.2 Results 128

5.2.1 WWP1 Crystallisation Efforts 128

5.2.1.1 Protein purification optimised using buffer screen *128*

5.2.1.2 Crystals of WWP1-L34H failed to diffract at high resolution. *132*

5.2.1.3 Novel C1 2 1 space group of WWP1-2L34H enables modelling of all WW domains. 135

5.2.2 WWP2 Crystallisation Efforts 138

5.2.2.1 Hanging drop crystallisation of WWP2-LH improved yield and quality.... 138

5.2.2.2 Ligand-soaking efforts of WWP2-LH were unsuccessful..... *142*

5.2.2.3 Co-crystallisation efforts of WWP2-LH were unsuccessful..... *145*

5.2.3 In Silico Molecular Docking against WWP1-2L34H and WWP2-LH 147

5.2.3.1 NSC-217913 analogues modelled into the active site pocket displayed a conserved heterocyclic orientation. *147*

5.2.3.2 Indole-3-carbinol derivatives modelled into the Ub exo site displayed improved interactions compared to I3C..... *152*

5.3 Discussion 156

5.3.1 Crystallisation efforts 156

5.3.1.1 Justification for WWP1 and WWP2 construct choice and crystallisation strategy. 156

5.3.1.2 Low-resolution WWP1-2L34H crystals may have been due to poor cryoprotection. 159

5.3.1.3 Ambiguous unoccupied density led to over-interpretation of the WWP2-LH ligand-soaked crystals. 160

5.3.2 In-silico Modelling Efforts 161

5.3.2.1 Justification for modelling programs and strategies used. 161

5.3.2.2 The active site pocket may indeed be NSC-217913's binding site. 162

5.3.2.3 The WW2 domain appears to be important for DIM and N-Tosyl-I3C interactions at the Ub exo site. 163

6 Chapter 6: Discussion 165

6.1 The Hit-identification Phase 166

6.1.1 The High-throughput Screening Strategy 166

6.1.1.1 Future high-throughput strategy improvements 166

6.1.2 Other Potential Small-molecule Inhibitors 167

6.1.3 Consideration of Proteolysis Targeting Chimeras 168

6.1.3.1 Possible ligands are available for PROTAC development 169

6.2 The Hit-to-Lead Phase 172

6.2.1 In Vitro and Silico SAR of NSC-2171913 172

6.2.1.1 Substrate and toxicity assays should be considered 173

6.2.2 In Vitro and Silico SAR of I3C and Stable Derivatives 173

6.2.2.1 Detergent sensitivity assays are required 174

6.2.3 The Structural Studies of Small Molecule Binding Modes 175

6.2.3.1 X-ray crystallographic studies were conducted too early 175

6.2.3.2 Mutational studies of the proposed binding sites for validation 175

6.2.3.3 Is the similarity of E3 ligase HECT domains a cause for concern? 176

6.3 Future Considerations 177

6.3.1 Are Covalent HECT Inhibitors a Better Strategy? 177

6.3.2 Future Therapeutic Applications 178

7 Appendix 179

7.1 STD-NMR of I3C against WWP1 and WWP2 180

7.2 DSF Trace of N-Tosyl-I3C against WWP2 181

7.3 Chemical Characterisation 182

7.3.1 NSC-217913 Hit Analogues 183

7.3.1.1 ***Compound S6/NSC-217913*** *– Ethyl 2-({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)-acetate 183*

7.3.1.2 ***Compound S7*** *– 2-({5,6-Dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)acetonitrile 186*

7.3.1.3 ***Compound S9*** *– 2-({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)acetylmorpholine 189*

7.3.1.4 ***Compound S11*** *– Ethyl 2-({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)butyrate 192*

7.3.1.5 ***Compound S13*** *– 3-({5,6-Dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)propionic acid 195*

7.3.1.6 ***Compound S15*** *– ({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)acetic acid 198*

7.3.2 I3C Hit Derivatives 201

7.3.2.1 ***DIM*** *– 3,3'-diindolylmethane 201*

7.3.2.2 ***Compound I13/1-Benzyl-I3C*** *– N-benzylindole-3-carbinol 203*

7.3.2.3 ***Compound I15/16*** *General Procedure 1: PTC Sulfonamide Formation ... 205*

7.3.2.4 ***Compound I15/16*** *General Procedure 2: N-(aryl)indole-3-carboxaldehyde Reduction 205*

7.3.2.5 ***Compound I15/N-Tosyl-I3C*** *– N-(4-methylbenzenesulfonyl)indole-3-carbinol 206*

7.3.2.6 ***Compound I16*** *– N-(4-trifluoromethylbenzenesulfonyl)indole-3-carbinol 208*

References 211

Abbreviations

AAD	– Active adenylation domain
ABPP	– Activity-based protein profiling
ADME	– Absorption, distribution, metabolism, and excretion
AIP2	– Atropin-1-interacting protein 2
AIP5	– Atropin-1-interacting protein 5
AML	– Acute myeloid leukaemia
Amp	– Ampicillin
APF-1	– ATP-dependent proteolysis factor 1
ATG8	– Autophagy-related 8
ATG12	– Autophagy-related 12
ARLD	– Armadillo domain
ATP	– adenosine triphosphate
BH3	– Bcl-2 homology 3 domain
Br-DIM	– Bis(5-bromo-1 H-indol-3-yl)methane
BRET	– Bioluminescence resonance energy transfer
BTK	– Bruton's tyrosine kinase
C2	– Calcium-binding domain
Cam	– Chloramphenicol
CD	– Circular dichroism
CI	– Confidence interval
Clog P	– Calculated n-octanol-water partition coefficient
CV	– Column volume
DIM	– 3,3′-diindolylmethane

DMSO	– Dimethyl sulfoxide
DNA	– Deoxyribonucleic Acid
DOS	– Diversity-oriented synthesis
DSF	– Differential scanning fluorimetry
DTT	– Dithiothreitol
DUB	– Deubiquitinating enzyme
DUBTAC	– Deubiquitinase-targeting chimaeras
E1	– Ubiquitin-activating enzyme
E2	– Ubiquitin-conjugating enzyme
E3	– Ubiquitin-ligating enzyme
EDTA	– Ethylenediaminetetraacetic acid
EGF	– Epithelial growth factor
ELISA	– Enzyme-linked immunosorbent assay
EM	– Electron microscopy
ErbB4	– Erb-B2 Receptor Tyrosine Kinase 4
ESI	– Electrospray ionization
F2	– Thrombin
FAT10	– F-associated transcript 10
FBDD	– Fragment-based drug discovery
FCCH	– First catalytic cysteine half
FDA	– Food and drug administration
FMDV	– foot-and-mouth disease virus 3C protease
GST	– Glutathione S-transferase
HDX	– Hydrogen/deuterium exchange
HECT	– Homologous to E6AP carboxyl terminus

HECW1	– HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1
HECW2	– HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1
HERC	– HECT and RLD domain containing E3 ubiquitin protein ligase
HPN	– Histidine, proline and asparagine
HRP	– Horseradish peroxidase
HUWE1	– HECT, UBA and WWE domain containing E3 ubiquitin protein ligase 1
I3C	– Indole-3-carbinol
IAD	– Inactive adenylation domain
IMAC	– Immobilized metal ion affinity chromatography
IPTG	– Isopropyl ß-D-1-thiogalactopyranoside
ITC	– Isothermal titration calorimetry
ITCH	– Itchy E3 ubiquitin protein ligase
ISG15	– Interferon-stimulated gene 15 kDa
JAMM	– JAB1/MPN+/MOV34 protease family
JNK1	– c-Jun N-terminal protein kinase 1
Kan	– Kanamycin
K_D	– Dissociation constant
K_{Inact}/K_I	– Covalent efficacy constant
KLF5	– Kruppel-like factor 5
LB	– Luria-Bertani
LPS	– Lipopolysaccharide
LTr-1	– 2-(indol-3-ylmethyl)-3,3′-diindolylmethane
MAPK	– Mitogen-activated protein kinase 1
MES	– 2-(N-morpholino)ethanesulfonic acid
MINDY	– Motif interacting with ubiquitin-containing novel DUB family

MJD	– Machado-Josephin family
MMS	– Multi-matrix seeding
MMT	– Malic Acid, MES and Tris
MR	– Molecular replacement
mRNA	– Messenger ribonucleic acid
MS	– Mass spectrometry
NCI	– National Cancer Institute
NEDD4	– Neuronal precursor cell-expressed developmentally downregulated 4
NEDD4-1	– NEDD4 protein 1
NEDD4-2	– NEDD4 protein 2
NEDD8	– Neuronal precursor cell-expressed developmentally downregulated 8
NMR	– Nuclear magnetic resonance
NSC	– National Service Centre
OTU	– Ovarian tumour family
PAGE	– Polyacrylamide gel electrophoresis
PAINS	– Pan-assay interference compounds
PBS	– Phosphate-buffered saline
PBST	– PBS containing 0.1% Tween
PDB	– Protein data bank
PEG	– Polyethylene glycol
PI3K	– Phosphatidylinositol 3-kinase
PMSF	– Phenylmethylsulfonyl fluoride
PPARγ	– Peroxisome proliferator-activated gamma receptor
*p*Ph-DIM	– 3,3’-(p-substituted phenyl)bis(1 H-indole)
PROTAC	– Proteolysis targeting chimaeras

PTEN	– Phosphatase and tensin homolog
PTM	– Post-translational modification
QSAR	– Quantitative SAR
RA	– Relative activity
RBR	– RING between RING
RBX1	– RING box protein 1
RCC1	– Regulator of Chromosome Condensation 1
RHOB	– RAS homolog family member B
RING	– Really interesting new gene
RLD	– RCC1-like domain
RNF11	– Ring finger protein 11
RNF213	– Ring finger protein 213
ROS	– Reactive oxygen species
SAE1	– SUMO-activating enzyme subunit 1
SAR	– Structure-activity relationship
SCCH	– Second catalytic cysteine half
SCF	– SKP, Cullin, F-box containing complex
SDS	– Sodium dodecyl-sulfate
SEC	– Size exclusion chromatography
SKP1	– S-phase kinase-associated protein 1
Smad2	– SMA and MAD family member 2
Smad7	– SMA and MAD family member 7
SMURF1	– SMAD-specific E3 ubiquitin protein ligase 1
SMURF2	– SMAD-specific E3 ubiquitin protein ligase 2
SOC	– Super optimal broth with catabolite repression

SPR	– Surface plasmon resonance
SPRY	– Spla and the ryanodine receptor domain
SSMPB	– Sulphosuccinimidyl 4-(p-maleimidophenyl)butyrate
STD	– Saturation transfer difference
SUMO	– Small ubiquitin-like modifier
TCEP	– Tris(2-carboxyethyl)phosphine
TGFβ	– Transforming growth factor beta
TIUL1	– TGIF-interacting ubiquitin ligase 1
Tm	– Midpoint melting temperature
Tm B	– Boltzmann derived Tm
Tm D	– First derivative Tm
TMB	– 3,3′, 5,5′ tetramethylbenzidine dihydrochloride
TR-FRET	– Time-resolved fluorescence resonance energy transfer assay
tRNA	– Transfer ribonucleic acid
TSA	– Thermal shift assay
TUBE	– Tandem ubiquitin-binding entities assay
TβR1	– TGFβ receptor type 1
Ub	– Ubiquitin
UBA	– Ubiquitin-associated domain
Uba1	– Ubiquitin-like modifier activating enzyme 1
Uba6	– Ubiquitin-like modifier activating enzyme 6
UBC	– Ubiquitin-conjugating catalytic domain
UBD	– Ubiquitin-binding domain
UBL	– Ubiquitin-like protein
UCH	– Ubiquitin C-terminal hydrolase family

UEV	– Ubiquitin E2 variant
UFD	– Ubiquitin-fold domain
UFM1	– Ubiquitin-fold modifier 1
URM1	– Ubiquitin-related modifier 1
URT	– Ubiquitin reference technique
UIM	– Ubiquitin-interacting motif
USP	– Ubiquitin-specific protease family
VHL	– Von Hippel-Lindau
WD40	– Beta-transducin repeat domain
WW	– Tryptophan-tryptophan domain
WWE	– Tryptophan-tryptophan-glutamate domain
WWP1	– WW-domain containing protein 1
WWP1-2L34H	– WW2-2,3-Linker-WW3-WW4-HECT WWP1 construct
WWP1-H	– HECT-only WWP1 construct
WWP1-L34H	– 2,3-Linker-WW3-WW4-HECT WWP1 construct
WWP2	– WW-domain containing protein 2
WWP2-FL	– Full-length WWP2
WWP2-H	– HECT-only WWP2 construct
WWP2-LH	– WW2-2,3-Linker-HECT WWP2 construct
ZnF-UBP	– Zinc finger ubiquitin-specific protease domain
ZUP1	– Zinc-finger containing ubiquitin peptidase family
3C	– PreScisson HRV-3C Protease
4HB	– Four-helix bundle

List of Figures

Figure 1-1: A schematic diagram of the drug discovery pipeline 3

Figure 1-2: A simplified overview of the ubiquitination system 5

Figure 1-3: Structures of ubiquitin and ubiquitin-like protein SUMO-1 8

Figure 1-4: Structures of Uba1 and Uba1 in complex with dual ubiquitin 10

Figure 1-5: Schematic diagram of the E2 conjugating enzyme classes 11

Figure 1-6: Structures of UbcH7b and UbcH5b in complex with ubiquitin 13

Figure 1-7: Structure of the RING finger domain 16

Figure 1-8: Structures of RING E3 ligases in complex with E2 and adaptor proteins..... 17

Figure 1-9: Domain architecture of the HECT-E3 ligase subfamilies 18

Figure 1-10: Structures of HECT domain movements and interactions 20

Figure 1-11: Schematic diagram of the NEDD4 family HECT autoinhibitory mechanisms..... 23

Figure 1-12: Schematic diagram of the E3 ligase processivity vs distributive mechanism 24

Figure 1-13: Chemical structure of Heclin and Indole-3-carbinol based inhibitors..... 29

Figure 1-14: Structure of compound 1 covalently attached to non-catalytic Cys627 of NEDD4-1 29

Figure 1-15: Chemical structure of other NEDD4 family inhibitors 31

Figure 1-16: Flow chart of thesis chapters and project strategy..... 32

Figure 3-1: Examples of Pan-assay interference compound (PAINS)..... 53

Figure 3-2: The basic principles of DSF 56

Figure 3-3: The WWP2 ELISA autoubiquitination assay..... 57

Figure 3-4: Domain architecture of the WWP ligase constructs used in Chapter 3..... 58

Figure 3-5: SDS-PAGE analysis of WWP1-L34H purification..... 60

Figure 3-6: SDS-PAGE analysis of WWP2-LH purification..... 61

Figure 3-7: High throughput DSF screen of the NCI Diversity Set VI compound library against WWP1 and WWP2 62

Figure 3-8: SDS-PAGE analysis of Uba1 and UbcH7 purifications 64

Figure 3-9: Auto-ubiquitination assay screen of PAINS-filtered DSF hit compounds against WWP1 (A) and WWP2 (B) 65

Figure 3-10: Dose-dependent auto-ubiquitination analysis of WWP1 and WWP2 against their respective single-shot hit compounds 67

Figure 3-11: Dose-dependent cross-selectivity analysis of WWP1 and WWP2 hit-compounds.68

Figure 3-12: Dose-dependent auto-ubiquitination assay of synthesised NSC-217913 against WWP1 and WWP2 70

Figure 3-13: PAINS analysis of NSC-2805 74

Figure 3-14: Structure of NSC-73735 (Redoxal) and its associated aggregate 80

Figure 3-15: Radical oxidation schemes of both Hydroquinone (A) and TMB (B) 81

Figure 4-1: Metabolic formation of I3C and DIM from Glucobrassicin 85

Figure 4-2: The basic principles of isothermal titration calorimetry 87

Figure 4-3: Domain architecture of the WWP ligase constructs used in Chapter 4 88

Figure 4-4: Structures of NSC-217913 (S) analogues under investigation with associated numbering 89

Figure 4-5: Structures of I3C (I) derivatives under investigation with associated numbering.... 90

Figure 4-6: SDS-PAGE analysis of WWP1-2L34H purification 91

Figure 4-7: DSF screen of NSC-217913 and its analogues against WWP1-2L34H (top) and WWP2-LH (bottom) 93

Figure 4-8: DSF screen of I3C, DIM and its derivatives against WWP1-2L34H (top) and WWP2-LH (bottom) 94

Figure 4-9: Single-shot Autoubiquitination screen of 27 NSC-217913 analogues against WWP1-L34H (top) and WWP2-FL (bottom) 96

Figure 4-10: Single-shot Autoubiquitination screen of I3C derivatives against WWP1 (top) and WWP2 (bottom) 97

Figure 4-11: Dose-dependent auto-ubiquitination assay of single-shot hit NSC-217913 analogues against WWP1-L34H 99

Figure 4-12: Dose-dependent auto-ubiquitination assay of single-shot hit NSC-217913 analogues against WWP2-FL 100

Figure 4-13: Dose-dependent auto-ubiquitination assay of single-shot hit I3C derivatives against WWP1-L34H..... 101

Figure 4-14: Dose-dependent auto-ubiquitination assay of single-shot hit I3C derivatives against WWP2-FL 102

Figure 4-15: SDS-PAGE analysis of His-tagged WWP2-H purification..... 104

Figure 4-16: SDS-PAGE analysis of StrepII-tagged WWP2-H purification. 106

Figure 4-17: Raw ITC traces of NSC-288387 against WWP2-H 107

Figure 4-18: Raw ITC traces of NSC-217913, I3C and its derivatives against WWP2-H 108

Figure 4-19: Structure of NSC-217913 analogue S11, with highlighted SAR features..... 112

Figure 4-20: Structures of DIM derivatives with therapeutic potential 115

Figure 4-21: Structure of OSU-A9..... 116

Figure 5-1: Potential sites of NEDD4 family HECT inhibition. 121

Figure 5-2: Structures of HECT-inhibitor/activator complexes in WWP1 and NEDD4-1 123

Figure 5-3: Crystal phase diagram mediated by a precipitant..... 124

Figure 5-4: The workflow of X-ray crystallography..... 126

Figure 5-5: Domain architecture of the WWP ligase constructs used in Chapter 5..... 127

Figure 5-6: DSF commercial buffer screens against His-tagged WWP1-2L34H 130

Figure 5-7: DSF custom buffer screen against His-tagged WWP1-2L34H..... 131

Figure 5-8: Crystal morphologies of WWP1-L34H..... 133

Figure 5-9: Crystal morphologies of WWP1-2L34H..... 135

Figure 5-10: Crystal structure of WWP1-2L34H (PDB ID: 9EQK)..... 136

Figure 5-11: Crystal morphologies of WWP2-LH..... 138

Figure 5-12: WWP2-LH crystal formation rates using hanging (blue) and sitting drop (orange) vapour diffusion..... 139

Figure 5-13: Crystal structure of WWP2-LH (PDB ID: 9EQH) 140

Figure 5-14: Structural changes between apo and ligand-soaked WWP2-LH crystals 143

Figure 5-15: Ligand poses of NSC-217913 analogues 6, 7, 9, 11, 13 and 15 within WWP1-2L34H 149

Figure 5-16: Ligand poses of NSC-217913 analogues 6, 7, 9, 11, 13 and 15 within WWP2-LH 150

Figure 5-17: Ligand poses of I3C, DIM and N-Tosyl-I3C with WWP1 153

Figure 5-18: Ligand poses of I3C, DIM and N-Tosyl-I3C with WWP2 154

Figure 5-19: WWP1-2L34H model comparison between PDB ID: 9EQK (left) and PDB ID: 6J1X (right) 158

Figure 5-20: Protein distribution within hanging and sitting drop vapour diffusion 159

Figure 6-1: Chemical structures of NSC-13151 and NSC-57103 168

Figure 6-2: Schematic diagram of the small-molecule PROTAC approach 169

Figure 6-3: Chemical structures of NSC-30260, NSC-107582, NSC-332670 and NSC-136513. 170

Figure 6-4: Structures of NSC-522131, HS-152 and other structurally similar DSF hits 171

Figure 6-5: SAR summary of NSC-217913 analogue S11 173

Figure 7-1. STD NMR binding epitope maps of I3C against WWP1 (top) and WWP2 (bottom) based on the normalized saturation transfer intensities (0-100%) 180

Figure 7-2. STD NMR binding experiments for Indole-3-carbinol (I3C) in complex with WWP1 and WWP2 181

Figure 7-3. DSF trace of N-Tosyl-I3C against WWP2-LH 181

List of Tables

Table 2-1: Plasmids, constructs, and their origins..... 34

Table 2-2: *E. coli* cell-lines and protein expression conditions..... 36

Table 2-3: Proteins and their respective buffers..... 39

Table 2-4: Lab-cast acrylamide gel recipes..... 42

Table 2-5: The His-tagged WWP1-2L34H DSF custom buffer screen. 47

Table 2-6: Protein crystallisation conditions..... 48

Table 3-1: Summary of potential inhibitors for WWP1 and WWP2 66

Table 3-2: Counter auto-ubiquitination screen of hit NCI Div VI compounds against UbcH7.. 69

Table 3-3. Summary of DSF hit compound PAINS filtering against WWP1..... 71

Table 3-4. Summary of DSF hit compound PAINS filtering against WWP2..... 72

Table 3-5. Aggregate risk rankings and associated descriptions from Aggregate Advisor. 72

Table 3-6: Summary of NSC-2805 analogues and their associated WWP2 IC_{50} values..... 75

Table 4-1: Uba1 and UbcH7 counter auto-ubiquitination screen of NSC-217913 analogues .. 103

Table 4-2: UbcH7 counter auto-ubiquitination screen of hit I3C (I) derivatives. 103

Table 4-3: Summary of NSC-217913 and analogue hits from DSF and autoubiquitination assay screening..... 111

Table 4-4: Summary of I3C and derivative hits from DSF and autoubiquitination assay screening..... 113

Table 5-1: NEDD4 family of E3 ligases solved by X-ray crystallography..... 120

Table 5-2: Summary of data collection statistics for WWP1-L34H. 134

Table 5-3: Summary of data collection and model statistics for WWP1-2L34H..... 137

Table 5-4: Summary of data collection and model statistics for WWP2-LH..... 141

Table 5-5: Summary of data collection and model statistics for WWP2-LH (unbound) I3C soak. 144

Table 5-6: Summary of data collection statistics for co-crystalised WWP2-LH with I3C (unbound). 146

Table 5-7: Cresset scoring functions and estimated binding free energies of hit NSC-217913 analogues against WWP1-2L34H and WWP2-LH. 148

Table 5-8: Glide scoring functions of hit I3C derivatives against WWP1-2L34H (PDB ID: 9EQK) and WWP2-LH (PDB ID: 6J1Z)..... 152

Chapter 1: Introduction

The advancement of modern medicine over the last century has revolutionised life as we know it, with some seemingly terminal diseases diminished to little more than a hiccup. Outside innovations in practical medicine, the uptake and development of the drug discovery pipeline is in part responsible for this success, designed to increase discovery speed whilst reducing costs. Despite this, the cost to develop a new drug can range from £125 million to £3.5 billion with an approval success rate between 10 – 20% once reaching clinical trials (Schlander *et al*., 2021; Yamaguchi, Kaneko and Narukawa, 2021). This pipeline can be segmented into the pre-discovery, drug-discovery, pre-clinical and clinical stages (Figure 1-1) (Hughes *et al.*, 2011). The pre-discovery stage involves the understanding of a disease on the molecular level to identify a treatment target. The drug discovery stage can be further separated into three phases; Hit-identification, Hit-to-Lead and Lead-optimisation. In the Hit-identification phase, high-throughput screening is undertaken to identify possible hit compounds/inhibitors of the therapeutic targets. In the Hit-to-Lead phase, the most promising hit compounds are selected with small steps taken to improve potency and selectivity whilst gaining an understanding of their chemical space. In the final Lead-optimisation phase, the lead compounds are optimised to maximise their potency as well as other features such as their absorption, distribution, metabolism, and excretion (ADME) properties. Lead compounds are taken through to the preclinical stage, whereby in vivo biological studies are undertaken, looking at properties such as toxicity and possible dosage. The final clinical trial stage involves patient participation and direct study of the targeted disease, separated into incremental three phases (I, II, II). Unfortunately, even after pre-discovery, this process can take up to 15 years, with the majority of lead compounds failing before and even after reaching the clinical trial phase, further promoting the high throughput nature of this pipeline (Sun *et al*., 2022).

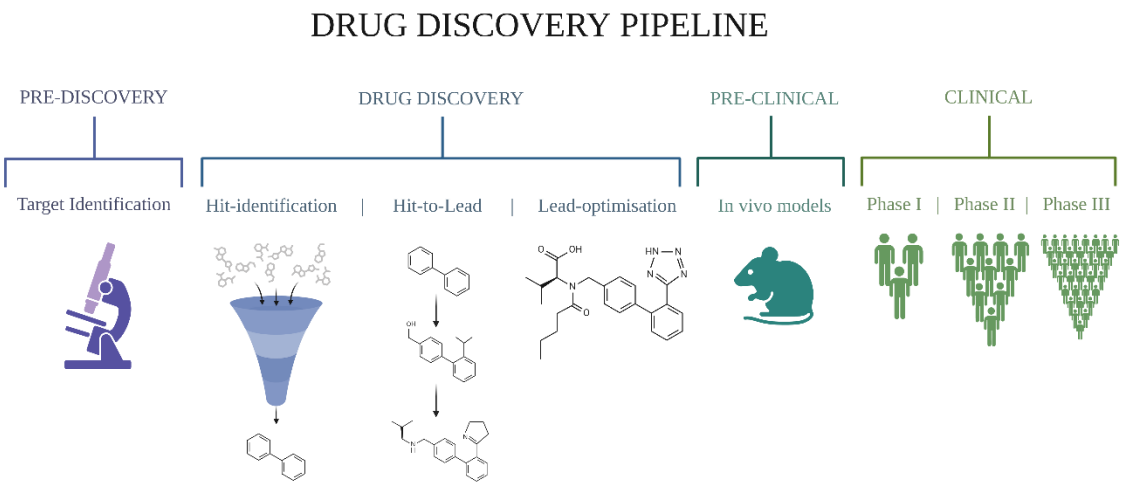


Figure 1-1: A schematic diagram of the drug discovery pipeline. Segregated into Pre-discovery (target identification), Drug discovery (Hit-identification, Hit-to-Lead and Lead-optimisation), Pre-clinical (In Vivo models) and Clinical (Phase I, II and III). Image created using BioRender.

The pre-discovery and early drug-discovery phases are the most applicable to academia, with the later stages taken up by the commercial sector. Disease can occur at any stage from the decoding of genetic material through to a biological function, termed the central dogma of biology. This process first involves the decoding of DNA into mRNA, called transcription. This mRNA is then further decoded into polypeptide chains using complementary tRNAs containing the sequence-specific amino acids, termed translation. These polypeptide chains are folded into their desired tertiary structures called proteins. Proteins are the essence of biological function and operate as molecular machines able to perform structural or catalytic functions classified as enzymes. Proteins are further processed through post-translational modifications, such as cleavages, phosphorylation and even the addition of other smaller polypeptides such as ubiquitin. Ubiquitination is an important feature involved in protein regulation and turnover, both through proteolytic and non-proteolytic functions (Callis, 2014). The two enzymes, WW-domain containing protein 1 (WWP1) and 2 (WWP2) are responsible for the ubiquitination of key tumour suppressor proteins, with their dysregulation linked to various diseases, notably prostate and breast cancer (Behera and Reddy, 2023; You *et al.*, 2024). However, there are only a limited number of small molecule inhibitors shown to target them, all with low potency.

This chapter highlights the pre-discovery stage of the WWP1 and WWP2 drug discovery process, introducing the process of protein ubiquitination at the molecular level, before finally examining both WWP1 and WWP2 involvement in disease and current therapeutic developments.

1.1 Ubiquitination

Ubiquitination was originally discovered as a post-translational modification (PTM) associated with protein degradation and was awarded the 2004 Nobel Prize in Chemistry shared between Aaron Ciechanover, Avram Hershko and Irwin Rose. Despite this renowned characteristic, it is now known that ubiquitination is a widespread regulatory mechanism conserved across eukaryotic organisms capable of defining the fate of all proteins beyond just that of simple degradation (Ciechanover, 2015). The process of ubiquitination can be defined by the covalent attachment of a 76 amino acid polypeptide to a target of interest. This small polypeptide, called ubiquitin (Ub) is loaded via a catalytic cascade of three distinct steps using Ub-activating (E1), Ub-conjugating (E2) and Ub-ligating (E3) enzymes (Komander, 2009; Ye and Rape, 2009; Callis, 2014). In the first 'activation' step, one of two E1 enzymes initially forms a high-energy thioester bond between its active site cysteine and the Ub glycine-76 residue, using an adenosine triphosphate (ATP) driven reaction. In the following 'conjugation' step, the activated Ub molecule is then transferred to one of ~40 E2 enzymes, forming a similar thioester bond. The final 'ligation' step utilises one of over 600 E3 enzymes to facilitate the Ub transfer from E2 to a target substrate. Canonically, an iso-peptide bond is formed between the Ub C-terminal carboxyl of glycine-76 and the ε-NH_2 group of the lysine residues located on a protein target (McDowell and Philpott, 2013). However, atypical ubiquitination has also been observed, including on the N-terminal residue via peptide formation (Ciechanover and Ben-Saadon, 2004), serine/threonine residues via hydroxyl ester formation (Shimizu, Okuda-Shimizu and Hendershot, 2010), and cysteine residues via thioester formation (Tait *et al*., 2007).

Although single monoubiquitylations are abundant, cascade cycles can form polyubiquitin chains of four or more Ub molecules either added sequentially to the target substrate or via the 'en bloc' method (Behrends and Harper, 2011) (Figure 1-2 top). The formation of Ub chains generally occurs using seven internal Ub lysine residues (K6, K11, K27, K29, K33, K48 and K63) as well as the N-terminal methionine (Met1), forming linear or branched chains, collectively termed the 'ubiquitin code' (Figure 1-2 bottom) (Komander and Rape, 2012). Much effort has been employed to decipher this code into discrete cellular signals, made more complex by the introduction of deubiquitinating enzymes (DUBs) able to not only remove and recycle Ub but further modify their linked chains (Clague, Urbé and Komander, 2019). Despite this, the DUB Ub chain binding specificity has made them a key deciphering tool. One of the earliest deciphers was lysine-48 (K48)-linked linear chains, now associated with downstream proteolysis mediated by the 26S-proteasome (Kwon and Ciechanover, 2017). Other Ub chain linkages have also been shown to regulate protein degradation functions independent of the 26S proteasome, as well as other non-degradation functions such as protein

trafficking and cellular signalling roles in DNA repair, programmed cell death, and the cell cycle (Kerscher, Felberbaum and Hochstrasser, 2006; Clague and Urbé, 2010).

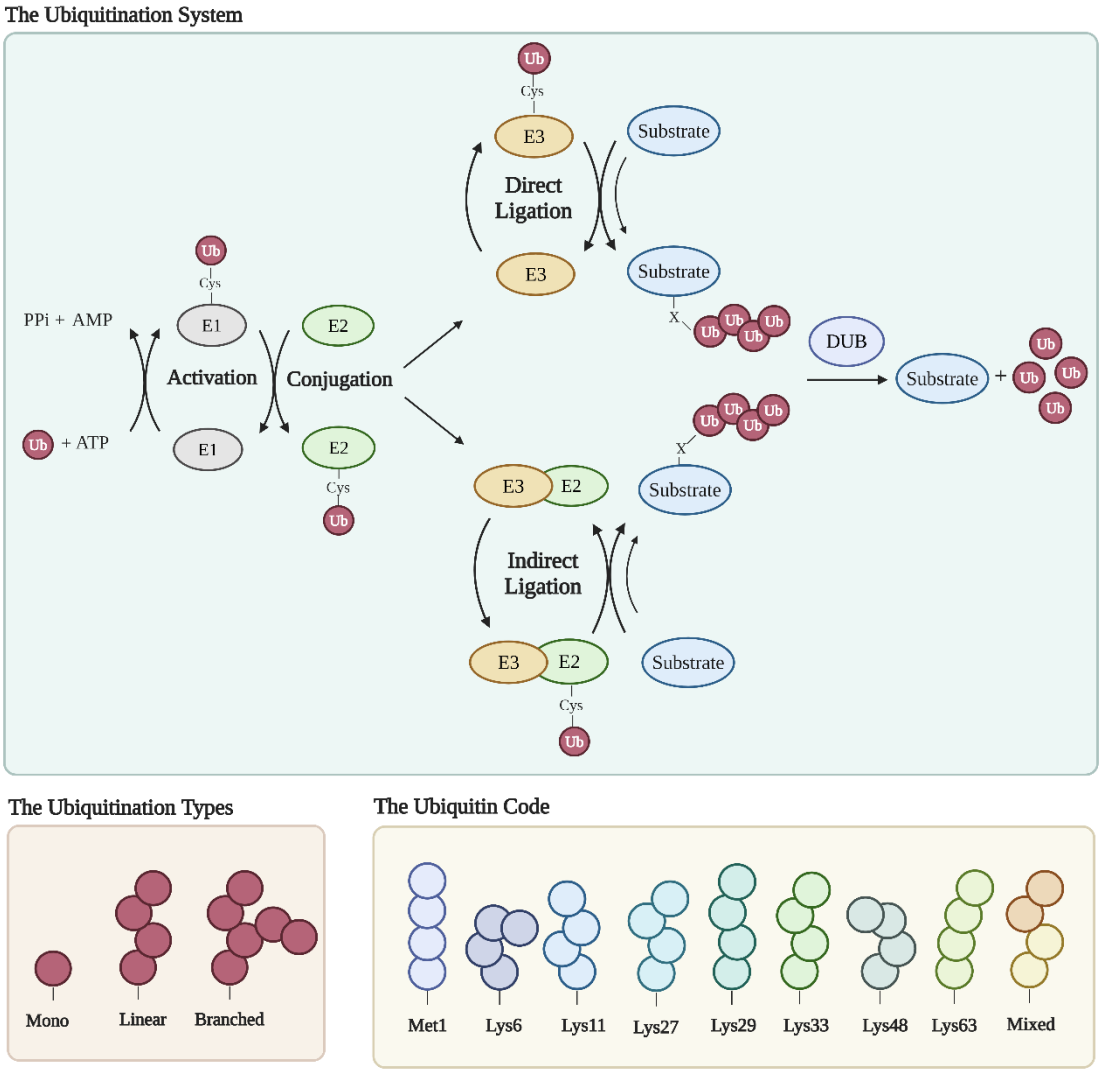


Figure 1-2: A simplified overview of the ubiquitination system. Ubiquitination occurs through three steps: Activation, Conjugation and Ligation. Ubiquitin (Ub, red) is first loaded onto the E1-activating enzyme (grey) forming a thioester bond, mediated by adenosine triphosphate (ATP). Ub is transferred to the E2-conjugating enzyme (green) forming another thioester bond. In the direct (HECT) ligation mechanism (top), Ub is first transferred to the E3 ligase enzyme (orange) before finally being transferred to the target protein (blue). In indirect (RING) ligation, the E3 ligase is used as a molecular scaffold to enable the E2:substrate transfer. The Ub is bound to the substrate via an isopeptide, peptide, hydroxyl ester or thioester bond dependent on the residue (X). Sequential cycles of this cascade can build Ub mono or poly chains (Ub types), via various covalent attachments (Ub Code). Ub can also be recycled by deubiquitinating enzymes (DUBs, purple) rescuing target proteins. The image was created using BioRender, adapted from Passmore and Barford (2004) and Damgaard (2021).

This code has recently become more complex with numerous studies also reporting alternative 'non-protein' substrates found ubiquitinated including nucleic acids (Zhu *et al*., 2024), sugars (Kelsall *et al*., 2022) and even lipids (Sakamaki *et al*., 2022). These substrates have been shown to have distinct signalling characteristics, for example, during *salmonella* invasion the E3 ubiquitin ligase ring finger protein 213 (RNF213) has been shown to ubiquitinate the bacterial lipopolysaccharide (LPS) generating a ubiquitin coat that mediates bacterial autophagy (Otten *et al*., 2021). Another level of complexity is also added when considering PTM crosstalk with the Ub system (E1, E2, E3 and DUBs) all subject to PTM regulation as well as Ub itself (Song and Luo, 2019). Ub-specific PTM's can significantly affect downstream Ub chain synthesis, whether that be through direct blocking of chain elongation in the case of K6 and K48 acetylation (Ohtake *et al*., 2015) or interfering with chain structure through phosphorylation commonly found on serine-65 (S65) preventing both E2 discharge but also chain disassembly by associated DUBs (Wauer *et al*., 2015).

The varying levels of the Ub system enable tight regulatory control and specificity, therefore it is of no surprise that its dysregulation has been associated with numerous detrimental diseases, for instance, cancer. However, such specificity also makes the Ub system an ideal target for therapeutics, with many of its control mechanisms analogous to that of other PTMs such as phosphorylation. Protein kinases have already demonstrated significant success as a drug target, and yet progress in Ub therapeutics has been limited, Why? One argument is that the reduced specificity between protein kinases is an advantage for therapeutic development despite the innate possibility of off-targets, with the same compound libraries able to be exploited across various types of kinases (Cohen and Tcherpakov, 2010). Most Ub therapeutics focus on the E3 ligase interactions with their specific targets and as such are non-transferable. Despite this, one could suggest future therapeutics should move towards having fewer off-targets to improve patient response, especially in cancer treatments where many have refused treatment due to the severity of side effects.

1.1.1 Ubiquitin and Ubiquitin-Like

Ub was first discovered in 1975, found to be conserved across all eukaryotic organisms as a small 76 amino acid polypeptide, ~8.6 kDa in size (Goldstein *et al.*, 1975). It was later characterised throughout the 1970 - 1980's as a PTM linked to proteasomal degradation, previously named ATP-dependent proteolysis factor 1 (APF-1) (Ciechanover, Hod and Hershko, 1978; Hershko, Ciechanover and Rose, 1979; Hershko *et al*., 1980). Ub is stable over a wide range of temperatures and pH, and is also resistant to digestion, attributed to its high degree of hydrogen bonding and tightly folded structure (PDB ID: 1UBQ) (Figure 1-3A). This structure consists of an α- helix and a 5 stranded β-sheet defined as the beta-grasp fold, alongside a short 3_{10} helix, as well as an exposed carboxy-terminal tail crucial for the isopeptide bond formation (Vijay-kumar, Bugg and Cook, 1987). As aforementioned, various polyubiquitin chains can form across 8 total positions (7 lysine residues, and the initial methionine), with the specific chain position and relative type resulting in numerous downstream signalling events beyond just that of degradation (Komander and Rape, 2012). Non-covalent interactions with Ub are also crucial to the complexities associated with Ub decoding, achieved through Ubiquitin-binding domains (UBDs), relating their specificity to Ub chain length and types through context-dependent multimeric interactions (Dikic, Wakatsuki and Walters, 2009). There are also several families of ubiquitin-like proteins (UBLs), involved in various similar PTMs having their distinct but smaller analogue cascades and signalling routes outside of protein degradation. These include autophagy-related 8 (ATG8), autophagy-related 12 (ATG12), neural precursor cell expressed developmentally downregulated protein 8 (NEDD8), small ubiquitin-related modifiers (SUMOs), human leukocyte antigen F adjacent transcript 10 (FAT10), interferon-stimulated gene 15 kDa (ISG15), ubiquitin-fold modifier 1 (UFM1), and ubiquitin-related modifier 1 (URM1) (Cappadocia and Lima, 2018). Interestingly, although SUMO-1 for example has less than 20% sequence identity to Ub, it is observed to have almost identical structural folds (Hay, 2005) (Figure 1-3B). As with other PTMs, Ub can also be the target of UBLs in processes such as SUMOylation resulting in regulation by allosteric inhibition (Tatham *et al*., 2011).

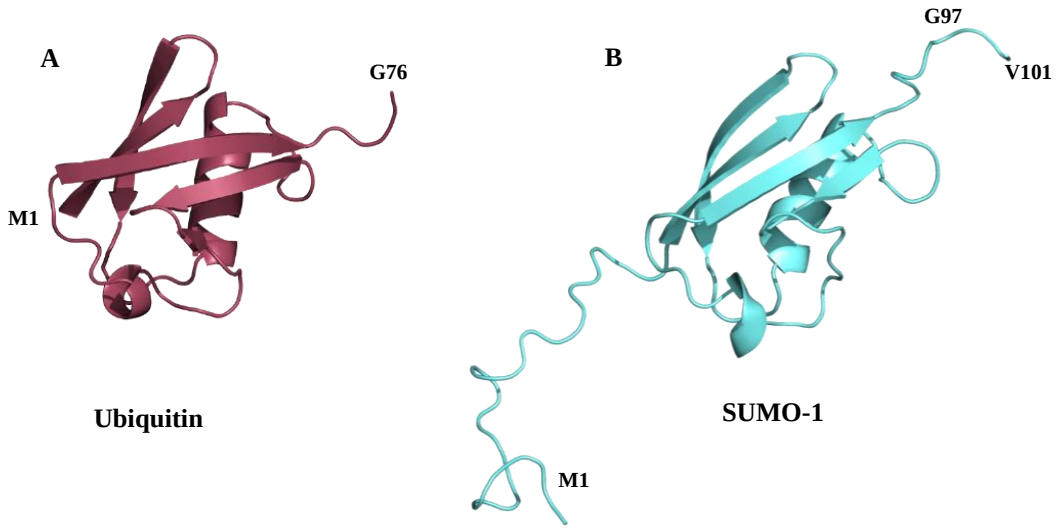


Figure 1-3: Structures of ubiquitin and ubiquitin-like protein SUMO-1. (A) Ubiquitin (red) (PDB ID: 1UBQ) was solved at 1.8 Å resolution using X-ray crystallisation (Vijay-kumar, Bugg and Cook, 1987). (B) SUMO-1 (blue) (PDB ID: 1A5R) was solved from NMR, with the best average structure represented from 10 models submitted (Bayer *et al.*, 1998). Both structures contain the characteristic ubiquitin features consisting of an α- helix, a 5-stranded β- sheet and a 3_{10} helix, as well as the exposed C-terminal tail containing the glycine-76 (G76) residue. SUMO-1 also contains a further exposed N-terminal tail, as well as a further exposed C-terminal with a terminal valine-101(V101) residue, containing the analogue glycine in position 97 (G97). Structures were generated in PyMOL as cartoon illustrations.

1.1.2 E1 Activating Enzymes

E1 enzymes are the gatekeepers of the Ub cascade vital in controlling its initiation using ATP through the formation of a Ub high-energy thioester bond. Among the 8 identified so far, only 2 are known to activate Ub, this being ubiquitin-like modifier activating enzyme 1 (Uba1) and ubiquitin-like modifier activating enzyme 6 (Uba6), with the remaining participating in the analogue UBL systems, such as SUMO-activating enzyme subunit 1 (SAE1) (Schulman and Wade Harper, 2009). Uba1 is renowned as the primary Ub system E1, with Uba6 only found expressed in vertebrates and sea urchins as well as having a dual specificity, able to activate the F-associated transcript 10 (FAT10), another UBL (Truongvan *et al.*, 2022). Uba1 can be found as two isoforms Uba1a (~118 kDa) or Uba1b (~114 kDa), with the former containing an additional nuclear locational signal sequence (Stephen *et al.*, 1997). Uba1 as with all E1 enzymes retains a highly conserved multidomain architecture consisting of an active and an inactive adenylation domain (AAD and IAD), a four-helix bundle (4HB), a C-terminal ubiquitin-fold domain (UFD) and a subdivided domain termed the first and second catalytic cysteine half's (FCCH and SCCH), with the latter containing the active site cysteine (Figure 1-4A) (Lv *et al.*, 2018). The dual catalytic activity of these E1 enzymes has been thoroughly studied and found to undergo various conformational changes referred to as active re-modelling. Initially, ATP, Mg^{2+} and Ub bind at the AAD site, before the Ub's glycine-76 is adenylated.

This causes a 100 – 130 ° rotation of the SCCH domain, transitioning the active site cysteine along with other structural elements into proximity to enable the thioester bond formation (Lv *et al.*, 2017). Upon formation, the SCCH and bound Ub rotate clear to enable another round of adenylation to a second Ub molecule, structurally represented in Figure 1-4B (Schäfer, Kuhn and Schindelin, 2014). A dual-loaded E1 has been found to improve E2 recruitment. This is achieved through the UFD domain, found to associate with a diverse range of E2 molecules and undergoes a conformational change that orientates them for thioester transfer from the SCCH domain (Olsen and Lima, 2013). Being at the apex of Ub/UBL systems, dysfunction of any E1, particularly Uba1, has detrimental effects. More recently, there has been strong links between Uba1 decreased expression and neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's disease (Groen and Gillingwater, 2015; Lambert-Smith, Saunders and Yerbury, 2020). Although Uba1 targeting has shown some success in anticancer therapeutics, these links alongside their apex role in the entire Ub system are a serious cause for concern for this type of therapeutic treatment (Barghout and Schimmer, 2021).

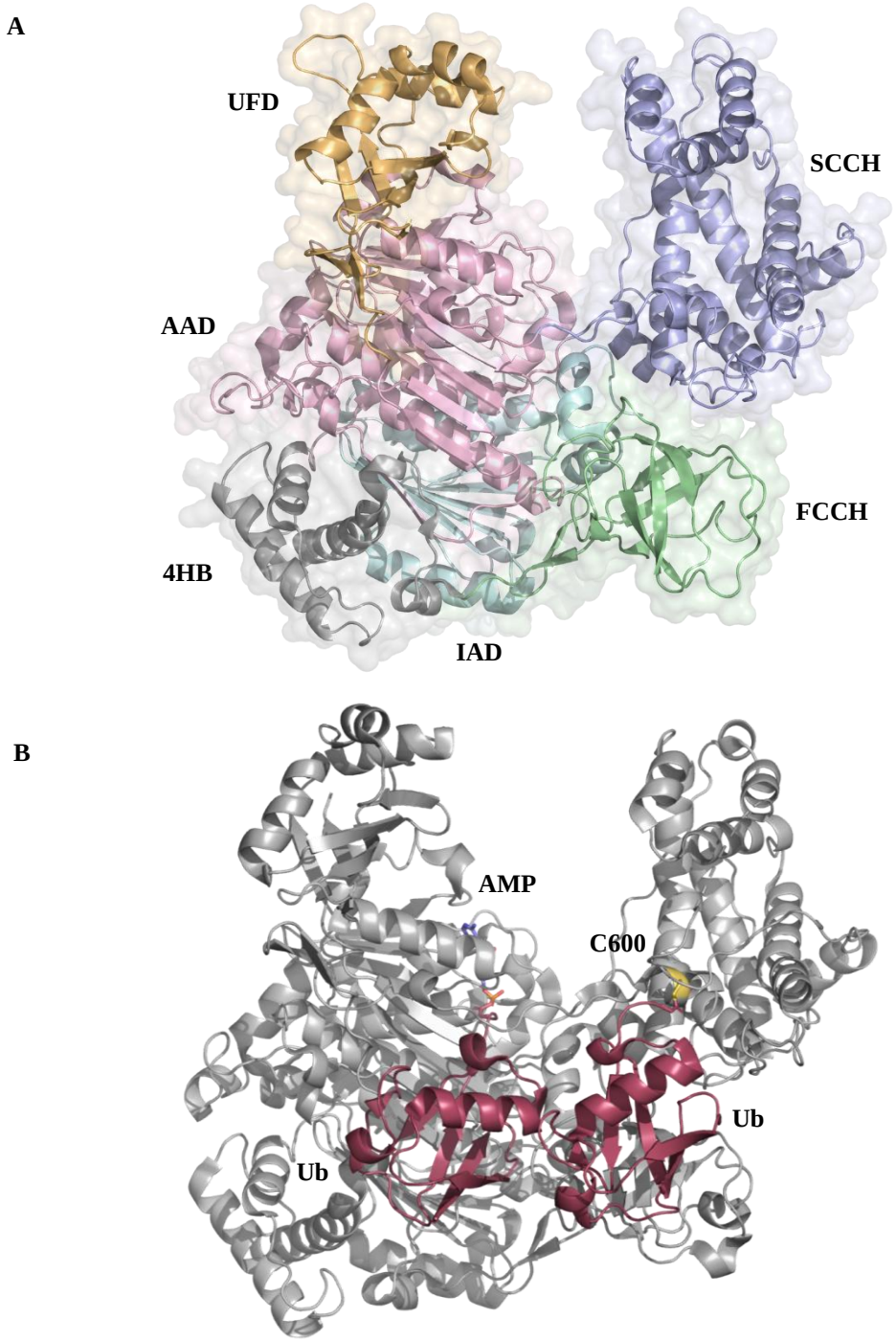


Figure 1-4: Structures of Uba1 and Uba1 in complex with dual ubiquitin. (A) The structure of Uba1 (PDB ID: 6DC6) was solved at 3.14 Å using X-ray crystallography (Lv *et al*., 2018). Domain architecture consists of an active (AAD, pink) and inactive adenylation domain (IAD, blue), a four-helix bundle (4HB, grey), a C-terminal ubiquitin-fold domain (UFD, yellow) and a subdivided domain called the first (FCCH, green) and second catalytic cysteine half (SCCH, purple). Ub bound at the adenylation site was removed. (B) The structure of Uba1 (from *S. cerevisiae*, grey) (PDB ID: 4NNJ) in complex with two molecules of Ub (red) was solved at 2.40 Å (Schäfer, Kuhn and Schindelin, 2014). The first complexed Ub-AMP (purple) is found bound at the AAD domain, with the second Ub shown bound to the active site Cys600 (yellow) via a thioester bond. Both structures are highly conserved with an RMSD = 0.975. Structures were generated in PyMOL as cartoon and surface illustrations.

1.1.3 E2 Conjugating Enzymes

E2 enzymes are responsible for the second step in the Ub cascade interacting with both E1 and E3 enzymes. In humans, there are ~40 E2 enzymes, able to have distinct interactions with numerous E3s and the final Ub substrates (Stewart *et al.*, 2016). Although E3 ligases are typically responsible for the substrate specificity (Section 1.1.4), E2 enzymes can directly influence the type of ubiquitination and therefore the fate of substrate (Wijk and Timmers, 2010). In most cases, E3s are incorporated to support the E2 Ub transfer to the target substrate rather than direct participation, with E2-Ub conjugates mostly inert without E3 interaction. However, more unique E2s such as Ube2L3 (UbcH7) are only able to transfer their bound Ub directly to HECT E3 ligases cysteines (Section 1.1.4.2) and not substrate or Ub lysines regardless of their ability to interact with both (Wenzel *et al.*, 2011). Despite their diversity, these relatively small enzymes all contain a conserved ubiquitin-conjugating catalytic (UBC) domain (14 – 16 kDa), containing the active site cysteine. This core domain consists of 4 α-helices, 4 stranded anti-parallel β-sheets and a short 3_{10} helix (Figure 1-6). The N-terminal helix forms part of the E1 binding region containing a three basic residue motif alongside an important E3 binding loop (Tokgöz *et al.*, 2012). Most UBCs contain a conserved His-Pro-Asn (HPN) tripeptide sequence crucial for isopeptide bond formation located in proximity to the active site cysteine. A minority of E2s also contain short N and/or C-terminal extensions aiding their functionality and enabling their categorisation: Class I – UBC domain, class II – N-terminal and UBC domain, class III – C-terminal and UBC domain, class IV – N-terminal, C-terminal and UBC domain (Wijk and Timmers, 2010) (Figure 1-5).

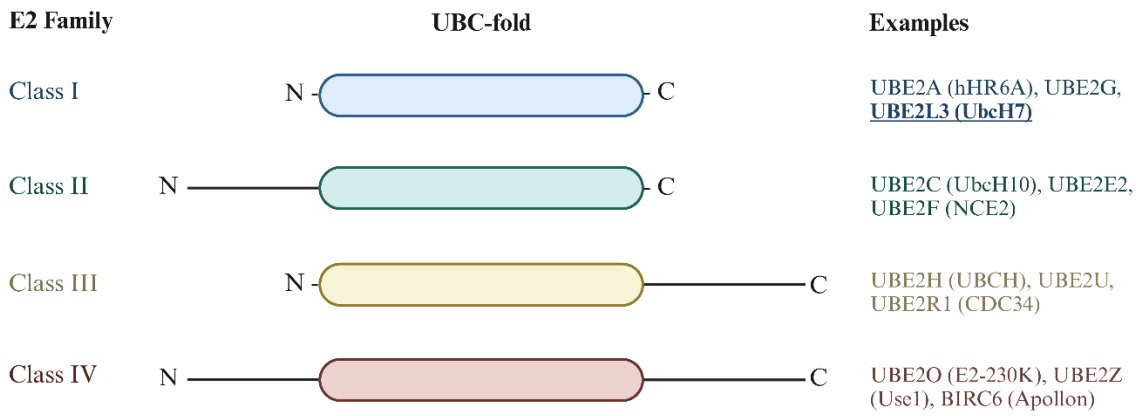


Figure 1-5: Schematic diagram of the E2 conjugating enzyme classes. The four classes are shown containing a ubiquitin-conjugating catalytic (UBC) domain, classified as either; Class I (Blue) containing just the UBC, Class II (Green) containing an N-terminal extension, Class III (Yellow) containing a C-terminal extension, or Class IV (Red) containing both C- and N-terminal extensions, with three examples shown for each. The Class I example UbcH7 has been highlighted as the E2 of choice used throughout this thesis project. The image was created using BioRender, adapted from Wijk and Timmers (2010).

With various E2s able to undergo a diverse range of intrinsic activity, dependent on lysine or cysteine Ub transfer, it is difficult to concisely discuss their structure-function relationship, further detailed elsewhere (Stewart *et al.*, 2016). However, what is apparent is the high flexibility of the E2-Ub conjugate able to transition between open and closed states and therefore sample the various orientations required for alternative E3 and substrate targets (Pruneda *et al.*, 2011). In some cases, E2's can function independently of E3's, for example, Ube2S is able to form K11-linkages through substrate-assisted catalysis dependent on the association of the donor Ub towards a Ube2S hydrophobic patch, positioning the donors glutamate-34 (E34) in proximity to the acceptors K11, lowering its pKa for isopeptide formation (Wickliffe *et al*., 2011). Further complexities exist with the introduction of ubiquitin E2 variants (UEVs), in essence, these E2s lack the catalytic cysteine residue and therefore are unable to directly conjugate with Ub. Instead, UEVs (originally identified as MMS2) typically act as further facilitators of E2 functionality through E2-UEV heterodimers (Broomfield, Chow and Xiao, 1998; Pastushok *et al.*, 2005). As targets, E2 enzymes do provide more potential, particularly due to their ability to define the fate of protein substrates. In the case of UbcH7 (HECT-specific E2), polymorphisms (single-nucleotide mutations) have been linked to numerous autoimmune diseases, cancers, and unsurprisingly Parkinson's disease, given its association with the Parkin E3 ligase (Alpi, Chaugule and Walden, 2016). Although targeting E2s may provide possible treatments, targeting the substrate-specific E3 ligase appears to provide the best potential in preventing off-targets.

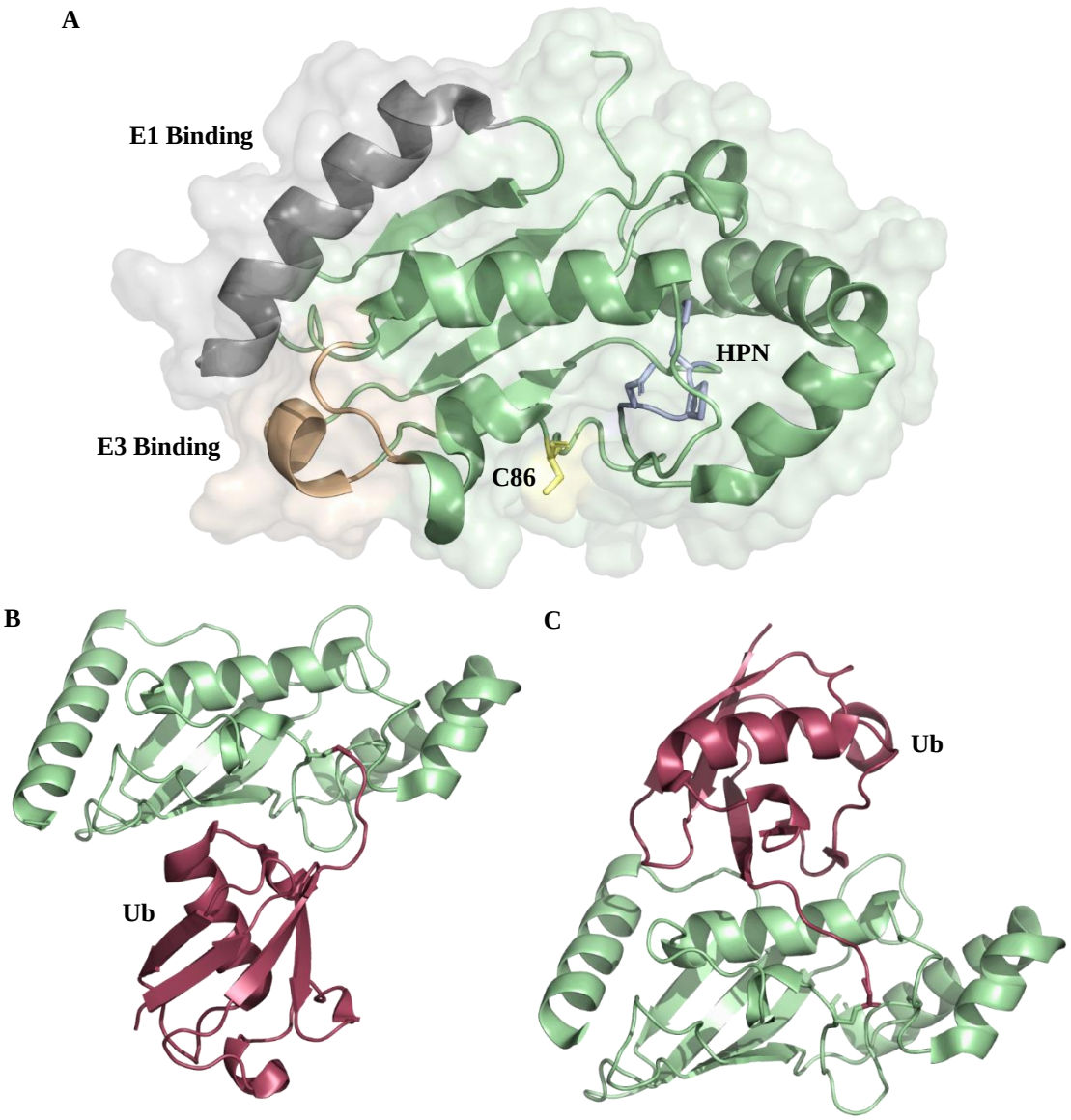


Figure 1-6: Structures of UbcH7b and UbcH5b in complex with ubiquitin. (A) The structure UbcH7b (PDB ID: 6XXU) was solved using NMR, with the best average structure represented from 21 submitted (Marousis *et al.*, 2020). The active site cysteine (C86, yellow) and HPN (purple) are shown alongside both the E1 (grey) and E3 (orange) binding regions. (B) The structure of UbcH5b (green) (PDB ID: 3A33) conjugated to Ub (red) was solved at 2.20 Å using X-ray crystallography (Sakata *et al.*, 2010). (C) The structure of UbcH5b (green) (PDB ID: 4AUQ) conjugated to Ub (red) and in complex with BIRC7 (not shown) was solved at 2.18 Å using X-ray crystallography (Dou *et al.*, 2012). The UbcH5b structures contain Ub conjugated to a serine instead of cysteine, a mutation required to prevent the instabilities of a catalytic active Ub conjugate. UbcH7b and UbcH5b UBC domains are conserved with an RMSD: 1.702. Structures were generated in PyMOL as cartoon, stick and surface illustrations.

1.1.4 E3 Ligating Enzymes

Being the final enzyme involved in the Ub cascade (excluding DUBs), over 600 human E3 ligases are utilised to guide the Ub substrate specificity. E3 ligases achieve this by facilitating Ub transfer from E2 to the Ub substrates through alternative mechanisms dependent on the E3 catalytic domain utilised. These domains categorise E3s into three subtypes: Really interesting new gene (RING), homologous to E6AP carboxyl terminus (HECT) and the more recently categorised RING between RING (RBR). The vast majority of E3 enzymes belong to the RING-type aiding Ub-transfer through the docking of both E2 and the substrate (Deshaies and Joazeiro, 2009). In contrast, the HECT subtype directly participates in the Ub-transfer, achieved by the formation of a new Ub thioester bond with the catalytic domain's active-site cysteine before then transferring Ub onto the substrate (Scheffner and Kumar, 2014). The final RBR subtype has been described as a RING-HECT hybrid, containing a similar RING cysteine/histidine scaffold arrangement alongside an active-site cysteine utilised in a direct Ub transfer mechanism from UbcH7 (Dove and Klevit, 2017). So far, 14 human RBR members have been identified including the Parkin E3 ligase known for its involvement in mitophagy, defined as the selective autophagy of damaged mitochondria. Dysregulation of Parkin is associated with neurodegenerative diseases such as Alzheimer's, Parkinson's and amyotrophic lateral sclerosis (Wade Harper, Ordureau and Heo, 2018). Despite the progress in understanding RBR's structure-function relationships investigated by Wang *et al*. (2023), only the more characterised RING and HECT E3 ligases are discussed further.

1.1.4.1 RING E3 ligases

The vast majority of E3 ligases belong to the RING-type at around 600 mammalian members, characterised by a conserved RING finger domain alongside various substrate recognition domains. This RING finger domain contains two centrally coordinated Zn^{2+} ions centred via a conserved cross-braced arrangement of eight cysteine and/or histidine residues defined as the CxHx signature motif (Figure 1-7). Canonically, RINGs have either one or two His in the linear arrangement of coordinating residues, denoted C3H2C3 or C3HC4, however, other variations exist (Metzger *et al.*, 2014). It should be noted there is a small subset of RING-like E3 ligases, termed U-box E3 ligases that also contain a C-terminal domain similar to a RING finger however lack the coordinated Zn^{2+} ions and are now becoming more accepted as their own class (Hu *et al.*, 2018). The RING domain also contains two crucial loop regions surrounding a central α-helix groove that provides a platform for E2 binding (Figure 1-8A). Although more simply viewed as a scaffold for E2-substrate interactions, RING E3s can activate the E2 Ub release by promoting an active E2 conformation (Deshaies and Joazeiro, 2009). Whilst many do

exist as single-chain enzymes, RING E3s tend to form homo, hetero or multi-subunit complexes, with various RING E3s only able to act through these oligomeric interactions, either lacking their specificity for the final Ub substrates or capabilities to interact with E2s, missing a vital α-helix required for binding (Metzger, Hristova and Weissman, 2012). These different oligomerisations enable the subdivision of RING families. The multi-subunit cullin RING ligase superfamily is a well characterised example of the complexities in the RING E3s function. This superfamily includes the SCF complex, consisting of S-phase kinase-associated protein 1 (SKP1), cullin and various F-box proteins alongside the E3 ligase RING box protein 1 (Rbx1) (Figure 1-8B). In this SCF complex, Rbx1/2 are responsible for E2 recruitment with the inter-changeable F-box proteins utilised for substrate recognition resulting in potentially 69 different mammalian SCF E3s (Nguyen, Wang and Xiong, 2017). Interestingly, these SCFs are only activated in the presence of NEDDylation, through an on/off switch mechanism. A conserved cullin lysine close to the RING-binding site must first be conjugated to NEDD8, resulting in a large conformation change of the C-terminal domain, repositioning Rbx1 to optimally recruit the E2-Ub conjugate (Duda *et al*., 2008; Scott *et al*., 2014). Unsurprisingly, being responsible for ~20% of ubiquitinated protein destruction via the 26S proteasome, dysregulation of these SCF complexes is involved in cancer (Wei and Sun, 2010). However, as with all RING E3 ligases, such therapeutic efforts require targeting of their protein-protein interactions whether this be to the E2-E3 complex or E3-substrates complex, notoriously difficult sites when considering the more accessible small molecule therapy, with peptide-based drugs having more target success but rarely make it through the discovery pipeline due to their various pharmacologic issues (Otvos and Wade, 2014).

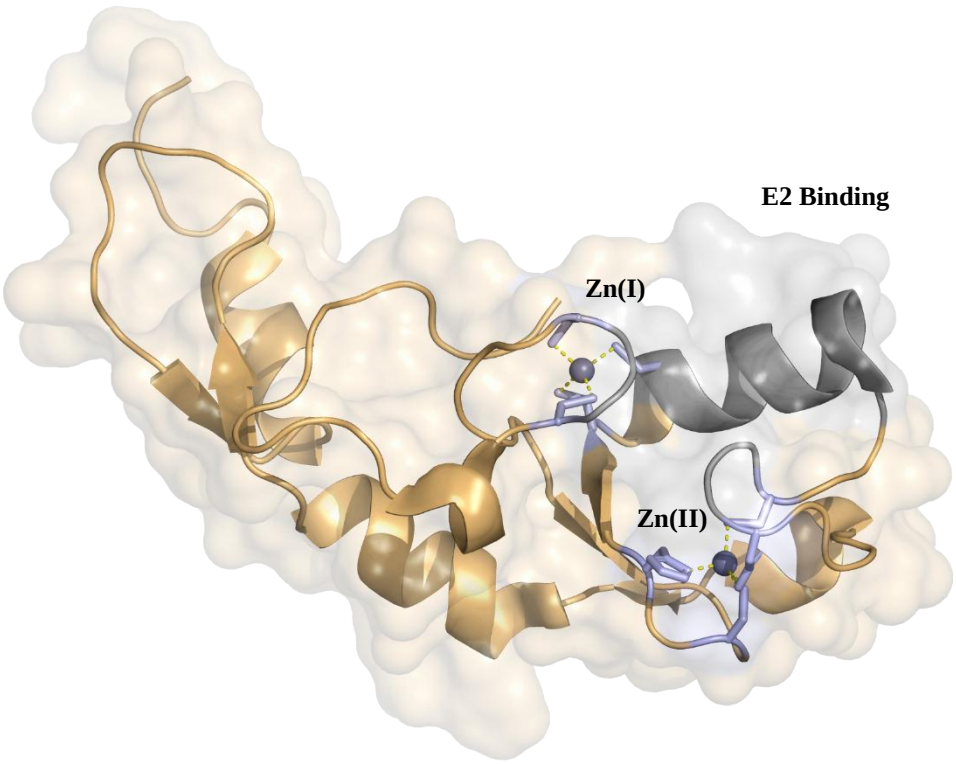


Figure 1-7: Structure of the RING finger domain. The structure of tumour necrosis factor receptor-associated factor 6 (TRAF6)'s RING domain (PDB ID: 3HCT) complexed with Ubc13 (not shown) was solved at 2.10 Å using X-ray crystallography (Yin *et al.*, 2009). The RING finger (orange) domain alongside the conserved dual coordinated Zn^{2+} ions to a cross-bridge of histidine and cysteine residues (Purple) in the CxHx arrangement C3HC3D, alongside the E2 binding central α-helix and loops (grey). Structures were generated in PyMOL as cartoon, stick, sphere and surface illustrations.

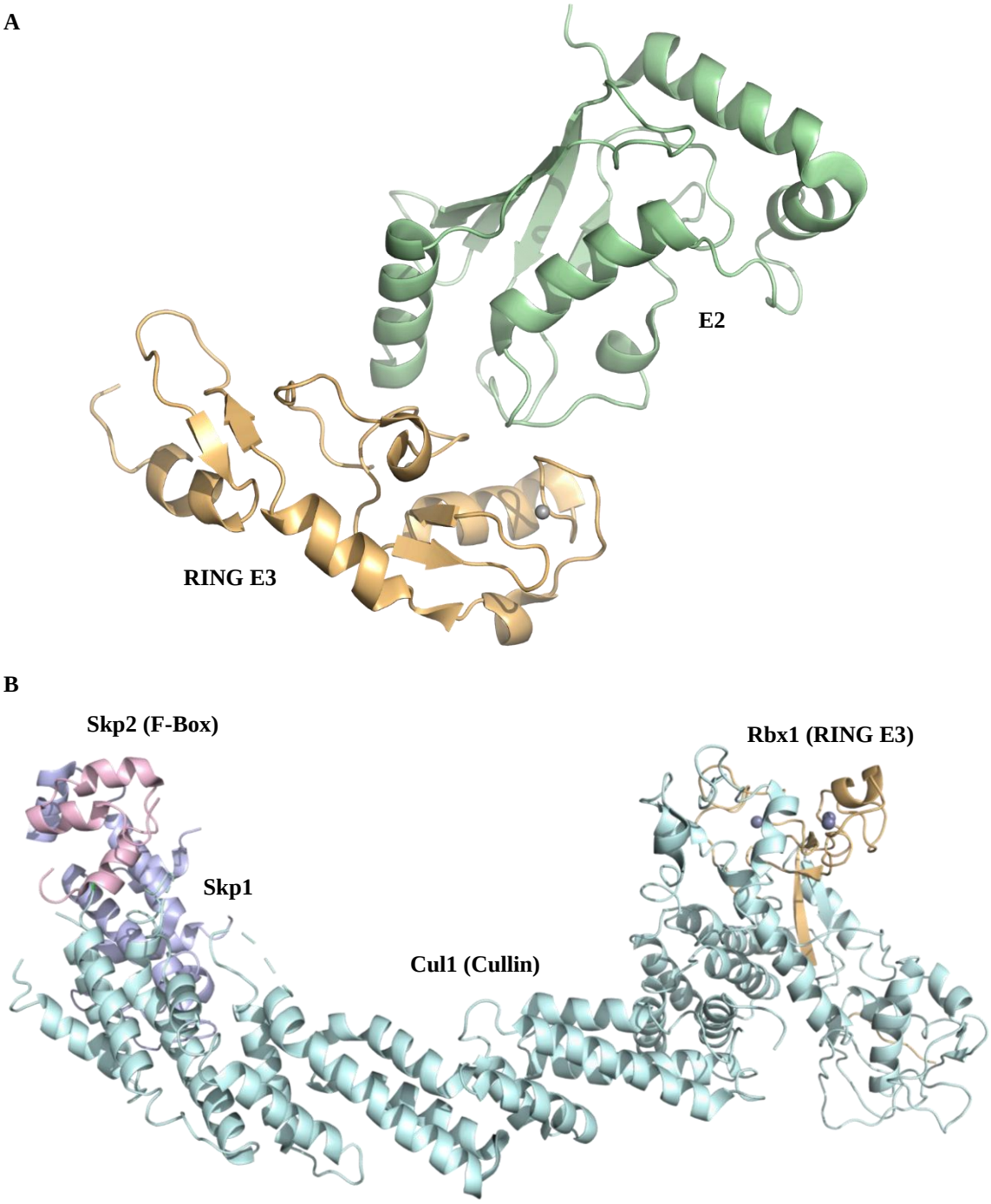


Figure 1-8: Structures of RING E3 ligases in complex with E2 and adaptor proteins. (A) The complete structure of TRAF6 (PDB ID: 3HCT) (orange) in complex with Ubc13 (green) solved at 2.10 Å using X-ray crystallography (Yin *et al.*, 2009). (B) The structure of the SCF complex (PDB ID: 1LDK) consisting of Cul1 (blue), Rbx1 (orange), Skp1 (purple) and F-Box of Skp2 (pink) was solved at 3.10 Å using X-ray crystallography (Zheng *et al.*, 2002). Substrate ubiquitination is achieved through the recruitment of Ub-E2 at the RING E3 Rbx1, alongside Skp2 F-box recruitment of substrates into the middle of the SCF complex. Structures were generated in PyMOL as cartoon and sphere illustrations.

1.1.4.2 HECT E3 ligases

HECT-E3 ligases were the first E3 subtype discovered, characterised by a C-terminal HECT domain (~350 amino acids) following an array of N-terminal extensions, with 28 mammalian members currently known (Huibregtse, *et al*., 1995). These N-terminal extensions contain defined domains that originally enabled the HECT E3 members to be classified into three subfamilies: The HERC family (6 members) containing RCC1-like domains (RLDs), the NEDD4 family (9 members) containing tryptophan-tryptophan (WW) domains, and the SI(ngle)-HECT family containing neither RLDs nor WW domains. The HERC family can be further subdivided into two groups, small HERCs (~100 kDa) and large HERCs (>500 kDa) containing more than one RLD domain (Scheffner and Staub, 2007) (Figure 1-9). However, considering the evolutionary trajectory of the HECT family members, it is clear that only the NEDD4 subfamily is supported, with a further 12 other subfamilies being suggested (Marin, 2010). Interestingly, this study indicated that the small and large HERC subfamily may have acquired the RCC1-like domains from two independent events as they are distantly related to their phylogenetic tree. In any case, the original classification is still maintained in the literature.

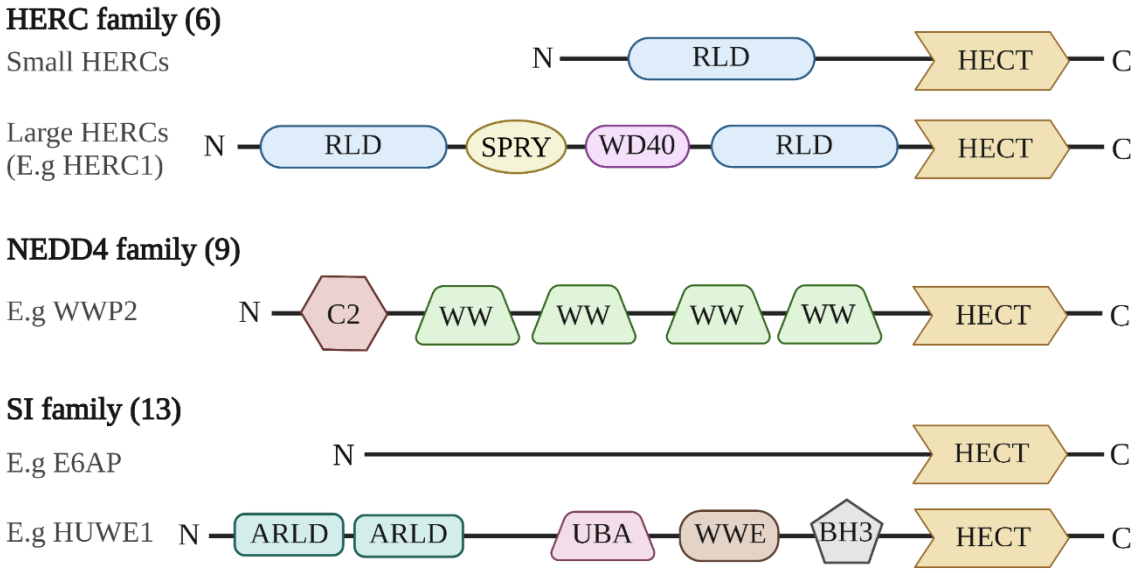


Figure 1-9: Domain architecture of the HECT-E3 ligase subfamilies. HECT-E3 ligases are classified into three subfamilies based on their N-terminal domain arrangements, all containing a C-terminal HECT domain (HECT – orange). The HERC family contain RCC1-like domains (RLD, blue) further subdivided into small and large HERCs, with large containing more than one RLD. HERC1 contains two RLDs, with a spla and the ryanodine receptor domain (SPRY, yellow) and a beta-transducin repeat (WD40, purple). The NEDD4 contain a Ca^{2+}-binding domain (C2, red) and 2-4 tryptophan-tryptophan domains (WW, green), shown in WWP2. The SI family have no specific N-terminal morphology. E6AP has no defined N-terminal domain arrangement, whereas HUWE1 has a Bcl-2 homology 3 domain (BH3, grey), a tryptophan-tryptophan-glutamate domain (WWE, brown), a ubiquitin-associated domain (UBA, pink), and two armadillo-repeats (ARLD, turquoise). The image was created using BioRender and adapted from (Rotin and Kumar, 2009).

The C-terminal HECT domain provides the catalytic function to all HECT-E3 families and is comprised of a large N-lobe utilised for both E2 and substrate docking, and a small C-lobe containing the catalytic cysteine. These lobes are connected via a flexible linker region that acts to juxtaposition the C-lobe cysteine to its functional targets. Dynamic transitioning between two conformations, described as the open L-shape and closed T-shape, is vital in the catalytic mechanism of these enzymes (Chen *et al.*, 2017) (Figure 1-10A and B). The Ub-E2 complex first binds at the N-lobe E2-docking sites, positioning the two active-site cysteines into proximity, and allowing the initial trans-thioesterification to take place. Intriguingly, the distance between these two cysteines is dependent on the Ub-bound E2 state, with the cysteine distances drastically increasing when Ub is not present and thus suggesting non-covalent interactions between the Ub and C-lobe are crucial (Scheffner and Kumar, 2014) (Figure 1-10C). Once Ub has been loaded onto the C-lobe, a conformation rotation of the C-lobe (120 °) forming the S-shape, enables the nucleophilic attack of the lysine residue and subsequent Ub-transfer to the substrate or Ub during chain formation. The HECT mechanism is tightly regulated to control both intrinsic activity and substrate specificity. Control over the HECT intrinsic activity is vital for both temporal and spatial restriction of the E3 enzymatic function with many mechanisms targeting the E2 N-lobe interaction. For example, members of the NEDD4 family can form an inactive conformation achieved using intramolecular interactions with their N-terminal regions, blocking the E2 site, whilst sterically locking the flexible hinge region. Other HECT members can achieve a similar inactive state, adopting to use of intermolecular interactions forming oligomers (Weber, Polo and Maspero, 2019). A fail-safe is also present, whereby a prolonged intrinsic activity can result in autoubiquitination and thus downstream self-destruction (Bie and Ciechanover, 2011). Extrinsic regulation is further utilised with many regulatory mechanisms controlled by post-translational modifications such as phosphorylation and SUMOylation, deubiquitination by DUBs as well as an array of other adaptor proteins generally found to mediate the actions of the HECT-E3 N-terminal regions which themselves dictate the substrate specificity (Sluimer and Distel, 2018; Shah and Kumar, 2021). With HECT E3 ligases found at the end of the Ub cascade, responsible for the substrate specificity and reliant on intrinsic cysteine activity shown to be closely regulated, they provide ideal targets for a Ub system-directed therapeutic approach. In particular, the NEDD4 family has gained much research attention given their responsibility in regulating the turnover of antitumour proteins, with more details found in Section 1.2.

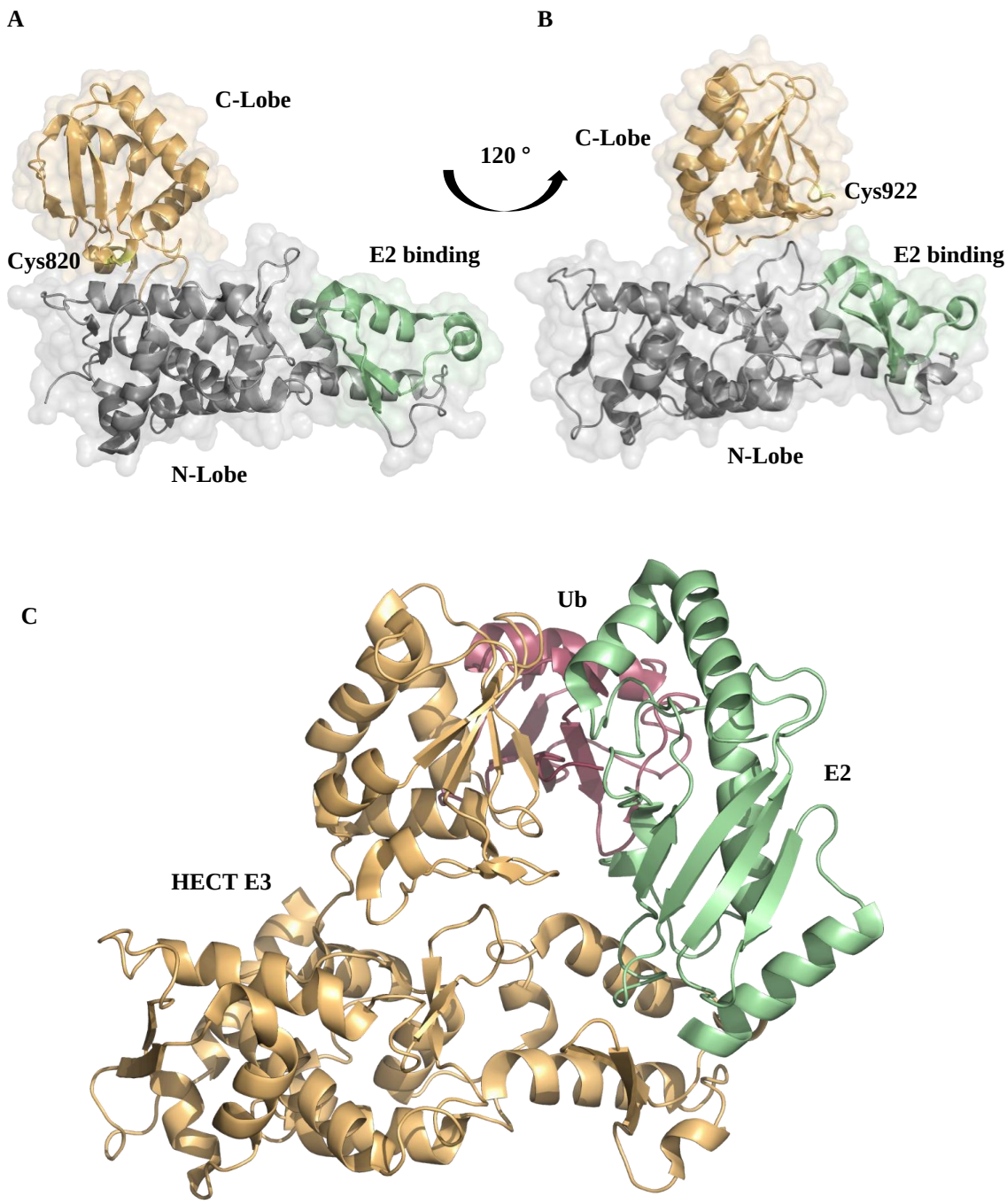


Figure 1-10: Structures of HECT domain movements and interactions. (A) Structure of E6AP (PDB ID: 1C4Z) in complex with UbcH7b (not shown) solved at 2.60 Å using X-ray crystallography (Huang *et al.*, 1999). The HECT domain is shown in the open L-shape, with catalytic Cys820 (yellow) and E2 binding region (green). (B and C). Structure of NEDD4-2 (PDB ID: 3JW0) in complex with UbcH5b solved at 3.10 Å using X-ray crystallography (Kamadurai *et al.*, 2009). The HECT domain (orange in C) is shown as both N-lobe (orange) and C-lobe (grey) with E2 binding region (green) highlighted. Both the open L-shape (A) with catalytic Cys820 (yellow) and closed T-shape (B) with catalytic Cys/Ser922 (yellow) are shown, with the 120 ° rotation of the flexible linker highlighted. This rotation positions the cysteine, mutated (C922S) to enable capture of the structure, in proximity (~7.4 Å) to the UbcH5b active site cysteine, also mutated (C85S). Structures were generated in PyMOL as cartoon, stick and surface illustrations.

1.1.5 Deubiquitinating Enzymes

DUBs play a fundamental regulatory role in the final stages of the Ub system responsible for the cleavage of conjugated Ub, disassembling chains and therefore downstream signalling. With ~100 human DUBs identified, this crucial recycling of Ub is achieved by cleaving isopeptide or peptide Ub bonds either at the distal chains ends (sequentially) or at the Ub-substrate site (whole chain removal) (Amerik and Hochstrasser, 2004). DUBs can also edit Ub chains by pruning branches, and therefore modify signalling responses. As such various classes of DUBs are found, more generally separated into two major classes: cysteine proteases and metalloproteases. Cysteine protease DUBs are characterised by a highly conserved catalytic triad composed of an active site cysteine and histidine, with most containing asparagine or aspartate. This class is further categorised into six families based on their domain architecture, these being the ubiquitin C-terminal hydrolase (UCH), ubiquitin-specific protease (USP), ovarian tumour (OTU), Machado–Josephin domain (MJD), zinc-finger containing ubiquitin peptidase (ZUP1), and the Motif interacting with Ub-containing novel DUB family (MINDY) (Lange, Armstrong and Kulathu, 2022). The zinc-binding JAB1/MPN+/MOV34 protease (JAMM) family is the only family associated with the metalloprotease class of DUBs, and instead functions through a catalytic serine and zinc cofactor (Shrestha *et al.*, 2014). Although these families have their unique folds and architecture, they typically contain a characteristic hydrophobic patch vital for Ub Ile44 interactions as well as an abundance of UBDs, including the zinc finger ubiquitin-specific protease (ZnF-UBP) and UBA domains, as well as the ubiquitin-interacting motif (UIM), particularly in the USP family, by far the largest at 58 members (Komander, Clague and Urbé, 2009). Distinguishing between Ub and UBLs is also important and achieved through their recognition and differences observed between their C-terminal tails, with most DUBs utilising the Arg72 and Arg74 to aid orientation of the scissile bond for cleavage (Sato *et al.*, 2008; Ye *et al.*, 2011). As expected and a common theme across the Ub system, dysregulation of DUBs is linked to diseases such as autoimmune disorders, chronic inflammation, oncology and neurodegeneration (Harrigan *et al.*, 2018). However, with only a relatively small number of DUBs compared to E3s and available substrates, many regulate a diverse range of targets with therapeutic efforts struggling with off-targets and non-selective inhibition given the similarities between their catalytic centres. Despite this, progress in high throughput screening technologies and cross-reactivity profiling provide promise for DUB therapeutics (Chan *et al.*, 2023).

1.2 The NEDD4 Family of HECT E3 Ligases

The NEDD4 family of HECT E3 ligases retains much research attention, believed as one of the best targets for the development of Ub-based cancer therapeutics. The NEDD4 family is comprised of nine members (NEDD4-1, NEDD4-2, HECW1, HECW2, WWP1, WWP2, SMURF1, SMURF2 and ITCH), all sharing a common architecture of three functional domains; an N-terminal C2 domain, 2-4 WW domains and a C-terminal HECT domain (Chen and Matesic, 2007). The N-terminal C2 domain (~120 amino acids) primarily functions in membrane localisation, utilising its Ca^{2+} binding capabilities to mediate domain interactions with both phospholipids as well as membrane-associated proteins (Rizo and Sudhof, 1998). This is particularly important given various NEDD4 family members target channel proteins such as NEDD4-2 (Goel, Manning and Kumar, 2015). More specifically, the C2 domain has been shown to interact with Annexin XIII used as a transport adaptor to co-localise NEDD4 to the apical plasma membrane (Plant *et al*., 2000). Intriguing this study found when Annexin XIII was no longer overexpressed, NEDD4 was observed in the cytosol suggesting such adaptors are required to specify NEDD4 localisation. The WW domains (35-40 amino acids) are a well-characterised domain type, defined by two conserved tryptophan residues spaced by 20-22 amino acids. Used in protein-protein interactions to target the Leu/Pro-Pro-X (any amino acid)-Tyr (L/PPXY) motif, the specific arrangement of these domains provides the major substrate specificity for each NEDD4 family member (Ingham *et al.*, 2005). The final C-terminal HECT domain has already been discussed in detail. However, the regulatory mechanisms associated with the HECT functionality do display some variance in the NEDD4 family when compared to other HECT E3 ligases, dependent on their intramolecular interactions from the C2 domain and WW domain regions. Using nuclear magnetic resonance (NMR) mapping, the C2 domain was shown to interact with a region overlapping a non-covalent Ub exosite on NEDD4-1, NEDD4-2 and SMURF2 and is suggested to act as a key regulator of HECT activity by preventing the trans-thioesterification step (Mari *et al.*, 2014). This autoinhibitory effect is not observed throughout the NEDD4 family, instead an alternative autoinhibitory role was played by the WW2-2,3-linker region. Using WWP2 as a model, Chen *et al*. (2017) showed that the WW2-2,3-linker forms an α-helix that interferes with both the Ub exosite and the flexible hinge region of the HECT domain, locking it in its T-shape. These intramolecular interactions have since been observed in ITCH and WWP1, with the WW4 domain also found to lock the 2,3-linker into a headband shape, binding to the HECT N-lobe at the adjacent side. Such intramolecular interactions are also mimicked by NEDD4-1, NEDD4-2 and SMURF2 between WW1 and the C2 domain (Figure 1-11) (Wang *et al.*, 2019). In ITCH, this autoinhibitory state is relieved via adaptor interactions or JNK1-mediated phosphorylation, highlighting additional activation pathways specific to the NEDD4 family (Zhu, *et al*., 2017).

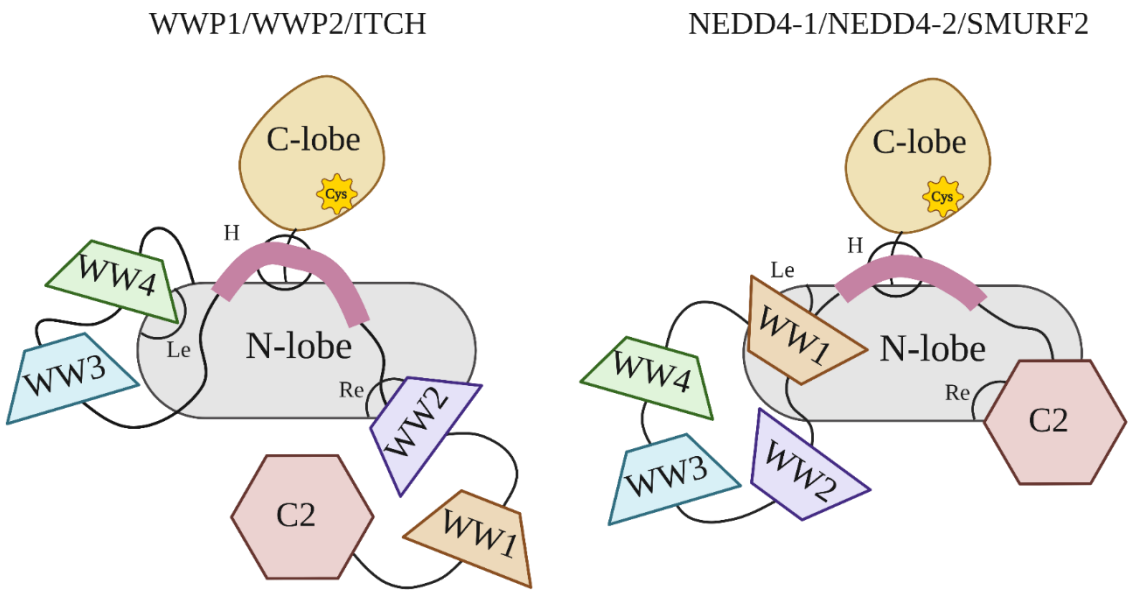


Figure 1-11: Schematic diagram of the NEDD4 family HECT autoinhibitory mechanisms. The HECT domain is shown in its T-shape, with a large N-lobe (grey) and catalytic C-lobe (yellow) connected via its flexible linker. The additional substrate domains including the C2 (red), WW1 (orange), WW2 (purple), WW3 (blue) and WW4 (green) also shown in their autoinhibitory arrangements dependent on either belonging to WWP1/WWP2/ITCH (left side) or NEDD4-1/NEDD4-2/SMURF2 (right side). The headband locations are illustrated as left ear (Le), right ear (Re) and the headband (H). Image created in BioRender, adapted from (Wang *et al.*, 2019).

The Ub exosite located at the base of the N-lobe was first discovered in the yeast E6AP homolog RSP5 before soon being identified in SMURF1, with both groups suggesting contrasting mechanisms as either a site to restrict or promote polyubiquitination, respectively (French, Kretzmann and Hicke, 2009; Ogunjimi *et al*., 2010). This site has since been alluded to play two related roles. Firstly, as an intramolecular autoinhibitory relief position, as indicated by Zhang and colleagues (2016), where Ub variants were shown to activate various NEDD4 members, having solved crystal structures with them bound at the Ub exosite. Secondly, as a processivity site for polyubiquitin chain elongation or otherwise better described as a Ub-substrate conjugate docking site, whereby solved crystal structures have shown that non-covalent Ub binding at the exosite enables E2-Ub conjugate accessibility (Maspero *et al*., 2011). This has been further supported by a covalent inhibitor binding at the Ub exosite that demonstrated a switch in the enzymatic behaviour of NEDD4 from a processive mechanism, whereby the substrate remains bound during processing, to a distributive mechanism, with substrate diffusion occurring between ubiquitin chain elongation by blocking the Ub exosite access (Figure 1-12) (Kathman *et al*., 2015). Other activation/regulatory mechanisms of the NEDD4 family exist in the form of endogenous isoforms as a result of alternative splicing events, with these isoforms observing an increased activity when lacking these autoinhibitory N-terminal domains (Soond *et al.*, 2013).

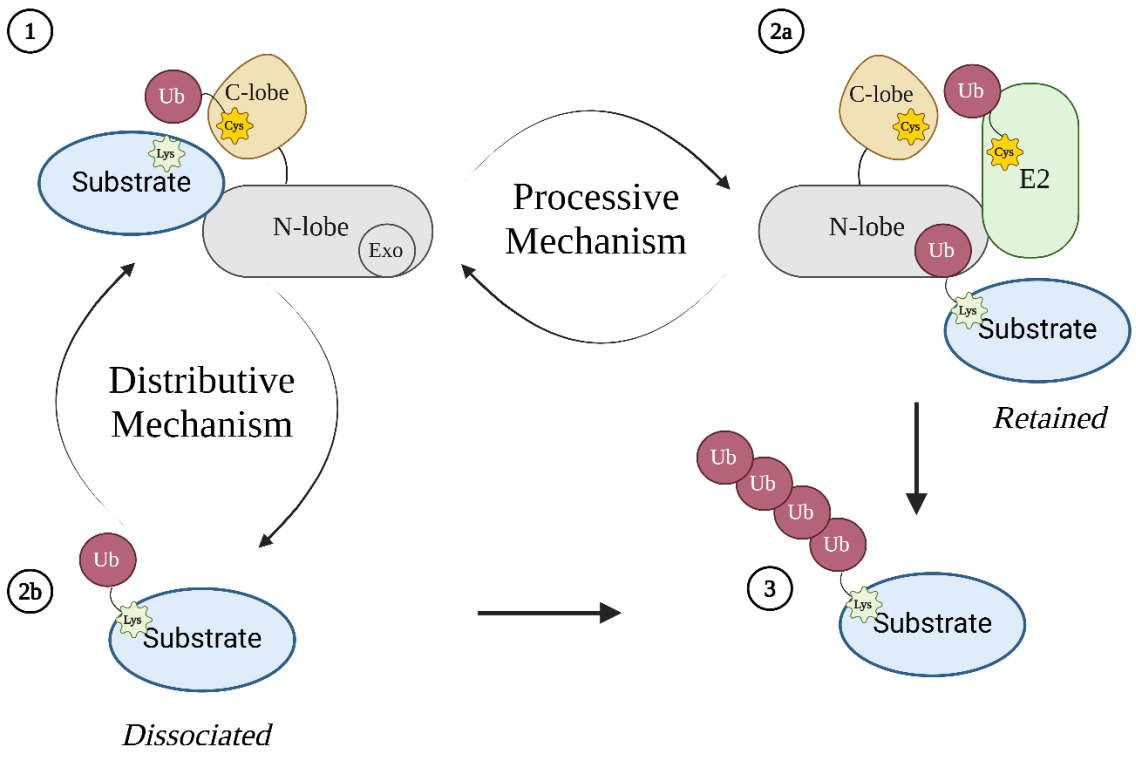


Figure 1-12: Schematic diagram of the E3 ligase processivity vs distributive mechanism. 1. The ubiquitin (red, Ub) charged E3 HECT domain (grey, N-lobe and orange, C-lobe) is shown transferring the Ub molecule to the substrate's (Blue) lysine (Lys) residue. 2. In a processive mechanism (2a), this ubiquitinated substrate is retained at the E3 HECT domain by association of Ub to the Ub exosite, whilst the next E2-E3 Ub transfer takes place, enabling Ub chain elongation within the complex. If the exosite is blocked, a distributive mechanism (2b) is used, whereby the ubiquitinated substrate can dissociate after mono-ubiquitination, or can reassociate to build the polyubiquitin chain (3). Image created in BioRender.

It is becoming increasingly evident that the abnormal expression of the NEDD4 family is related to a variety of pathologies, including inflammatory, cardiovascular and neurological diseases (Zhang *et al.*, 2020; Conway, Kinsman and Kramer, 2022; Xu *et al.*, 2024). But what is most prevalent and arguably the reason behind their substantial characterisation is their association with various malignancies (Tian *et al.*, 2023). For example, the expression of NEDD4 is commonly found elevated in tumours using immunohistochemistry including breast, lung, gastric, melanoma, pancreatic and bladder cancer (Sun *et al.*, 2014; Wen *et al.*, 2017; Weng *et al.*, 2017; Shao *et al.*, 2018; Wan *et al.*, 2019; Cheng *et al.*, 2021). In vitro overexpression in hepatocellular carcinoma cells by transfection also promoted cell proliferation and migration whilst inhibiting apoptosis (Zheng *et al.*, 2018). These characteristics follow through to the majority of the NEDD4 family, with two members, the WW domain-containing E3 ligase 1 (WWP1) and 2 (WWP2), being of particular interest due to their targeting of the tumour suppressor protein phosphatase and tensin homolog (PTEN) as well as various other

tumour suppressors and transcription factors (Maddika *et al.*, 2011; Lee *et al.*, 2019). To no surprise, both WWP1 and WWP2 dysregulation has been directly linked to oncogenic diseases mainly acute myeloid leukaemia (AML), prostate and breast cancer as well as more specific cardiovascular, osteogenic and even infectious diseases including COVID-19 via viral budding (Zhi and Chen, 2012; Novelli *et al.*, 2021; Wang *et al.*, 2024). However, below we will focus on their oncogenic properties.

1.2.1 WWP1 and WWP2 E3 Ligases and Cancer

First discovered as Atropin-1-interacting protein 5 (AIP5), and TGIF-interacting ubiquitin ligase 1 (TIUL1), the now-known WWP1 is located on the human chromosome 8q21 and encompasses 26 exons (Wood *et al.*, 1998; Seo *et al.*, 2004). The 70% sequence homolog WWP2 was also earlier named atrophin-1-interacting protein 2 (AIP2) and is located on the human chromosome 6q22.1 and has 24 exons (Pirozzi *et al.*, 1997). Both these genes can undergo alternative splicing events forming no less than six and five isoforms for WWP1 and WWP2 respectively (Flasza *et al.*, 2002; Chantry, 2011). The full-length E3 ligases are at 922 and 870 amino acids in size and follow the characteristic NEDD4 architecture consisting of a C2 domain, four WW domains, and catalytic HECT domain with an active site Cys890 and Cys828, respectively. The HECT domains of these E3 ligases are even more similar at 83% sequence identity, with WWP1's solved in 2003 (PDB ID: 1ND7), and the WWP2 HECT structure solved much later in 2015 (PDB ID: 4Y07) (Verdecia *et al.*, 2003; Gong *et al.*, 2015). The various WWP1 and WWP2 isoforms display modified enzymatic activity, through relieved autoinhibitory mechanisms, as well as substrate recognition and therefore downstream signalling, with some isoforms also found to be more prevalent in disease (Chantry, 2011; Wahl *et al.*, 2019).

WWP1 is arguably the better-characterised WWP homolog, regulating the turnover of a variety of proteins (such as PTEN, p63, p27, TβR1, Smad2, ErbB4, and RNF11) involved in the regulation of cell growth and survival processes (Hu *et al*., 2021). To no surprise, high expression of WWP1 has been associated with numerous cancers including breast, bowel, oral, liver, gastric, thyroid, and bone cancers (Chen *et al*., 2007; Lin *et al*., 2013; L. Zhang *et al*., 2015; Wu *et al*., 2015; X. F. Zhang *et al*., 2015; Chen and Zhang, 2018; Wang *et al*., 2022). More notably, the overexpression of WWP1 appears to always relates to poor patient prognosis. For example, in a study of 149 patients with hepatocellular carcinoma, the overall survival rates were dramatically contrasted from ~20 % to ~70 % when comparing the high to low WWP1 expression in tissue samples (X. F. Zhang *et al*., 2015). However, the picture is not clear, with other studies suggesting WWP1 can act as a tumour suppression, for example through the Ub-mediated degradation of Kruppel-like factor 5 (KLF5), a known breast cancer antagonist,

suggesting WWP1 properties may be context-dependent (Zhao *et al*., 2012). Such complexity is evident in WWP1's interactions with the p53 family transcription factor p63. This apoptotic regulator can be expressed as two alternative isoforms, TAp63 and ΔNp63, able to sensitise or desensitise cells to apoptosis, respectively (Wu *et al.*, 2003). This effect can be observed during WWP1 knockdown studies; knockdown in MCF-10A breast cancer epithelial cells led to an increase in ΔNp63, whereas knockdown in HCT-116 colon cancer cells led to an increase in TAp63 levels (Li, Zhou and Chen, 2008). Despite displaying opposing roles, WWP1 is still very much considered a promising target for a variety of cancer treatments, including for AML, being described as a potential biomarker. This is mainly due to WWP1's downregulation of the tumour suppressor p27kip1 involved in G0/G1 cell cycle arrest, observed during AML (Sanarico *et al.*, 2018). WWP1 has also been shown as a negative regulatory of PTEN, with shown to cause 'inactivation' of PTEN dimerization through K27-ubiquitination and therefore initiate tumour growth through the phosphatidylinositol 3-kinase (PI3K) signally pathway (Lee *et al.*, 2019). A later study by Lee and colleagues (2020) discovered three prevalent mutations, R86H, K740N, and N745S, in patients demonstrating PTEN-mutation-free Cowden-like syndrome. Cowden syndrome is directly related to mutations in PTEN, leading to a high prevalence of sporadic cancers. An in vitro analysis of these mutations led to a gain-in-function, significantly down-regulating PTEN compared to wild type. In particular, the K740N mutation located on the catalytic HECT domain was shown to interfere with the autoinhibitory interactions of the WW2-WW3 linker. WWP1 also has major roles in other oncogenic signally pathways, namely the epithelial growth factor (EGF), involved in regulating ErbB4, with overexpression correlated to breast cancer cell lines (Li *et al.*, 2009). Other signalling pathways also include the ominous transforming growth factor beta (TGFβ) pathway, whereby WWP1 overexpression plays a dysfunctional role in downregulating the TβR1 receptor as well as a range of SMADs, resulting in poor prostate cancer prognosis (Chen *et al.*, 2007).

Although WWP2's oncogenic roles are not as vast as its sequence homolog, it still provides a promising target, having well-documented oncogenic functions including dysregulation in cancers such as lung, oral, gastric, and uterine (Fukumoto *et al*., 2014; Clements *et al*., 2015; Yang *et al*., 2016; Wang *et al*., 2020). For instance, the dysfunction of the TGFβ pathway is also observed in WWP2 malignancies, whereby the increased WWP2 targeting of Smad7, a key negative regulator of the TGFβ pathway, has been shown to promote late-stage tumours to undergo metastasis (Soond and Chantry, 2011). Malignant cells lacking the WWP2-mediated destruction of OCT4 have also been observed to regain their pluripotency, promoting the initiation of tumours (Qian *et al.*, 2012). Finally, the WWP2-mediated destruction of PTEN (alternative to WWP1 inactivate) was found to increase PI3K-Akt signalling, leading to malignant cells becoming resistant to stress-induced cell death (Maddika

et al., 2011). Other pathogenic roles of WWP1 and WWP2 are further reviewed elsewhere (Behera and Reddy, 2023; You *et al.*, 2024).

1.3 Targeting the NEDD4 Family of HECT E3 Ligases

The FDA approval of Bortezomib (Velcade), a reversible proteasome inhibitor used in the treatment of multiple myeloma originally established the therapeutic potential of targeting the Ub-proteasome system and was found to induce apoptosis in cancer cells with a higher sensitivity over their native counterparts (Daviet and Colland, 2008; Kazi *et al*., 2009). Since then, there has been considerable research effort to target more specific members of the Ub system. However, only a handful of RING E3 ligase inhibitors have gained FDA approval, based on Thalidomide previously used as a sedative for pregnant women but was found to cause birth defects due to off-targeting the E3 ligase Cereblon (Ito *et al*., 2010). Celeblon is a key regulator of ion transport, cell proliferation, apoptosis, and most importantly growth factors involved in limb outgrowth during embryo development (Higgins *et al*., 2008). This led to the development of second-generation inhibitors Lenalidomide and Pomalidomide, having higher potency in the treatment of multiple myeloma through the induced degradation of lymphoid transcription factors IKZF1 and IKZF3 by blocking Celeblon's endogenous substrates (Fischer *et al*., 2014; Krönke *et al*., 2014). There are various other E3 ligase inhibitors now in clinical trials, including the novel small molecules termed proteolysis targeting chimaeras (PROTACs), that use an E3-specific targeted protein degradation strategy. Given the NEDD4 family's prevalent involvement in disease and their primary HECT catalytic cysteine, they provide an attractive target for the development of E3 ligase inhibitors. The NEDD4 family's HECT-E3 functional mechanisms also provide various opportunities for small molecules to interfere with the HECT function. This includes interfering with the regulatory sites (Ub exosite, phosphorylation sites and the flexible linker region) as well as external recognition surfaces (E2, adaptors and target protein) however these would more likely require a peptide-based strategy, having their own set of challenges. Here the current small molecule targeting of the NEDD4 family is reviewed.

1.3.1 Small Molecule Inhibitors

Small Molecules HECT inhibitors are a relatively new drug design approach, with the first small molecule inhibitor only recently discovered by Mund *et al*. (2014). This reversible inhibitor named Heclin (HECT ligase inhibitor), was derived from bicyclic peptides that interfere with the E2 recognition site of SMURF2, NEDD4-1, and WWP1 originally identified in a phage library screen (Figure 1-13). Intriguingly, Heclin does not interfere with the E2 recognition site but instead acts to expose the catalytic cysteine, resulting in its oxidation and

thus loss of HECT activity. With IC_{50} values in the low µM range, Heclin offers promising results; for example, a study using WWP1 as a biomarker for AML observed a decrease in leukemic blast growth when in the presence of Heclin (Sanarico *et al.*, 2018). Although encouraging, the target versatility of Heclin alongside its undesirable reactive groups (a metabolically unstable furan ring and acrylamide Michael acceptor) has limited its therapeutic success unless these factors can be averted (Chen, Gehringer and Lorenz, 2018).

The discovery of Heclin and in vitro success promoted further research on HECT inhibitors, highlighting the catalytic cysteine as a major target. A covalent tethering approach provided a possible method, adopted by Kathman and colleagues (2015) in work on the NEDD4-1 ligase. To their surprise, this group identified an indole compound referred to as compound 1 (Figure 1-13), that selectively reacted with a 'non-catalytic' cysteine (Cys-627) located in the Ub exosite (PDB ID: 5C91) (Figure 1-14). These compounds prevented Ub binding to the exosite and thus prevented Ub-chain formation. One compound in particular displayed high selectivity against other proteins within the ubiquitination system containing a catalytic cysteine (E1, E2, DUBs etc.) as well as other HECT domains, including E6AP (with no Ub exosite) and WWP1 (with an Ub exosite), with only NEDD4-2 also displaying binding. Whether these compounds react with off-target proteins in vivo has still not been determined, a common problem associated with covalent inhibitors. Indole-3-carbinol (I3C) (Figure 1-13), a metabolite released during the breakdown of cruciferous vegetables, may provide an alternative non-covalent option for targeting this Ub exosite region. Already known for its antiproliferative properties in cancer, I3C has been linked to the inhibition of NEDD4-1 and WWP1, stabilising the tumour suppressor protein PTEN (Lee *et al.*, 2019). More recently, I3C has displayed an anti-viral effect when targeting both NEDD4-1 and WWP1 in the treatment against COVID-19, interfering with Ub-mediated viral budding (Novelli *et al.*, 2021). In silico modelling has indicated that I3C may interact near the Ub exosite site, suggesting a similar mechanism may occur as seen by the covalent tethered compounds (Aronchik *et al.*, 2014). Although at the early stages of development, I3C-derived HECT inhibitors may provide future success, with N-benzyl derivates already showing increased potency during the examination of I3C and NEDD4-1 structure-activity relationship (SAR) (Quirit *et al.*, 2017) (Figure 1-13).

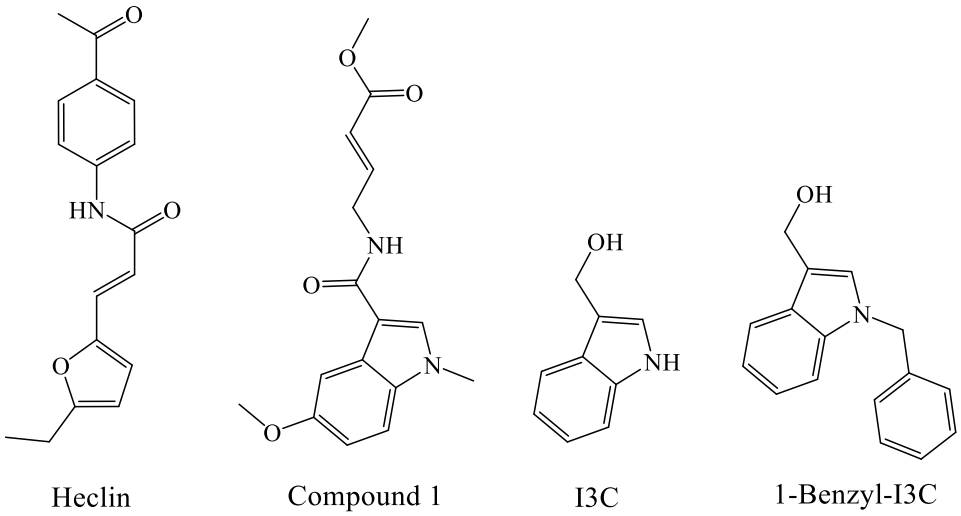


Figure 1-13: Chemical structure of Heclin and Indole-3-carbinol based inhibitors. The commercially available Heclin has been shown to target SMURF2, NEDD4-1, and WWP1 (IC_{50} of 6.8, 6.3, and 6.9 µM, respectively) (Mund *et al*., 2014). Compound 1 was shown to inhibit and bind to the Ub exosite of NEDD4-1 (K_{Inact}/K_I of 0.089 $M^{-1}S^{-1}$) (Kathman *et al.*, 2015). Indole-3-carbinol (I3C) has been shown to target NEDD4-1 (IC_{50} of 284 µM) and WWP1, alongside 1-Benzyl-I3C (IC_{50} of 12.3 µM) having improved potency towards NEDD4-1 (Quirit *et al.*, 2017; Lee *et al.*, 2019).

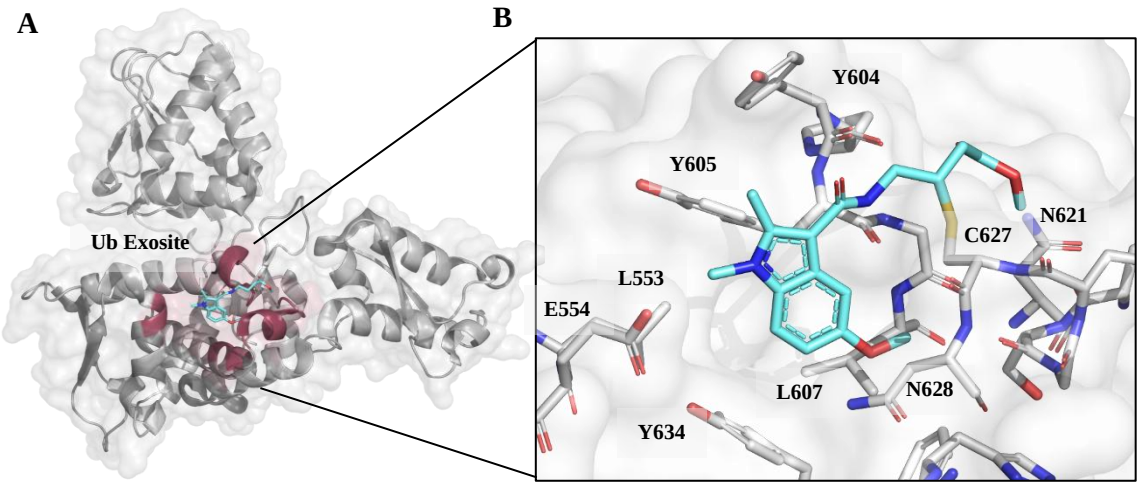


Figure 1-14: Structure of compound 1 covalently attached to non-catalytic Cys627 of NEDD4-1. Structure of NEDD4-1 in complex with an I3C-derived covalent inhibitor (PDB ID: 5C91) solved at 2.44 Å using X-ray crystallography (Kathman *et al.*, 2015). (A) Overview of compound 1 (blue) binding at the Ub exosite (Red). (B)

Close-up of Ub exosite binding pocket, with labelled residues (grey) interacting with compound 1 (blue) covalent bound to Cys627. Structures were generated in PyMOL as cartoon, stick and surface illustrations.

To overcome the lengthy process of drug development, some studies attempt to re-purpose old FDA-approved drugs, screening them against their proposed targets. Such a process discovered that the antidepressant Clomipramine also has inhibitory properties selective to the HECT activity of E3 ligases, specifically preventing the trans-thioesterification step (Figure 1-15). This inhibitory effect was observed on the ITCH ligase, blocking both in vitro autoubiquitination and ITCH-specific p73 ubiquitination (Rossi *et al.*, 2014). Another study by Tian and colleagues (2019) designed a high throughput cell-based assay to increase their discovery turnover. This assay took advantage of a dual-luciferase assay alongside the Ub reference technique (URT) whereby a K48R mutated Ub molecule (to prevent degradation) is sandwiched between the target proteins. Ub-specific cleavage by Ub proteases yields the two target proteins in equal amounts preventing the need to correct for cell variation (Varshavsky, 2005). By fusing *Renilla* luciferase to the N-terminal Ub (K48R), followed by the SMURF1 substrate RAS homolog family member B (RHOB) with C-terminal Firefly luciferase, this assay was able to measure the SMURF1 ubiquitin-mediated degradation of RHOB-Firefly using the *Renilla*-Ub as a reference point and in turn, compare this to small-molecule treated cells. This screening effort was used to identify a novel inhibitor they termed HS-152 (Figure 1-15); A urea-containing compound suggested to hinder SMURF1 stimulated TGFβ:PAR6-mediated tumour metastasis, highlighting the potential success of such a screening assay. In the case of WWP1 and WWP2, there are only a limited number of potential small-molecule inhibitors. For WWP1, small molecule inhibitors have been limited to non-selective Heclin with an IC_{50} of 6.9 μM and I3C along with its derivatives only indicated to inhibit through cell proliferation studies although was shown to interact with an Ub exosite in through point mutations (Lee *et al.*, 2019; Zheng *et al.*, 2023). Only a handful of potential WWP2 inhibitors have been highlighted in an in vitro autoubiquitination screen of the National Cancer Institute (NCI) Diversity Set V compound library reported by the Chantry group (Watt *et al.*, 2018). This paper focussed on NSC-288387 with an IC_{50} of 2.3 μM, confirmed to interact via NMR epitope mapping. However, the fragment-sized di-phenol NSC-2805 was found to be the most potent with an IC_{50} of 0.38 μM (Figure 1-15).

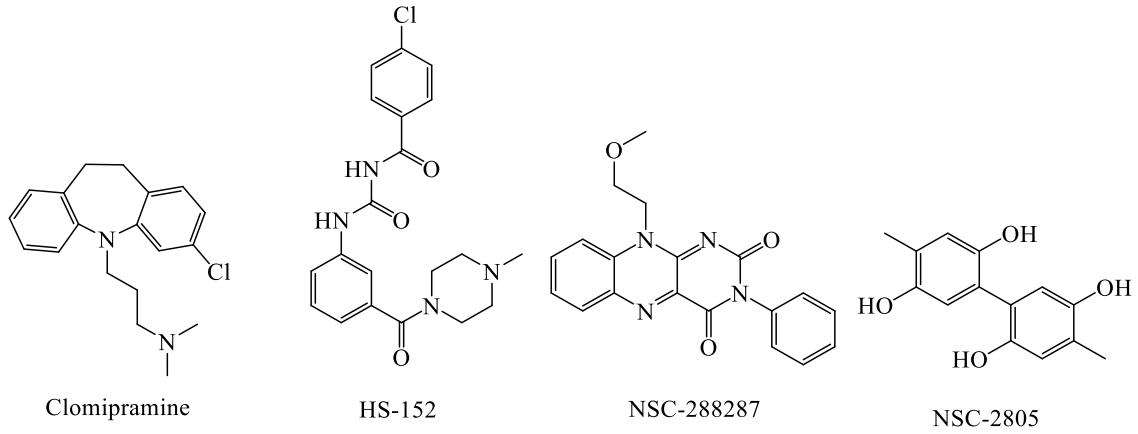


Figure 1-15: Chemical structure of other NEDD4 family inhibitors. The FDA-approved antidepressant clomipramine was found to inhibit ITCH p73 ubiquitination (Rossi *et al.*, 2014). HS-152 (IC_{50} of 3.2 µM) was found to inhibit SMURF1 RHOB ubiquitination during a cell-based assay (Tian *et al.*, 2019). Both NSC-288387 (IC_{50} of 2.3 µM) and NSC-2805 (IC_{50} of 0.38 µM) compounds were found to inhibit WWP2 from a high throughput screen (Watt *et al.*, 2018). Drawn in ChemDraw.

1.4 Thesis Aims

It is clear that despite their involvement in various malignancies, there is a limited selection of lead candidates for future drug development against WWP1 and WWP2. This research project aims to expand this small molecule inhibitor space by identifying new molecular scaffolds for future lead development. This will be achieved through a hit-identification and hit-to-lead development strategy, as outlined across three experimental chapters (3, 4 and 5) (Figure 1-16). Chapter 3 will expand upon the previous work of the Chantry group to screen the 'next in series' NCI compound library to question; Are there any other small molecules that can effectively modulate the activity of WWP1 and WWP2? This approach will specifically utilise two high-throughput assays as part of the hit-identification phase, these being differential scanning fluorimetry (DSF) to reduce the library size by ligand screening and an in vitro ELISA auto-ubiquitination assay to determine if any of these compounds are inhibitory. Chapter 4 will investigate the hit compound(s) alongside the natural metabolite I3C, previously highlighted in the literature, by answering the question; What is the structure-activity relationship of these potential small molecules against WWP1 and WWP2? This will be achieved by both an SAR by catalogue and synthesis approach in collaboration with the organic chemist Jake Rigby, from the UEA School of Chemistry, by generating an analogue library. The previous high-throughout assays will be further used to characterisation the biochemical activity of the analogue library to determine their SAR, with isothermal titration calorimetry to be introduced to assess their binding contributions. Chapter 5 looks to structurally investigate both the parent small molecules as well as various other analogues to question; Where are the binding sites and how

do they interfere with WWP1 and WWP2 function? The first approach will be to reproduce the X-ray crystallisation of apo WWP1 and WWP2 from the literature before undertaking ligand-soaking and co-crystallisation experiments with the hit small molecules in an effort to solve the first WWP1 and WWP2-ligand bound complexes. In silico molecular docking will also be investigated to aid a structure-informed approach to the SAR analysis.

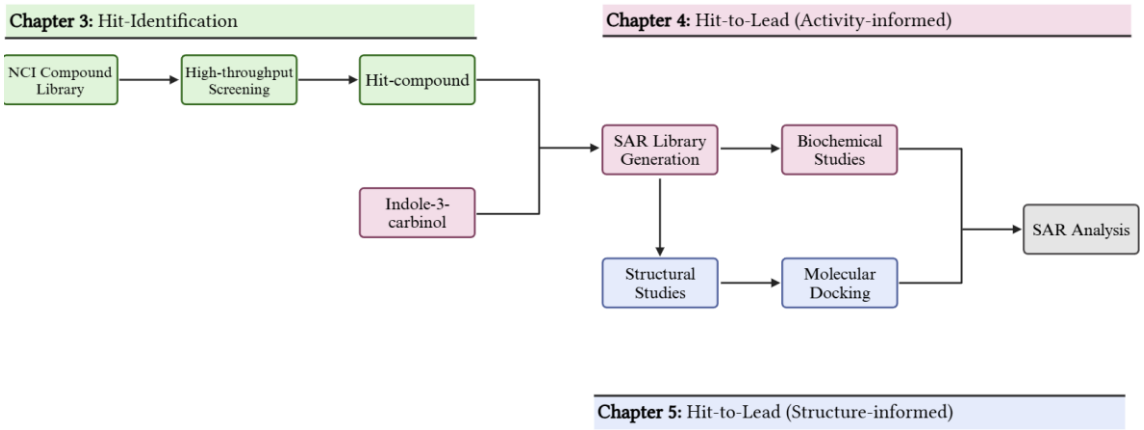


Figure 1-16: Flow chart of thesis chapters and project strategy. Relationships between chapters 1 (green), 2 (pink) and 3 (blue) are highlighted with arrows.

Chapter 2: Material and Methods

2.1 DNA Techniques

2.1.1 Plasmids

All plasmids were either purchased or gifted from various sources (Table 2-1).

Table 2-1: Plasmids, constructs, and their origins

Plasmid	Construct	Origin
pET3a- His-hUba1	Wildtype	Gifted by Dr Titia Sixma
pGEX4T-1- GST-F2-hUba1	Wildtype	Gifted by Dr Arthur L Haas's Lab
pET3a- His-UbcH7	Wildtype	Gifted by Professor Martin Scheffner
pGEX4T-1- GST-F2-UbcH7a	Wildtype	Gifted by Dr Arthur L Haas's Lab
pOPINF- His-F2-WWP1-FL	Wildtype	Dr Andrew Chantry's Lab
pET32a-His-Trx-3C-WWP1-2L34H	WW2-2,3-linker-WW3-WW4-HECT	Gifted by Dr Wenyu Wen
pET32a-His-Trx-3C-WWP1-L34H	2,3-linker-WW3-WW4-HECT	Gifted by Dr Wenyu Wen
pGEX2T- GST-F2-WWP2-FL	Wildtype	Dr Andrew Chantry's Lab
pGEX6p-2-GST-3C-WWP2-LH	WW2-2,3-linker-HECT	Gifted by Professor Philip Cole
pET28a-His-F2-WWP2-H	HECT	Dr Andrew Chantry's Lab
pET52b+-StrepII-WWP2-H	HECT	Purchased from GenScript

2.1.1.1 Plasmid regeneration

Plasmid stocks were regenerated using either the QIAprep Spin Miniprep kit or the QIAprep Spin Midiprep kit.

2.1.2 Transformations

Under sterile conditions, plasmids were transformed into their respective competent *E. coli* strains via heat-shock or electroporation before spreading 200 µL and incubating for 12-18 hours at 37 °C on LB agar plates containing respective antibiotics (Table 2-2).

2.1.2.1 Generation of chemically competent cells

Selected cell lines were inoculated from overnight starter cultures into 50 mL LB media containing appropriate antibiotics and incubated at 37 °C, 180 rpm until an OD^{600} of 0.3 – 0.4. Cultures were then placed on ice for 15 minutes before spinning down at 2,000 g, 4 °C for 5 minutes. Pelleted cells were resuspended in 30 mL of 0.1 M $CaCl_2$ and placed on ice for a further 30 minutes before re-pelleting. The cultures were finally resuspended in 4 mL of 0.1 M $CaCl_2$ containing 30% glycerol before being snap-frozen and stored at -80 °C.

2.1.2.2 Heat shock

The selected chemically competent cells were thawed on ice and separated into 50 µL aliquots. To these, 1 µL of the desired plasmid was added and gently mixed before incubating on ice for 30 minutes. Samples were then 'heat shocked' at 42 °C in a water bath for 45 – 60 seconds and immediately placed on ice for 10 minutes. Samples were then incubated for 1 hour at 180 rpm, 37 °C in 350 µL of LB before being plated

2.1.2.3 Electroporation

Selected cell lines were inoculated and grown to stationary phase (1.5 – 2.0 OD^{600}) in LB containing suitable antibiotics. A 3 mL aliquot was pelleted at 6000 g, 20 °C for 2 minutes before washing three times in 1 mL of 10% glycerol. Pellets were resuspended in 70 µL of 10% glycerol before adding 1 µL of the desired plasmid. These were loaded into a 0.1 cm electroporation cuvette and pulsed at 1.20 mV (BioRad MicroPulser) before being resuspended into 1 mL SOC media and recovered at 37 °C, 180 for 2 hours before plating.

2.2 Protein Purification Techniques

All protocols were performed on ice or at 5 °C unless stated otherwise.

2.2.1 Protein Expression

LB media containing associated antibiotics were inoculated with *E. coli* from the transformed LB-Agar plates/glycerol stocks and incubated overnight at 37 °C, 180 rpm. The starter culture was diluted (1: 80) in the desired volume and grown until reaching an OD^{600} between 0.6 – 1.0. The desired recombinant proteins were expressed, inducing with IPTG and incubating at protein-specific conditions (Table 2-2). Cells were finally pelleted by centrifugation (Beckman Coulter J20, JLA 8.1000 rotor) at 4,000 g, 4 °C for 30 minutes and stored at -20 °C.

Table 2-2: *E. coli* cell-lines and protein expression conditions.

Recombinant Protein	*E. coli* cell-line	[IPTG] (mM)	Temp (ºC)	Antibiotics†
His-**hUba1**	BL21-Star (DE3)	1.0	25	Amp
GST-F2-**hUba1**	BL21 (DE3)	1.0	25	Amp
His-**UbcH7**	BL21-CodonPlus RP	1.0	25	Amp, Cam
GST-F2-**UbcH7a**	BL21 (DE3)	0.4	20	Amp
His-F2-**WWP1-FL**	BL21-CodonPlus RP	0.5	16	Amp, Cam
His-Trx-3C-**WWP1-2L34H**	BL21-CodonPlus RP	0.5	16	Amp, Cam
His-Trx-3C-**WWP1-L34H**	BL21-CodonPlus RP	0.5	16	Amp, Cam
GST-3C-**WWP2-FL**	BL21-CodonPlus RP	0.75	30	Amp, Cam
GST-3C-**WWP2-LH**	BL21-CodonPlus RP	0.5	16	Amp, Cam
His-F2-**WWP2-H**	BL21-Rosetta1	1.0	25	Kan, Cam
StrepII-**WWP2-H**	BL21-Rosetta2	0.1	25	Amp, Cam

† *50 ug.mL^{-1} Amp (Ampicillin or Carbenicillin) and Kan (Kanamycin), 34 ug.mL^{-1} Cam (Chloramphenicol).*

2.2.2 Cell Lysis and Clarification

Pelleted cells were resuspended in their respective resuspension buffers containing 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and/or a Roche protease inhibitor cocktail tablet. Lysis was achieved by one of three methods dependent on the purification scale. For small scale (10-20 mL), sonication was achieved using a 4710 series ultrasonic homogenizer CP50 (Cole-Parmer) at 50% amp for 10 seconds on, and 10 seconds off for a total of 6 minutes. For medium scale (25 – 50 mL), sonication or french press was utilised by either sonicating at 50-60% amp pulses (2 seconds on, 1 second off) for a total of 8 minutes (Sonics Vibracell VCX 130) or pressed at 16,000 psi using a pre-cooled pressure cell (Thermo French Press). These were repeated 2-3 times with incubations on ice to prevent heat damage. For large scale (>100 mL), cell disruption was achieved at 18,000 psi using the Microfluidizer LM20 with 2 – 3 cycles. Lysed cells were clarified by centrifugation (Beckman Coulter J26, JA 25.50 Rotor) at 40,000 g, 4 ºC for 45 – 60 minutes before filtering through a 0.22 or 0.45 µm Millipore™ Membrane Filter.

2.2.3 Affinity Chromatography

Affinity columns (Cytiva Life Sciences) were installed onto a bench-top peristaltic pump (Parnachia Biotech) at 20 °C or AKTA pure 2 system (Cytiva Life Sciences) at 5 °C. These were run at maximum speed, 4 or 5 column volumes (CV) per minute unless otherwise stated, and pre-washed with filtered H_2O before equilibrating in selected buffers absent any reducing agent. Samples were eluted up-flow where possible and monitored by UV absorbance at 280 nm, confirming peak fractions using SDS-PAGE analysis.

2.2.3.1 His-tagged proteins

Clarified lysed cells containing 10 – 20 mM imidazole, were loaded onto pre-packed 1 or 5 mL HisTrap™ HP or FF columns. These were washed with 5 CV of resuspension buffers or wash buffers before being eluted straight for 6 CV, stepwise (25, 50, 75, 100%) at 2 CV per step or by linear gradient for 8 CV in elution buffers containing 250 – 500 mM imidazole.

2.2.3.2 GST-tagged proteins

Clarified lysed cells were loaded onto pre-packed 1 or 5 mL GSTrap™ HP or FF columns. These were washed with 5 CVs of resuspension buffers or wash buffers before being eluted with 6 CVs of elution buffers containing 50 mM reduced glutathione.

2.2.3.3 StrepII-tagged Proteins

Clarified lysed cells containing 1 unit of avidin per litre of culture (BioLock IBA Life Sciences), were loaded onto pre-packed 5 mL StrepTrap™ XT at 0.5 CV per minute. This was washed with 7 CVs of resuspension buffer before being eluted with 8 CVs of 50 mM Biotin at 2.5 CVs per minute. Columns were refreshed with 25 mM NaOH.

2.2.4 Dialysis and Cleavage

Samples were buffer exchanged into appropriate dialysis buffers either by passing through a HiPrep™ 26/10 Desalting column (Cytiva Life Sciences) equilibrated with the desired buffer or by dialysis using a 3 or 10 kDa molecular weight cut-off SnakeSkin (Thermo Fisher Scientific) incubated in a 2 - 5 L reservoir, refreshing at least once in 24 hrs. For samples that required de-tagging, PreScission Protease (3C) or Thrombin (F2) were directly added to the dialysis solution at 1 – 2 units per mg of protein-dependent on protease required, or for on-column cleavage were loaded onto the column at 10 units per mg and incubated for 12 – 24 hrs.

2.2.4.1 Protease removal

The PreScission protease and thrombin cleaved samples were loaded onto a 1 mL GSTrap™ HP and HiTrap™ Benzamidine FF column (Cytiva Life Sciences) respectively and flow through was collected. Both columns were refreshed in 5 CV of either 50 mM reduced glutathione in PBS or glycine pH 3.0. All column use was followed as stated in Section 2.2.3.

2.2.5 Size Exclusion Chromatography

The samples for size exclusion chromatography (SEC) were concentrated (see Section 2.2.6) to 2 – 10 mL and filtered through a 0.22 µm membrane before being loaded into a 2 mL sample loop or 10 mL super loop installed onto the AKTA Pure 2. Using the Unicorn 7.0 software, a program was run of 2 mL injections per 120 mL elution onto either a HiLoad 16/600 Superdex 75 pg or 200 pg column (Cytiva Life Sciences) equilibrated with the selected gel filtration buffer. The elution was run at 0.5 – 1 mL per min and separated into 2 mL fractions whilst being monitored by UV absorbance at 280 nm, confirming peak fractions using SDS-PAGE analysis.

2.2.6 Concentration and Storage

Sample concentration was carried out through spin or pressure-based concentrating. For spin concentrating, a 5 or 10 kDa MW cut-off Vivaspin protein concentrator (GE Healthcare) was centrifuged (Beckman Coulter J-15R, JS-4.750 rotor) at 4,000 g for 10 to 20 minutes per spin with mixing. For pressure concentrating, a 50 – 300 mL stirrer cell (Amicon®) was used with 3 or 10 kDa MW cut-off Ultrafiltration discs (Ultracel®) under 50 psi of compressed air with stirring. Once purified, samples were stored for short term at 4 °C and for long-term samples were stored at – 80 °C after ‘snap-freezing’ by plunging into liquid nitrogen.

2.2.7 Protein-Specific Protocols

Many different complex constructs that were purified had poor stability and solubility issues; therefore, a detailed description of each purification strategy is given here, alongside their buffers (Table 2-3).

Table 2-3: Proteins and their respective buffers.

Recombinant Protein	Buffer	Contents
His-**hUba1**	Resuspension/Wash	20 mM Na_2PO_4 pH 7.4, 500 mM NaCl, 20 mM Imidazole, 1 x Roche tablet
	Elution	20 mM Na_2PO_4 pH 7.4, 500 mM NaCl, 500 mM Imidazole
	Dialysis	50 mM Tris.HCl pH 7.4, 1 mM DTT
GST-F2-**hUba1**	Resuspension/Wash	10 mM Na_2HPO_4 pH 7.4, 2 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl, 1 x Roche tablet
	Elution	10 mM Na_2HPO_4 pH 7.4, 2 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl, 50 mM Reduced
	Dialysis	10 mM Na_2HPO_4 pH 7.4, 2 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl
His-**UbcH7**	Resuspension/Wash	20 mM Na_2PO_4 pH 7.4, 500 mM NaCl, 20 mM Imidazole, 1 x Roche tablet
	Elution	20 mM Na_2PO_4 pH 7.4, 500 mM NaCl, 500 mM Imidazole
	Dialysis	50 mM HEPES pH 7.4, 150 mM NaCl, 1 mM DTT
GST-F2-**UbcH7a**	Resuspension/Wash	10 mM Na_2HPO_4 pH 7.4, 2 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl, 1 x Roche tablet
	Elution	10 mM Na_2HPO_4 pH 7.4, 2 mM KH_2PO_4, 2.7 mM KCl, 500 mM NaCl
His-Trx-3C-**WWP1-2L34H**	Resuspension/Wash	100 mM Tris pH 8.0, 250 mM NaCl, 10 mM imidazole, 1 mM PMSF
	Elution	100 mM Tris pH 8.0, 250 mM NaCl, 250 mM imidazole
	Dialysis	100 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 25 mM Maltose
	SEC	50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA
His-Trx-3C-**WWP1-L34H**	Resuspension	50 mM Tris pH 8.0, 500 mM NaCl, 10 mM imidazole, 1 mM PMSF
	Wash	50 mM Tris pH 8.0, 500 mM NaCl, 30 mM imidazole
	Elution	50 mM Tris pH 8.0, 500 mM NaCl, 250 mM imidazole
	Dialysis & SEC	50 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA
	Buffer Exchange	50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT
	Second SEC	50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA
GST-3C-**WWP2-LH**	Resuspension	25 mM Tris.HCl pH 8.0, 250 mM NaCl, 1 mM PMSF and 1 x Roche tablet.
	Wash	25 mM Tris.HCl pH 8.0, 250 mM NaCl, 0.1% Triton X-100
	Elution	25 mM Tris.HCl pH 8.0, 250 mM NaCl, 50 mM reduced glutathione.
	Dialysis	25 mM Tris.HCl pH 8.0, 250 mM NaCl, 5 mM DTT
	SEC	25 mM Tris.HCl pH 7.5, 150 mM NaCl, 5 mM DTT or 3 mM TCEP
His-F2-**WWP2-H**	Resuspension	50 mM Tris.HCl pH 8.0, 500 mM NaCl, 2 mM DTT, 10 mM Imidazole, 5 % Glycerol, 1 mM PMSF
	Wash	50 mM Tris.HCl pH 8.0, 500 mM NaCl, 2 mM DTT, 25 mM Imidazole, 5 % Glycerol
	Elution	50 mM Tris.HCl pH 8.0, 500 mM NaCl, 2 mM DTT, 300 mM Imidazole, 10 % Glycerol
	Dialysis	50 mM Tris.HCl pH 8.9, 150 mM NaCl, 5 mM DTT
	SEC	50 mM Tris.HCl pH 8.9, 150 mM NaCl, 1 mM TCEP
StrepII-**WWP2-H**	Resuspension/Wash	100 mM Tris.HCl pH 8.0, 300 mM NaCl, 2 mM DTT, 1 mM EDTA, 1 mM PMSF
	Elution	100 mM Tris.HCl pH 8.0, 300 mM NaCl, 2 mM DTT, 1 mM EDTA, 50 mM Biotin
	SEC	25 mM Tris.HCl pH 8.0, 150 mM NaCl, 3 mM TCEP
	Dialysis	10 mM Tris.HCl pH 8.0, 20 mM NaCl

2.2.7.1 His-tagged hUba1 and UbcH7

Pelleted cells from 1 L culture were resuspended in 35 mL buffer before being lysed, clarified, and purified through a 5 mL HisTrap™ FF column via straight elution. These were dialysed at 10 kDa cut-off in a 5 L reservoir overnight before spin concentrating and snap freezing in 50 µL aliquots.

2.2.7.2 His-tagged WWP2-H

Pelleted cells from 1 L culture were resuspended in 35 mL buffer before being lysed, clarified, and purified through a 5 mL HisTrap™ FF column, washing for 10 CV before finally eluting via a gradient. To chelate leaked nickel, 1 mM EDTA was added before dialysis. The sample was then dialysed at 10 kDa cut-off in a 5 L reservoir overnight before spin concentrating to 2 mL for gel filtration on a SEC 75 pg column. Eluted fractions were selected, and further spin was concentrated for snap freezing and storage.

2.2.7.3 His-tagged WWP1-L34H

Pelleted cells from 8 L culture were resuspended in 70 mL buffer before being lysed, clarified, and purified through a 5 mL HisTrap™ FF column, washing for 20 CV before finally eluting stepwise. The sample was then dialysed at 10 kDa cut-off in a 5 L reservoir overnight, filtering any precipitation before adding 5% glycerol and spin concentrating (10 °C) to 2 mL for gel filtration on a SEC 75 pg. PreScission protease was added at 2 units per mg to the eluted fractions and incubated overnight before being removed as stated in Section 2.2.4.1. Buffer exchange was performed using the HiPrep™ 26/10 Desalting column to remove EDTA, followed by a reverse HisTrap™. The flow through was spin concentrated to 2 mL and a second gel filtration was carried out on the SEC 75 pg. Eluted fractions were spin-concentrated and either immediately plated for crystallography or snap-frozen and stored.

2.2.7.4 His-tagged WWP1-2L34H

Pelleted cells from 8 L culture were resuspended in 70 mL buffer before being lysed, clarified, and purified through a 5 mL HisTrap™ FF column, washing for 20 CV before finally gradient eluting over 10 CV. The sample was dialysed at 10 kDa cut-off in a 5 L reservoir overnight and filtered to remove any precipitation. PreScission protease was added at 1.5 units per mg and incubated during further dialysis. The protease was removed as stated in Section 2.2.4.1, before the sample was spin concentrated to 2 mL and loaded onto SEC 200 pg. Eluted fractions were spin-concentrated and either immediately plated for crystallography or snap-frozen and stored.

2.2.7.5 GST-tagged WWP2-LH

Pelleted cells from 4 L culture were resuspended in 35 mL buffer before being lysed, clarified, and purified through a 5 mL GSTrap™ FF column on a benchtop, collecting the straight elution on ice. PreScission protease was added at 1 unit per mg, and incubated during dialysis at 10 kDa cut-off in a 5 L reservoir overnight, before being refreshed for a further 4 hrs. A reverse GSTrap was used as in Section 2.2.4.1, but further removing uncleaved and free GST. The flow-through sample was spin-concentrated to 2 mL for gel filtration on a SEC 75 pg column. Eluted fractions were selected and again passed through the reverse GSTrap before adding 5% glycerol. Samples were kept at 20 °C for further spin concentrating and either immediate use in crystallography or snap-frozen and stored at -80 °C.

2.2.7.6 GST-tagged hUba1

Pelleted cells from 1 L culture were resuspended in 35 mL buffer before being lysed, clarified, and purified through a 1 mL GSTrap™ HP column via straight elution. The sample was then dialysed at 3 kDa cut-off in a 2 L reservoir overnight. Aliquots of 20 µL were snap-frozen and stored.

2.2.7.7 GST-tagged UbcH7

Pelleted cells from 2 L culture were resuspended in 35 mL buffer before being lysed, clarified, and loaded onto a 1 mL GSTrap™ HP column. On-column cleavage was achieved by incubating with 20 units. mL^{-1} of thrombin overnight. Untagged UbcH7 was collected by washing the column with 7 CVs of high salt PBS (500 mM NaCl) before passing through a 1 mL HiTrap Benzamidine column (Cytiva), removing thrombin. The aliquoted sample was snap-frozen and stored.

2.2.7.8 StrepII-tagged WWP2-H

Pelleted cells from 4 L culture were resuspended in 35 mL buffer containing 4 units of avidin before being lysed, clarified, and purified through a 5 mL StrepTrap™ XT column via straight elution. Selected fractions were pressure concentrated to 10 mL, loaded into a superloop, and sequentially injected onto a SEC 75 pg column. Eluted fractions were selected, and pressure concentrated before dialysing overnight for immediate use in ITC.

2.2.8 SDS-PAGE Analysis

Various samples were collected at relevant purification stages and diluted to the appropriate concentration for SDS-PAGE analysis. Coomassie InstantBlue (Expedeon) was used to stain gels overnight before de-staining in H_2O. Images were collected using a BioRad Universal Hood II Gel Doc System, and enhanced using the Image Lab software (BioRad) for optimal contrast.

2.2.8.1 Lab-cast acrylamide gels

A 20 µL aliquot of each diluted sample was prepared by boiling with Laemmilli at an equivalent volume, containing 5% glycerol and 10 mM DTT. Analysis was achieved on 10, 15 and 20% acrylamide gels (Table 2-4) run in a Mini-PROTEAN tank using a 1% SDS running buffer at 180 V, for approximately 45 minutes.

Table 2-4: Lab-cast acrylamide gel recipes.

Materials	6 % Stacking	10 % Separating	15 % Separating	20 % Separating
	Volume added (µL)			
H_2O	2900	3800	2800	1800
40% Acrylamide	750	2000	3000	4000
1.5 M Tris pH 8.8	-	2000	2000	2000
0.5 M Tris pH 6.5	1250	-	-	-
10% SDS	50	80	80	80
10% APS	50	80	80	80
TEMED	5	8	8	8

2.2.8.2 Bolt™ pre-cast acrylamide gels

A 19.5 µL aliquot of each diluted sample was prepared by heating in a thermal cycler at 75 °C for 10 minutes in the presence of 7.5 µL 1 × LDS Sample Buffer and 1 × Sample Reducing Agent before 20 µL was loaded onto pre-cast 4 – 12% BIS-Tris Plus acrylamide gels in an Invitrogen Mini Gel Tank using 1 × MES SDS Running Buffer at 165 V for 30 minutes.

2.3 Biochemical and Biophysical Techniques

2.3.1 In Vitro Autoubiquitination Assays

All assays and plate coating were carried out at 20 °C (room temperature) unless stated otherwise.

2.3.1.1 Glutathione coated 96-well plates

Protocols follow methods created by Murray *et al*. (1998). Corning™ Costar™ EIA 96-well plates were coated with 100 µL of 2% bovine haemoglobin solution containing 50 mM Na_2CO_3 and incubated overnight at 4 °C. The wells were washed three times with PBS before adding a heterobifunctional cross-linker sulphosuccinimidyl 4-(p-maleimidophenyl)butyrate (SSMPB) to a final concentration of 0.1 mM in 50 µL PBS per well, incubating for 1 hour. The wells were again washed three times with PBS before adding 50 µL of 10 mM reduced glutathione in 10 mM Na_2HPO_3 pH 7.4, 150 mM NaCl, 1 mM EDTA and incubated overnight. The plates were finally washed three times with PBST (PBS containing 0.1% Tween) and stored in this solution at 4 °C.

2.3.1.2 GST-immobilised WWP2 autoubiquitination assay

GST-tagged WWP2-FL pelleted cells from 50 mL culture were resuspended in 15 mL PBS and lysed by sonication. The cells were then centrifuged (Heraeus Megafuge 1.0R) at 4,000 g, 4 °C for 15 minutes before adding 100 µL of clarified lysate to the 96-well glutathione-coated plates and incubated for 1 hour. During which His-hUba1, His-UbcH7, FLAG-ubiquitin and ATP at 10 ng, 150 ng, 60 ng and 1.25 mM per well, respectively, were incubated together in 25 mM Tris pH 8.0, 100 mM NaCl, 4 mM $MgCl_2$, for 40 minutes to form the reaction mixture. The plate was then washed three times with PBST and tapped dried before 2 µL of the compound was added at the desired concentration followed by 18 µL of the reaction mixture. This was then incubated for 2 hours with 0% and 100% controls. The plate was again washed three times with PBST before 100 µL of anti-FLAG M2-Peroxidase HRP (1:10000 PBST) was added to each well and incubated for 1 hour. The plate was finally washed a further three times with PBST before 100 µL of 1 × TMB substrate solution (Invitrogen) was added to each well and incubated for up to 10 minutes until sufficient blue colour change was observed. To stop the reaction, 100 µL of 1 M HCl was added.

2.3.1.3 His-immobilised WWP1 autoubiquitination assay

His-tagged WWP1-L34H pelleted cells from 50 mL culture were resuspended in 15 mL PBS and lysed by sonication. The cells were then centrifuged (Heraeus Megafuge 1.0R) at 4,000 g, 4 °C for 15 minutes before adding 100 µL of clarified lysate to the 96-well Nickel-coated plates (Pierce™ Thermo Fisher) and incubated for 1 hour. Plates were then washed three times with PBST containing 15 mM imidazole. Plate blocking was carried out with the addition of 1% Bovin Serin Albumin (BSA) in PBST after incubating for 1 hour. During which GST-hUBa1, untagged-UbcH7, FLAG-ubiquitin and ATP at 3 ng, 15 ng, 50 ng and 1.25 mM per well, respectively, were incubated together in 25 mM Tris pH 8.0, 100 mM NaCl, 4 mM MgCl2, for 40 minutes to form the reaction mixture. The plate was then washed three times with PBST containing 15 mM imidazole and tapped dried before 2 µL of the compound was added at the desired concentration followed by 18 µL of the reaction mixture. This was then incubated for 2 hours with 0% and 100% controls. All other steps followed as the 2.3.1.2 GST-immobilised autoubiquitination assay.

2.3.1.4 Uba1 counter-assay

GST-hUba1 pelleted cells from 50 mL culture were resuspended in 15 mL PBS and lysed by sonication. The cells were then centrifuged (Heraeus Megafuge 1.0R) at 4,000 g, 4 °C for 15 minutes before adding 100 µL of clarified lysate to the 96-well glutathione-coated plates and incubated for 1 hour. The plate was then washed three times with PBST and tapped dried before adding 20 µL reaction mixture containing 50 ng/well FLAG-ubiquitin, 1.25 mM ATP and 1 mM 1% DMSO compound in 25 mM Tris pH 8.0, 100 mM NaCl, 4 mM $MgCl_2$. This was then incubated for 2 hours to determine the activity of hUba1-Ub conjugation with 0 % (No hUba1 lysate) and 100 % (DMSO only, no compound) controls. All other steps followed as the 2.3.1.2 GST-immobilised autoubiquitination assay.

2.3.1.5 UbcH7 counter-assay

The reaction mixture containing 3 ng/well GST-hUba1, 200 ng/well His-UbcH7, 60 ng/well FLAG-ubiquitin, 1 mM ATP and 1 mM 1% DMSO compound was incubated in 25 mM Tris pH 8.0, 100 mM NaCl, 4 mM $MgCl_2$ for 1 hour. Clear 96-well Nickel-coated plates (Pierce™ Thermo Fisher) were rinsed three times with PBST, before loading 20 µL of reaction mixtures and incubated for a further 1 hour to determine the activity of UbcH7-Ub conjugation with 0 % (His-UbcH7 only) and 100 % controls (DMSO only, no compound). Plates were washed three times with PBST containing 15 mM imidazole, with all other quantification steps followed as the 2.3.1.2 GST-immobilised autoubiquitination assay.

2.3.1.6 Plate reading and data processing

The absorbance of the activity assays was read at 450 nm using a Hidex sense microplate reader, with the raw results processed in Excel. These were first normalised to a background noise taken from the 0 % control before being standardised to the 100 % control.

$$Enzyme\ Activity\ (\%) = \frac{Observed\ (A) - Background\ (A)}{Control\ (A) - Bacground\ (A)} X\ 100$$

The IC_{50} non-linear regression curves were then calculated in GraphPad from the standardised activity percentages, plotted against the Log10 ligand concentration using a variable slope model (four-parameter dose-response curve) with the bottom constrained to 0 %.

$$Enzyme\ Activity\ (\%) = Bottom\ (0) + \frac{Top - Bottom\ (0)}{1 + 10^{(LogIC50(\mu M) - \mathrm{Log}[Ligand](\mu M)) \times Hillslope}}$$

2.3.2 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) was used in both ligand-binding analysis and buffer stability screening against various proteins.

2.3.2.1 Ligand-binding analysis

A 96 well-plate (MicroAmp Optical) was loaded with 18 µL of 2.5 µM WWP2-LH, 1 µM WWP1-L34H and 3.8 µM WWP1-2L34H in their respective final buffers (Table 2-3) containing 5 × SYPRO orange dye. A 2 µL aliquot of compound was added to a final concentration ranging from 10 – 100 µM containing 0.1 % DMSO before the plate was sealed (MicroAmp Clear Adhesive Film). Both non-protein and DMSO controls were also generated. The plates were briefly centrifuged before the assay was run using either an ABI 7500 standard or fast RT-PCR machine following the melt curve using ROX™ (575 nm) as the 'pre-set' fluorescence dye. A standard thermal profile of 25 – 70 °C, rising at 0.5 °C per minute was used. The mid-point melting temperature (Tm) was calculated from either the first derivative or Boltzmann fit to the fluorescence curve using the Protein Thermal Shift Software v1.4 (Thermal Fisher). Results were further processed and graphed in Excel.

2.3.2.2 Buffer stability screening

Four commercial screens (RUBIC buffer, RUBIC additive, Durham salt and Durham pH) were purchased from Molecular Dimensions and their protocols were followed to assess the thermal

stability of WWP1-2L34H. The WWP1-2L34H custom screen can be found in Table 2-5, and used the following steps in 2.3.2.1.

2.3.3 Isothermal Titration Calorimetry

A 96 deep-well plate (NUNC) was loaded with 20 – 30 µM WWP2-$H^{StrepII}$ and 200 – 500 µM compounds in separate wells, both solutions having matching dialysis buffer containing equivalent DMSO (0.2 – 0.5 %). Plates were covered with slit seal plate film (Malvern Panalytical) and briefly centrifuged before loading into the MicroCal PEAQ-ITC Automated (Malvern Panalytical) pre-chilled to 10 °C. A standard 13 – 19 injection protocol was followed at 25 °C using 2 – 3 µL ligand injections over 4 seconds, stirring at 750 rpm with a spacing of 150 seconds. Between runs, the cell was pre-rinsed with buffer containing equivalent DMSO and the syringe was cleaned following the automated method. Thermal traces were analysed using the MicroCal PEAQ-ITC Analysis Software 1.41v.

Table 2-5: The His-tagged WWP1-2L34H DSF custom buffer screen.

	1	2	3	4	5	6	7	8	9	10	11	12
A	**100 mM NaH_2PO_4 pH 7.0**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl	**150 mM NaH_2PO_4 pH 7.0**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl
B	**100 mM Na_2HPO_4 pH 7.4**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl	**150 mM Na_2HPO_4 pH 7.4**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl
C	**50 mM Tris pH 8.0**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl	**100 mM Tris pH 8.0**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl
D	**50 mM HEPES pH 7.5**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl	**100 mM HEPES pH 7.5**	50 mM NaCl	100 mM NaCl	150 mM NaCl	250 mM NaCl	500 mM NaCl
E	**100 mM NaH_2PO_4 pH 7.0, 250 mM NaCl** / DTT 1 mM	5 mM DTT	1 mM EDTA	5 mM EDTA	25 mM Sucrose	25 mM Maltose	50 mM L-Arginine	100 mM L-Arginine	500 mM L-Arginine	5% Glycerol	10% Glycerol	Water (Control)
F	**100 mM Na_2HPO_4 pH 7.4, 250 mM NaCl** / DTT 1 mM	5 mM DTT	1 mM EDTA	5 mM EDTA	25 mM Sucrose	25 mM Maltose	50 mM L-Arginine	100 mM L-Arginine	500 mM L-Arginine	5% Glycerol	10% Glycerol	Water (Control)
G	**50 mM Tris pH 8.0, 250 mM NaCl** / DTT 1 mM	5 mM DTT	1 mM EDTA	5 mM EDTA	25 mM Sucrose	25 mM Maltose	50 mM L-Arginine	100 mM L-Arginine	500 mM L-Arginine	5% Glycerol	10% Glycerol	Water (NPC)
H	**50 mM HEPES pH 7.5, 250 mM NaCl** / DTT 1 mM	5 mM DTT	1 mM EDTA	5 mM EDTA	25 mM Sucrose	25 mM Maltose	50 mM L-Arginine	100 mM L-Arginine	500 mM L-Arginine	5% Glycerol	10% Glycerol	Water (NPC)

2.3.4 X-ray Crystallography

2.3.4.1 Protein crystallisation

Crystals were grown at 16 °C in either a 96-well MRC 2 sitting drop (SWISSCI™) or a 15-well EasyXtal hanging drop plate (QIAGEN™). For the sitting drop, the dual 1 µL drops were plated using the OryxNano robotic dispenser (Douglas Instruments) programmed to mix 0.4 - 0.6 µL protein to 0.4 – 0.6 µL reservoir with the adjacent well containing 100 µL of reservoir solution. Plates were sealed using ClearVue™ sheets (Molecular Dimensions). For hanging drop, DG X-Seal crystal supports were loaded with 2 – 4 µL drops, manually mixing in equivalent volumes of protein to the reservoir, before securing into wells containing 500 µL of reservoir solution. Protein-specific crystallisation conditions (Table 2-6) were roughly screened with varying pH, crystallisation agent and salt concentration.

Table 2-6: Protein crystallisation conditions.

Protein	Concentration (mg/mL)	Protein Buffer	Reservoir Solution	Cryo Protectant
WWP2-LH	1 – 2	25 mM Tris.HCl pH 7.5, 150 mM NaCl, 5 mM DTT	100 mM MMT pH 6.0, 25% PEG1500	20% Glycerol (mother liquor)
WWP2-LH (I3C)	1 – 2 5 mM I3C, 2.5% DMSO	25 mM Tris.HCl pH 7.5, 150 mM NaCl, 5 mM DTT	200 mM sodium acetate trihydrate, 100 mM Bis-Tris propane pH 7.5, 20% PEG 3350.	20% Glycerol (mother liquor)
WWP2-LH (I4) †	1 – 2 1 mM I4, 1% DMSO	25 mM Tris.HCl pH 7.5, 150 mM NaCl, 5 mM DTT	100 mM MES pH 6.0, 150 mM $CaCl_2$, 16% PEG6000	30% Glycerol (mother liquor)
WWP1-L34H	3 – 10	50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA	100 mM sodium malonate pH 5.0, 12% PEG3350	25% Glycerol (mother liquor)
WWP1-2L34H	3 - 4	50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA	100 mM Tris.HCl pH 6.5, 17.5% reagent alcohol (90% EtOH, 5% MeOH, 5% IPA)	30% Glycerol (Reservoir × 5%)

† *Small-molecule Crystal*

2.3.4.2 Co-crystallisation, commercial screening and seeding

Co-crystallisation was achieved by first incubating the selected compound with protein (Table 2-6) for 1 – 3 hours at 16 °C before centrifuging at 20,000 g for 5 minutes to remove any precipitation. Subsequent supernatant was then plated against commercial screens following stated protocol (2.3.4.1). Various commercial screens were used to search for new crystallisation conditions consisting of Structure 1 and 2 ECO, LFS ECO, JCSG-*plus*™, MIDAS-*plus*™, PACT *premier*™, Morpheus® and LMB screen™ (Molecular Dimensions) as well as Index™, PEG/Ion Screen™ and PEGRx™ screens (Hampton Research). To improve crystals with poor morphology, crystal seeding was achieved using the Seed Bead™ kit (Hampton Research) following the standard protocol, seeding either directly at various Log10 dilutions or using Multi-Matrix Seeding (MMS) via the OryxNano robotic dispenser.

2.3.4.3 Crystal harvesting and ligand soaking

Single crystals were harvested at 16 °C using 0.1 – 0.2 mm round LithoLoops (Molecular Dimensions). Apo crystals were placed into a 1 µL drop of cryoprotectant (Table 2-6) before snap-freezing in liquid nitrogen and loading into a UniPuck ready for transportation. For ligand-soaking, harvested crystals are placed into a 1 µL drop of mother liquor containing either 1, 2.5 or 25 mM compound (1 or 2.5% DMSO) and incubated at either 3, 24 hours or 7 days. Cryo protectant was then either added directly to the drop or crystals further transferred to a fresh cryo protectant drop containing equivalent compound before snap-freezing for transport.

2.3.4.4 X-ray data collection and processing

Protein crystals were sent to the synchrotron at Diamond Light Source for data collection using either the I04 or I24 beamline. The auto-processed data from Xia2 (Winter, 2010), DIALS (Winter *et al.*, 2022) or AutoProc (Vonrhein *et al.*, 2011) pipelines were downloaded for subsequent molecular replacement, refinement and ligand fitting using a combination of the CCP4 software suit v8.0 (Agirre *et al.*, 2023), PHENIX (Liebschner *et al.*, 2019) and COOT (Emsley *et al.*, 2010), with integrated PHASER (McCoy *et al.*, 2007), and REMAC (Vagin *et al.*, 2004) programs, with final models represented using PyMOL v2.5 (Delano, 2002) or ChimeraX v1.9 (Goddard *et al*., 2018; Pettersen *et al*., 2021; Meng *et al*., 2023).

2.3.5 In Silico Molecular Modelling

2.3.5.1 Cresset flare docking

Molecular docking was performed using Cresset Flare v8.0 utilising the Lead Finder program for docking calculations (Cheeseright *et al*., 2006; Stroganov *et al*., 2008). For protein and ligand preparation the refined structures of WWP1-2L34H (PDB ID: 9EQK) and WWP2-LH (PDB ID: 9EQH) were imported into Cresset Flare. Protein preparation was carried out using the default settings at pH 7.4. The molecular structures of the ligands identified in this project were energy-minimized with charges at pH 7 before docking in Flare (Cheeseright *et al.*, 2006). The grid box for WWP1-2L34H was defined according to the proposed binding site cavity close to the active site cysteine (Cys890), via picking of amino acids Tyr639, Cys640, Ser679, Ser698, Ile699, Thr851, Gly852, Thr853, Thr889 and Cys890. The grid box for WWP2-LH was similarly defined surrounding the active site cysteine, Cys838, via picking of key amino acids Tyr587, Cys588, Thr627, Ser646, Ile-647, Thr-799, Gly-800, Thr-801, Thr-837 and Cys-838. Generated poses were remodelled using PyMOL v2.5 (Delano, 2002).

2.3.5.2 Schrödinger glide docking

Molecular docking was performed using the Schrödinger Suite 2020-3. The protein structures of NEDD4 HECT (PDB ID: 5C91)(Kathman *et al.*, 2015), as well as WWP1 (PDB ID: 9EQK) and WWP2 (PDB ID: 6J1Z) (Wang *et al.*, 2019) both containing the HECT and the WW2 domains, were prepared using the Schrödinger's Protein Preparation Wizard module (Epik v5.5, Impact v8.8) (Madhavi Sastry *et al.*, 2013; Johnston *et al.*, 2023). I3C, DIM and I15/N-Tosyl-I3C were prepared using LigPrep v5.5 (Epik v5.3)(Johnston *et al.*, 2023). Default settings were used for both proteins and ligands at pH 7.0 ± 0.2, removing all waters and adding hydrogen atoms. Both WWP1 and WWP2 structures and ligands were aligned to NEDD4 and its covalent I3C analogue, before performing minimisation to both the ligand and residues surrounding an 8 Å radius. This was achieved using the OPLSe force field in MacroModel v12.9 at a default 2500 iterations (Jorgensen and Tirado-Rives, 1988; Mohamadi *et al.*, 1990; Jorgensen, Maxwell and Tirado-Rives, 1996; Shivakumar *et al.*, 2010; Watts *et al.*, 2014; Harder *et al.*, 2016; Roos *et al.*, 2019). The ligands were then re-docked into the minimised pseudo bound structures using the Glide SP v8.8 program with grids generated from the individual minimised ligand positions (Friesner *et al.*, 2004; Halgren *et al.*, 2004). Default settings were used with the top five poses generated, enabling post-ligand minimisation before being ranked and binding affinity given as GlideScore. Figures were created in 2D using the Schrodinger Ligand Interaction Diagram module, with 3D molecular models generated using PyMOL v2.5 (Delano, 2002).

Chapter 3: Discovery of Novel WWP1 and WWP2 Inhibitors

3.1 Introduction

The NEDD4 family of HECT E3 ligases are known to play fundamental roles in tumour initiation and progression, and as such are considered to be promising therapeutic targets (Chen and Matesic, 2007). Despite this, there is only a limited number of small-molecule inhibitors currently identified, with even less known to inhibit WWP1 and WWP2 (Chen, Gehringer and Lorenz, 2018). For WWP1, the commercially available Heclin is the only small molecule to demonstrate in vitro inhibition recording an IC_{50} of 6.9 µM. Yet, as mentioned Heclin is a non-specific HECT inhibitor, having similar inhibition to a handful of other HECT E3 ligases (Mund *et al.*, 2014). The only other small molecule known to potentially inhibit WWP1 is the natural metabolite I3C, originally found to inhibit NEDD4-1; this is further discussed in Chapter 4. Although found to interact, various groups have only indicated its potential inhibition through in vivo cell proliferation studies (Lee *et al.*, 2019; Zheng *et al.*, 2023). For WWP2, a handful of potential small molecule inhibitors have been highlighted in a screen of the National Cancer Institute (NCI) Diversity Set V compound library reported previously by the Chantry group (Watt *et al.*, 2018). NSC-288387 with an IC_{50} of 2.3 µM was the main focus of the paper, with various assays including saturation transfer difference (STD) NMR highlighting its interactions with WWP2. However, the most potent compound reported was NSC-2805 with an IC_{50} of 0.38 µM.

It was clear that the inhibitor space for WWP1 and WWP2 required expansion, so a hit-discover strategy was undertaken. When screening for novel small-molecule inhibitors studies tend to use compound libraries ranging from thousands to hundreds of thousands of compounds requiring high throughput screening techniques. These libraries are designed to screen a variety of chemical space, either completely diversified to maximise the volume searched or targeted to the enzyme of choice (Harris *et al*., 2011). Some libraries can be designed around other characteristics such as known natural products, FDA-approved compounds, or fragment-sized compounds for a fragment-based drug discovery (FBDD) approach (Lahlou, 2013; Shi and von Itzstein, 2019; Wan *et al.*, 2021). In either case, the compounds are typically designed to fulfil the Lipinski rule of 5 used to evaluate a compound's drug-likeness for increased oral activity (Lipinski *et al*., 2001). These general rules include a molecular weight of less than 500 Da, 5 or fewer hydrogen bond donors, 10 or fewer hydrogen bond acceptors, and a calculated n-octanol-water partition coefficient (Clog P) of less than 5. Chosen for the previous WWP2 screen, the NCI Diversity compound library is derived from 140,000 compounds from the National Institutes of Health NCI Developmental Therapeutics Program repository, with each set (~2000) deduced based on a variation of chemical characteristics. Given their variation and sample size the NCI Diversity sets are good starting points for academic screening efforts whereby

automated processes are not available (Zweifach, 2020). The 'Next in Series' NCI Diversity Set VI (1584 compounds) was designed to screen over 1,000,000 possible three-dimensional pharmacophore arrangements selected to be amenable for structure-based approaches following specific selection rules; compounds must contain 5 or more new pharmacophore centres, must be relatively rigid with 5 or fewer rotatable bonds and have a tendency to be planar with 1 or less chiral centres. Given we wanted to later move into X-ray crystallographic studies, this set was ideal.

One issue with the NCI library is the significant presence of Pan-assay interference compounds (PAINS), broadly defined as compounds containing weak electrophilic or nucleophilic moieties that can result in off-target interactions (Baell and Holloway, 2010; Baell and Nissink, 2018). Examples of these properties are shown in Figure 3-1. As the PAINS annotate indicates, compound screening often faces these adversities leading to false positives that can later result in a significant amount of wasted labour and materials in the lead development stage. To avoid this, screening strategies require these moieties to be filtered out before or early in the screening process, typically using virtual libraries such as ZINC 15 (Sterling and Irwin, 2015).

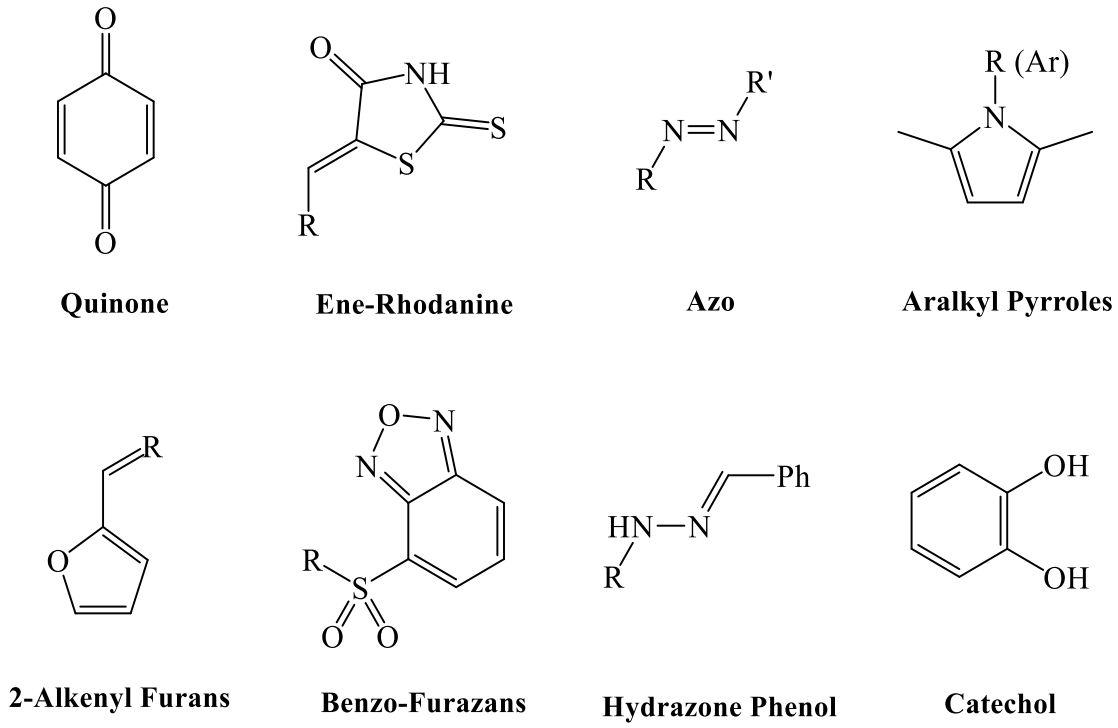


Figure 3-1: Examples of Pan-assay interference compound (PAINS). Images created using ChemDraw.

The screening assay of choice is also important, and dependent on a range of factors including output, cost, and sensitivity requirements as well as the specific interactions being assayed (binding or activity). For HECT E3 ligases, high-throughput studies have mainly focussed on an array of autoubiquitination assays as a standard protocol for measuring E3 ligase activity. These assays can be commercially purchased and include western blot, enzyme-linked immunosorbent (ELISA)-based assays, as well as tandem ubiquitin-binding entities (TUBEs), and time-resolved fluorescence resonance energy transfer (TR-FRET)-based assays with each type having their own benefits and limitations (Marblestone *et al.*, 2010, 2012; Lin, Zhang and Xie, 2023).

Western blotting is a commonly utilised assay outside of ubiquitination, however, can be used to specifically measure the change in the Ub state of a target protein as observed through a band-shift in SDS-PAGE. The target protein itself is identified through antibody detection, enabling its specific Ub state to be compared after a Ub reaction (E1, E2 and E3) in both the absence and presence of potential inhibitors. Such an approach benefits from its ability to be used both in vitro and in vivo through affinity pulldowns such as immunoprecipitation (Sigismund and Polo, 2016). However, this approach is limited by its low throughput and minimal quantitative analysis as well as the requirement for the generation of target-specific antibodies, an expensive process. Despite this, several antibodies have been suggested to target specific Ub linkages, adding the benefit of substrate Ub linkage determination (Matsumoto *et al*., 2018).

This chain linkage specificity is also demonstrated by TUBEs, an array of engineered UBDs that can bind in the nanomolecular affinity to polyUb chains, either as pan-selective TUBEs (non-selective) or as chain-selective TUBEs (Hjerpe *et al*., 2009). As such, TUBEs can be used as an alternative to antibody detection in a Western blotting-like approach but are also used for high-throughput strategies in pulldown assays, where TUBEs can be pre-coated onto multi-well plates (Kadimisetty *et al*., 2021). Although sometimes beneficial that TUBEs can distinguish between mono- and polyUb chains, this is commonly seen as a limitation with ~50% of ubiquitination found in the mono-Ub state and thus many of these events are undetected due to their low affinity towards mono-Ub (Mattern *et al.*, 2019).

As in the name, ELISA-based assays exploit antigen-antibody interactions, whereby antibodies are typically recombinantly linked to a catalytic enzyme. In the case of ubiquitination assays, the antigen is typically flag-tagged Ub, whereby after antibody incubation the addition of the recombinant enzyme-substrate results in a measurable colour change that then can be associated with the amount of ubiquitination that has taken place, see section 3.1.1.2 (Alhajj, Zubair and Farhana, 2023). This assay benefits from high-throughput capabilities, as well as its colourimetric analysis, and can further utilise TUBEs as an alternative for the selection of PolyUb chains. However, this assay is limited to its involvement of an alternative catalytic reaction, which in itself can be affected by small molecule inhibitors.

Despite this, the main limitation of these assays is that they detect ubiquitination once it has been completed, potentially missing key information such as rate changes. TR-FRET assays can be used to overcome this, whereby the Ub chains are either conjugated to alternating fluorescence acceptors (e.g. fluorochrome) and donators (e.g. terbium), or fluorescent donor TUBEs, leading to a strong FRET signal as PolyUb chains are formed in real-time (Madiraju *et al*., 2012). In the case of small molecule inhibitor screening, FRET interference can be an issue and is a limitation of such an assay as well as the expense associated with the conjugated Ub.

To overcome expense issues, cheaper less robust assays can first be used to reduce a compound library size. DSF is commonly utilised in this manner and has proven successful in a study of the E3 ligase NEDD4-1, looking at the binding Indole-3-carbinol-based inhibitors before using the more traditional ELISA-based autoubiquitination assay to screen for inhibition (Quirit *et al.*, 2017). This assay is further detailed below in section 3.1.1.1. Our work looked to follow this strategy, having proved successful against another member of the NEDD4 family, known to have similar homology to both WWP1 and WWP2. Furthermore, the classical ELISA-based assay had already been optimisation for both WWP1 and WWP2 in the previous screen reported by the Chantry group (Watt *et al.*, 2018).

3.1.1 Biophysical Techniques

3.1.1.1 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF), previously known as Thermofluor is a simple and widely used biophysical technique for identifying ligand interactions with target proteins. Generally known as a type of thermal shift assay (TSA), this technique measures changes in the thermal stability of the protein in the presence of varying conditions (Huynh and Partch, 2015). Although this can be ligands/inhibitors, other conditions such as buffers, pH and even mutations can be investigated. In DSF, a fluorogenic dye enables the thermal denaturation of a protein to be traced upon unfolding and exposure of its hydrophobic surfaces. The binding of the dye to these hydrophobic regions results in a characteristic fluorescence and when read against increasing temperature leads to a sigmoidal curve as seen in Figure 3-2. The term ‘melting temperature’ (Tm) is the midpoint of this curve and is determined in one of two ways; either the peak of the curve’s first derivative or non-linear fitting to the Boltzmann equation. In the case of ligand identification, DSF assays exploit the idea that interacting molecules can stabilise or in truth destabilise the thermal stability of target proteins resulting in a shift in Tm (Pantoliano *et al.*, 2001).

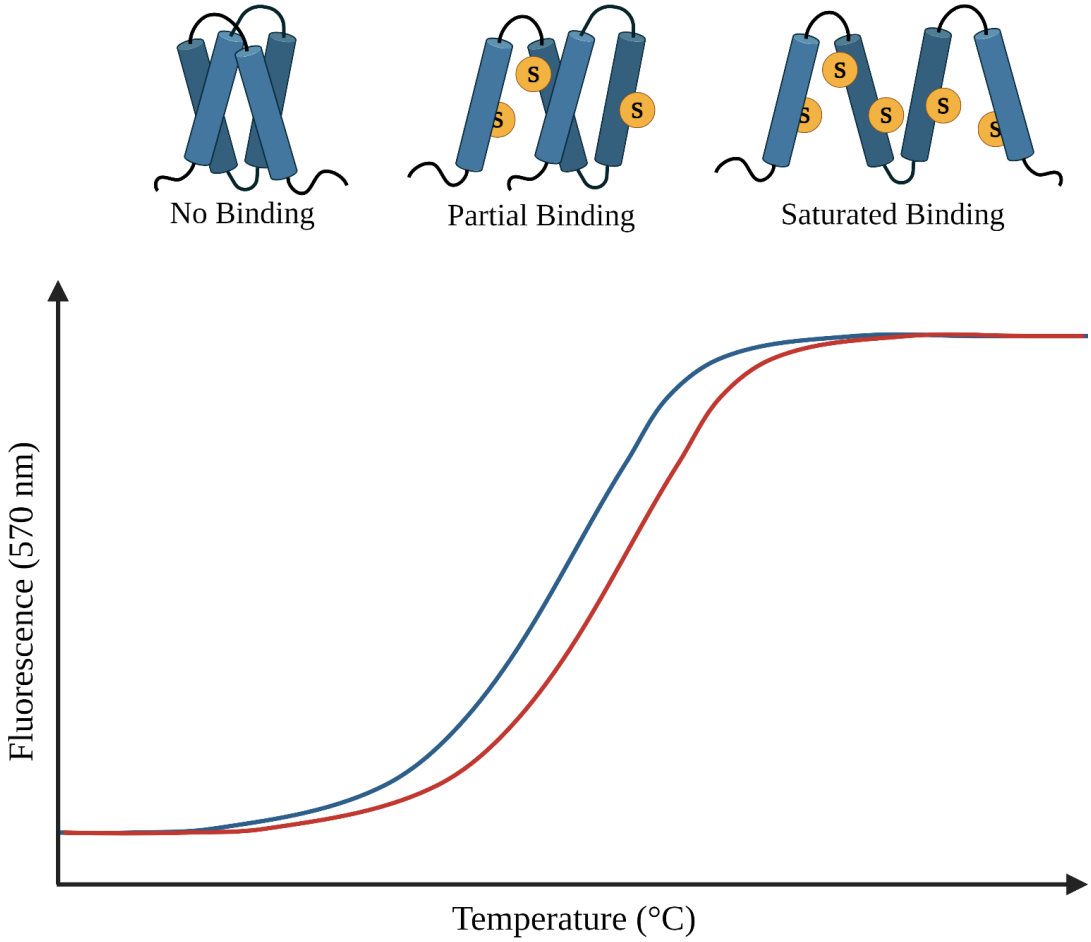


Figure 3-2: The basic principles of DSF. Schematic diagram of Sypro orange (S circles) saturating the protein as it begins to thermally denature (top), and the thermal fluorescence profile of a protein in the absence (blue) and presence (red) of a ligand (bottom).

3.1.1.2 Enzyme-linked immunosorbent autoubiquitination assay

The ELISA autoubiquitination assay is one of the few high-throughput methods utilised to measure the ubiquitination activity of E3 ligases. In the case of autoubiquitination assays, the antigen is recombinant flag-tagged Ub, with the associated anti-flag antibody linked with horseradish peroxidase (HRP). The addition of 3,3',5,5'-tetramethylbenzidine (TMB) results in its oxidation forming a blue colour before acidification stops the reaction and produces a characteristic yellow fluorescent at 450 nm (Josephy, Eling and Mason, 1982). This assay takes advantage of the E3 ligase autoinhibitory mechanism whereby overactivity results in their self-ubiquitination and therefore downregulation in the cell (Bie and Ciechanover, 2011). By first immobilising an E3 ligase to a 96-well plate, a ubiquitination cascade can be induced with the associated components (E1, E2, Flag-tagged Ub, ATP). Simply washing off these components after a set period leaves the E3 ligase and covalently attached flag-tagged Ub chains as part of

the autoubiquitination process. The assay therefore can associate TMB's fluorescence to the amount of autoubiquitination that has taken place, and as such can also be compared in the absence and presence of potential inhibitors. A visual representation of this process can be seen below in Figure 3-3 using GST immobilised WWP2 as an example.

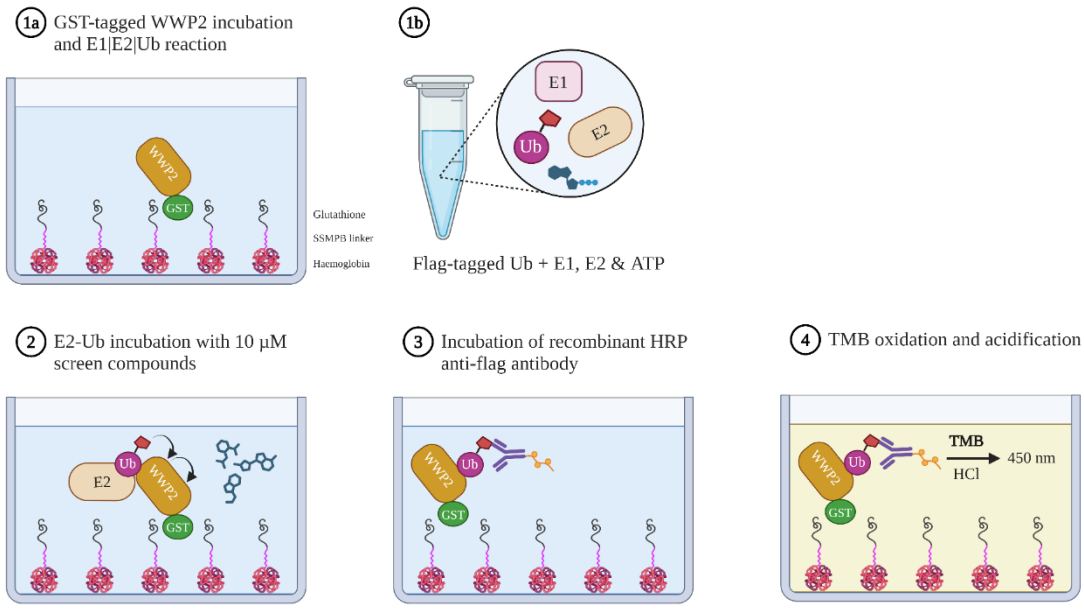


Figure 3-3: The WWP2 ELISA autoubiquitination assay. (1a) Immobilisation of GST-Tagged WWP2 on glutathione coated plates. The GST-tagged (green) WWP2-FL (orange) is bound to glutathione (black), further linked via an SSMPB linker (pink) bound to haemoglobin (red). (1b) Reaction mixture of His-E1, His-E2 and flag-tagged Ub. Flag-tagged (red) Ub (dark pink) is first charged to Uba1/E1 (light pink) using ATP (blue) and then transferred to UbcH7/E2 (light brown). (2) Ub charged E2 incubation with immobilised WWP2. UbcH7/E2 transfers flag-tagged Ub to WWP2 before WWP2 auto-ubiquitination occurs in the absence or presence of compounds (blue). (3) Incubation of recombinant HRP anti-flag antibody with auto-ubiquitinated WWP2. Anti-flag antibody (purple) containing recombinant HRP (orange) associates with the flag-tagged Ub. (4) HRP-mediated oxidation of TMB. Single-electron oxidation of TMB by HRP in the presence of H_2O_2 creates first a blue colour change, before acidification results in a yellow colour change (450 nm) stopping the reaction.

3.1.2 Experimental Aims

This chapter aims to answer the question; Are there any other small molecules that can effectively modulate the activity of WWP1 and WWP2? To answer this, the screening strategy previously utilised by the Chantry group is expanded upon, combining DSF and in vitro ELISA autoubiquitination assays in a high throughput screen of both WWP1 and WWP2 using their various constructs (Figure 3-4) alongside the 'next in series' NCI Diversity Set VI compound library to identify novel hit candidates. This comprises three subsections, with the first section focusing on identifying ligand interaction using the DSF assay to reduce the initial library size by screening against the complex constructs WWP1-L34H and WWP2-LH. The second section will look to identify small molecule inhibitors from these DSF hits using the ELISA autoubiquitination assay and various counter screens against WWP1-L34H and WWP2-FL. Finally, the third section highlights the all-important PAINS analysis, through both in silico ZINC15 filtering, as well as giving an example of the problems associated with PAINS compounds using NSC-2805, previously highlighted as a potent WWP2 inhibitor. The compound(s) highlighted in this chapter are later taken through to the next Hit-to-Lead phase of the drug discovery pipeline, in Chapter 4.

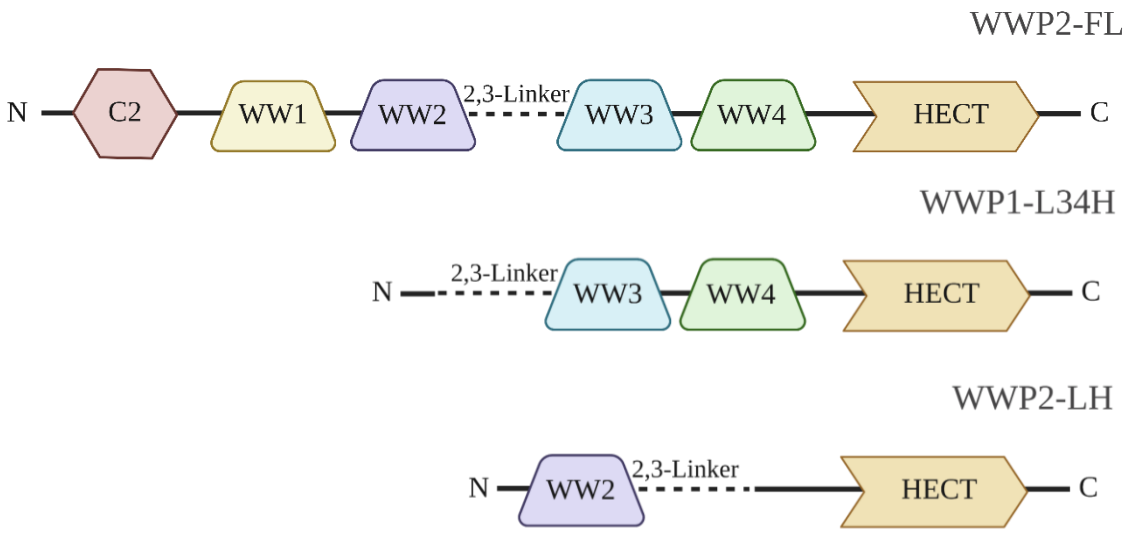


Figure 3-4: Domain architecture of the WWP ligase constructs used in Chapter 3. The full-length WWP2 (WWP2-FL) E3 ligase is composed of a Ca^{2+}-binding domain (C2, red) and the four tryptophan-tryptophan (WW) domains, WW1 (yellow), WW2 (purple), WW3 (blue) and WW4 (green), as well as the autoinhibitory 2,3-linker region (dashed line), before finally the catalytic HECT (orange) domain. The WWP1-L34H construct, previously reported by Wang *et al.* (2019), consists of the WW3, WW4 and HECT domains as well as the 2,3-linker located at the N-terminus. The WWP2-LH construct, previously reported by Chen *et al.* (2017), consists of the WW2 and HECT domains separated by the 2,3-linker region.

3.2 Results

The data presented in this chapter was collected in collaboration with Dr Gregory Hughes, and in part has been submitted previously for the MSci Master Degree of Ashley Dudey. The material breakdown is as follows: The DSF screen (Section 3.2.1) was completed by both Dr Gregory Hughes and Ashley Dudey, previously submitted; The autoubiquitination screen was completed by Dr Gregory Hughes (Sections 3.2.2.2 and 3.2.2.3), with the counter and resynthesized assays original to this project (Sections 3.2.2.4 and 3.2.2.5); The PAINS analysis is original to this project, with only a few NSC-2805 analogues submitted previously by Ashley Dudey. The protein purification protocols given have also been optimised throughout this project.

3.2.1 DSF screening of NCI Diversity Set VI against WWP1 and WWP2

As a cost-effective high throughput screening technique, DSF was utilised to complete the screening efforts of the NCI diversity set VI compound library with the intention of identifying new ligands for both WWP1 and WWP2.

3.2.1.1 DSF assay: construct selection and purification

Given the dynamic nature of wildtype WWP1 and WWP2, the more stable complex constructs of WWP1-L34H and WWP2-LH reported by Wang *et al*. (2019) and Chen *et al*. (2017), were chosen taking advantage of the autoinhibition effects of the 2,3-linker region preventing T – L shape transitions of the HECT domain. Preliminary work by Dr Gregory Hughes (2019) showed these constructs give a stable thermal profile compared to the HECT-only variants, reducing the noise-to-signal ratio required for small molecule ligand identification. The construct DNA was kindly gifted by the corresponding groups pre-cloned into pET32a and pGEX6p-2 plasmids respectively and transformed into BL21-codon plus RP cells expressing with respective His-Trx and GST tags. Purification strategies for both WWP1-L34H and WWP2-LH are stated in Section 2.2.7.3 and 2.2.7.5 with SDS-PAGE analysis shown in Figure 3-5 and Figure 3-6, respectively.

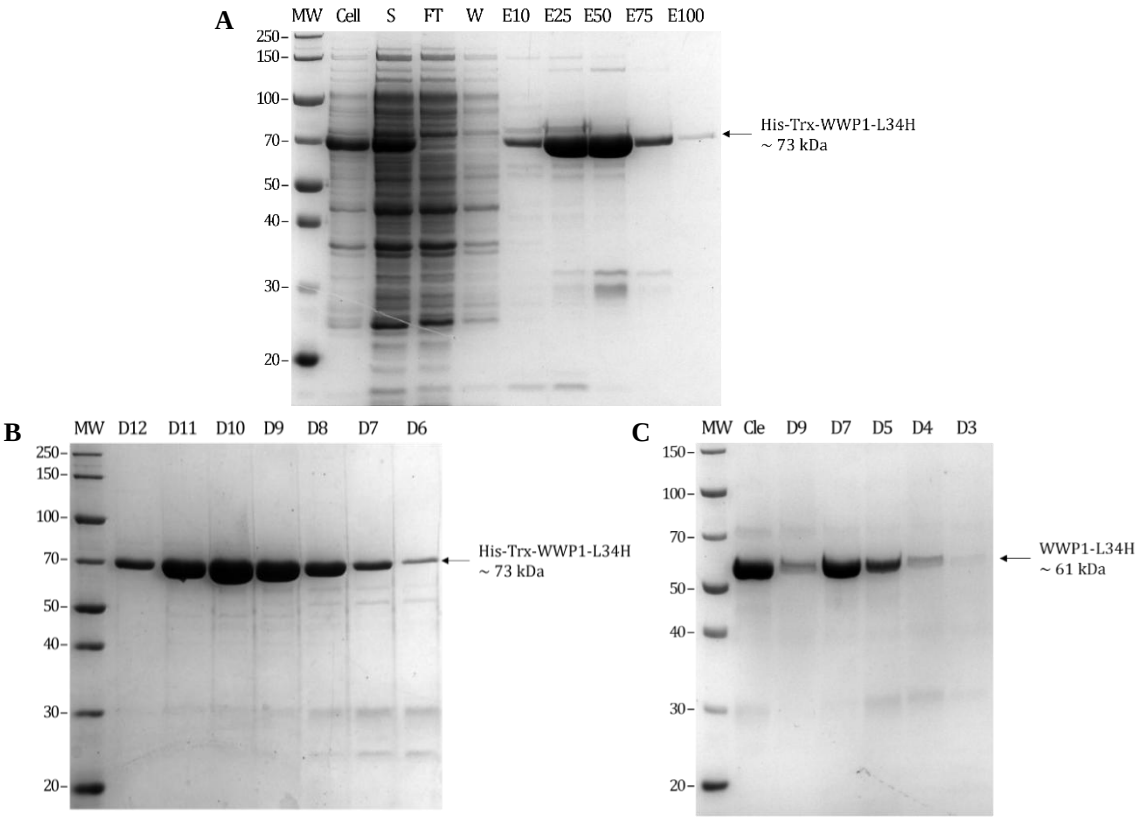


Figure 3-5: SDS-PAGE analysis of WWP1-L34H purification. (A) Stepwise IMAC isolation. Intense band at ~73 kDa indicating His-Trx tagged WWP1-L34H. Lane labels: Ladder (MW), whole cell (Cell), soluble lysate (S), column flow-through (FT), column wash (W), 10% elution (E10), 25% elution (E25), 50% elution (E50), 75% elution (E75), 100% elution (E100). All fractions were pooled. (B) First SEC purification. His-Trx tagged WWP1-L34H identified as shown by a band at ~73 kDa. Lanes labelled as column fraction positions excluding ladder (MW). (C) Cleavage and second SEC purification. Intense band at ~61 kDa as untagged WWP1-L34H. Lanes labelled as fraction positions other than ladder (MW) and cleavage sample (Cle). Fractions D9 – D4 were pooled. All gels were run on pre-cast 4 – 12 % Bis-Tris gels. The figure includes data contributions by A. Dudey.

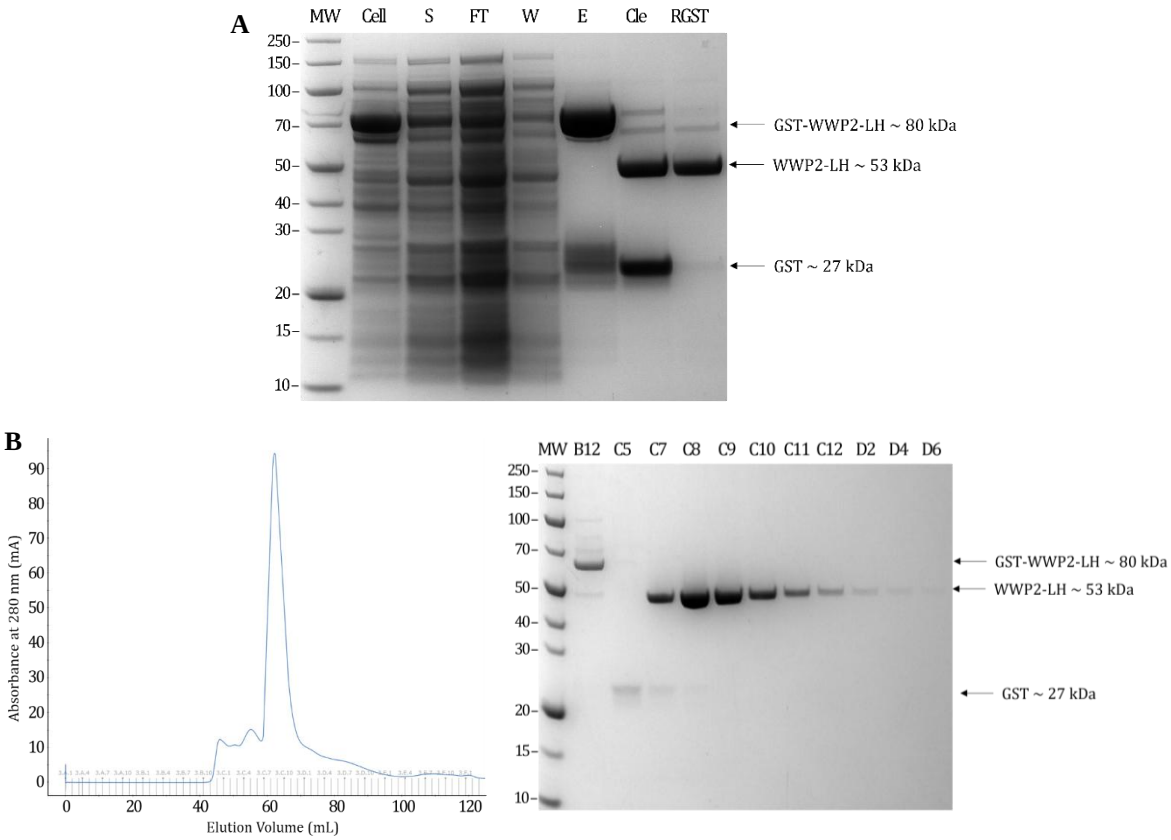


Figure 3-6: SDS-PAGE analysis of WWP2-LH purification. (A) Straight GSTrap isolation and cleavage. Isolation and cleavage of GST-tagged WWP2-LH are shown by an intense band at ~80 kDa and ~53 kDa respectively, with free GST shown at ~27 kDa band. Lane labels: Ladder (MW), whole cell (Cell), soluble lysate (S), column flow-through (FT), column wash (W), column elution (E), cleavage sample (Cle) and reverse GSTrap flow-through (RGST). (B) SEC purification. The AKTA absorbance trace at 280 nm is shown alongside SDS-PAGE. Untagged WWP2-LH identified by intense bands at ~53 kDa. Lanes labelled as fraction positions, with fractions C7 – D6 pooled. All gels were run on pre-cast 4 – 12 % Bis-Tris gels. The figure includes data contributions by A. Dudey.

3.2.1.2 High-throughput DSF compound screen discovered 24 compounds of interest for both WWP1 and WWP2

The NCI Diversity Set VI compound library, consisting of 1584 NSC compounds, was screened at 10 µM (0.1% DMSO) following the experimental procedure previously optimised by Dr Gregory Hughes (Section 2.3.2.1) (Hughes, 2019). The complete single-shot screen is displayed in Figure 3-7A. The Tm temperatures were determined from the first derivative (Tm D) of the respective melting curves with the preliminary hits selected based on a threshold of three times the standard derivation of the respective WWP1 and WWP2 controls. We considered this to be an appropriate threshold as it takes into account large variations seen between plates. A total of 209 and 264 compound hits were initially identified for WWP1 and WWP2, respectively. These initial screens were then performed in triplicate with 24 NSC compounds observed to maintain an average ΔTm D beyond the threshold for both targets (Figure 3-7B).

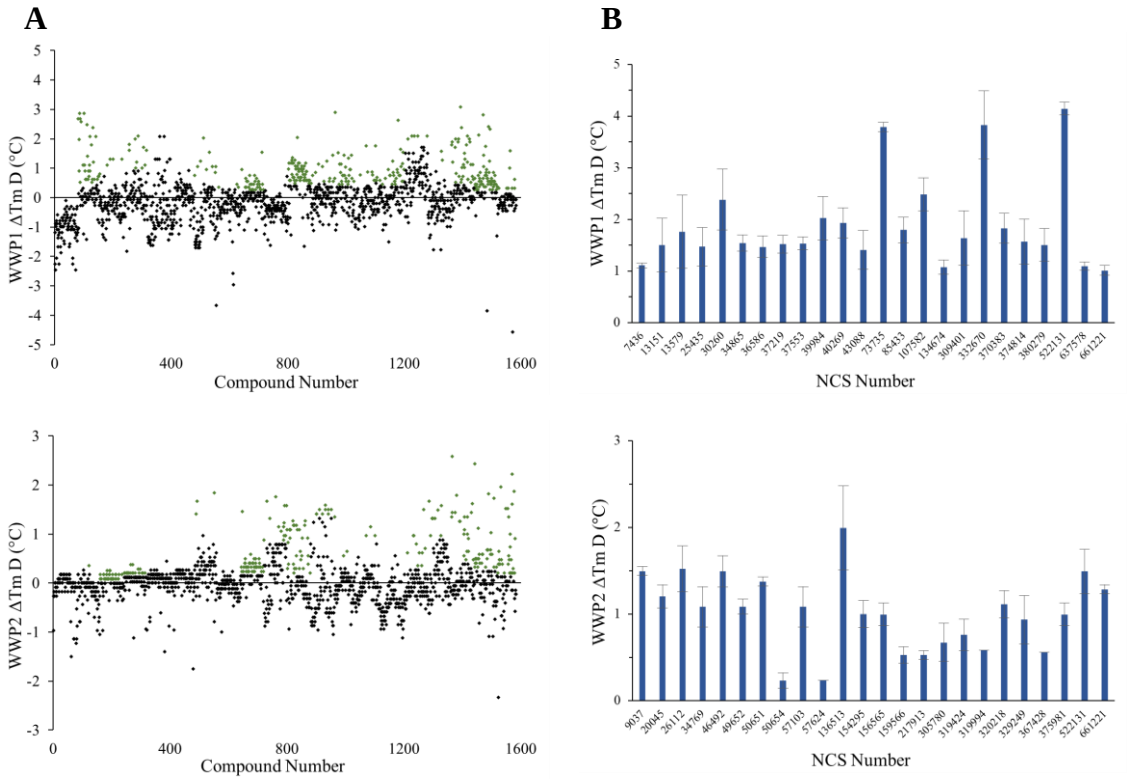


Figure 3-7: High throughput DSF screen of the NCI Diversity Set VI compound library against WWP1 and WWP2. Assays were carried out using 10 µM of compound, containing 0.1% DMSO and normalized by ΔTm calculated from controls. Positive hits (green) displayed had an average ΔTm and associated errors above a threshold of three times the standard deviation of their respective controls. (A) Single-shot DSF screen of WWP2-LH (top) and WWP1-L34H (bottom) against NCI Diversity Set VI compounds. (B) Triplicate DSF screen of 24 confirmed hit compounds for WWP1-L34H and WWP2-LH. The figure includes data contributions by G. Hughes and A. Dudey.

3.2.2 Auto-ubiquitination Screening of DSF Hits Against WWP1 and WWP2

The selected compounds from the DSF assay were first passed through a PAINS filter (Section 3.2.3.1) before screening for inhibitory potential against WWP1 and WWP2's E3 ligase activity. This was achieved using an auto-ubiquitination assay optimised by Dr Jessica Watt (Section 2.3.1.2 and 2.3.1.3) to measure compound inhibition by their ability to interfere with autoubiquitination and thus its intrinsic E3 ligase activity.

3.2.2.1 Auto-ubiquitination assay: construct selection and purification

Initially, both full-length constructs of WWP1 and WWP2 were to be selected, however, WWP1-FL showed minimal activity. This problem has been previously highlighted by Wang *et al*. (2019) where wildtype WWP1 was found in an autoinhibited state. The WWP1-L34H construct, missing the C2-WW1-WW2 domains, was therefore chosen having regained increased activity by relieving these inhibitory interactions (Wang *et al.*, 2019). Trx-His-tagged WWP1-L34H and GST-tagged WWP2-FL were expressed (Table 2-2) and immobilised onto respective nickel NTA and glutathione-coated plates from the lysate. For the WWP1 assay, fresh GST-tagged Uba1 and untagged UbcH7 were prepared and purified by strategies stated in Sections 2.2.7.6 and 2.2.7.7. Untagged UbcH7 was required due to the GST tag causing steric inhibition. Once untagged, any freeze-thawing resulted in a significant loss in activity therefore fresh enzyme was always used. For the WWP2 assay, His-tagged Uba1 and UbcH7 were prepared and purified by strategies stated in Section 2.2.7.1. The SDS-PAGE analysis of these purifications is shown in Figure 3-8. All other proteins included in the autoubiquitination assays were purchased commercially.

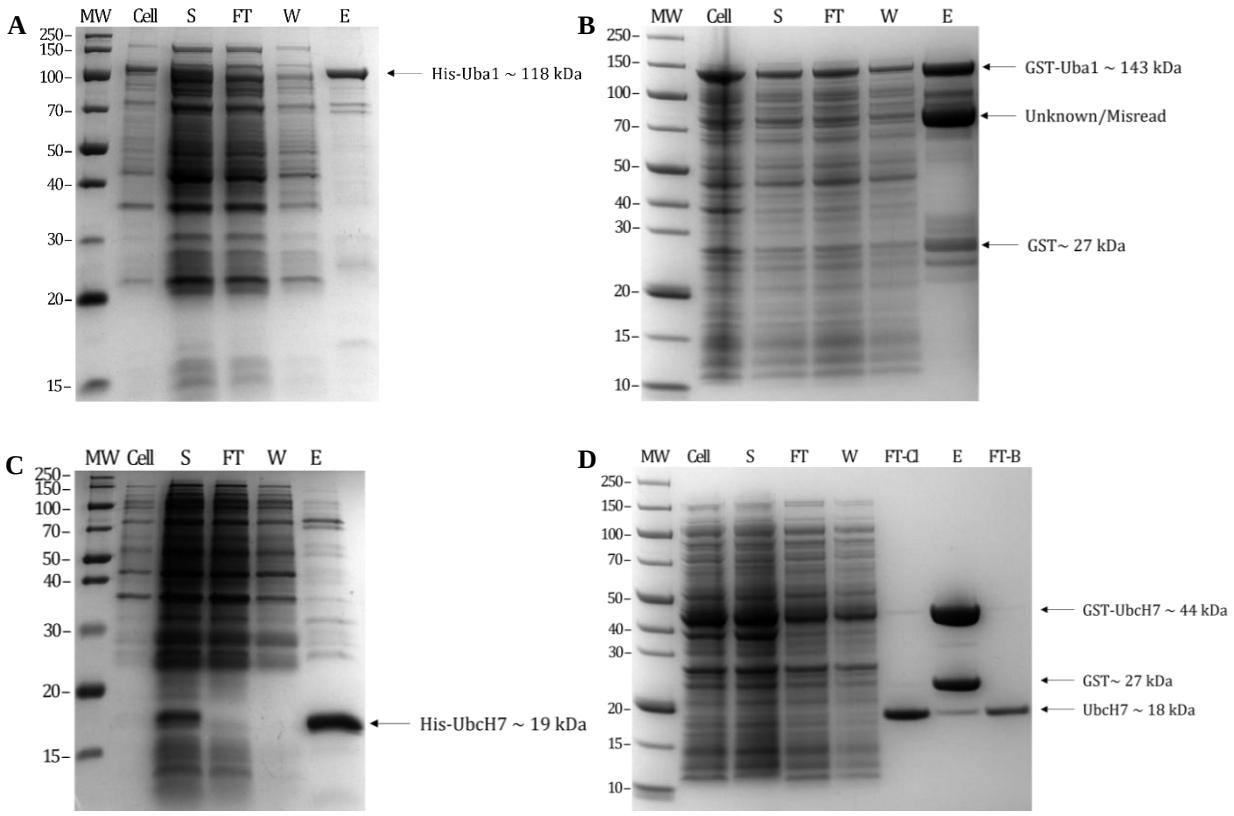


Figure 3-8: SDS-PAGE analysis of Uba1 and UbcH7 purifications. Lane labels: Ladder (MW), whole cell (Cell), soluble lysate (S), column flow-through (FT), column wash (W), column elution (E), on-column cleavage flow-through (FT-Cl) and Benzamidine column flow-through (FT-B). A. Straight IMAC isolation of Uba1. Intense band at ~118 kDa for His-tagged Uba1. B. Straight GST affinity isolation of Uba1. Intense bands at ~ 143 kDa for GST-tagged Uba1, as well as possible misread at ~80 kDa. C. Straight IMAC isolation of UbcH7. Intense band at ~19 kDa for His-tagged UbcH7. D. Straight GST affinity isolation of UbcH7 and cleavage. Intense bands at ~44 kDa for GST-tagged UbcH7, ~ 27 kDa for free GST and 18 kDa for untagged UbcH7. All gels were run on pre-cast 4 – 12 % Bis-Tris gels. The figure includes data contributions by A. Dudey.

3.2.2.2 The single-shot inhibitor screen discovered four hit-compounds for WWP1 and three for WWP2.

The compounds were first evaluated in a single-shot assay at 500 µM (1% DMSO), measuring inhibition against their relative auto-ubiquitination activity following the protocols for WWP1 and WWP2 as stated in Sections 2.3.1.2 and 2.3.1.3, respectively. Positive hits were chosen by a threshold of less than 50% relative activity when compared to controls. Four NSC compounds for WWP1 (13151, 73735, 85433 and 637578) and three NSC compounds for WWP2 (26112, 57103 and 217913) were initially selected as shown in Figure 3-9.

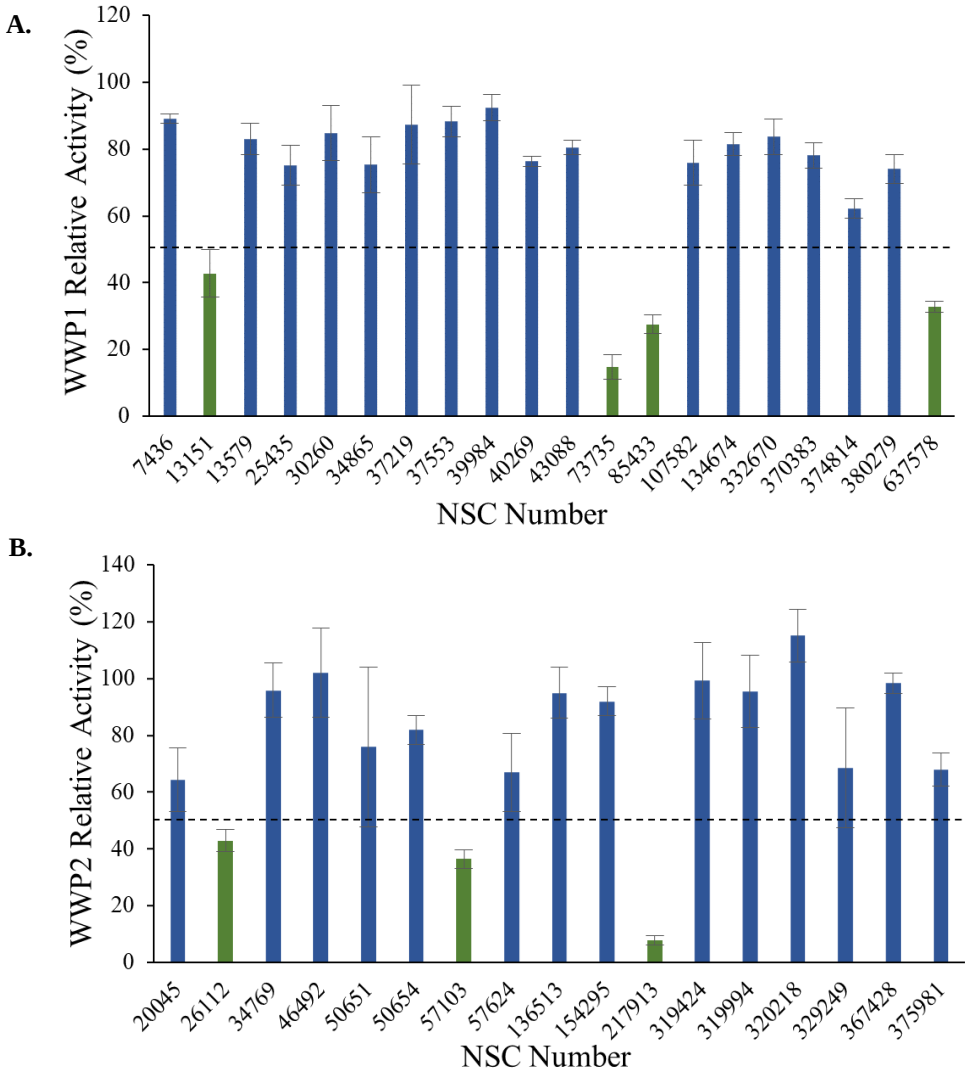
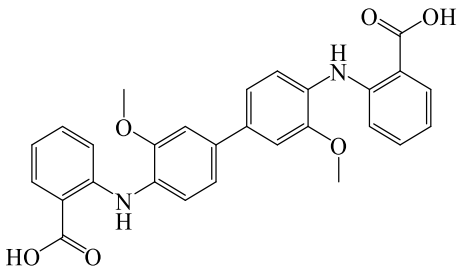
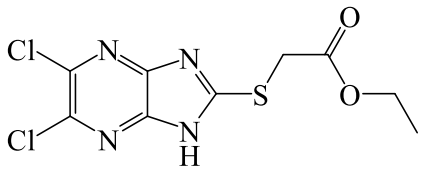
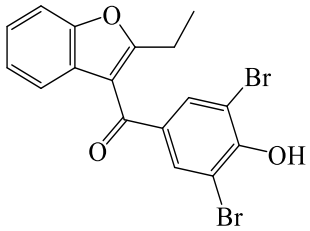


Figure 3-9: Auto-ubiquitination assay screen of PAINS-filtered DSF hit compounds against WWP1 (A) and WWP2 (B). Enzyme inhibition was measured at a compound concentration of 500 µM (1% DMSO), with a hit threshold (hits shown in green) of less than 50% relative activity, normalised to their respective 0% and 100% WWP1-L34H and WWP2-FL controls. The figure includes data contributions by G. Hughes.

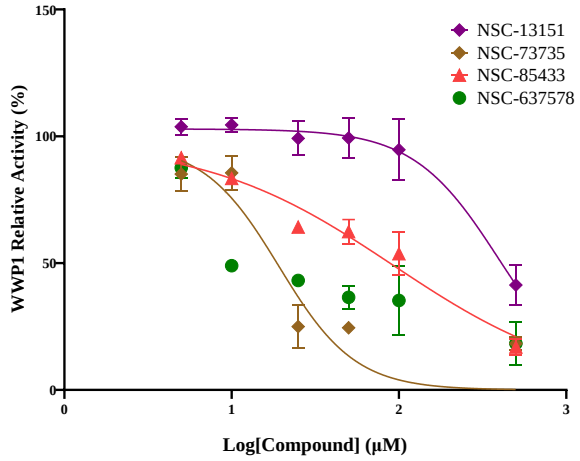
3.2.2.3 Dose-response analysis of hit-compounds highlighted three compounds with a IC_{50} below 500 µM for WWP1 and WWP2.

The single-shot hit compounds were further tested in dose-dependent assays from 500 – 5 µM with the log concentration plotted to enable their IC_{50} values to be calculated as shown in Figure 3-10. This data was provided by the Chantry group and as such was not originally designed for a logarithmic plot. Upon reflection, the use of a logarithmic scale would have better aided the fitting of a standard curve. Despite that, three NSC compounds for WWP1 (13151, 73735 and 85433) and only two NSC compounds for WWP2 (57103 and 217913) were selected as they displayed a concentration dependency. The most potent NSC compounds (73735, 85433 and 217913), having a perceived IC_{50} below 100 µM, were selected for cross-reactivity analysis against the alternative homolog target (i.e. WWP1 or WWP2) (Figure 3-11). A logarithmic concentration scale was used for the WWP2 cross-reactivity screen from 1 mM – 10 nM, providing a better fit for the standard curve. All three compounds displayed a greater selectivity towards WWP1, including NSC-217193, a hit compound from the WWP2 screen. A summary of the compound structures and IC_{50} values can be found below in Table 3-1.

Table 3-1: Summary of potential inhibitors for WWP1 and WWP2

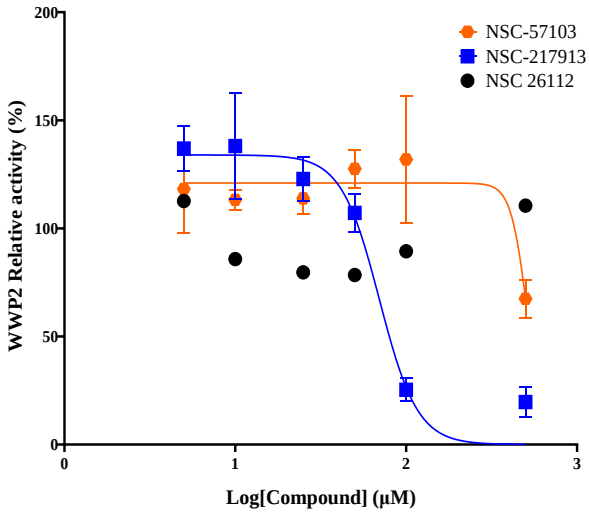
NSC Number	Structure	WWP1 IC50 (µM)	WWP2 IC50 (µM)
73735		19.2	49.1
217913		33.3	69.8
85433		92.4	256.4

A



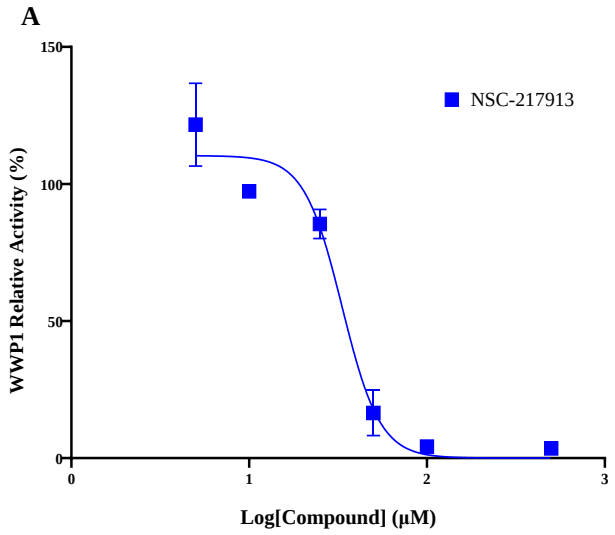
NSC Number	IC_{50} (µM)	95 % CI (µM)
13151	396.9	314.8 – 485.4
73735	19.2	??? – 30.01
85433	92.4	27.1 – 153.4
637578	N/A	N/A

B

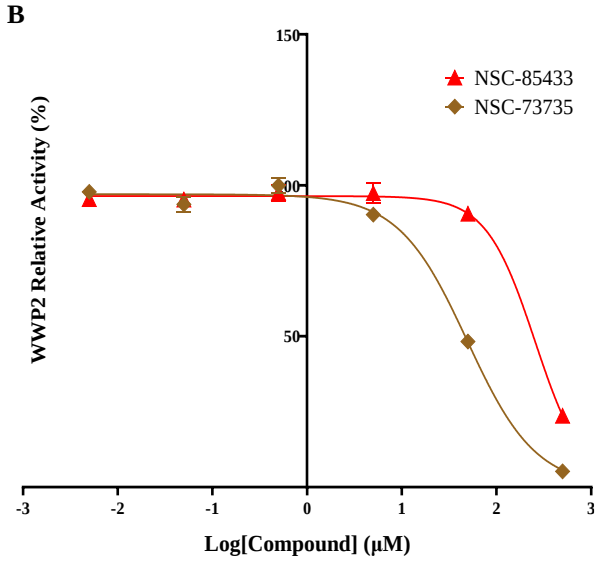


NSC Number	IC_{50} (µM)	95 % CI (µM)
57103	513.0	429.5 – ???
217913	69.8	59.1 – 82.9
26112	N/A	N/A
97920	N/A	N/A

Figure 3-10: Dose-dependent auto-ubiquitination analysis of WWP1 and WWP2 against their respective single-shot hit compounds. Plotted at log concentration of 500 – 5 µM (1% DMSO) against WWP1/2 relative activity, standardised to their respective controls. IC_{50} curves were plotted using Prism with error bars representing standard deviation. Calculated IC_{50} values are shown in adjacent tables, with associated 95 % confidence interval (CI). (A) WWP1-L34H dose-response of NSC-13151 (purple), NSC-73735 (brown), NSC-85433 (red) and NSC-637578 (green). Unable to fit a curve to NSC-637578. (B) WWP2-FL dose-response of NSC-57103 (orange), NSC-217913 (blue) and NSC-26112 (black). Unable to fit a curve to NSC-26112. The figure includes data contributions by G. Hughes.



NSC Number	IC50 (µM)	95 % CI (µM)
217913	33.3	29.1 – 38.5



NSC Number	IC50 (µM)	95 % CI (µM)
85433	256.4	231.0 – 287.1
73735	49.11	43.3 – 55.4

Figure 3-11: Dose-dependent cross-selectivity analysis of WWP1 and WWP2 hit-compounds. IC_{50} curves were plotted using Prism with error bars representing standard deviation. Calculated IC_{50} values are shown in adjacent tables, with associated 95 % confidence interval (CI). (A) WWP1-L34H auto-ubiquitination assay against WWP2 hit-compound. Dose-response of NSC-217913 (blue) ranging from 500 – 5 µM (1% DMSO) and plotted as a logarithmic scale against WWP1 relative activity. (B) WWP2-FL auto-ubiquitination assay against WWP1 hit-compounds. Dose-response of NSC-85433 (red) and NSC-73735 (brown) ranging from 1 mM – 10 nM (1% DMSO) and plotted as a logarithmic scale against WWP2 relative activity. The figure includes data contributions by G. Hughes.

3.2.2.4 NSC-217913 was the sole-hit compound after counter screening.

We wanted to further test whether these compounds were inhibiting their intended target or were interfering with Uba1 and UbcH7 earlier on in the ubiquitination cascade, and as such demonstrating a false-positive WWP1 or WWP2 inhibitory effect. Counter-assays were performed in the presence of these compounds for both Uba1 and UbcH7 as stated in Section 2.3.1.4 and 2.3.1.5 with results shown in Table 3-2. These counter-assays measure the effect of these compounds on the activity of both Uba1 and UbcH7, following the same principle as the autoubiquitination assay, comparing the amount of Ub-conjugation after a set incubation period, in essence, ending the assay before the final E3 ligation step. Given the UbcH7 assay utilises both Uba1 and UbcH7, any signs of inhibition during the counter-assay would be sufficient to suggest a possible PAINS-like property, i.e. off-targeting against these downstream E1 and E2 enzymes. NSC-217913 was the only compound to show no significant signs of interference with Uba1 and UbcH7, with both NSC-73735 and NSC-85433 reducing UbcH7's relative activity to ~60% and ~8% respectively.

Table 3-2: Counter auto-ubiquitination screen of hit NCI Div VI compounds against UbcH7.

NSC Number	UbcH7 Activity (%)
73735[a]	62.5 ± 2.4
85433	7.5 ± 0.6
217913[a]	92.2 ± 1.1

[a]synthesised

3.2.2.5 Synthesised NSC-217913 demonstrated reduced inhibition for both WWP1 and WWP2

To ensure the results observed from NSC-217913 inhibition were repeatable, the compound was freshly synthesised by our collaborators in Dr Thomas Storr's research group, in particular, Jake Rigby and retested for dose dependency against WWP1 and WWP2 using the log concentration range from 1 mM to 10 nM compound. Synthesised NSC-217913 had a significantly reduced potency with a measured IC_{50} of 158.3 µM against WWP1 and with WWP2 inhibition too weak to determine accurately (Figure 3-12).

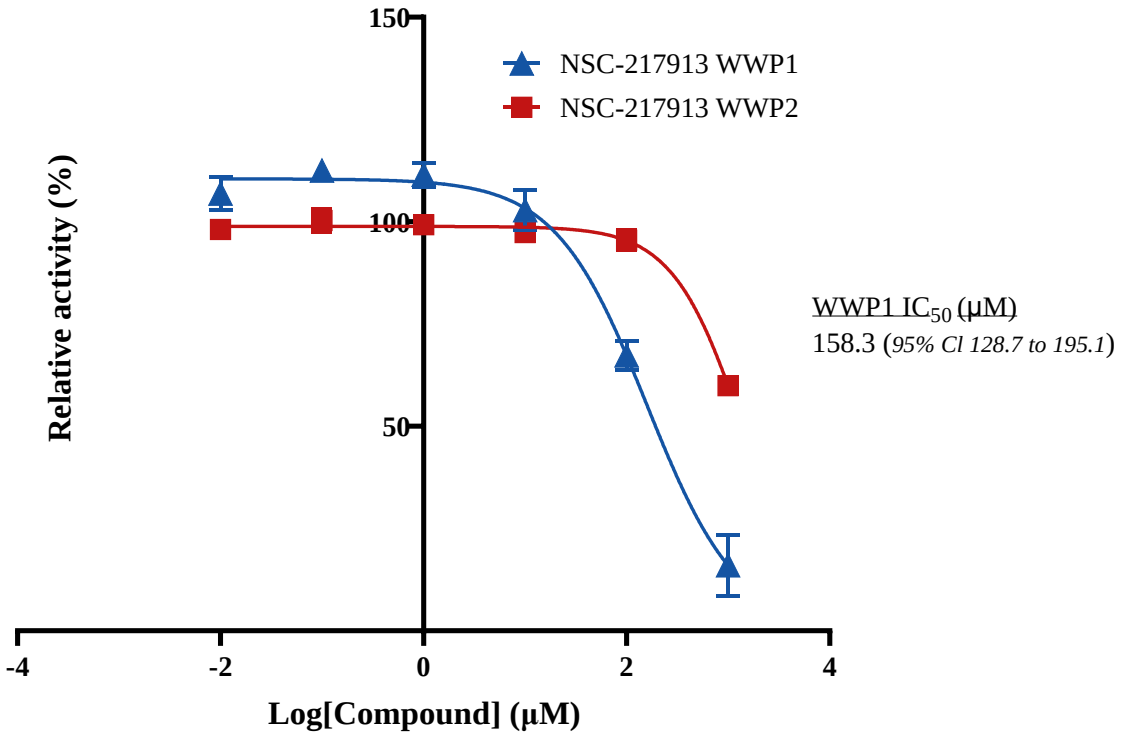


Figure 3-12: Dose-dependent auto-ubiquitination assay of synthesised NSC-217913 against WWP1 and WWP2. Compound inhibition was measured on a log scale from 1 mM – 10 nM (1% DMSO), normalised to 0% and 100% WWP1-L34H (blue) and WWP2-FL (red) controls. IC_{50} values were calculated from non-linear regression curves fitted in GraphPad. The figure includes data contributions by A. Dudey.

3.2.3 PAINS Analysis

3.2.3.1 In silico PAINS screening was used to remove problematic compounds

After the initial DSF screening of the NCI Diversity Set VI compound library, the hit compounds were also virtually screened through a PAINS filter initially developed by Baell and Holloway (2010) and later incorporated into ZINC 15's chemical catalogue (Sterling and Irwin, 2015). The ZINC 15 system initially highlighted six problematic moieties that could give false positive results. A further six compounds were removed based on other properties including reactive species, impurities, aggregators, and radical quenchers that could interfere with our ELISA-based assays shown in Table 3-3 for WWP1 and Table 3-4 for WWP2 screens. Aggregators were determined using aggregate advisor, an in-silico screen that calculates the LogP values and associates known aggregators to compounds with similar structures (Irwin *et al.*, 2015).

Table 3-3. Summary of DSF hit compound PAINS filtering against WWP1.

NSC Number	PAINS Filter[(a)]	Other PAINS	Aggregate Risk[(c)]	Omitted
7436	-	-	Very Low	N
13151	-	-	Medium	N
13579	-	-	High	N
25435	-	-	Medium	N
30260	-	-	High	N
34865	-	-	Very Low	N
36586	-	Aggregate	Very High	Y
37219	-	-	Medium	N
37553	-	-	High	N
39984	-	-	High	N
40269	-	-	Low	N
43088	-	-	High	N
73735	-	-	High	N
85433	-	-	High	N
107582	-	-	High	N
134674	-	-	Low	N
309401	Anil_no_alk(40)	-	Very Low	Y
332670	-	-	High	N
370383	-	-	Very Low	N
374814	-	-	High	N
380279	-	-	Very Low	N
522131	-	Radical Quencher[(c)]	High	Y
637578	-	-	Medium	N
661221	Quinone(370)	-	Medium	Y

[(a)] ZINC15 Filter (Sterling and Irwin, 2015). [(b)] Ranking based on descriptions from Aggregate Advisor (Table S3) (Irwin *et al*., 2015) .

[(c)] Auto-ubiquitination assay-specific PAINS (Berggren *et al*., 2019).

Table 3-4. Summary of DSF hit compound PAINS filtering against WWP2.

NSC Number	PAINS Filter[a]	Other PAINS	Aggregate Risk[b]	Omitted
9037	Catechol_A(92)	-	High	Y
20045	-	-	High	N
26112	-	-	Medium	N
34769	-	-	Very Low	N
46492	-	-	High	N
49652	-	Reactive	High	Y
50651	-	-	High	N
50654	-	-	High	N
57103	-	-	High	N
57624	-	-	High	N
136513	-	-	High	N
154295	-	-	Very Low	N
156565	Quinone(370), Anthranil_one_A(38)	-	Very Low	Y
159566	Quinone(370)	-	Low	Y
217913	-	-	Very Low	N
305780	-	EtOH Impurity	High	Y
319424	-	-	High	N
319994	-	-	High	N
320218	-	-	High	N
329249	-	-	High	N
367428	-	-	Very Low	N
375981	-	-	High	N
522131	-	Radical Quencher[c]	High	Y
661221	Quinone(370)	-	Medium	Y

[a] ZINC15 Filter (Sterling and Irwin, 2015). [b] Ranking based on descriptions from Aggregate Advisor (Table S6) (Irwin *et al*., 2015).

[c] Auto-ubiquitination assay-specific PAINS (Berggren *et al*., 2019).

Table 3-5. Aggregate risk rankings and associated descriptions from Aggregate Advisor.

Risk	Descriptions from Aggregate Advisor
Very Low	This molecule does not look like one that has been previously observed to aggregate.
Low	This molecule is not similar to a known aggregator, but with a high logP, is a candidate for aggregation.
Medium	This molecule is somewhat similar to a molecule that has been previously observed to aggregate, AND, with a high logP, is suspicious.
High	This molecule is very similar to a molecule that has been previously observed to aggregate.
Very High	This molecule has previously been observed to aggregate.

3.2.3.2 NSC-2805 was found to be a PAINS compound

The aforementioned NSC-2805 was highlighted as a promising potential inhibitor for WWP2 from a similar screen seen here using the NCI Diversity Set V compound library by Watt *et al*. (2018). Preliminary STD-NMR and X-ray crystallography data provided by the chantry group also suggested interaction but required further exploration. We wanted to explore the SAR of NSC-2805 to not only reinforce this data but to further increase its inhibitory potential. This was achieved through collaboration with Dr Thomas Storr's research group, in particular Dr Joshua Hall, whereby ~60 analogues were synthesised in a series of every 10 – 20 compounds based on autoubiquitination assay observations from the previous series, with 23 compounds retaining inhibition shown in Table 3-6. Despite our best efforts to design analogues towards moieties that improved the apparent inhibition, no clear pattern could be observed other than the presence of the mono- or di-hydroquinone group. There were concerns that this group was demonstrating a PAINS property, influencing our analogue exploration. To investigate this various PAINS analysis was undertaken. Firstly, NSC-2805 was added to HRP and TMB independent of the autoubiquitination assay, displaying a clear interference. Although indicative of a PAINS property, this was not conclusive as the assay contained various plate-washing steps, therefore the compound should not be present at the HRP/TMB stage. To verify this effect, NSC-2805 was incubated during and after the autoubiquitination reaction as shown in Figure 3-13A. This still displayed possible interference however to a lesser degree when compared to NSC-2805 incubated during the reaction, potentially indicating inhibition of WWP2 was still taking place. Two compounds, NSC-2805 analogue 40 and 41 supported this by lacking the problematic hydroquinone moieties and yet exhibited a much higher IC_{50} range of 265.7 and 393.0 µM respectively. However, after undertaking a Uba1 counter assay, it was clear NSC-2805 was also interfering with other enzymes in the ubiquitination cascade, with 500 µM almost removing all Uba1 activity.

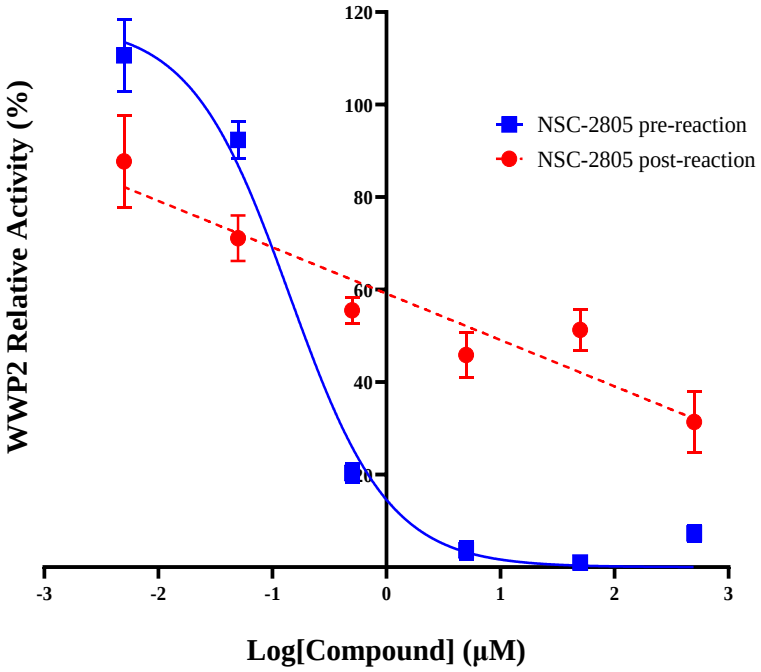
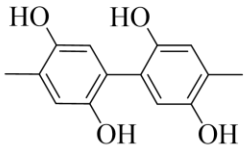
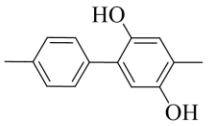
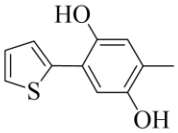
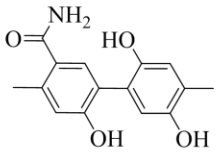
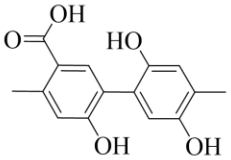
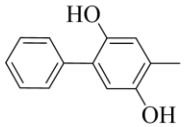
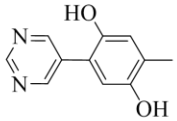
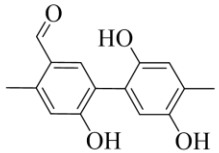
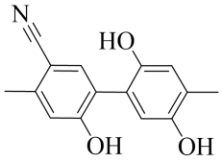
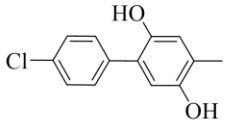
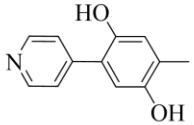
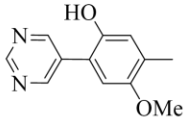
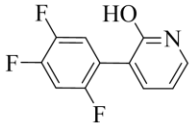
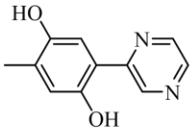
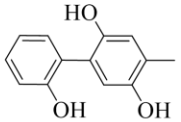
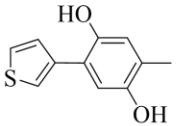
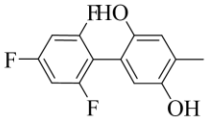
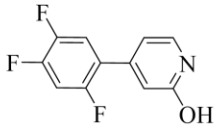
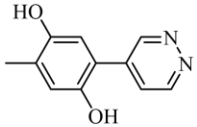
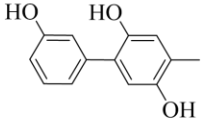
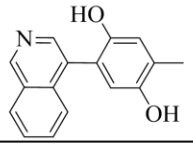
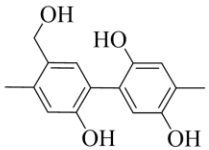
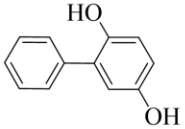
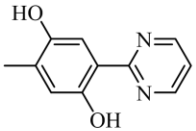


Figure 3-13: PAINS analysis of NSC-2805. NSC-2805 against the WWP2 auto-ubiquitination assay. NSC-2805 was incubated during (blue) and after the auto-ubiquitination reaction (red) at a concentration range from 500 µM – 5 nM (1% DMSO) and plotted as a logarithmic scale against WWP2 relative activity. Standard curves were plotted using Prism with error bars representing standard deviation. The figure includes data contributions by A. Dudey.

Table 3-6: Summary of NSC-2805 analogues and their associated WWP2 IC_{50} values

N.o	Structure	IC_{50} (µM)	N.o	Structure	IC_{50} (µM)	N.o	Structure	IC_{50} (µM)	N.o	Structure	IC_{50} (µM)	N.o	Structure	IC_{50} (µM)
	NSC-2805  **IC_{50} = 0.33 µM**		8		1.61 ±1.29	18		15.25 ±2.12	26		2.95 ±0.65	52		0.79 ±0.15
			10		1.78 ±0.03	21		0.29 ±0.05	27		2.59 ±0.31	53		0.82 ±0.22
2		7.21 ±6.73	12		1.61 ±0.41	22		11.1 ±1.56	40		265.7 ±134.6	57		0.68 ±0.42
4		0.88 ±0.30	14		19.78 ±5.64	24		1.16 ±0.21	41		393.0 ±65.7	58		0.88 ±0.46
6		0.96 ±0.85	16		5.17 ±4.37	25		6.68 ±0.71	49		0.67 ±0.52	59		19.30 ±14.27

3.3 Discussion

This chapter aimed to address the shortage of small molecule inhibitors available for the HECT E3 ligases WWP1 and WWP2 by screening against a small but diverse compound library. A DSF assay was used to screen the NCI Diversity Set VI library, resulting in a total of 24 compounds highlighted to interact with each target. A virtual PAINS filtering process reduced this total to 20 and 17 compounds, before using an ELISA autoubiquitination assay to screen for inhibition. From these, 4 and 3 compounds displayed inhibition for WWP1 and WWP2 respectively, with 3 and 2 compounds maintaining a dose response. The most potent compounds NSC-217913, NSC-73735 and NSC-85433 all displayed activity towards both WWP1 and WWP2, with a preference for WWP1. Counter screening highlighted NSC-217913 as the sole hit small molecule inhibitor, and once re-synthesized demonstrated an IC_{50} of 158.3 µM against WWP1, unable to record such dose dependency against WWP2 (Figure 3-12). This chapter also confirmed NSC-2805, a hit inhibitor from the previous WWP2 screen, as a prevalent PAINS compound, and highlighted the problems associated when such a compound is carried forward into the SAR stage. Here the reasons behind protein construct and assay choice are explained, before discussing the highlighted ligands and NSC-217913 inhibitor in relation to the literature. Finally, the discovered NSC-2805 PAINS property is further argued concerning a possible assay-specific PAINS mechanism.

3.3.1 Justification for Construct and Assay Choice

In the vast topic of drug discovery, a diverse range of high throughput screening techniques are available, all having their advantages and disadvantages. The choice of assay is not only dependent on the demands of the topic and its application but also the accessibility of individual research groups, whether restricted financially or to more mundane parameters such as manpower. Other factors also come into effect when choosing the most appropriate screening assays with the biological targets themselves able to have a significant influence. At least when working in vitro, the selected targets must be cloned, expressed, and purified into sufficient volumes to reach assay demands, not always so easily achieved. This becomes more complex when considering variations of expressed constructs, with many proteins comprised of numerous domains not directly linked to their activity. These constructs can vary in activity, solubility, and stability resulting in changes to expression levels and purification challenges as well as affecting their possible applications.

Following these restrictions, the DSF thermal shift assay was used for the initial screening efforts due to its relative accessibility and cost efficiency, vital for any academic research. Although not as sensitive as other screening techniques, this approach also required

less protein and so was appropriate for when using the WWP1-L34H and WWP2-LH constructs, normally expressed at low levels (~0.956 and 2.5 mg/ L of culture). Belonging to the NEDD4 family, WWP1 and WWP2 characteristically contain a multitude of domains (C2, WW and HECT) able to alternate between non- and autoinhibitory states, resulting in a high degree of dynamic movement (Chen *et al.*, 2017). Given our search for inhibitors, any construct used was required to at least contain the HECT domain, responsible for the activity. Our first choice was the HECT-only WWP1 and WWP2 constructs, given any indication of binding would be more likely to associate with an inhibitory role, and such an approach had been used previously for NEDD4-1 (Quirit, 2017). Despite this, these constructs were difficult to isolate and shown to be highly unstable, as discussed later (seen in Section 4.2.4.1). Full-length WWP1 and WWP2 were also considered, however, previous work by Dr Gregory Hughes (2019) had shown that the isolation of these constructs was at this stage unsuccessful, despite showing activity from lysate pulldowns. In the literature, two constructs for WWP1 (Wang *et al.*, 2019) and a single construct for WWP2 (Chen *et al.*, 2017) had already been isolated to a purity level that was appropriate for DSF, these being the WWP2-LH, WWP1-2L34H and WWP1-L34H constructs. However, at this stage in the project, we were unable to successfully isolate WWP1-2L34H, and therefore the slightly more active WWP1-L34H alongside WWP2-LH was chosen. We theorised that an increase in the dynamic stability as seen in these complex constructs would result in reduced background noise and thus potentially increase the sensitivity of the DSF assay. Despite this, these constructs did demonstrate possible limitations relating to their WW2-2,3-linker intermolecular interactions locking them in an autoinhibitory T-shape. As such, compound interactions may have been blocked by the WW2-2,3-linker in the case of WWP2-LH, leading to false negatives or in contrast, compounds may have been interacting due to artificial binding surfaces generated from the construct design itself.

These WWP1 and WWP2 construct choices were further expanded into the ELISA autoubiquitination assay, except with a focus on ubiquitination activity and therefore requirement for unimpeded dynamic movement. As noted, the full-length WWP2 demonstrated high autoubiquitination activity and had been optimised previously (Watt *et al*., 2018). In the case of WWP1, efforts to use the full-length constructs resulted in diminished activity, a characteristic observed by Wang *et al*. (2019). The full-length WWP1 is natively found in an autoinhibitory state, only once relieved as in the case of the WWP1-L34H construct (missing the C2-WW1-WW2 domains), does the ubiquitination and as such autoubiquitination activity significantly increase. As mentioned earlier in Section 3.1, this specific ELISA-based autoubiquitination assay had been previously optimised by the chantry group for both WWP1 and WWP2, and given this project was already time-restricted due to issues such as the COVID-19 pandemic, it was maintained as the activity assay of choice. However, there were some fundamental limitations of this assay, primarily based on the use of the traditional HRP-TMB

reaction. Beyond the susceptibility of this reaction to PAINS compounds, as discussed in Section 3.3.3, the main limitation was the requirement for the reaction to be stopped with the objective timing of this being dictated by an appropriate 'blue' colour change. This timing was inherently different between each assay, changing due to differences in plate coating, enzymatic activity (Uba1, UbcH7 and the WWP ligase), and even batches of TMB solution, and therefore a time-specific incubation could not be used. As the recombinant HRP is a catalyst for the oxidation of TMB, if left to incubate indefinitely or for too long as with cases omitted in this study, there would be differences in the perceived inhibition as the excess TMB would be completely oxidised in both the sample and control. This may have been resolved by monitoring the absorbance of the control and stopping the reaction after a specific value, however, without specialist equipment able to dispense automatically to each 96-well plate, the time taken to remove the plates and manually dispense would negate this effect. In the future, a similar study would benefit from using a non-reversible reaction or even in this specific case, fluorescently tagged Ub such as Fluorescein-Ubiquitin™ available commercially, having been used elsewhere in TR-FRET (Madiraju *et al*., 2012).

3.3.2 High-throughput Screen Analysis

The NCI diversity set VI compound library screen was not as successful as the prequel WWP2 screen, identifying a single small molecule inhibitor as 'NSC-217913' for both WWP1 and WWP2. Again, when re-synthesised this compound demonstrated a marked reduction in inhibition displaying an IC_{50} of 158.3 µM against WWP1, and unable to record such dose dependency against WWP2 (Figure 3-12). Our chemist, Jake Rigby did look to pinpoint the reasons for this, highlighting possible signs of stock oxidation when analysing the NMR trace, however issues around water being present in the DMSO solvent led to inconclusive results, a common issue with freeze-thaw cycles. These stocks were provided directly from the NCI depository and may have contained possible metal impurities we were unaware of at the time. In either case, a single hit inhibitor is most likely a reasonable return when considering that commercial libraries use hundreds of thousands of compounds and although the NCI library is designed to screen a diverse range of chemical space, a library size of 1584 compounds is particularly small. Furthermore, given the more vigorous filtering approach taken firstly by highlighting ligand interactions before screening for activity, as well as the ZINC 15 PAINS filtering, the chances of false positives were also decreased. We also filtered out less potent inhibitors to reduce the sample sizes to more manageable volumes given at this stage in the project it was only myself and Jake Rigby available for future lead development.

3.3.2.1 Literature analysis of hit compounds highlighted possible aggregation properties of NSC-73735.

NSC-217913, or in its IUPAC name, ethyl 2-[(5,6-dichloro-1H-imidazo[4,5-b]pyrazin-2-yl)sulfanyl]acetate, appears to show no significant inhibitory interactions with any other targets. Given its use in various libraries, there is an array of bioassay data available on the PubChem servers under CID 5358027 (National Center for Biotechnology Information, 2024). Out of the 339 different bioassays tested including both in vivo and in vitro screens, only 7 (2%) assays found 'inconclusive' evidence of inhibition (showing inhibition but having a poor dose-response), with the remaining bioassays deeming NSC-217913 as inactive, this includes various in vivo proliferation screens against Leukaemia cell-lines. Both NSC-85433 and NSC-73735, although appeared to inhibit WWP1 and WWP2, were later shown to interact with Uba1 and UbcH7 in the counter screens (Table 3-2).

NSC-73735, otherwise known as Redoxal, demonstrated more uncertainty in these interactions having a relative activity of 62.5% against UbcH7, and yet displayed potent inhibition against WWP1 and WWP2 with an IC50 of 19.21 and 49.11 µM respectively (Table 3-1). This ambitious nature is apparent when analysing the available bioassay PubChem data under CID 72571 (National Centre for Biotechnology Information, 2024), revealed from 894 screens, 116 (13%) were deemed active with concentrations ranging from 0.5 – 100 µM, with another 75 bioassays shown to be inconclusive (8%). This ranging activity is generally indicative of possible PAINS interactions. However, given that NSC-73735 contains no known PAINS moieties, this possible assay interference property may be associated with the aggregation of the compound in question (Auld, Inglese and Dahlin, 2017). Compound aggregation is a significant source of nonspecific bioactivity found in high throughput screening, with one study finding that 90% of their hit compounds were detergent sensitive; a common additive used to account for aggregation (Jadhav *et al*., 2010). Unfortunately, aggregation was not accounted for in our assay design, so a virtual screen was run using Aggregative Advisor (Table 3-5). This program associates known aggregate structures and estimates the octanol-water partition coefficient (LogP) of the screened compounds to determine the likelihood of aqueous aggregation (Irwin *et al*., 2015). NSC-73735 was labelled as having a high risk of aggregation, given its high LogP: 6.25 and similarities to 2-(4-methylphenyl)-6-naphthalen-2-ylpyridine-4-carboxylic acid that has been previously observed to aggregate (McGovern *et al*., 2002) (Figure 3-14). By comparison, NSC-217913 was deemed as low risk having no associated aggregate structures and has an estimated LogP: 2.31.

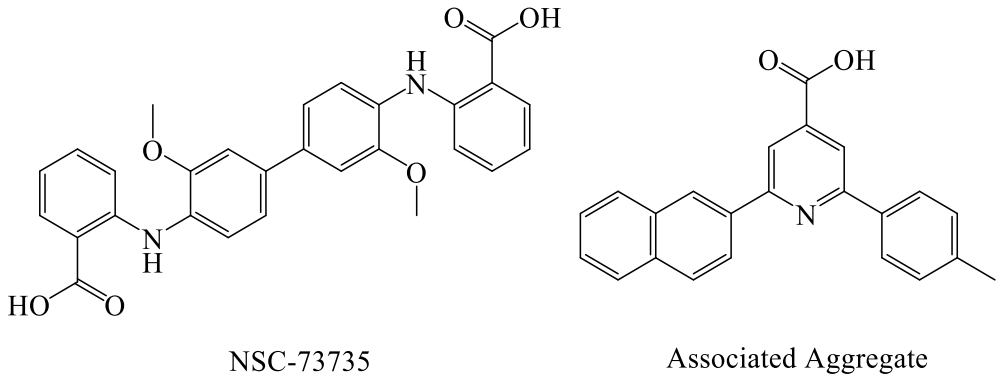


Figure 3-14: Structure of NSC-73735 (Redoxal) and its associated aggregate. 2-(4-methylphenyl)-6-naphthalen-2-ylpyridine-4-carboxylic acid was associated using Aggregate Advisor. Structures are drawn in ChemDraw.

3.3.3 An assay-specific PAINS

After hit validation, future development of the hit compound(s) requires analysis of their SAR to improve both binding and/or inhibition against WWP1 and WWP2. Such evolution efforts were originally sampled here using the NSC-2805 compound in a SAR by synthesis approach against WWP2 (Table 3-6). As seen, our compound evolution efforts failed after no apparent pattern could be distinguished beyond the inclusion of mono- or di-hydroquinone groups, suggesting a PAINS interference. Although NSC-2805 was shown to significantly inhibit Uba1, our PAINS analysis also highlighted its direct effect on the TMB/HRP reaction without the presence of the target protein. Such effect was believed to be due to a possible radical quenching mechanism associated with the oxidation of hydroquinone groups to quinone via a semi-quinone radical intermediate (Figure 3-15A). As described the ELISA autoubiquitination assay, utilises a common immunoassay approach measuring ubiquitination through the fluorescent reaction of TMB with HRP conjugated to an anti-flag antibody (Figure 3-3). This reaction between TMB and HRP occurs via a radical intermediate forming the oxidised TMB product before acidification enables the characteristic fluorescent at 450 nm (Josephy, Eling and Mason, 1982) (Figure 3-15B). Within the assay setup, the intensity of this fluorescence directly correlates with autoubiquitination, and therefore activity. The requirement for this intermediate may have led to NSC-2805 directly inhibiting the formation of the oxidised TMB product with a semi-quinone conversion acting as a free radical sink. In essence, the HRP enzyme would be oxidising the hydroquinone groups rather than TMB, resulting in a dose-dependent response. It is interesting however that the NSC-2805 analogues 40 and 41, without the presence of the hydroquinone moieties still showed signs of inhibition with an IC_{50} of 265.7 and 393.0 µM respectively. These may have simply been interacting with the Uba1 or UbcH7 enzymes earlier in the cascade as no counter screens were completed for these analogues. If this proved wrong it

would suggest that the core scaffold may have some inhibitor potential. To avoid the observed PAINS properties scaffold hopping may be of future interest, whereby core structures of a molecule are replaced with improved bioactive structures whilst maintaining important target interactions (Sun, Tawa and Wallqvist, 2012). However, such a feat was outside the scope of this project and thus all work with NSC-2805 was discontinued, with the learnt lessons carried forward in the lead development of NSC-217913 (Chapter 4).

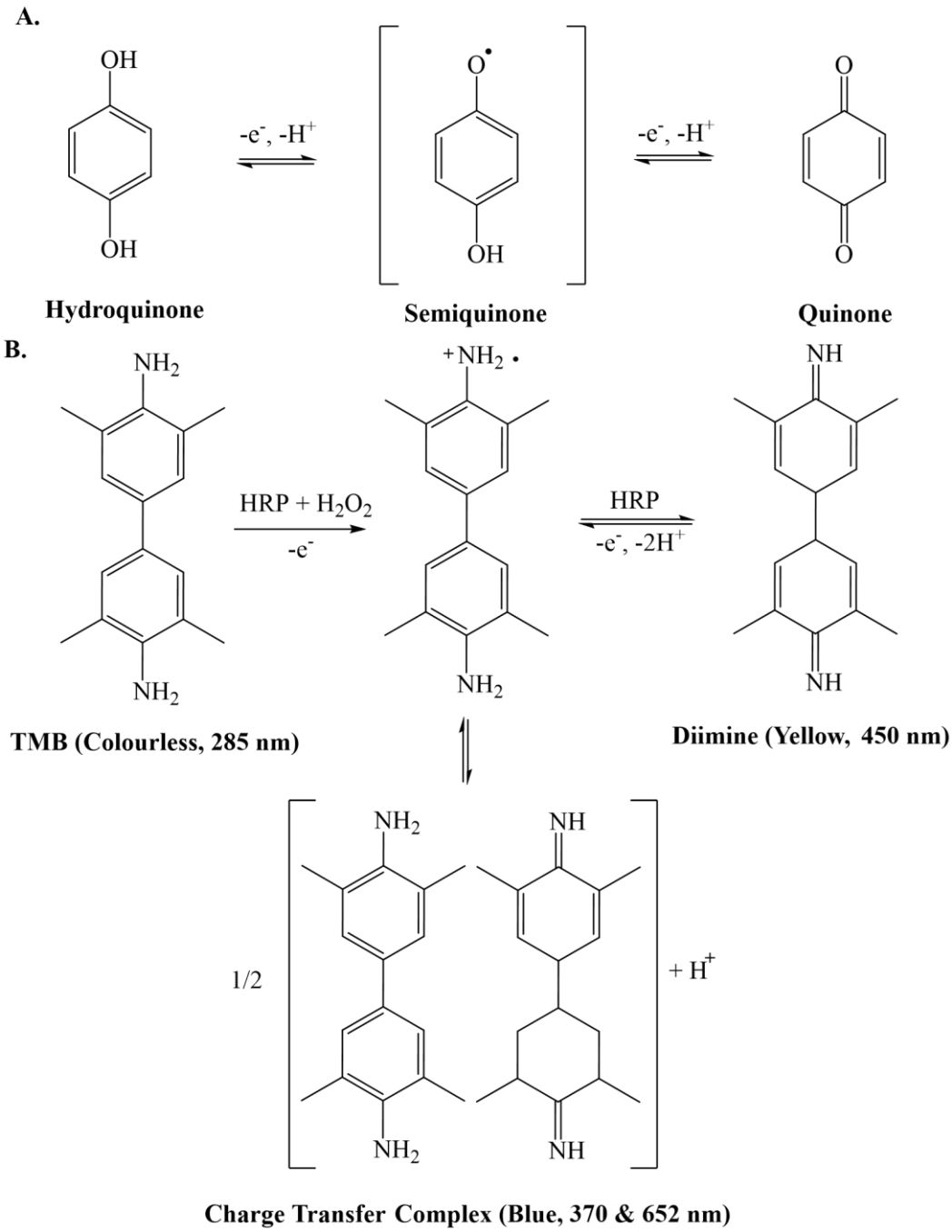


Figure 3-15: Radical oxidation schemes of both Hydroquinone (A) and TMB (B). Created using ChemDraw.

Chapter 4: The Development of WWP1 and WWP2 Inhibitors

4.1 Introduction

After the discovery of a bioactive compound, the next stage in the drug discovery pipeline is termed lead identification or the Hit-to-Lead phase (Hughes *et al.*, 2011). This is where the compound in question is validated through more intensive assays and optimised to improve its biological activity as well as other crucial factors including physiochemical, pharmacokinetic, and toxicity properties. In the earliest stages of a Hit-to-Lead phase, the determination of a small SAR is vital in validating a hit's potential as a lead compound and whether continued optimisations of other pharmacologic properties are worthwhile.

By definition, SAR is the relationship between the chemical structure of a compound to the biological activity of the target studied. In such studies, a series of analogues are generated with the aim of probing the available chemical space in a ligand-receptor site, with the observed SAR utilised to guide the next generated series. Although activity assays are the typical focus of a SAR as a means to increase a hit's potential potency, other biophysical approaches also provide valuable information about ligand interactions that can aid optimisations and improve the overall understanding of a SAR. These biophysical assays range from the continued use of high throughput Hit-identification tools such as the previously mentioned TSA (Section 3.1.1.1) to more sensitive validation tools such as mass spectrometry (MS) used to determine ligand stoichiometry or in more advanced cases such as hydrogen/deuterium exchange (HDX)-MS, ligand binding locations, otherwise termed epitope maps (Ishii, Noda and Uchiyama, 2016). As a non-atomic resolution alternative to the classical structural techniques described later, HDX-MS measures protein mass changes during hydrogen-to-deuterium exchange after a sample is incubated in a deuterated buffer, with the rate of this exchange altered at the interface site by small molecule binding (Marciano, Dharmarajan and Griffin, 2014). This has successfully guided the drug design of SR1164, a novel antidiabetic candidate for type 2 diabetic treatment targeting the peroxisome proliferator-activated gamma receptor (PPARγ) (Choi *et al*., 2011).

For mechanistic characterisations, complementary techniques such as isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR) are used to determine either thermodynamic or kinetic contributions. By measuring heat exchange during molecular interaction, ITC is used to quantify ligand thermodynamic contributions and as such dissociation constraints (K_D) (Section 4.1.2.1), and benefits from its simplicity. This technique does not require any form of label or immobilisation, however, can lack sensitivity to interactions that display minimal heat exchange and is limited by being sample expensive (Krishnamoorthy *et al*., 2020). SPR is used to quantify kinetic parameters such as K_{on}/K_{off} rates through changes in the refractive index of a sensor surface when bound by an immobilised target interacting with its ligand (Holdgate *et al.*, 2019). This technique is much more sensitive

to weaker interactions and is ideal for real-time measurements, yet it can be limited by the potential blocking of binding sites due to the requirement for immobilisation.

The gold standard of ligand characterisations is the ability to gain tangible structural information, enabling a structure-informed approach that can drastically improve understanding of a lead's SAR. Three structural techniques are available in this regard, including X-ray crystallography, NMR and cryo-electron microscopy (Cryo-EM), only more recently becoming applicable in small molecule drug design (Lees, Dias and Han, 2021). However, the demanding nature of structural studies both in resources and manpower better suits these techniques for characterisation of the original/parent hit compound and final lead rather than a continued tool for SAR, as discussed in Section 5.1. In reality, a combined use of these characterisation tools as well as in some cases more complex or target-specific assays are required to validate and optimise hit compounds, the choice of which also has to take into account the accessibility and expertise in such assays. In the case of this study, ITC was chosen due to its simplicity and ability to determine $K_{D,}$ thus aiding the characterisation and understanding of NSC-217913 SAR against WWP1 and WWP2. The previous high-throughput assays, DSF and the ELISA autoubiquitination assay were also utilised. The basic principle behind ITC is described in Section 4.1.2.1. With the expertise of Professor Andrew Hemmings, crystallographic studies of potential inhibitor complexes were also selected to aid a structure-informed approach, with these efforts described in Chapter 5.

The Hit-to-Lead phase of drug discovery is renowned for its poor returns with many hit compounds originally believed to have high therapeutic potential failing. With only NSC-217913 passing the initial high throughput screening process, we wanted to expand our candidate selection to avoid a potential dead-end. Interestingly, earlier work by Gregory Hughes using STD-NMR, demonstrated that the known WWP1 interactor and potential inhibitor I3C had similar interactions with WWP2, although substantially weaker (Appendix Figure 7-1). In this regard, I3C was also selected for Hit-to-Lead optimisations against both WWP1 and WWP2, with further background information found below (Section 4.1.1).

4.1.1 Indole-3-carbinol as a NEDD4 Family Inhibitor

Cruciferous vegetables have long been associated with health benefits, from compounds with high nutritional value to bioactive phytochemicals (Ağagündüz *et al.*, 2022). A significant proportion of the pro-health phytochemicals have been largely accredited to the enzymatic breakdown of glucosinolates (Hayes, Kelleher and Eggleston, 2008), in particular glucobrassicin known to produce the metabolic product I3C (Prado *et al.*, 2022). I3C has been extensively studied for its broad therapeutic potential shown to not only display anti-tumour but also anti-inflammatory, anti-oxidative, anti-viral and anti-microbial properties (Centofanti *et al.*, 2023).

However, these diverse properties may also be attributed to I3C's acid-catalysed condensation products, mainly 3,3′-diindolylmethane (DIM) rapidly produced in acidic environments such as those found in the stomach (Grose and Bjeldanes, 1992) (Figure 4-1). To no surprise, both I3C and DIM can be purchased as health supplements and are under various clinical trials, mainly associated with their effects on breast and prostate cancer (Reyes-Hernández *et al.*, 2023).

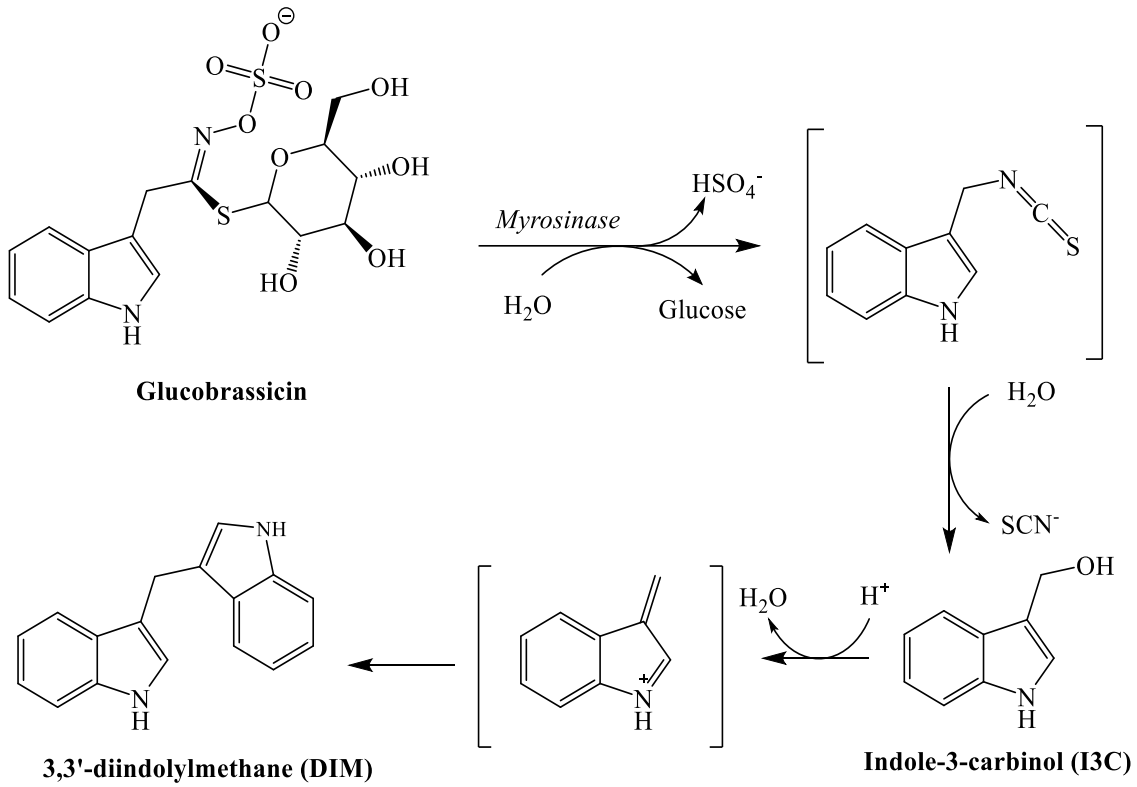


Figure 4-1: Metabolic formation of I3C and DIM from Glucobrassicin. Image created using ChemDraw, adapted from Błoch-Mechkour *et al*. (2010)

Current investigations into I3C's protein interactions have so far identified three targets. Inhibition of the serine protease Elastase was the first to be discovered and resulted in the disruption of NF-kB signalling causing cell cycle arrest and apoptosis in breast cancer cell lines (Nguyen *et al.*, 2008; Aronchik, Bjeldanes and Firestone, 2010; Aronchik *et al.*, 2012). More recently, I3C has been found to target two members of the NEDD4 family of HECT E3 ligases, NEDD4-1 and WWP1. These studies have not only discovered an anti-carcinogenic characteristic of I3C through interfering with NEDD4-1/WWP1 mediated ubiquitination of tumour suppresser protein PTEN but also have highlighted an anti-viral property shown to prevent viral budding in COVID-19 (Aronchik *et al.*, 2014; Lee *et al.*, 2019; Novelli *et al.*, 2021). With NEDD4-1 and WWP1 already known to be promising therapeutic targets, various studies have looked to characterise their I3C interactions, aiming to increase both their acid

stability and potency for future lead development. It has been proposed that I3C binds to a Ub exo site and likely prevents ubiquitin chain elongation (Maspero *et al.*, 2011). This was discovered from an I3C-derived covalent inhibitor found bound to NEDD4-1's non-catalytic Cys627 (PDB ID: 5C91) (Kathman *et al.*, 2015). This site has since been utilised to dock I3C as a means to develop various more potent and stable inhibitors, including 1-Benzyl-I3C shown to decrease I3C's IC_{50} from 284 to 12.3 µM against NEDD4-1. This derivative was also highly potent during MCF-7 breast cancer cell growth studies when compared to I3C (Quirit *et al.*, 2017). However, this is likely a result of a dual action with 1-Benzyl-I3C also shown to target Elastase with increased potency (Aronchik *et al.*, 2012). The I3C Ub exo-site is further supported against WWP1, whereby site mutations into the proposed binding pocket resulted in both a reduced K_D and loss of I3C sensitivity in cell proliferation assays (Lee *et al.*, 2019). Interestingly, this study also demonstrated that cells with WWP1 deletions over NEDD4-1 were more resistant to I3C suggesting WWP1 is the direct target. Thermal shift has also been used to confirm I3C's interactions with the HECT domain, demonstrating an impressive 3.5 °C shift (Zheng *et al.*, 2023). Despite this, in vitro inhibition of I3C, its bioactive condensation product DIM or other stable derivatives on WWP1 have not been stated.

4.1.2 Biophysical Techniques

4.1.2.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is in principle a very simple label-free technique measuring the heat released during the binding of a ligand to a receptor. In practice, the target receptor/protein is added to a sample cell contained within an adiabatic shield. A twinned reference cell is also found within this shield, either filled with water or sample buffer. The chosen ligand is then titrated into the sample cell in small aliquots before the resulting reaction temperature is recorded by a change in feedback power required to maintain the same temperature as the reference cell (Figure 4-2). As the assay itself is performed at a constant temperature and pressure, the associated heat is directly related to the enthalpy (ΔH) and entropy (ΔS) of the system, linked through the Gibbs free energy (G) relationships:

$$\Delta G = -RT\ln K_{eq} \text{ and } \Delta G = \Delta H - T\Delta S$$

Where, R = gas constant and T = temperature (K)

Once conditions have been optimised to enable both sufficient heat release and saturation of the associated binding sites, several factors can be determined such as dissociation constants (K_D), reaction enthalpy (ΔH), and even binding stoichiometry (n) (Bastos *et al.*, 2023). If run at varied

temperatures, the heat capacity (Δ*Cp*) of binding can also be determined with large values associated with hydrophobic interactions.

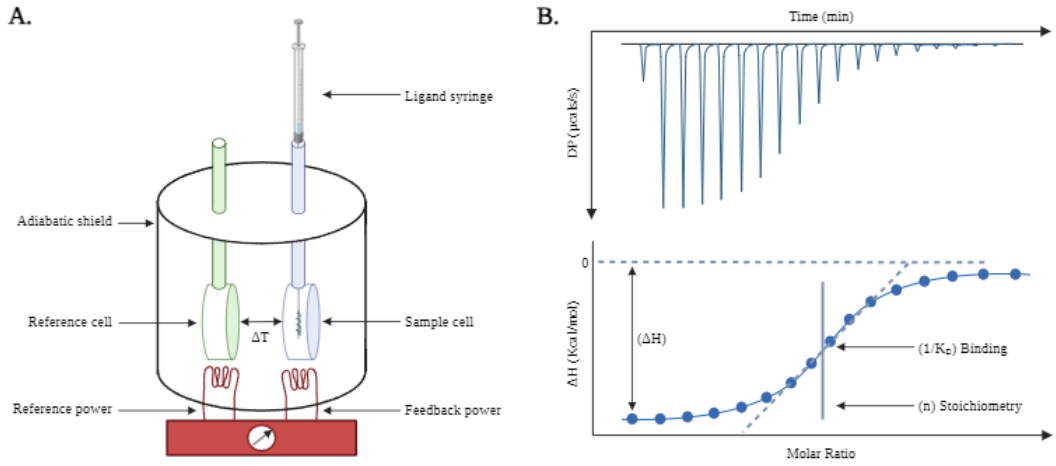


Figure 4-2: The basic principles of isothermal titration calorimetry. (A) Schematic illustration of an ITC instrument. The ligand is injected into the sample through a syringe containing a stirrer at the end. This is done sequentially at a specified volume over a set time. The temperature between the sample and reference cells (water or buffer) is maintained by a heater (red), recording the power required. The sample and reference cells are contained in an adiabatic shield to prevent heat exchange with the external environment. (B) Example experimental data (top) with evaluation features (bottom).

4.1.3 Experimental Aims

In continuation of the previous chapter and the next Hit-to-Lead phase in the drug discovery pipeline, this chapter aims to explore the surrounding chemical space of both NSC-217913 (discovered from the NCI Diversity VI compound library) and I3C (highlighted in literature) against a range of WWP1 and WWP2 constructs (Figure 4-3) to answer the question; What is the structure-activity relationship of these small molecules against WWP1 and WWP2? In collaboration with Jake Rigby from the School of Chemistry UEA, an SAR by synthesis approach will be undertaken for NSC-217913 generating a small analogue series with a focus on chemical accessibility. In contrast, given its commercial presence, a SAR by catalogue approach will be used for I3C, with only a few derivatives synthesised that have been highlighted in the literature. Comprising of three subsections, the first section will focus again on identifying target interactions using DSF assays for both parent series by screening against WWP1-2L34H and WWP2-LH constructs. The second section will focus on target activity using the ELISA autoubiquitination assay using the WWP1-L34H and WWP2-FL constructs, further validating through counter screens. The final section looks to introduce ITC as a means to gain a more detailed binding characterisation, first focusing on optimising the assay towards the WWP2-H construct.

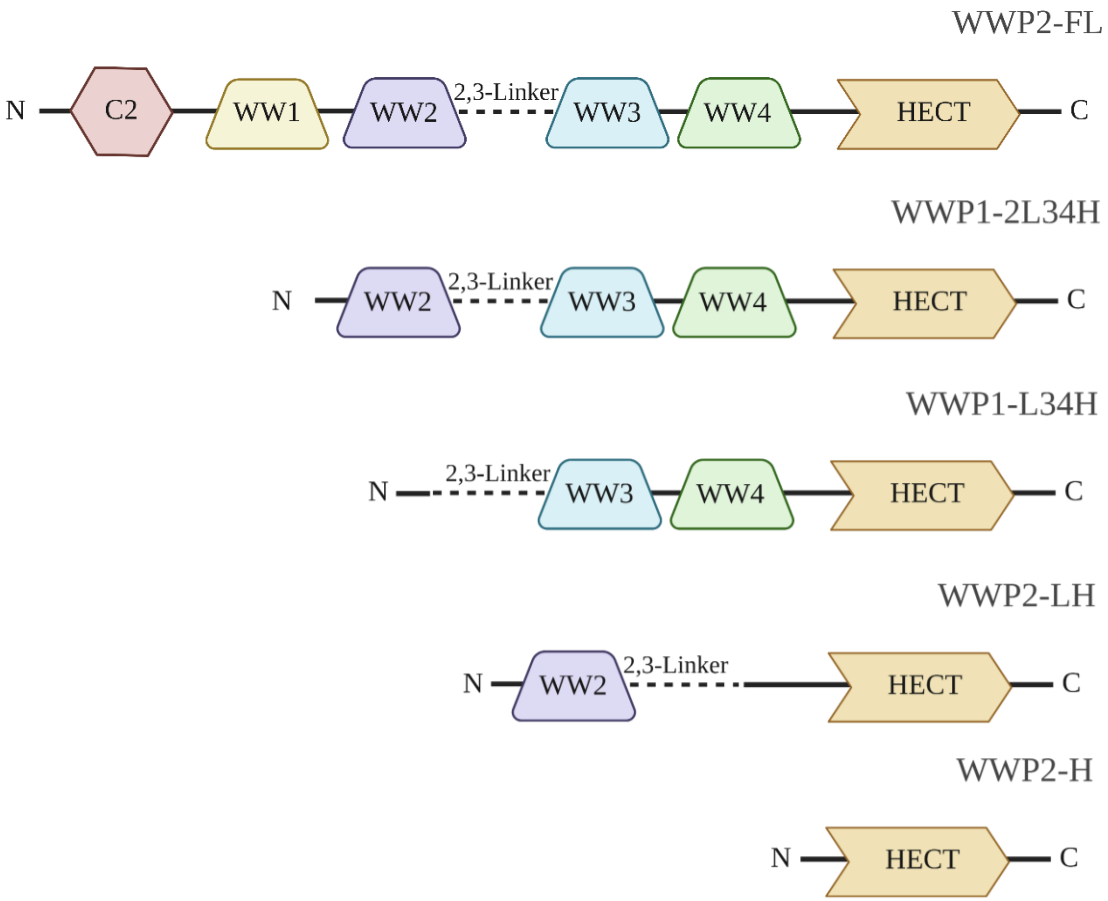


Figure 4-3: Domain architecture of the WWP ligase constructs used in Chapter 4. The full-length WWP2 E3 ligase is composed of a Ca^{2+}-binding domain (C2, red) and the four tryptophan-tryptophan (WW) domains, WW1 (yellow), WW2 (purple), WW3 (blue) and WW4 (green), as well as the autoinhibitory 2,3-linker region (dashed line), before finally the catalytic HECT (orange) domain. The WWP1-L34H and WWP1-2L34H construct, previously reported by Wang *et al*. (2019), consists of the WW3, WW4 and HECT domains as well as the 2,3-linker located at the N-terminus, with WWP1-2L34H also containing an N-terminal WW2 domain. The WWP2-LH construct, previously reported by Chen *et al*. (2017), consists of the WW2 and HECT domains separated by the 2,3-linker region. The WWP2-H construct, generated in this study, consists of only the catalytic HECT domain.

4.2 Results

4.2.1 NSC-217913 and I3C Analogue Generation

In collaboration with Dr Thomas Storr's research group, namely PhD researcher Jake Rigby from the School of Chemistry UEA, a series of analogues were generated for both NSC-217913 and I3C. A total of 27 compounds were synthesised for NSC-217913 with a focus on chemical accessibility. Analogues are represented with the letter S, including NSC-217913 itself labelled as S6 (Figure 4-4). For I3C, a total of 21 compounds were either synthesised or purchased, including DIM and other derivatives represented with the letter I to aid distinction (Figure 4-5). As well as those commercially available, various stable derivatives were synthesised based on those observed in literature to possibly interact with WWP1 or NEDD4-1 (Weng *et al.*, 2007; Quirit *et al.*, 2017; Zheng *et al.*, 2023). All chemical information regarding synthesis strategies for the provided compounds can be found in the Appendix and/or shown in Jake Rigby's Thesis (2024).

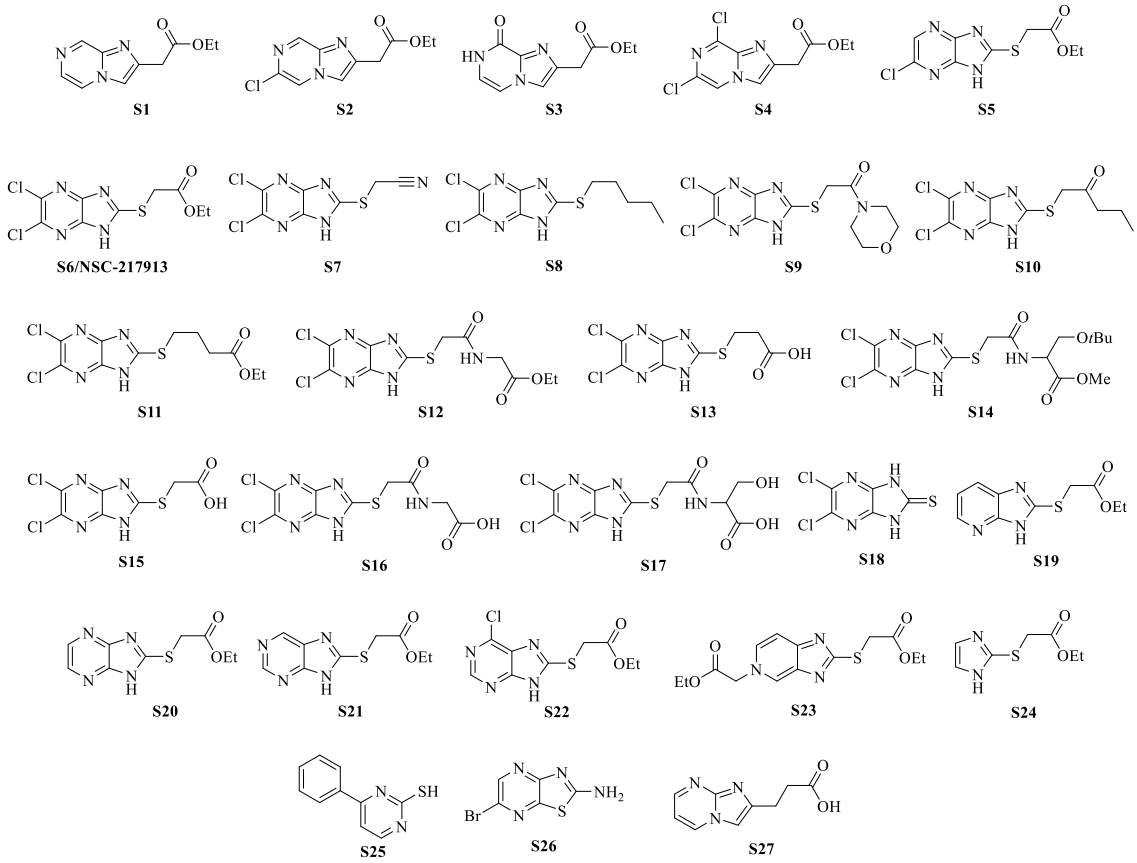


Figure 4-4: Structures of NSC-217913 (S) analogues under investigation with associated numbering. Images created in ChemDraw.

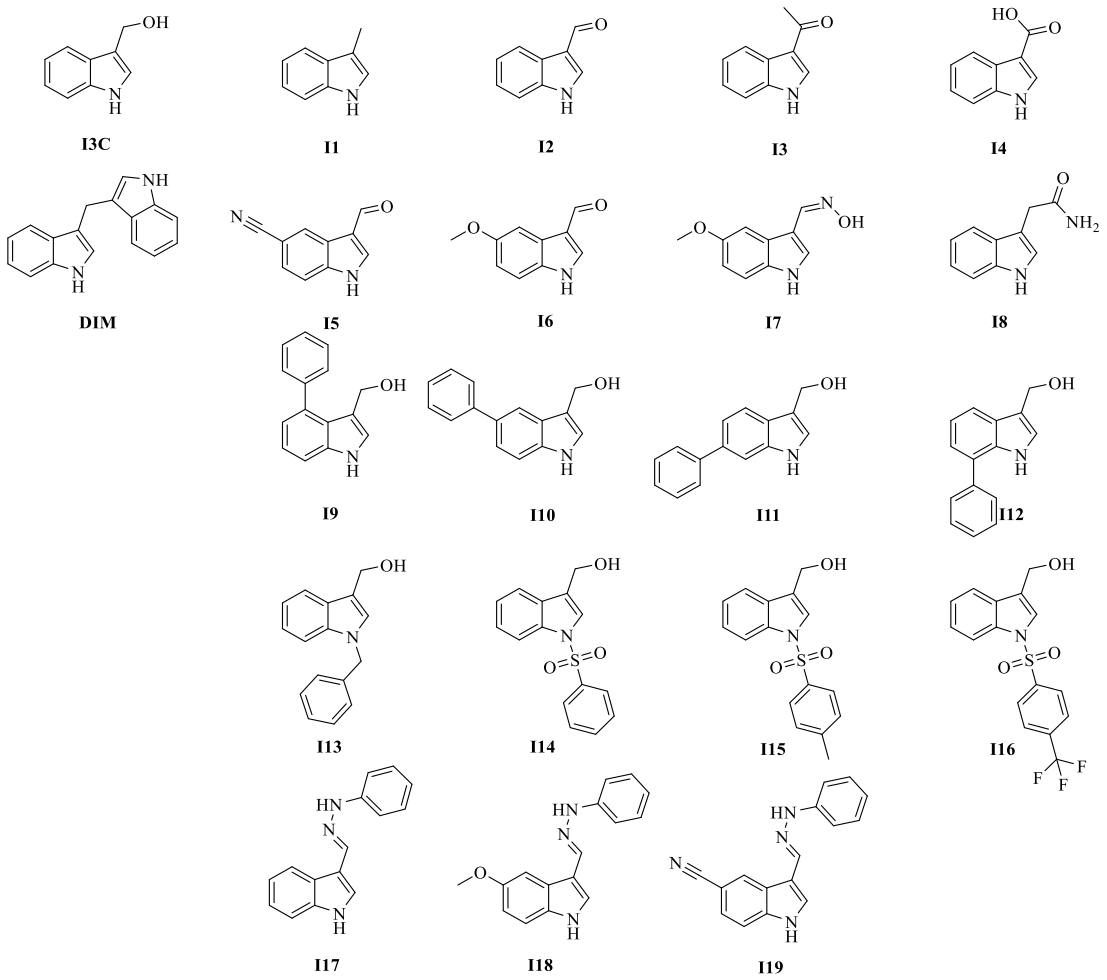


Figure 4-5: Structures of I3C (I) derivatives under investigation with associated numbering. Images created in ChemDraw.

4.2.2 DSF Screening of NSC-217913 and I3C Analogues Against WWP1 and WWP2

Given the success of the high throughput screening strategy to identify NSC-217913 (Section 3.2.1), DSF was also used to assess possible interactions of analogue series with WWP1 and WWP2.

4.2.2.1 DSF assay: construct selection and purification

At this stage in the project a slightly larger construct of WWP1, WWP1-2L34H, had been successfully purified for crystallography (See Section 5.2.1) and so was chosen for future DSF screening (Figure 4-3). This construct contains the WW2 domain, in comparison to the previous WWP1-L34H construct, and was theorised it may provide additional stabilising interactions associated with the WW2 domain binding at the N-lobe (Wang *et al.*, 2019). Purification strategies for both WWP1-2L34H and WWP2-LH are stated in Sections 2.2.7.4 and 2.2.7.5 with SDS-PAGE analysis shown in Figure 4-6 and Figure 3-6, respectively.

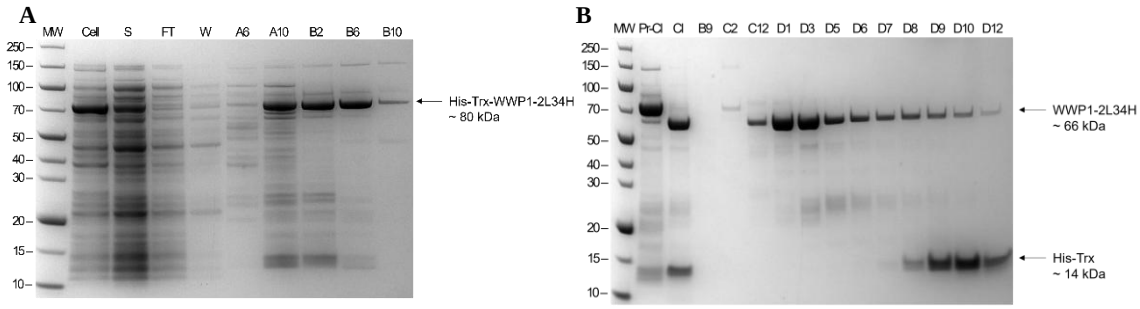


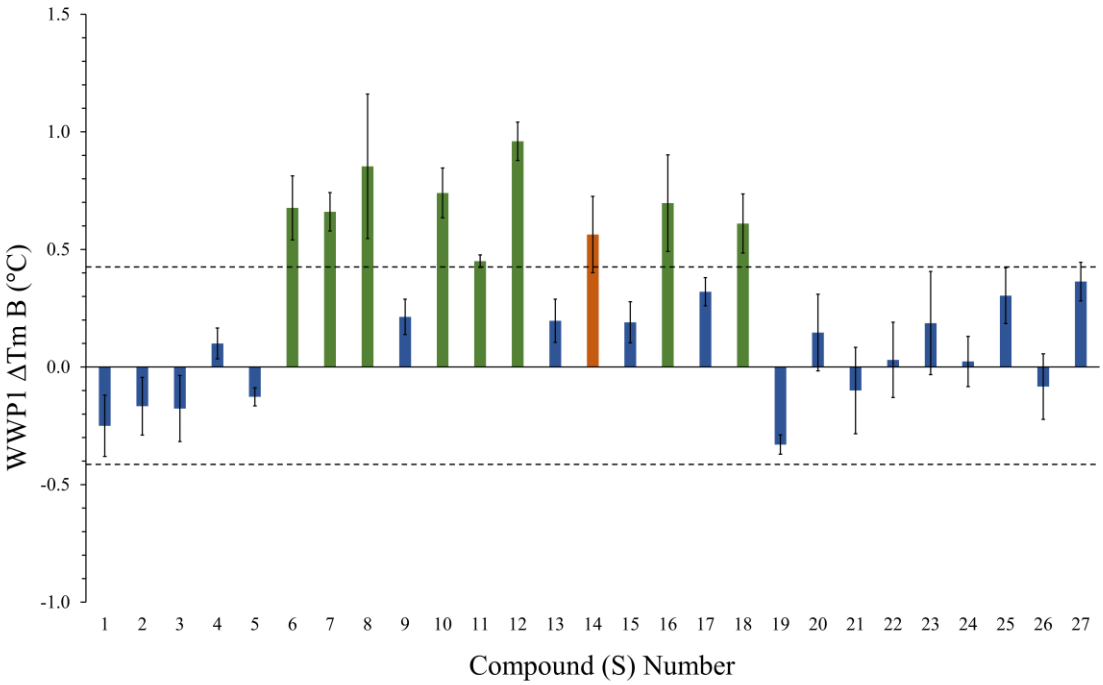
Figure 4-6: SDS-PAGE analysis of WWP1-2L34H purification. (A) Gradient IMAC isolation. Intense bands at ~80 kDa indicating Trx-His tagged WWP1-2L34H. Lane labels: Ladder (MW), whole cell (Cell), soluble lysate (S), column flow-through (FT), column wash (W), elution fractions positions (A6 – B10). Fractions A10 – B10 were pooled. (B) SEC purification. Untagged WWP1-2L34H identified by intense bands at ~66 kDa. Lanes labelled as fraction positions other than ladder (MW), pre-cleavage (Pr-Cl) and cleavage sample (Cl). Fractions C12 – D6 were pooled. All gels were run on pre-cast 4 – 12 % Bis-Tris gels.

4.2.2.2 DSF screen of NSC-217913 and I3C analogues discovered various stabilising and destabilising compounds

The analogue series of NSC-217913 and I3C were screened at 100 µM (0.1% DMSO) following experimental procedures (Section 2.3.2.1), including those optimised for WWP1-2L34H during buffer screening as later described in Section 5.2.1.1. Conversely, to the previous high throughput screen, the Tm for each compound was determined using a Boltzmann fit (Tm B) to the respective melting curve. This was found to give more consistent results being able to minimise noise created from various compounds and remove the effect of peak shift seen in the first derivative model. Positive hits were selected based on a threshold of three times the standard derivation of the respective WWP1 and WWP2 controls including both 'stabilising' and 'destabilising' interactions. Borderline hits were considered as compounds displaying a Tm over the threshold but have an error that spans it.

For the NSC-217913 analogues, 9 compounds including **S6**/NSC-217913 were identified as stabilising WWP1 hits, with S14 as borderline (Figure 4-7 top). Only 7 compounds displayed WWP2 stabilising interactions, 3 of which were borderline hits. Interestingly this did not include S6/NSC-217913. Furthermore, S13 and S17 were observed to destabilise WWP2, showing no signs of interaction with WWP1 (Figure 4-7 bottom). Compounds S7, S12 and S16 were the only analogues to show consistent interaction across WWP1 and WWP2.

For I3C derivatives, 6 compounds including both I3C and DIM were observed as having stabilising WWP1 interactions, with 2 of these being borderline hits. Derivative I14 was the only compound to display destabilising WWP1 interactions, although only as a borderline hit, with this feature repeated against WWP2 (Figure 4-8 top). For the other WWP2 interactions, 4 derivatives each were found to stabilise and destabilise WWP2 (Figure 4-8 bottom). Of note, DIM, and I7 both displayed significant ΔTm B shifts compared to other compounds at around 1.0 °C. I3C itself did not show any signs of WWP2 interaction. Derivatives with consistent interactions across WWP1 and WWP2 include DIM, I7, I14, and I19 with I11 showing alternative stabilising and destabilising interactions.



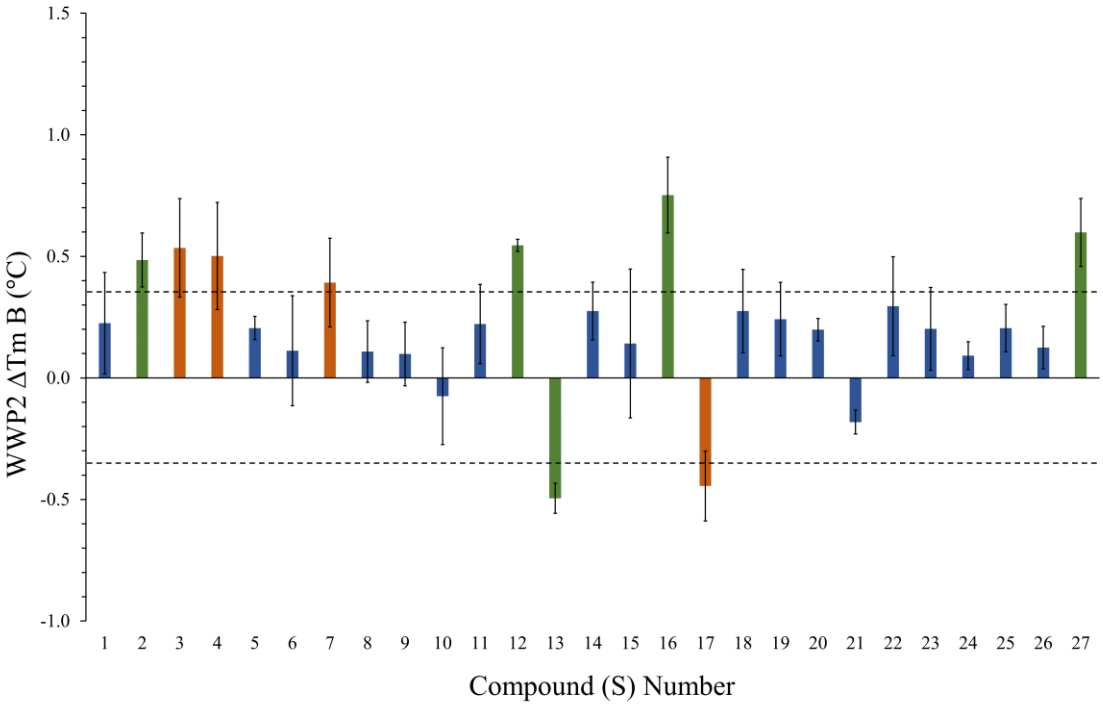


Figure 4-7: DSF screen of NSC-217913 and its analogues against WWP1-2L34H (top) and WWP2-LH (bottom). Assays were carried out using 100 µM of compound, containing 0.1% DMSO and normalized by ΔTm calculated from controls. Positive hits (green) displayed had an average ΔTm and associated errors above a threshold of three times the standard deviation of their respective controls, with borderline hits (orange) having associated errors spanning the threshold. NSC-217913 is represented as compound S6.

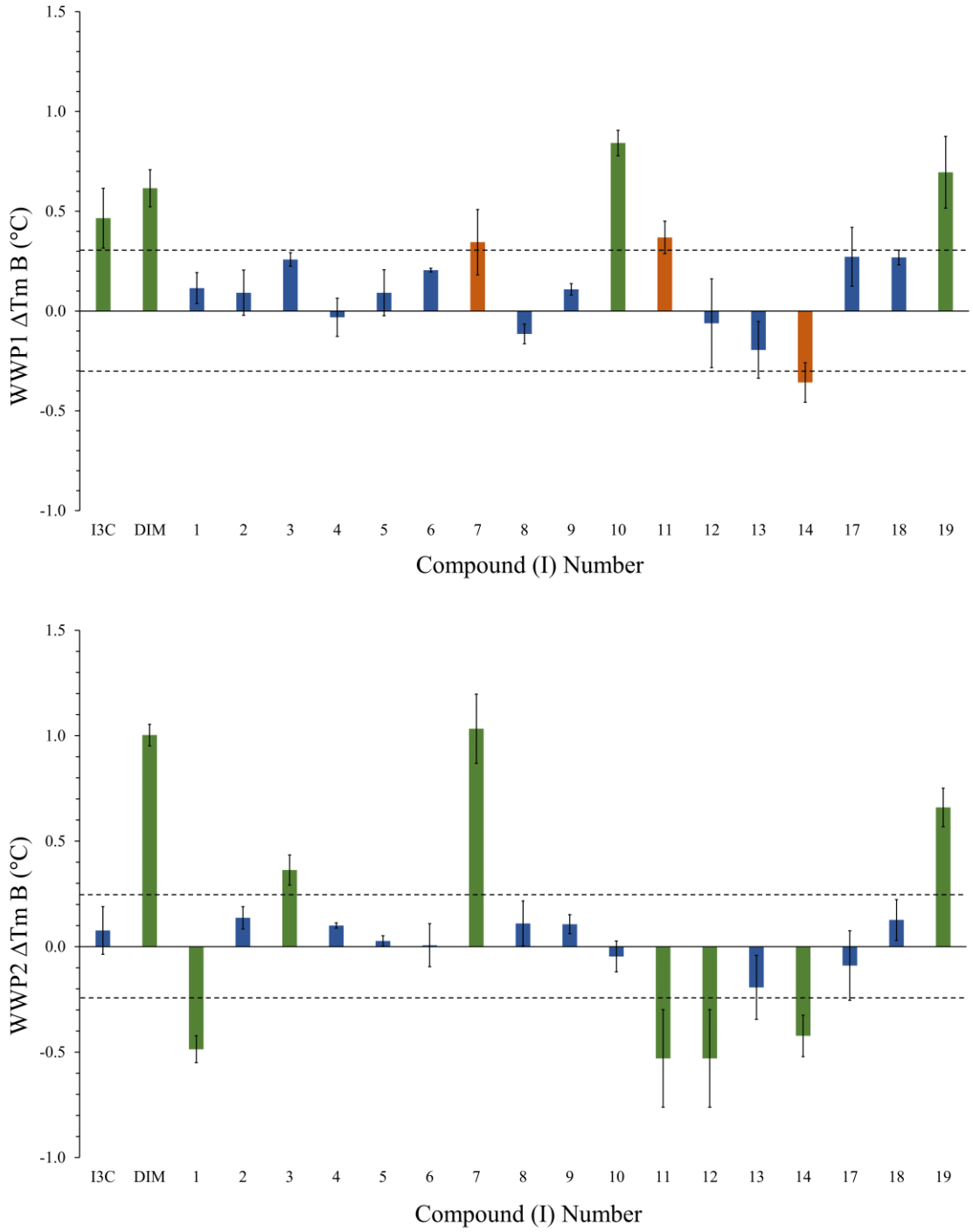


Figure 4-8: DSF screen of I3C, DIM and its derivatives against WWP1-2L34H (top) and WWP2-LH (bottom). Assays were carried out using 100 µM of compound, containing 0.1% DMSO and normalized by ΔTm calculated from controls. Positive hits (green) displayed had an average ΔTm and associated errors above a threshold of three times the standard deviation of their respective controls, with borderline hits (orange) having associated errors spanning the threshold. Compound I15/N-Tosyl-I3C and compound 16 were omitted due to unusual thermal traces.

4.2.3 Screening of NSC-217913 and I3C Analogues Against WWP1 and WWP2 Autoubiquitination Assays

The NSC-217913 and I3C analogue series were next screened for inhibition again following the same autoubiquitination strategy as observed in Section 3.2.2. All compounds in the series were screened irrespective of the previous DSF results.

4.2.3.1 Single-shot auto-ubiquitination assay highlighted various inhibitory compounds from both NSC-217913 and I3C analogues

The analogue series was first evaluated in a single-shot assay at 1 mM (1% DMSO) measuring inhibition against their relative auto-ubiquitination activity following the protocols for WWP1 and WWP2 as stated in Sections 2.3.1.2 and 2.3.1.3, respectively. Positive hits were chosen by a threshold of less than 50% relative activity (RA) when compared to 0 % and 100 % controls.

For the NSC-217913 analogues, 6 compounds including S6/NSC-217913 were observed to be below the threshold for WWP1, with 8 compounds excluding S6/NSC-217913 selected as WWP2 hits (Figure 4-8). The WWP2 inhibition was overall less potent with only S18 displaying less than 25% RA, interestingly this compound essentially removed all WWP1 activity at 1 mM displaying a value below the negative control (- 2.0% RA). Compounds S7, S11, S15, and S18 all showed inhibition towards both WWP1 and WWP2.

For the I3C derivatives, 7 compounds including DIM were selected as WWP1 hits, with only 5 compounds displaying WWP2 inhibition below the threshold (Figure 4-10). Intriguingly, although I3C did display inhibition towards WWP1 with a relative activity below the threshold, its associated error led to it being deemed as a borderline hit. In WWP2, neither I3C nor DIM reached the threshold, however, DIM's increased inhibition capacity over I3C observed for WWP1 was carried through to WWP2, with DIM having a RA ~ 60% at 1 mM. Compounds I15, I16, I18 and I19 all displayed inhibition to both WWP1 and WWP2 within the threshold.

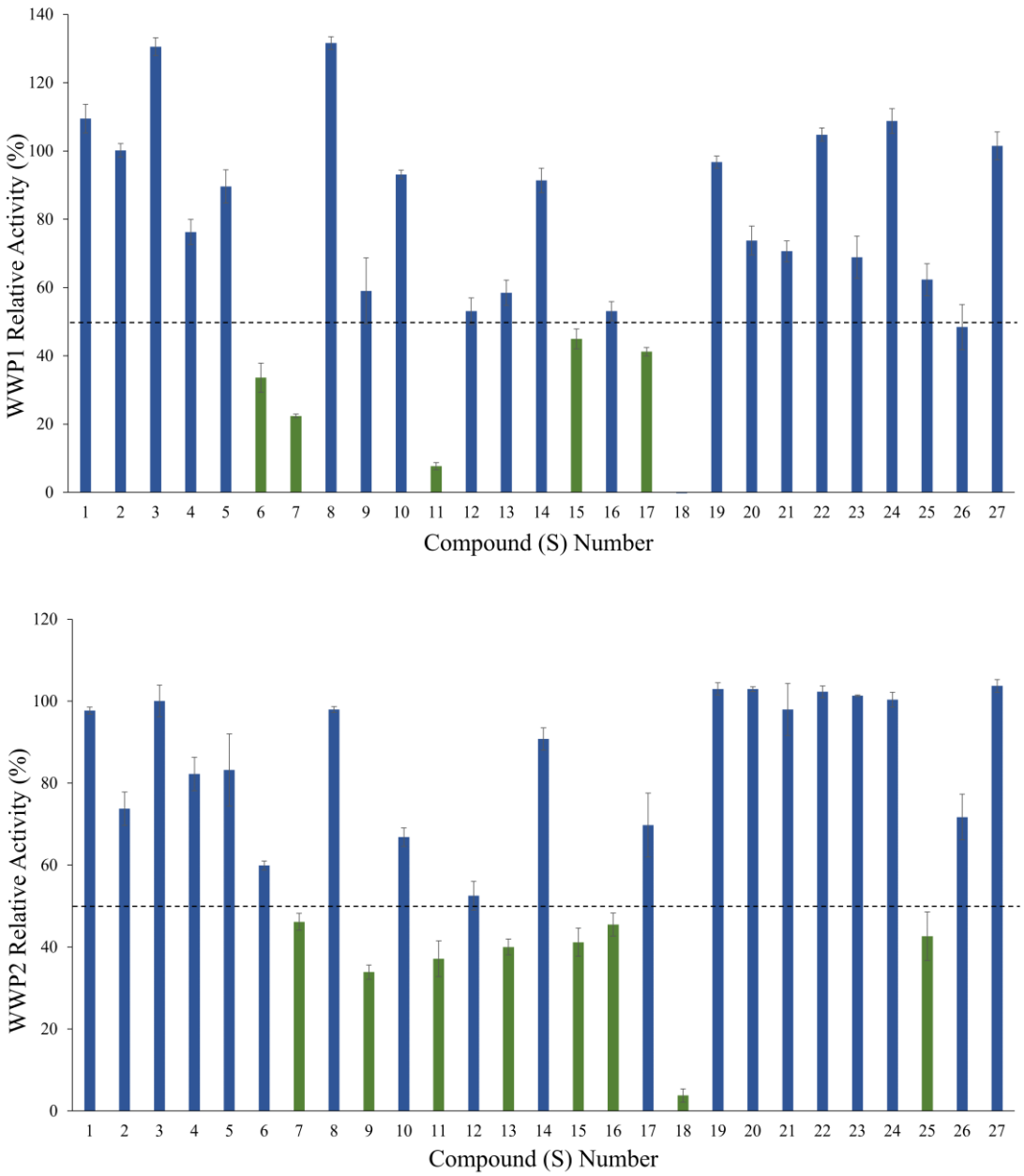


Figure 4-9: Single-shot Autoubiquitination screen of 27 NSC-217913 analogues against WWP1-L34H (top) and WWP2-FL (bottom). Compound inhibition was measured at 1 mM (1% DMSO), with a hit threshold (green) of less than 50% relative activity (dashed line), normalised to their respective 0% and 100% WWP1-L34H and WWP2-FL controls.

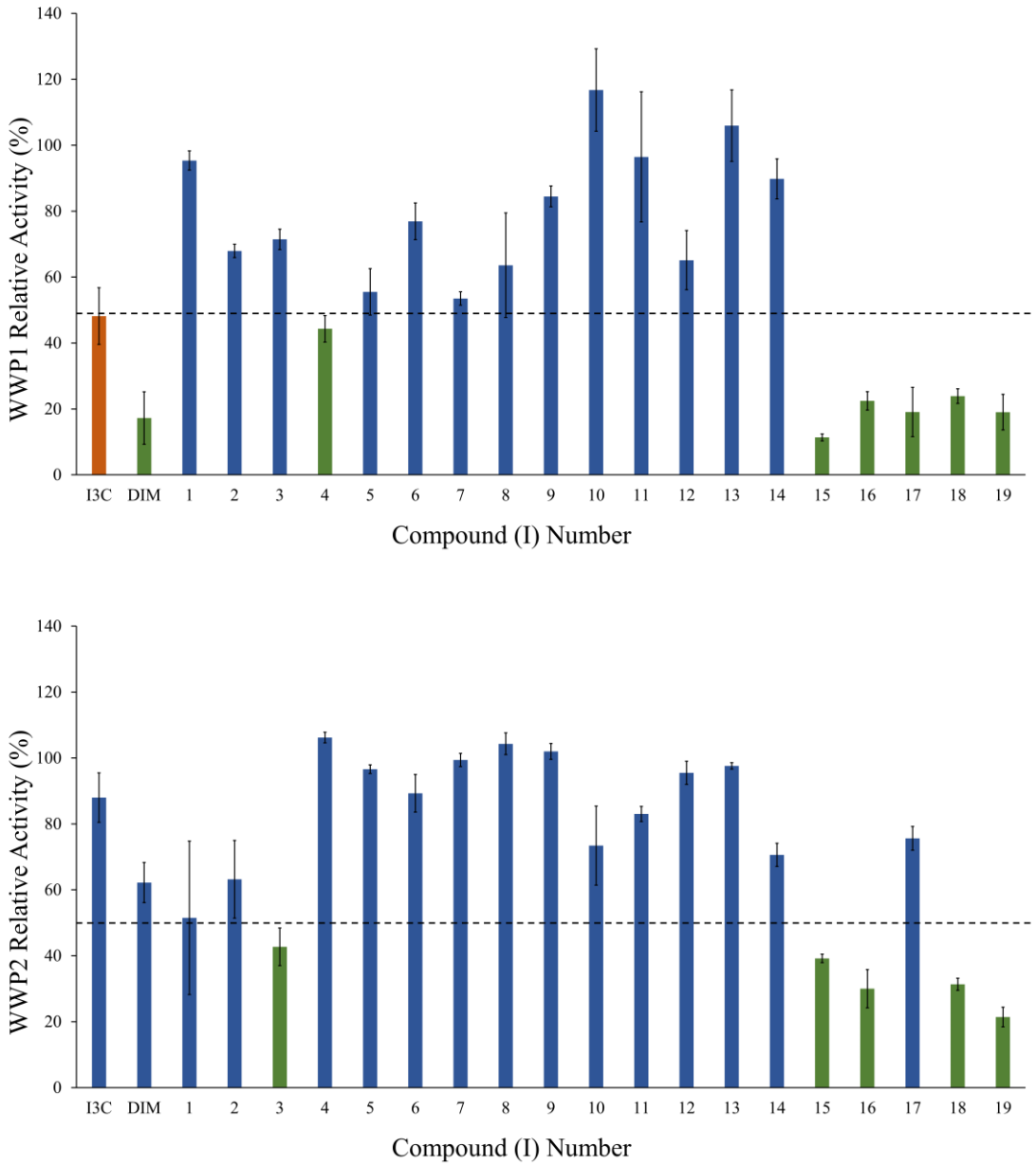


Figure 4-10: Single-shot Autoubiquitination screen of I3C derivatives against WWP1 (top) and WWP2 (bottom). Compound inhibition was measured at 1 mM (1% DMSO), with a hit threshold (green) of less than 50% relative activity (dashed line), normalised to their respective 0% and 100% WWP1-L34H and WWP2-FL controls. Borderline hits (orange) are displayed with a relative activity below the threshold but have associated errors above.

4.2.3.2 Dose-response screening discovered analogues from both the NSC-217913 and I3C libraries had increased potency for WWP1 and WWP2.

The single-shot hit compounds were further tested in dose-dependent assays using a logarithmic scale from 1 mM – 10 nM, following the lessons learned from our previous high throughput screen. The log concentration was plotted against relative activity to enable IC_{50} values to be calculated. Please note that the WWP1 dose dependency screens were completed after both WWP2 and counter-assay screens, so various compounds were not carried forward from the single-shot assays.

For the NSC-217913 analogues, NSC-217913 itself was not included having already been screened earlier in the process (Figure 3-12). For the WWP1 dose dependency assay, 4 compounds were screened (Figure 4-11). Compound S11 displayed the most potent inhibition with an IC_{50} of 32.74 µM. Unfortunately, although S17 did display the second most potent inhibition at an IC_{50} of 110.4 µM, closer inspection revealed a purity issue (82%) with the submitted compound. Compounds S7 and S15 did show dose dependency at IC_{50} of 220.5 and 375.5 µM, respectively. However, it should be noted that the 10 nM values were omitted to enable an IC_{50} curve to be fit. All 8 hit compounds were tested for dose dependency against WWP2 (Figure 4-12). Compound S25 had an unusual trace and therefore was omitted at this stage. The most potent inhibition was observed by S18 with an impressive IC_{50} of 1.19 µM. S11 again retained a dose response against WWP2 as with WWP1, although less potent with an IC_{50} of 269.2 µM. All other tested compounds recorded IC_{50}'s over 600 µM, including the WWP1 hits S7 and S15 again being less potent with an IC_{50} of 856.3 and 666.4 µM respectively.

For the I3C derivatives, all 7 hit compounds were screened for dose dependency against WWP1, with all apart from I4 displaying a dose response (Figure 4-13). The most potent inhibition was observed from DIM with an IC_{50} of 111.2 µM. Three other compounds were also mildly potent (IC_{50} below 250 µM), I18, I19 and I15 with respective IC_{50} of 180.5, 207.8 and 218.3 µM. All 5 single shot-hit compounds were screened against WWP2 (Figure 4-14). Compound I15 displayed the most potent inhibition with a recorded IC_{50} of 223.7 µM, similar to that observed against WWP1, albeit with a slightly larger than ideal 95% confidence interval. This similarity between WWP1 and WWP2 inhibition was repeated for I19 with an IC_{50} of 245.6 µM. The other three derivatives had IC_{50} over 250 µM, including I3 with no stable curve able to be calculated at the concentrations used.

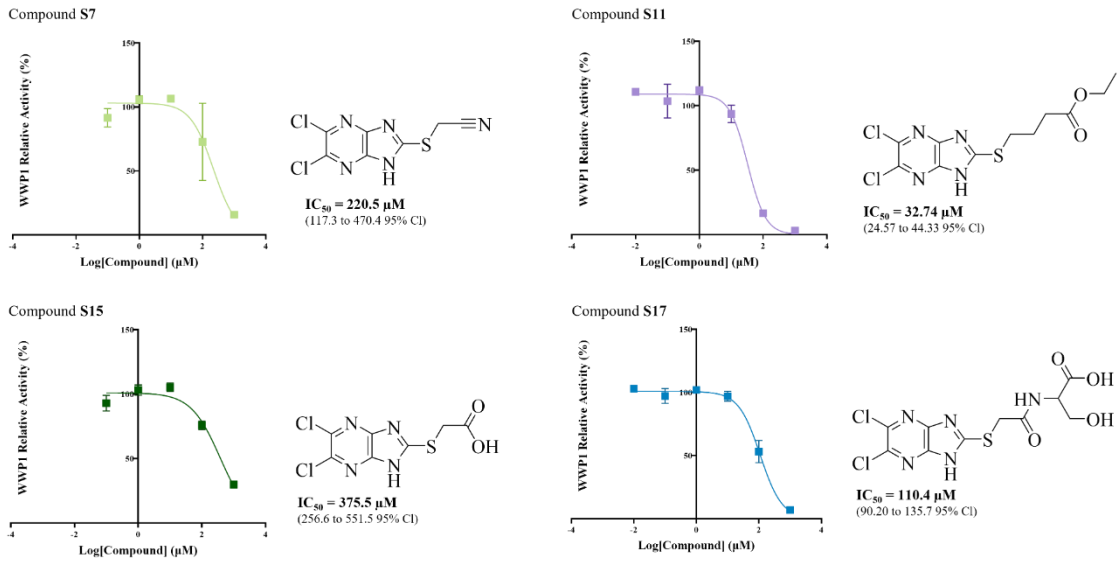


Figure 4-11: Dose-dependent auto-ubiquitination assay of single-shot hit NSC-217913 analogues against WWP1-L34H. Compound inhibition was measured on a log scale from 1 mM – 10 nM (1% DMSO), normalised to 0% and 100% controls. IC_{50} values were calculated from non-linear regression curves fitted in GraphPad software. The 95% confidence interval (CI) is given in brackets. The 10 nM values of compounds S7 and S15 were omitted to enable curve fitting.

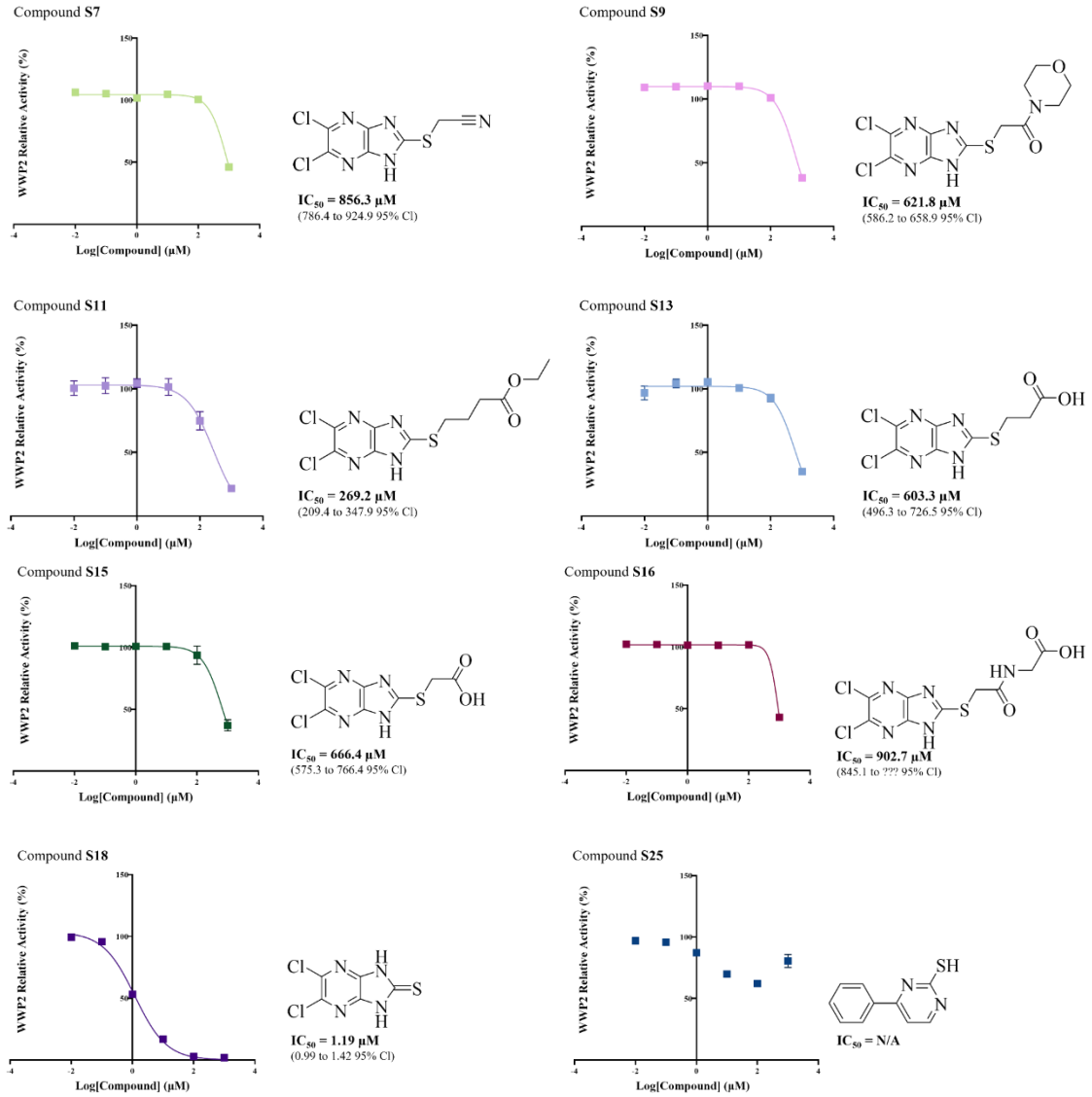


Figure 4-12: Dose-dependent auto-ubiquitination assay of single-shot hit NSC-217913 analogues against WWP2-FL. Compound inhibition was measured on a log scale from 1 mM – 10 nM (1% DMSO), normalised to 0% and 100% controls. IC_{50} values were calculated from non-linear regression curves fitted in GraphPad software. The 95% confidence interval is given in brackets.

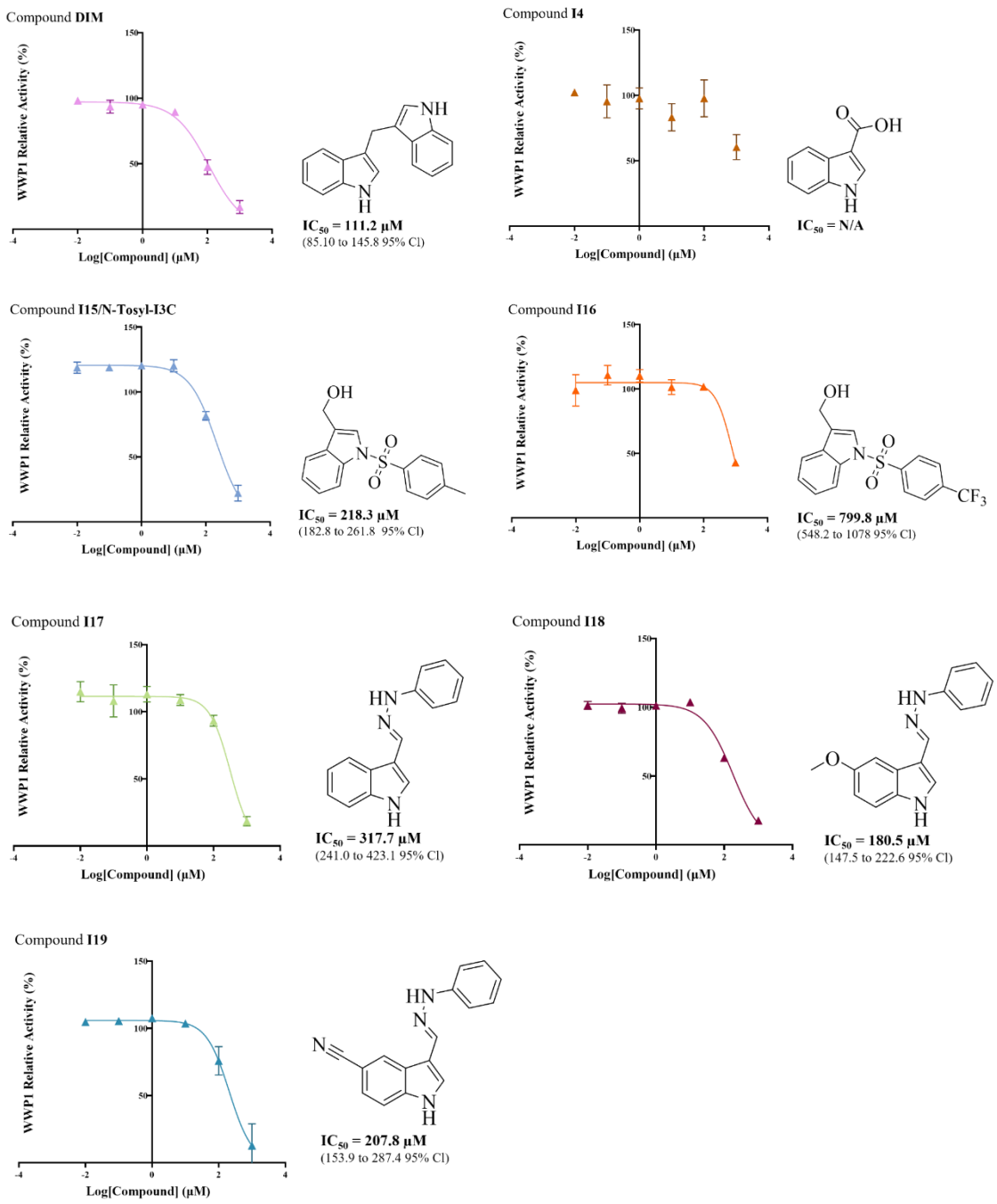


Figure 4-13: Dose-dependent auto-ubiquitination assay of single-shot hit I3C derivatives against WWP1-L34H. Compound inhibition was measured on a log scale from 1 mM – 10 nM (1% DMSO), normalised to 0% and 100% controls. IC_{50} values were calculated from non-linear regression curves fitted in GraphPad software. The 95% confidence interval is given in brackets.

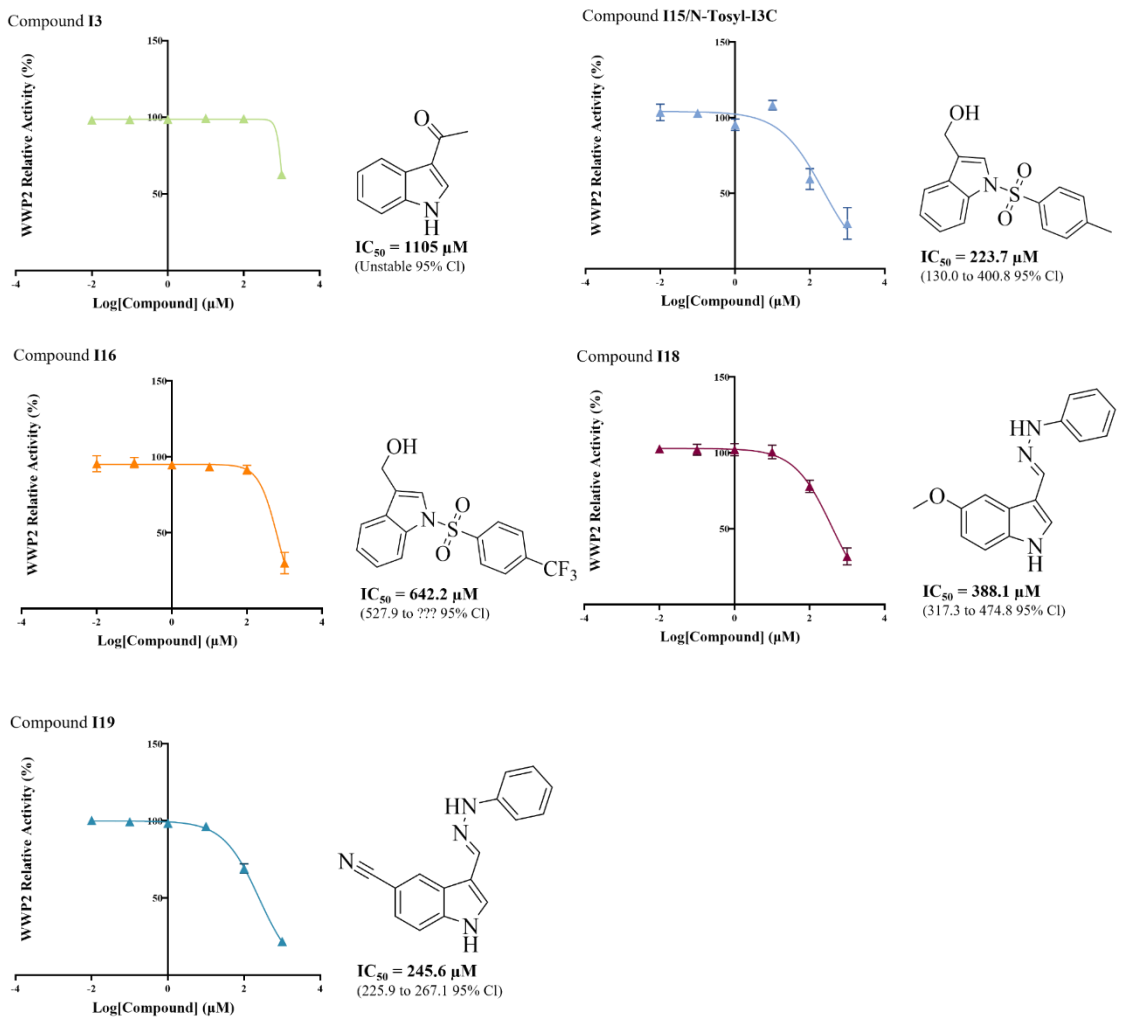


Figure 4-14: Dose-dependent auto-ubiquitination assay of single-shot hit I3C derivatives against WWP2-FL. Compound inhibition was measured on a log scale from 1 mM – 10 nM (1% DMSO), normalised to 0% and 100% controls. IC_{50} values were calculated from non-linear regression curves fitted in GraphPad software. The 95% confidence interval is given in brackets.

4.2.3.3 Counter-assay analysis of hit analogues showed the most potent NSC-217913 and I3C analogues were acting off-target

Our final screening strategy was to check whether the most potent compounds (below 250 µM) were inhibiting their intended targets or were simply interfering with Uba1 and UbcH7. Counter-assays were performed at 1 mM compound (1% DMSO) following procedures stated in Sections 2.3.1.4 and 2.3.1.5, with most only requiring UbcH7 activity given its use of both Uba1 and UbcH7 as previously discussed.

For the NSC-217913 analogues, compounds S7, S11, S15, S17 and S18 were selected for the counter assays (Table 4-1). Except for S18, all other compounds displayed minimal interference, including S17 previously mentioned to be impure.

For the I3C derivatives, DIM, I15, I16, I17, I18 and I19 were selected for the counter-assays (Table 4-2). Both DIM and I15 displayed the least interference above 80% RA, with I16 only just below this cutoff. The other three derivatives all display varying amounts of interference, with I18 inhibiting the UbcH7 activity by almost half.

Table 4-1: Uba1 and UbcH7 counter auto-ubiquitination screen of NSC-217913 analogues

Compound (S)	Uba1 Activity (%)	UbcH7 Activity (%)
7	-	94.2 ± 0.8
11	-	88.8 ± 0.2
15	-	93.0 ± 1.0
17	-	88.8 ± 0.7
18	22.1 ± 4.1	-

Table 4-2: UbcH7 counter auto-ubiquitination screen of hit I3C (I) derivatives.

Compound (I)	UbcH7 Activity (%)
DIM	83.3 ± 2.0
15/N-Tosyl-I3C	86.3 ± 1.8
16	79.7 ± 1.5
17	61.2 ± 1.6
18	52.2 ± 3.2
19	64.4 ± 1.5

4.2.4 Efforts Towards ITC Ligand Characterisation

To further characterise interactions between the identified small molecule inhibitors with WWP1 and WWP2, we looked to introduce ITC to determine their binding constants (K_D). With the project focus initially on WWP2, a significant proportion of time was used in attempting to purify a high-yielding construct to account for ITC's renowned sample demands.

4.2.4.1 WWP2 HECT purification and optimisation required the use of a StrepII-tag

WWP2-H was chosen as the highest-yielding construct and the most likely to interact with the inhibitors given the HECT domain's requirement for activity. The initial choice was His-tagged WWP2-H originally utilised by the Chantry group for STD-NMR screening and previously crystallised by Gong *et al*. (2015)(PDB ID: 4Y07). The pET28a vector containing His-tagged WWP2-H was expressed in BL21-Rosseta cells, at 1 mM IPTG, 25 °C, overnight and purified following the same protocol used for STD-NMR preparation as described in Section 2.2.7.2, with SDS-PAGE analysis shown in Figure 4-15. Unfortunately, to obtain samples of the purity level required for ITC, the final yield was significantly reduced (~34 mg to 4 mg/L culture). This was likely due to WWP2's HECT domain directly attaching to the Ni-NTA matrix, as seen in Figure 4-15A, whereby imidazole-mediated elution occurs at all fractions. This could have also been through soluble precipitants of misfolded WWP2-H non-specifically interacting with the column, supported by a significant proportion of WWP2-H lost in the void volume (B11-C1) during SEC purification (Figure 4-15B). Alternative attempts at purifying the His-tagged WWP2-H construct resulted in worse outcomes including on-column cleavage and anion exchange as utilised in the original literature (Gong *et al*., 2015). A large-scale anion exchange may have been more successful, but the equipment costs were deemed too high a risk.

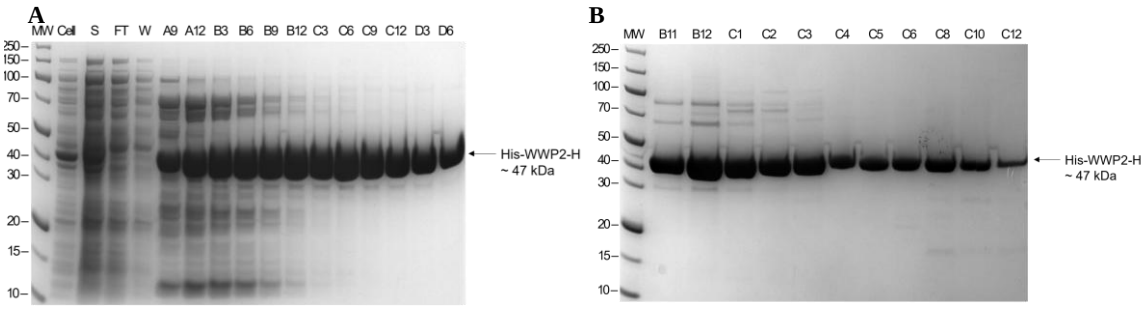


Figure 4-15: SDS-PAGE analysis of His-tagged WWP2-H purification. Intense bands at ~44 kDa indicating His tagged WWP2-H. (A) Gradient IMAC isolation. Lane labels: Ladder (MW), whole cell (Cell), soluble lysate (S), column flow-through (FT), column wash (W), elution fractions positions (A9 – D6). Fractions C6 – D6 were pooled. (B) SEC purification. Lanes labelled as fraction positions other than ladder (MW). Fractions C4 – C12 were pooled. All gels were run on pre-cast 4 – 12 % Bis-Tris gels.

To overcome the purity and yield issues, we purchased a WWP2-H construct cloned with the more modern StrepII tag. This enabled the use of a Strep-Tactin XT affinity column known for its higher purity and binding capacity (Schmidt and Skerra, 2007). We hoped this would also remove the possible metal-binding interactions as well as instabilities caused by a longer tag length associated with a previously unused thrombin cleavage site since removed. This construct was provided on a pET52b+ vector and transformed into BL21-Rosseta-2 cells before undergoing expression trials at ranging IPTG concentrations (Figure 4-16A). Unfortunately, due to time restraints both temperature (25 °C) and induction time (overnight) were not optimised. Similar to his-tagged WWP2-H, a significant proportion was found to be insoluble and thus was assumed as being misfolded and/or expressed as inclusion bodies. With no noticeable difference in soluble yield, conditions at 25 °C, and 0.1 mM were chosen, and purification was attempted. First attempts utilised standard StrepII buffers leading to a successful purification with high yield (26 mg/L of culture). Despite this, it was clear optimisation was required with precipitation occurring during concentration steps, and a relatively poor isolation for a StrepII purification (Figure 4-16B, C). Given we wanted to use WWP2-H for ITC we were also concerned with the presence of Tris-based buffer and DTT, both known to interfere with ITC traces. During our optimisation stage, various other buffer systems were attempted including HEPES pH 7 – 7.5, however, these resulted in significant precipitation and therefore the Tris.HCl-based buffer system was carried forward, exchanging DTT for TCEP. To aid the initial isolation affinity steps a weak detergent (0.1% Triton-X100) was first attempted but was found to adversely affect the folded protein yield. Most successes came with the addition of 1 unit of Avidin (Biotin blocker) per litre of culture, found to significantly reduce the initial co-purification of biotinylated proteins. To overcome observed precipitation during concentrating, an Amicon stirrer was utilised preventing the local high concentrations as seen in the more traditional spin concentrators. To further prevent the requirement for high sample concentrations the SEC's were also run in batches exploiting a 10 mL superloop. These optimisations, as more neatly described in Section 2.2.7.8, led to both higher purity and yields required for ITC with SDS-PAGE analysis shown in Figure 4-16D.

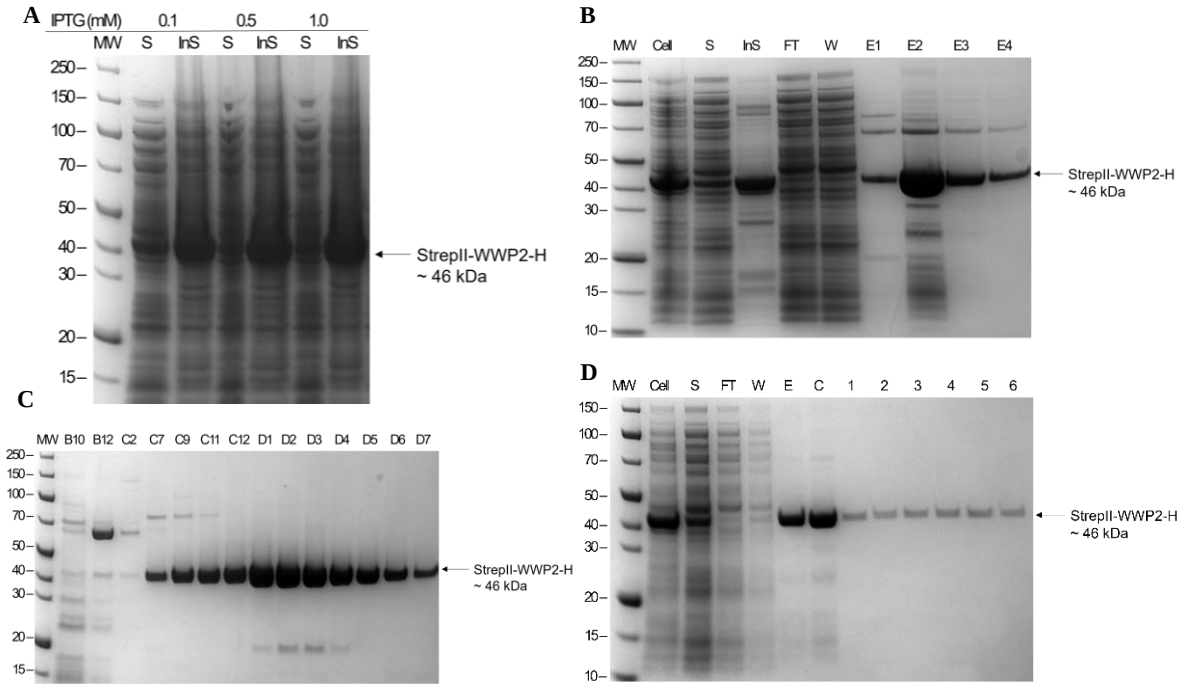


Figure 4-16: SDS-PAGE analysis of StrepII-tagged WWP2-H purification. StrepII-tagged WWP2-H identified by intense band ~ 46 kDa. (A) IPTG expression trials. BL21-Rosetta2 cells were induced with 0.1 – 1 mM of ITPG, overnight at 25 ºC. Lane labels: Ladder (MW), soluble lysate (S) and insoluble material (InS) at each concentration. (B) First successful gradient affinity isolation. Lane labels: Ladder (MW), whole cell (Cell), soluble lysate (S), insoluble material (InS), column flow-through (FT), column wash (W), biotin elution fractions (1 – 4). All fractions were pooled. (C) First successful SEC purification. Lanes labelled as fraction positions other than ladder (MW). Fractions C12 – D7 were pooled. (D) Optimised StrepII-tagged WWP2-H purification. Lane labels: Ladder (MW), whole cell (Cell), soluble lysate (S), column flow-through (FT), column wash (W), straight elution (E), concentrated sample (C), peak fraction collected from each SEC batch (1 – 6). Pooled all UV peak fractions from each SEC batch. All gels were run on pre-cast 4 – 12 % Bis-Tris gels.

4.2.4.2 ITC optimisation efforts did not achieve a positive result

Following standard ITC optimisation protocols shown in Section 2.3.3, we initially looked at running a 1: 10 (protein: ligand) molar ratio, using 20 µM WWP2-H and 200 µM (0.2% DMSO) of a control small molecule inhibitor, NSC-288387, previously been highlighted to target WWP2 through STD-NMR (Watt *et al.*, 2018). As standard, the buffer (25 mM Tris.HCl pH 8, 150 mM NaCl, 3 mM TCEP) was used directly from the concentrator flow through for ligand dilutions, accounting for DMSO by ensuring equivalent concentrations were present in all samples. Unfortunately, these first efforts highlighted that the buffer system was most likely causing issues, resulting in significant 'heats of dilution' preventing an accurate measurement from being determined (Figure 4-17A). This was even more apparent during the standard ligand-to-buffer control, as seen in (Figure 4-17B).

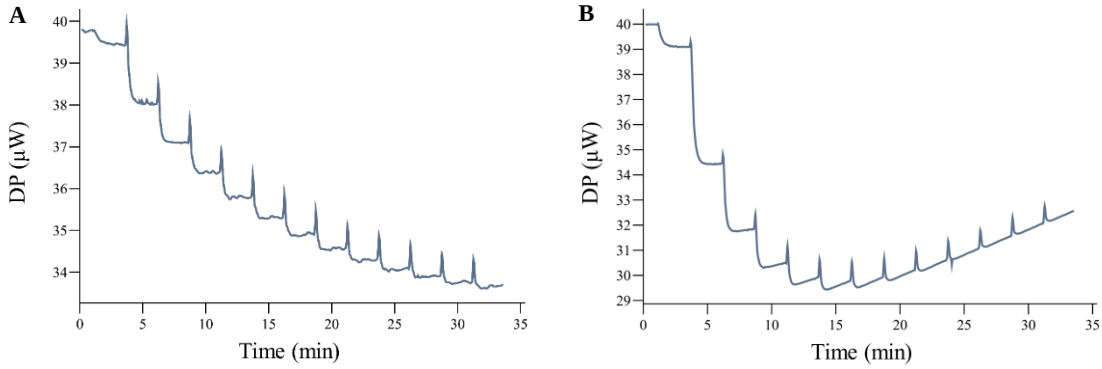


Figure 4-17: Raw ITC traces of NSC-288387 against WWP2-H. No baseline was fitted. **(**A) NSC-288387 (200 µM) was injected into WWP2-H (20 µM). (B) NSC-288387 (500 µM) was injected into the buffer (25 mM Tris.HCl pH 8, 150 mM NaCl, 3 mM TCEP, 0.5% DMSO). Figures created in MicroCal PEAQ-ITC analysis software.

To account for this, purified strepII-tagged WWP2-H was dialysed into a low tris and low salt buffer (10 mM Tris.HCl pH 8, 20 mM NaCl) and was surprisingly stable. This aimed to overcome both the high noise effects of Tris buffers whilst minimising 'heats of dilution' through ligand ionisation effects. High salt concentrations have also been shown to interfere with small molecule protein interactions and therefore this negated that effect. To this regard, reducing agents were also removed as our previous inhibition assays were run with oxidised samples and therefore binding interactions may require the active site cysteines to also be oxidised. A selection of the synthesised small molecule WWP2 inhibitors or those at least thought to inhibit at this stage in the project (not all screening steps had been completed yet) including I3C, NSC-217913, DIM, I15 and I19 were screened at 500 µM (0.5% DMSO) against 30 µM WWP2-H (Figure 4-18). This was the maximum concentration of WWP2 that could be

reached without causing further precipitation issues, to increase the signal-to-noise ratio. Despite our best efforts, none of the screened compounds showed any signs of interaction or reaching saturation as to be expected with a high molar ratio of ~1: 17. Furthermore, all but I19 displayed no substantial heats of reactions (differential power (DP) > 1 µW). The only interesting/possible signs of some interaction are the slow feedback peaks observed from both DIM and I19. Due to time restraints, no further screening was undertaken.

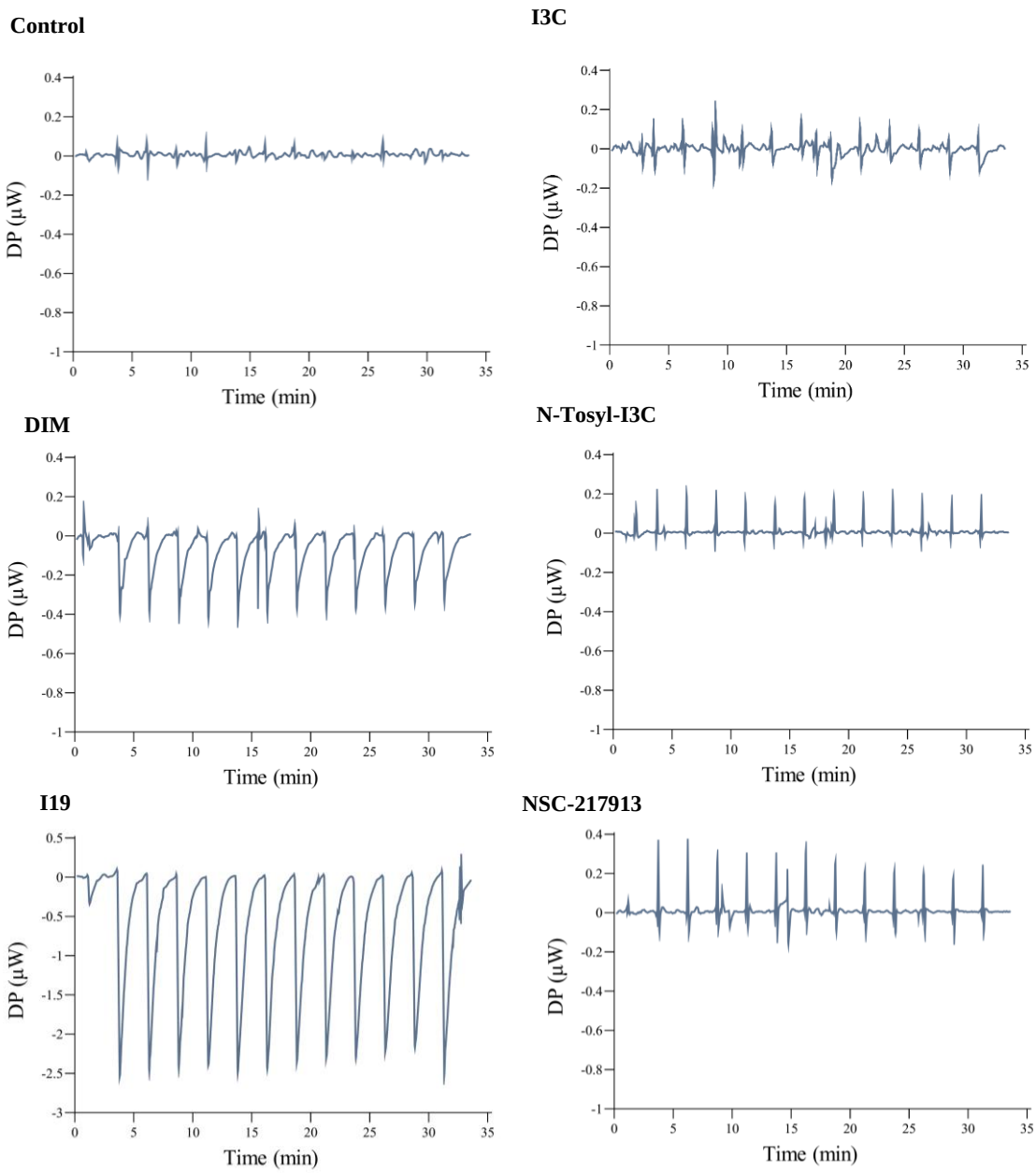


Figure 4-18: Raw ITC traces of NSC-217913, I3C and its derivatives against WWP2-H . 500 µM (0.5% DMSO) of NSC-217913, I3C, DIM, N-Tosyl-I3C and I19 as well as buffer (10 mM Tris.HCl pH 8, 20 mM NaCl, 0.5% DMSO) were injected into 30 µM of WWP2-H. Buffer-to-buffer and ligand-to-buffer controls were also completed. Plots were normalised by subtracting fitted baselines. Figures created in MicroCal PEAQ-ITC analysis software.

4.3 Discussion

This chapter aimed to explore the SAR of both NSC-217913 and I3C against the HECT E3 ligases WWP1 and WWP2, using a combination of synthesised and purchased analogue series. In total the two series consisted of 27 and 21 analogues for NSC-217913 (S) and I3C (I) respectively, synthesised through collaboration with Jake Rigby from the School of Chemistry UEA (Figure 4-4 and Figure 4-5). These small SAR libraries were screened using a DSF and ELISA autoubiquitination assay to highlight both interaction and inhibition of the target enzymes, with the idea of exploring these further using ITC. However, the ITC optimisation efforts were unfortunately unsuccessful and therefore no further mechanistic information could be determined. To aid the reader, the DSF, single-shot and dose dependency results for the most significant NSC-217913 and I3C analogues are shown in Table 4-3 and Table 4-4 later in this section.

Assessment of a small molecule SAR can be a complicated task. In theory, one looks to find strong trends relating to characteristic chemical groups, otherwise termed bioisosteres, where a small change in chemical moiety, either those similar physically or chemically, will result in a similar biological activity (Patani and LaVoie, 1996). However, this typically follows the assumption that molecules of similar structure have similar activities. This is not always the case and requires other factors to be considered such as solubility and reactivity. In the later stages of the drug discovery pipeline, these complexities are overcome using quantitative SAR (QSAR) computational models (Tropsha *et al.*, 2024). In the case of 2D-QSAR, a machine learning program is trained against an extensive analogue library using not only activity data but also pharmacological ADME properties. This can be made more complex in 3D-QSAR, taking into account 3D-dimension space through implicating force fields. Once appropriate restraints are placed, such programs can comprehend significantly more complexity than any optimistic researcher and can even begin to predict activities for future development. However, such computational tools only become useful with large data sets, not so easily achieved in academia. Despite this, at the earlier stages of the Hit-to-Lead phase, even weak SARs hold significance as a means of validating a target compound.

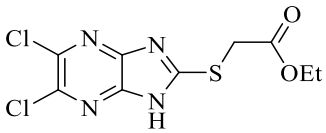
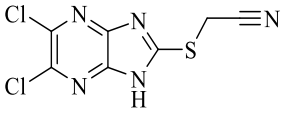
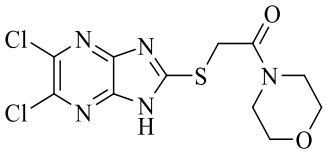
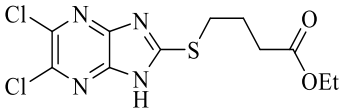
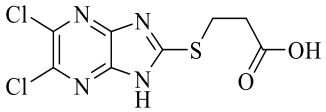
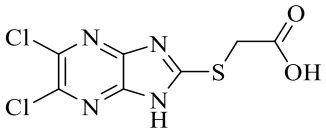
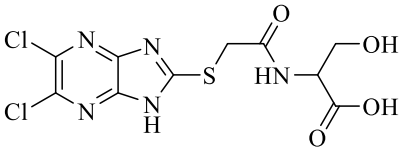
4.3.1 SAR analysis of NSC-217913

The NSC-217913 SAR analysis against the target enzymes WWP1 and WWP2 was successful, with the best overall hits summarised in Table 4-3. Following the previous chapter, NSC-217913 targeted WWP1 having both positive Tm B shift, as well as recording an IC_{50} of 158.3 µM. Interestingly, although originating from the WWP2 high throughput screen, there was no significant indication of this parent compound interacting with WWP2. Compound S11 provided the best-observed activity for both WWP1 IC_{50} of 32.74 µM and WWP2 IC_{50} of

269.2 µM, an improvement over NSC-217913, further supported by a positive DSF shift for WWP1. The nitrile derivative S7 provided intermediary activity against WWP1 and very low activity against WWP2, however, it gave a positive DSF shift for both. The NSC-217913 hydrolysed variant S15 was shown to provide worse activity against WWP1 but did demonstrate weak inhibition against WWP2. Compounds S9 and S13 demonstrated improved activity against WWP2 however failed to reach the activity threshold during the WWP1 single-shot assay. Interestingly, S13 also displayed a negative DSF shift, suggesting a destabilising of the WWP2-LH construct. This was repeated by S17, itself showing no activity for WWP2, but a potent IC_{50} of 110.4 µM for WWP1. Although there were concerns over the purity of S17 (~80%), no cross-reactivity was discovered suggesting such activity is most likely related to this structure. That being said, the presence of metal impurities or off-target compounds could also be responsible and is a key limitation of the ELISA based-assay as previously mentioned in Section 3.3.1.

Given that no structural information was available when synthesising the NSC-217913 library, it was based on chemical tractability rather than an informed design, limiting the possible SARs that could be deciphered. A better approach would have been to at minimum have an indication of the possible epitope map, either using a similar approach as used for I3C such as STD-NMR (Section 7.1), or a technique such as HDX-MS, enabling a partially structure-informed approach that could rely on QSAR molecular modelling as shown in Chapter 5. Despite this, some SAR trends were discovered as highlighted below.

Table 4-3: Summary of NSC-217913 and analogue hits from DSF and autoubiquitination assay screening.

Compound	Structure	WWP1			WWP2		
		ΔTm B (°C)	RA (%)	IC_{50} (µM)	ΔTm B (°C)	RA (%)	IC_{50} (µM)
NSC-217913		0.68	33.6	158.3	0.11	59.9	-
S7		0.66	22.4	220.5	0.39	46.2	856.3
S9		0.21	59.0	-	0.09	33.9	621.8
S11		0.45	7.7	32.74	0.22	37.2	269.2
S13		0.20	58.5	-	-0.50	40.0	603.3
S15		0.19	45.0	375.5	0.14	41.2	666.4
S17		0.32	41.2	110.4	-0.45	69.8	-

Colour Key: For DSF and single shot; Green – Threshold hits, Orange – Borderline hits. For IC_{50}; Green - below 250 µM, Orange - below 500 µM.

4.3.1.1 A dichlorinated heterocycle was required for high potency

Although our analogue libraries were small, in the case of NSC-217913 we did indeed observe possible SAR patterns as shown in Figure 4-19. Generally, the activity against WWP1 was lower than that against WWP2 and the inclusion of polar functionalities on the thioether chain is seemingly required for this activity. When comparing to the heterocyclic variants S1 – 5 and S19 – 27 also synthesised in this project, it seems that the common dichlorinated heterocycle is

typically required for lower activity against WWP1 and WWP2. Compound S5, containing the 5-chloroimidazo[4,5-*b*]pyrazine core, S4 containing 6,8-dichloroimidazo[1,2-*a*]pyrazine core, S2 with the less chlorinated 6-chloroimidazo[1,2-a]pyrazine core and S26 with the 5-bromothiazolo[4,5-*d*]pyrazine system provide some decrease in activity during the single-shot assays, yet compared to these, the non-halogenated heterocycles effectively did not. As a possible bioisostere, these chlorides could have been exchanged with another halogen as a means to increase or decrease their electrostatic effect on the heterocyclic core (Wheeler *et al.*, 2010). Unfortunately, this SAR pattern was not observed until after the study had concluded and so was not explored further.

4.3.1.2 A longer ethyl ester chain improved potency

A tentative trend between NSC-217913 and the best-hit analogue S11 shows improved inhibitory activity when increasing the ethyl ester chain length. Several other compounds (**S8,** S10 and S14) bearing the 5,6-dichloroimidazo[4,5-*b*]pyrazine heterocycle provided a slight decrease in the relative activity of both WWP1 and WWP2 in the single-shot assay. The thioether chains on S8 and S10 were rather undecorated, and whilst S14 was larger and more functionalised, it bears a bulky O^{t}Bu group which may have obstructed binding when compared to S17. In further comparison, S12 and S16, which contain ethyl acetamidoacetate or acetamido-ethanoic acid thioether chains provided better single-shot data against both WWP1 and WWP2. This along with the data obtained from the IC_{50} auto-ubiquitination assays points towards more hydrophilic interactions being preferred on linear thioether groups deeper within their binding sites. Further work surrounding the ethyl butyrate side chain, along with further investigation of the imidazo[4,5-*b*]pyrazine core and potential functionalisation of the chlorine atoms should lead to more interesting and potentially selective inhibitors of WWP1 and WWP2.

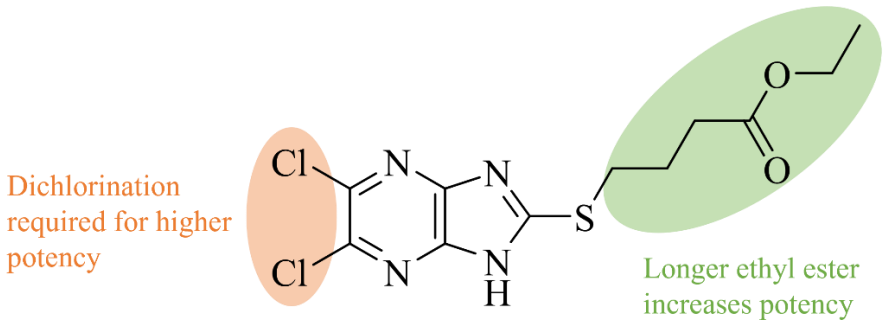
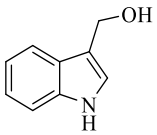
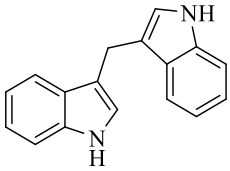
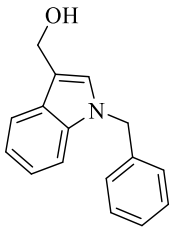
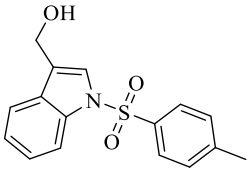
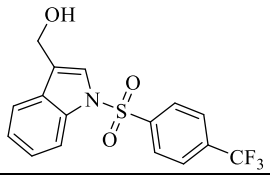


Figure 4-19: Structure of NSC-217913 analogue S11, with highlighted SAR features. The most potent WWP1 and WWP2 analogues, at IC_{50} of 32.74 and 269.2 µM respectively. Dichlorination (orange) was found to be important to maintain potency, with the length and hydrophilicity of the thioether chain (green) shown to increase it.

4.3.2 SAR analysis of I3C

The SAR by catalogue approach for I3C against WWP1 and WWP2 was not particularly successful. This was most likely due to the size of this series, with only the first 12 derivatives commercially purchased. The remaining compounds were either synthesised based on derivatives found to target NEDD4-1 in literature, or based around an azo-moiety, later found to be a PAINS property (I17-19). These literature-based derivatives did indeed show inhibition, as summarised in Table 4-4.

Table 4-4: Summary of I3C and derivative hits from DSF and autoubiquitination assay screening.

Compound	Structure	WWP1			WWP2		
		ΔTm B (°C)	RA (%)	IC_{50} (µM)	ΔTm B (°C)	RA (%)	IC_{50} (µM)
I3C		0.47	48.2	-	0.08	88.0	-
DIM		0.62	17.2	111.2	1.00	62.2	-
I13/ 1-Benzyl-I3C		-0.19	105.9	-	-0.19	97.6	-
I15/ N-Tosyl-I3C		-	11.4	218.3	-	39.2	223.7
I16		-	22.5	799.3	-	30.0	642.2

Colour Key: For DSF and single shot; Green – Threshold hits, Orange – Borderline hits. For IC_{50}; Green - below 250 µM, Orange - below 500 µM.

I3C itself displayed minimal inhibition towards WWP1 (48.2% RA) with no significant inhibition observed towards WWP2 (88% RA), and so no accurate IC_{50} could be determined. This differs to the NEDD4-1-I3C inhibition observed (IC_{50} of 284 µM), and counters the suggestion of WWP1 being I3C's direct target (Quirit *et al.*, 2017). Interestingly, the ketone derivative I3 demonstrated the opposite picture with WWP1 and WWP2, displaying a positive thermal shift and increased inhibition towards WWP2 (43 %), with no signs of interaction with WWP1. In further contrast, compound I13 otherwise reported as 1-Benzyl-I3C showed no signs of binding or inhibition towards WWP1 and WWP2 despite being highlighted as a potent inhibitor against NEDD4-1 (Quirit *et al.*, 2017). This demonstrates a possible degree of selectivity towards the HECT E3 ligases and is most likely a result of various point mutations between NEDD4-1, WWP1 and WWP2 in the proposed Ub binding site. This includes NEDD4-1's non-catalytic cysteine (Cys627 mutation to Ile649/Ile597), found reintroduced on the adjacent chain (Gly606 mutation to Cys629/Cys577). This study would have benefited from including NEDD4-1, as such selectivity could have been highlighted using the same principal assays, and in fact preliminary work was underway before time restraints led to this work being discontinued. Despite this, two I3C analogues demonstrated some interesting results against WWP1 and WWP2 as discussed below.

4.3.2.1 DIM is a promising candidate displaying improved potency towards WWP1

I3C's condensation product DIM was considerably more potent towards WWP1 with an IC_{50} of 111.2 µM, and even demonstrated strong binding (1.00 °C ΔT_m) and inhibition towards WWP2, although substantially weaker (62.2% RA). Given I3C's poor stability in both acidic and to a lesser extent neutral conditions, its anti-proliferation properties at least through WWP1 may indeed be a result of its conversion to the more potent DIM (Grose and Bjeldanes, 1992; Bradlow and Zeligs, 2010). This is supported through DIM's targeting of the Akt-PTEN signalling pathway, also associated with WWP1 and WWP2's malignant properties (Banerjee *et al.*, 2011; Maddika *et al.*, 2011; Lee *et al.*, 2020). Therapeutic trials of 'I3C' supplements have even demonstrated DIM's conversion and maintained presence in studied tissues (Anderton *et al.*, 2004). DIM itself is known to undergo acid condensation, however, even these products demonstrate various anti-proliferative properties including 2-(indol-3-ylmethyl)-3,3′-diindolylmethane (LTr-1) found to inhibit the growth of breast cancer (Chang *et al.*, 1999; Amare *et al.*, 2020). Various other synthetic analogues of DIM have also been shown to have increased in vivo anti-proliferative potency including 5-bromo (Br-DIM) and *para*-substituted phenyl additions (*p*Ph-DIM) (Vanderlaag *et al.*, 2006; Omar *et al.*, 2009). Such derivatives would provide good starting points for future compound development against WWP1, with many already shown to have minimal effect on cell toxicity (Figure 4-20).

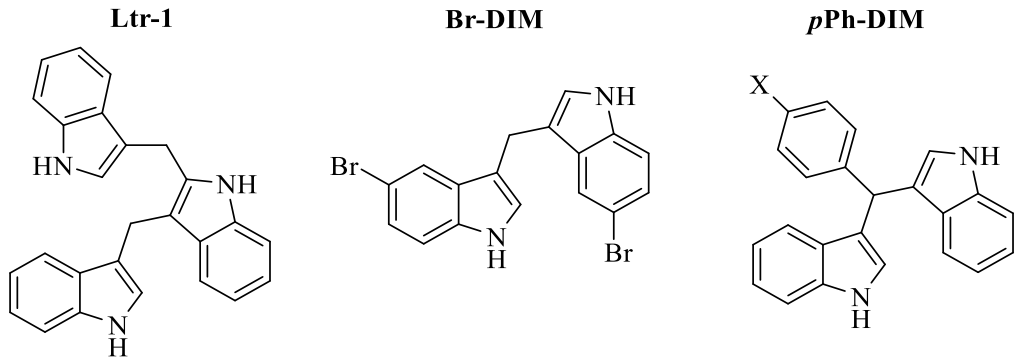


Figure 4-20: Structures of DIM derivatives with therapeutic potential. Ltr-1 – 3,3'((1*H*-indole-2,3-diyl)bis(methylene))bi(1*H*-indole), Br-DIM – bis(5-bromo-1 H-indol-3-yl)methane and *p*Ph-DIM – 3,3'-(p-substituted phenyl)bis(1 H-indole).

4.3.2.2 N-Toysl-I3C had similar potency for both WWP1 and WWP2

Compounds I15/N-Tosyl-I3C and I16 displayed a non-selective inhibition towards WWP1 and WWP2, both containing a stabilising N-tosyl addition to the nitrogen position of I3C, with a respective 4-methyl or 4-trifluoromethyl substituent. N-Tosyl-I3C was more potent with an IC_{50} of 218.3 µM WWP1 and 223.7 µM WWP2. However, no accurate thermal shift could be determined, with both compounds appearing to interfere with fluorescence. Another possible reason for this observation is that both WWP1 and WWP2 are unstable in the presence of N-Tosyl-I3C, independent of temperature. This highlights a fundamental concern of the approach used throughout this study, with the DSF assay being carried out using a lower concentration of compound compared to activity assays. As such, the drop-in activities may be due to instability of the target protein rather than inhibition. In future, other quality control assays could be used to confirm this when incubating in the presence of the compounds at 1 mM, such as circular dichroism (CD), used to measure the absorption of polarised light affected by changes in secondary structure (Greenfield, 2007). Despite this, it should be noted that I14, containing an unsubstituted N-Tosyl addition, did not correlate with the observed activity demonstrating no signs of inhibition and yet did display significant destabilising DSF shifts with both WWP1 and WWP2. Such observations may be simply due to the loss of the methyl substituent, a common addition utilised to increase hydrophobic interactions, termed the methyl effect (Leung *et al.*, 2012). The three analogues I14-16 were based on OSU-A9, an I3C analogue primarily designed to overcome the I3C's acid instability and was shown to demonstrate a 100-fold increase in its anti-proliferative properties (Weng *et al*., 2007) (Figure 4-21). Various studies have also highlighted OSU-A9 targeting of the Akt-NF-κB signalling pathway, leading to suppression of various cancers through inhibiting angiogenesis and facilitating reactive oxygen species (ROS)-mediated apoptosis (Omar *et al*., 2009; Weng *et al*., 2009; Bai *et al*., 2013). Unfortunately,

OSU-A9 itself was never screened against WWP1 and WWP2 but highlights a clear next stage in I3C's future SAR analysis.

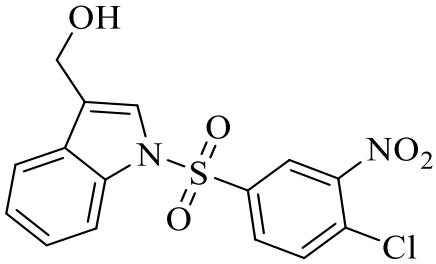


Figure 4-21: Structure of OSU-A9. An acid-stabilised I3C derivative first synthesised by Weng *et al*. (2007).

4.3.3 ITC Optimisation and the Problems Encountered

Although reasonable SARs were recognised from the DSF and autoubiquitination assays, our analysis was limited due to DSF's poor sensitivity. These problems may not have been encountered if the target interactions were substantially more potent, and so larger stabilising or even destabilising effects could have been observed. As this was not the case, DSF was used to support our activity observations rather than give any meaningful SARs. For this reason, we did look to introduce ITC, a considerably more sensitive assay in determining binding interactions. However, despite significant effort, including changing purification tags and varying constructs to cope with the high yields required for such an assay, the results were inconclusive. One such issue was the difficulties in purifying a HECT-only WWP2 construct, particularly using a Ni-NTA column. Interestingly, a recent paper looking at helicon peptide inhibition of WWP1 and WWP2 illustrated a zinc-binding site adjacent to the active site Cys838 in their X-ray crystal models (PDB ID: 8EI7) (Tokareva *et al.*, 2023). Unfortunately, they did not specify why a zinc ion was selected for this metal site but one could assume it was observed through X-ray anomalous scattering. In either case, in the non-restricted HECT-only construct, this metal binding site would be exposed and able to interact with a Ni-NTA column. Such an interaction could explain the unusual observations seen including the binding of the detagged HECT protein as well as a purple discolouration of the column at the site of binding. In the less elaborate sense, the formation of soluble aggregates could also relate to some of these purification issues. The WWP2-H construct is fairly unstable, a problem noted by the original author in their attempts at crystallisation and a characteristic we observed in both the His and StrepII-tagged versions (Gong *et al.*, 2015). Unfortunately, time pressures resulted in the protein fold characterisation being left to more ambiguous techniques such as SEC. These instabilities of the WWP2-H construct may have led to downstream problems such as the inconclusive ITC results whereby the target protein was simply misfolded. The major issue was not having a

known ligand to help optimise the system so such a problem could not be detected. The only indication would be through DIM demonstrating no saturation of WWP2, given it had such as significant ΔTm B in the DSF assay. Any future ITC efforts should consider starting with WWP1, whereby binding of both DIM and I3C is all but confirmed as well as now other potent inhibitors such as the NSC-217913 S11 analogue. Our final hopes of having more significant SAR data relied on structural analysis of these ligands complexed with their target WWP1 and WWP2 ligases, found in the next chapter.

Chapter 5: Structural Analysis of WWP1 and WWP2 Inhibitor-Bound States

5.1 Introduction

Structural biology was introduced to the wider scientific community in the 1950s and 1960s when the first high-resolution protein structures were solved using X-ray crystallography (Van Montfort and Workman, 2017). It soon became clear that a deeper understanding of protein structure and function would significantly impact future medicinal capabilities and was not long before these structures were being used to aid therapeutic discoveries, such as the development of slow-acting insulin that revolutionised the treatment of diabetes (Adams *et al*., 1969; Blundell *et al*., 1972; Thomas *et al*., 2017). Nowadays, structure-based drug design is an essential tool in the drug discovery pipeline, with the three major techniques, protein NMR, macromolecular X-ray crystallography and cryo-EM, able to solve protein structures of varying sizes in their complexes with small molecules, as well as peptides, nucleotides and other macromolecules. In the case of small molecules, these techniques can reveal their binding sites and orientations, otherwise termed binding mode, as well as their possible inhibitory mechanisms through an understanding of the protein functions. These binding modes can then be exploited in Lead-optimisation to enhance the inhibitory mechanism and/or increase the interaction points within a binding pocket. Being the only macromolecular structure technique available for decades before protein NMR and cryo-EM were developed, X-ray crystallography is the most widely used structural technique (Section 5.1.1.1) able to provide high-resolution structures (1.0-2.5 Å) ideal for drug design. However, the requirement for crystallisation does limit this technique to well-ordered proteins of small to medium size (up to 150 kDa) and can result in structural artefacts due to crystal packing. As protein NMR works in solution, it is ideal for providing insights into protein dynamics and flexibility, however, is limited to small proteins (~30-50 kDa) (Shi and Zhang, 2021). CryoEM, which has advanced to near-atomic resolution, allows for the study of large complexes without the need for crystallization, making it ideal for large macromolecules and membrane proteins. Most recently, structures down to 40k Da have now been solved, with studies now looking to utilise the CryoEM in drug design (Zhang *et al*., 2019). Despite this, CryoEM is currently resource-limited, making it highly expensive. In the case of this study, X-ray crystallography was the most appropriate structural technique, not only being the supervision of Prof Andrew Hemmings, an expert, but structures for both WWP1and WWP2 had already been solved alongside various others for the NEDD4 family members (Table 5-1).

Table 5-1: NEDD4 family of E3 ligases solved by X-ray crystallography.

Protein	Construct	Complex	Resolution (Å)	PDB	Reference
NEDD4	HECT	Apo, Ub	2.50, 2.68	2XBF, 2XBB	Maspero *et al*., 2011
	HECT	Small molecule	2.44	5C91	Kathman *et al*., 2015
	HECT	Ub:Ub	2.51	4BBN	Maspero *et al*., 2013
	HECT	UbV	3.00	5C7J	Zhang *et al*., 2016
WWP1	HECT	Apo	2.1	1ND7	Verdecia *et al*., 2003
	HECT	UbV, UbV & UbCH7	2.05, 2.84	5HPS, 5HPT	Zhang *et al*., 2016
	HECT	Helicon	2.43	8EI4	Tokareva *et al*., 2023
	L34H, 2L34H	Apo	2.55, 2.30	6J1Y, 6J1X	Wang *et al*., 2019
WWP2	HECT	Apo	2.51	4Y07	Gong *et al*., 2015
	HECT	Helicon	2.60, 3.62, 2.22, 2.90	8EI5, 8E16, 8E17, 8E18	Tokareva *et al*., 2023
	2LH	Apo	2.60, 2.30, 2.75	5TJ7, 5TJ8, 5TJQ	Chen *et al*., 2017
NEDD4L	HECT	Ub & UbcH5b	3.30, 3.10	3JVZ, 3JW0	Kamadurai *et al*., 2009
	HECT	UbV	2.43	5HPK	Zhang *et al*., 2016
ITCH	HECT	UbV	3.03	5C7M	Zhang *et al*., 2016
	12L34H	Apo	2.60	5XMC	Zhu *et al*., 2017

When it comes to protein-ligand structures using X-ray crystallography, two approaches are used, these being co-crystallisation, whereby the ligand is crystallised alongside the target protein, or ligand soaking, whereby the pre-crystallised protein is soaked with the ligand of choice. This technique was vital in the development of the FBDD strategy, enabling the synthetic linking of various fragment-size ligands that are found to bind in proximity (Hartshorn *et al.*, 2005). More recently, it is also being used in high throughput screening as part of the XChem strategy employed at Diamond (Douangamath *et al.*, 2021). With over 220,000 structures submitted to the PDB, we have an extensive understanding of macromolecule and small molecule interactions, so much so that we are now able to model these through in silico

methods. These molecular docking techniques are another vital toolkit in the drug discovery process and are generally utilised to streamline the SAR development of lead compounds once a binding site is located, able to virtually screen thousands of alternative chemical structures (Chang *et al.*, 2023). Possible binding locations can also be determined by in silico methods; however, this is much more ambiguous and struggles to consider the dynamic nature of proteins. These in silico methods are further described below in Section 5.1.1.2.

When considering the chemical structures of NSC-217913 and I3C, it is important to first understand the structure-function relationship of the NEDD4 family E3 ligases to comprehend the possible mechanisms of inhibition that could be achieved through their binding. As mentioned previously, the HECT domain contains the primary active site cysteine residue and functions by 120 ° rotation of a flexible hinge region positioned between a large N-lobe and catalytic C-lobe (Verdecia *et al.*, 2003). This process is autoinhibited by the WW and C2 domain contacts dependent on the NEDD4 family member. In WWP1 and WWP2, the 2,3-linker is found to block the hinge region preventing the T – L shape transition, and is positioned in a headband conformation with WW4 and WW2 domain interacting at the Le and Re positions as highlighted in Chapter 1 (Figure 1-11). The Re position is located at the Ub exo site, shown to interact with WW2. The binding of Ub to the exosite relieves the 2,3-linker inhibition and enables a Ub-charged E2 to bind at the N-lobe before undergoing the thioesterification of Ub to the active site cysteine (Weber, Polo and Maspero, 2019; Tokareva *et al.*, 2023). The unrestricted WW domains are able to recruit the target proteins, which can then be ubiquitinated after the rotation of the C-lobe. Taking these structure-function relationships into account, there are four possible sites which can inhibit the HECT domain activity with most already being utilised by the native autoinhibitory mechanisms: these being the Ub exo site, the E2 binding site, the active site cysteine and the flexible hinge region (Figure 5-1).

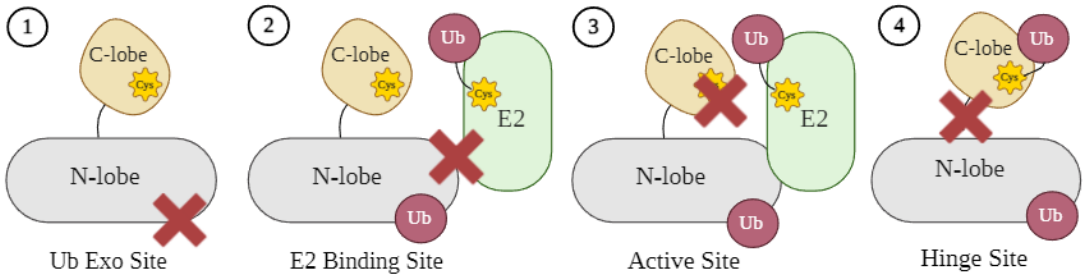


Figure 5-1: Potential sites of NEDD4 family HECT inhibition. The HECT domain's N-lobe (grey) and C-lobe (yellow) with active site Cys (dark yellow) are shown with Ub (red) and E2 (green)**.** (1) Blocking of the Ub exo site position. (2) Blocking of the E2 binding site. (3) Direct interference with active site cysteine. (4) Blocking of the flexible hinge region.

These four potential inhibitory sites have all been validated by either small or peptide-based molecules. In terms of small molecule inhibitors, occupation of the Ub exo site is the only position that has been structurally observed, this being an I3C derivative covalently bound to the non-catalytic residue Cys627 of NEDD4 (PDB ID: 5C91, Figure 5-2A) (Kathman *et al*., 2015). This Ub exo site position has been suggested to play various roles, believed to be a switch between processive and distributive ubiquitination, a potential loading site for Ub and/or ubiquitinated target proteins. However, more recently it has been identified as an autoinhibitory relief position discovered through Ub variant probes (Kim *et al.*, 2011; Maspero *et al.*, 2011). In particular, the UbV P2.3 probe interacting at the Ub exosite led to the activation of WWP1 (PDB ID: 5HPS, Figure 5-2B) (Zhang *et al.*, 2016). The same study demonstrated Ub variant probe (P1.1) also displayed inhibitory potential by blocking the E2 binding site (PDB ID: 5HPT, Figure 5-2C/D). Such inhibition has also been demonstrated by various bicyclic peptides, originally attempted by Mund and colleagues (2014). However, in their efforts to translate these peptides into a small molecule they discovered Heclin, now known to target the active site cysteine through an oxidative mechanism. A more recent study demonstrated significant success using constrained α-helical peptides termed Helicons (Tokareva *et al.*, 2023). As well as various E2 blocking Helicons (H302 and H301), this study also highlighted two WWP2 inhibitors (H305 and H308) that bound at the flexible hinge region mimicking interactions of the 2,3-linker (PDB ID: 8EI6 and 8EI8, Figure 5-2E/F). Beyond the HECT domain itself, the WW domains can also be targeted to look at interfering with target-specific interactions such as the SMURF-1 Smad interactions by mimicking the PY motifs (Cao *et al.*, 2014). However, given these domains are abundant within cellular signalling pathways, cross-selectivity would be of concern with off-targeting other WW domain-containing proteins. In summary, both the Ub exo site and active site cysteine have been targeted by small molecules, with peptide-based strategies utilised for the other positions, most likely due to their larger surface interactions.

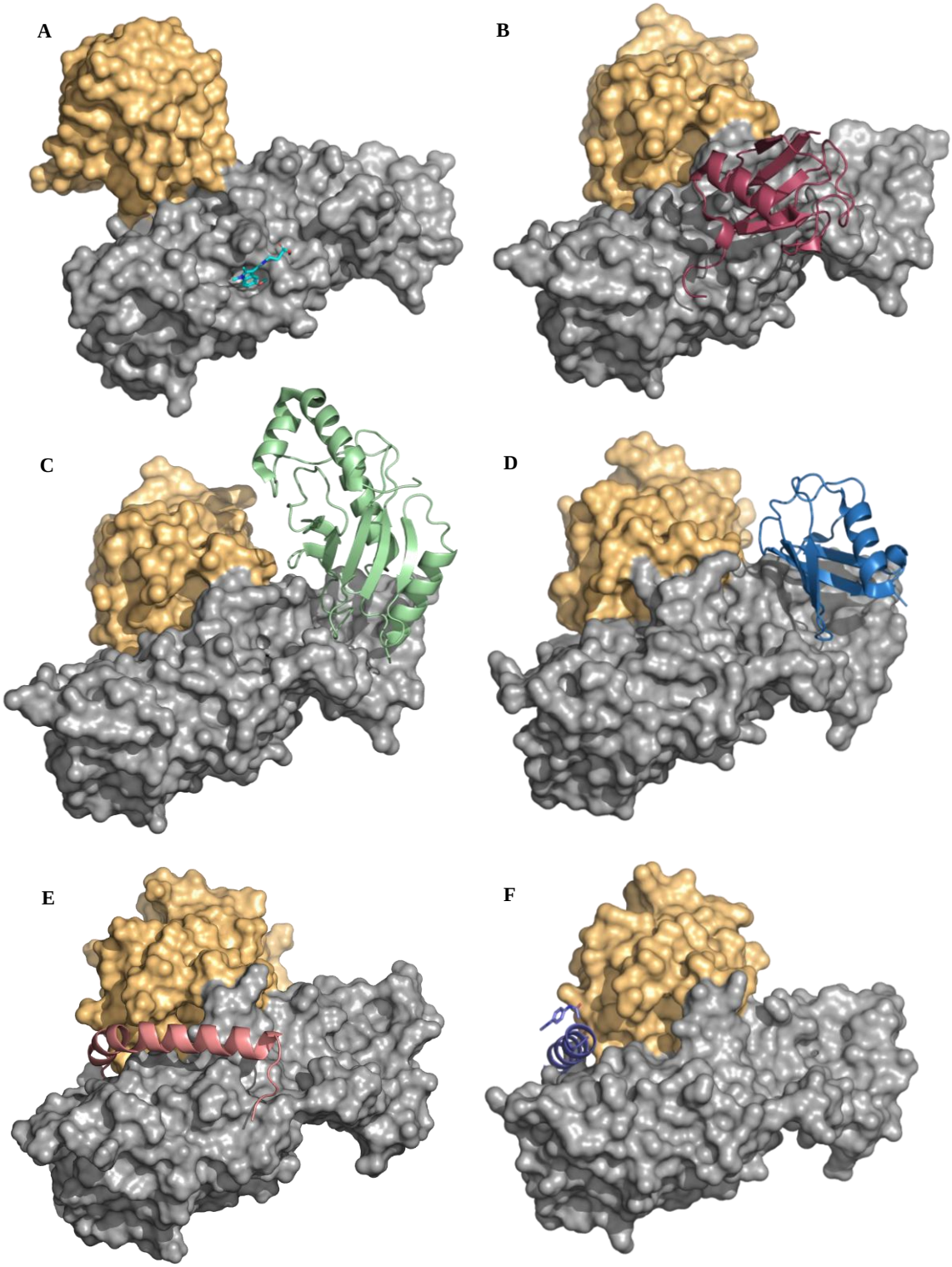


Figure 5-2: Structures of HECT-inhibitor/activator complexes in WWP1 and NEDD4-1. The HECT domains N-lobe (grey) and C-lobe (yellow) surfaces are represented. (A) I3C-derived covalent inhibitor (blue) bound to NEDD4-1 at Ub exo site, PDB ID: 5C91. (B) UbV P2.3 variant (red) bound to WWP1 at Ub exo site, PDB ID: 5HPS. (C) E2/UbcH7 (green) bound to WWP1 at the E2 site, PDB ID: 5HPT. (D) UbV P1.1 variant (blue) bound to WWP1 at the E2 site, PDB ID: 5HPT. (E) The 2,3-linker (pink) bound at the flexible hinge region in WWP2, PDB ID: 9EQH. (F) Helicon H308 (purple) bound at the flexible hinge region in WWP2, PDB ID: 8EI8. Images created using PyMOL.

5.1.1 Biophysical Techniques

5.1.1.1 X-ray crystallisation

X-ray crystallography is a complex, labour-intensive technique, first requiring the formation of highly ordered protein crystals, secondly, recording the intensities of X-ray reflections arising from diffraction from said crystals, before finally using computational methods to convert this information via phasing into a comprehensible 3D protein structure (Rhodes, 2006).

To obtain proteins in a crystalline state, the most common vapour diffusion methods use an aqueous buffer containing a precipitant at a concentration below that required to cause precipitation. The concentration of both the protein and precipitate is then slowly increased by controlled solvent evaporation until supersaturation is reached. At this solubility limit of the macromolecule, the sample can either precipitate or crystalise (McPherson and Gavira, 2014). Crystals are formed by two key stages, *nucleation* and *growth*, as depicted in the phase diagram below (Figure 5-3). Nucleation is the point of molecular cluster formation in the transition from an unordered aqueous to an ordered solid-state phase (García-Ruiz, 2003). If successful, the formation of these crystals reduces the protein concentration, and with careful management of the conditions will move the crystals into the growth phase.

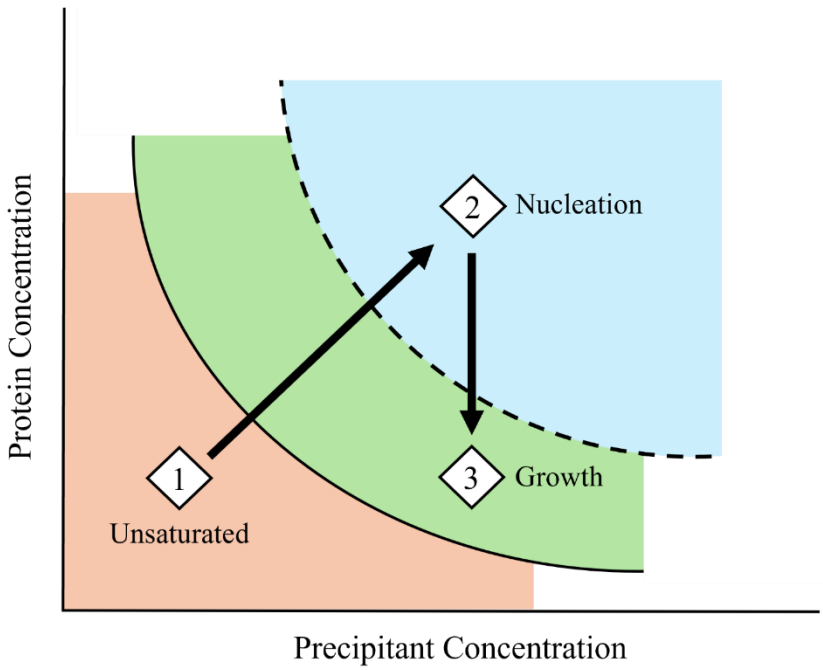


Figure 5-3: Crystal phase diagram mediated by a precipitant. A graph of protein against precipitant concentration representing the regions of unsaturated aqueous protein (orange) and supersaturated protein where crystal nucleation (blue) and crystal growth (green) occur. The numbering from 1 – 3, is the ideal strategy for growing large protein crystals. Figure adapted from Rhodes (2006).

In practise, the crystallization conditions are screened by varying proteins concentrations as well as buffers, salt, pH and precipitant concentrations. In more complicated cases, crystal seeding can also be implemented, in essence using smaller crystals to aid the initial nucleation point (Espinosa *et al.*, 2016).

Once grown to sufficient size, the crystals are harvested, cryo-protected and plunged into liquid nitrogen before being mounted on a goniometer for X-ray data collection. The resulting X-ray beams are diffracted upon encountering the electrons contained within the crystal lattice and result in a characteristic diffraction pattern of varying intensities called reflections (Rhodes, 2006). Following Braggs law, these reflections correspond to parallel planes of the crystal lattice, and only after stepped rotation of the crystal is sufficient information obtained (Bragg and Thomson, 1913). These reflections are indexed, enabling the unit cell in real space to be calculated and, in turn, the lattice symmetry. Nowadays such calculations are automated, termed autoindexing, and use programs such as DIALS and AutoProc (Vonrhein *et al.*, 2011; Winter *et al.*, 2022). To convert the reflection intensities into a 3D electron density map by means of a Fourier synthesis, structure factors need to be determined. Unfortunately, although the structure factor amplitudes are directly determined, the phase contribution is lost. This is commonly known as the phase problem (Taylor, 2003). The simplest method to solve the phase problem is called molecular replacement (MR) whereby a similar structure, termed a search model, is used and oriented within the unit cell. Following this step, structure factor phases calculated from the search model can then be combined with the experimental amplitudes to calculate an electron density map. This process has become more straightforward with the introduction of Alphafold2, a computational protein structure prediction program that can generate search models when no structures of close homologues are available (McCoy, Sammito and Read, 2022). The successful merging of the experimentally determined amplitudes and molecular replacement phases, alongside the determined unit cell parameters, can generate the electron density map. Again, this stage is typically automated through programs such as PHASER (McCoy *et al.*, 2007). Finally, the positioned and oriented search model is refined by minimizing the differences between experimental (observed) and calculated structure factor amplitudes, whilst taking into account the geometry of the molecule through indicators such as Ramachandran angles, bond lengths, side chain conformations and clashes, as well as the contributions to scattering from any additional molecules within the until cell such as water (ordered and bulk phase), crystallization agents, cryoprotectants and small molecules. These processes are done through computational programs such as REFMAC (Vagin *et al.*, 2004), but also require more manual assistance using molecular graphics programs such as COOT (Emsley *et al.*, 2010). This complete workflow is illustrated below in Figure 5-4.

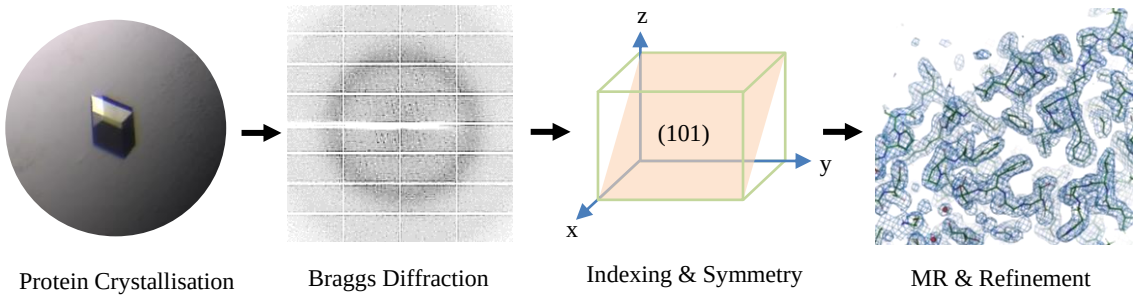


Figure 5-4: The workflow of X-ray crystallography. The target protein is first crystallised into an ordered, high-quality crystal lattice. This crystal is then placed in a monochromatic X-ray beam with diffraction recorded as Bragg reflections on a detector. These reflections are indexed in reciprocal space, allowing the unit cell to be determined in real space. Reflection intensities are then converted to structure factor amplitudes. Using molecular replacement (MR), a search model and structure factor amplitude information are used to generate approximate structure factor phases. A Fourier synthesis is used to convert these structure factor amplitudes and phases into an electron density map. In an iterative process, the search model is remodelled and refined to convergence against the experimental structure factor amplitudes to provide final atomic coordinates and temperature factors.

5.1.1.2 In silico molecular docking

Molecular docking is a computational method that investigates possible conformations and orientations, otherwise known as poses, of small molecules in target protein binding sites. This is achieved using searching algorithms to generate various poses that are then ranked using scoring functions (Meng *et al.*, 2012; Torres *et al.*, 2019). In practice, one must first obtain the target coordinates, typically from the PDB or experimentally. Currently, computation methods such as AlphaFold are still unable to accurately model the atomic details required, however, the field is ever improving with Alphafold3 most recently introduced (Abramson *et al.*, 2024). These target structures normally require some fixes, including re-building sidechains and loops as well as energy minimisations, particularly in the case of X-ray structures where crystal contacts can alter conformations (Eyal *et al.*, 2005). The charge and protonation states of the protein amino acids may also require changing to suit a preferred condition, and this also applies to the ligand structures. Next, a receptor grid is generated at the proposed binding site to minimise the required docking calculations and therefore computational demands. Various algorithms can be utilised depending on the software chosen, however, these are often classified as either: systematic, whereby the ligands degree of freedom is explored incrementally; stochastic, whereby the ligands degree of freedom is randomly explored; and deterministic, whereby the subsequent iteration is dependent on the free energy of the previous pose (Brooijmans and Kuntz, 2003). Some strategies incorporate a combination of these, as well as molecular dynamic simulations to consider the flexibility of the receptor site as well as the ligand itself. Finally, the docked poses are then ranked using scoring functions, themselves

further divided into either: force-fields, based on a sum of bonded (bond length, angle etc) and non-bonded (electrostatic and van der Waals) energy terms; empirical, based on a sum of binding energies (hydrogen bonding, ionic, hydrophobic and entropy); and knowledge, based on the statistical frequency of observed contacts in known ligand-protein complexes (Guedes, de Magalhães and Dardenne, 2014). Again, these functions can also be merged referred to as consensus scoring.

5.1.2 Experimental Aims

This final chapter aims to structurally characterise the hit NSC-217913 and I3C derivatives against the target WWP1 and WWP2 E3 ligases, questioning; Where are the binding sites and how do they interfere with WWP1 and WWP2 function? These efforts will focus on identifying the inhibitor binding modes and provide possible explanations of inhibition whilst enabling a structure-informed approach to aid the understanding of their SARs. The chapter is broken into two sections; the first section looks to reproduce the high-resolution X-ray crystallographic structures of WWP1-L34H, WWP1-2L34H and WWP2-LH (Figure 5-5) according to literature conditions, before attempting to solve their ligand complexes with the parent small molecules NSC-217913 and I3C, and later expanding this work to include their more potent analogues. These efforts first focus on ligand soaking techniques before screening for co-crystallization conditions. The second section focuses on in silico molecular modelling targeting potential binding sites of the compounds, before docking the most successful analogues as a means to aid the interpretation of their SAR.

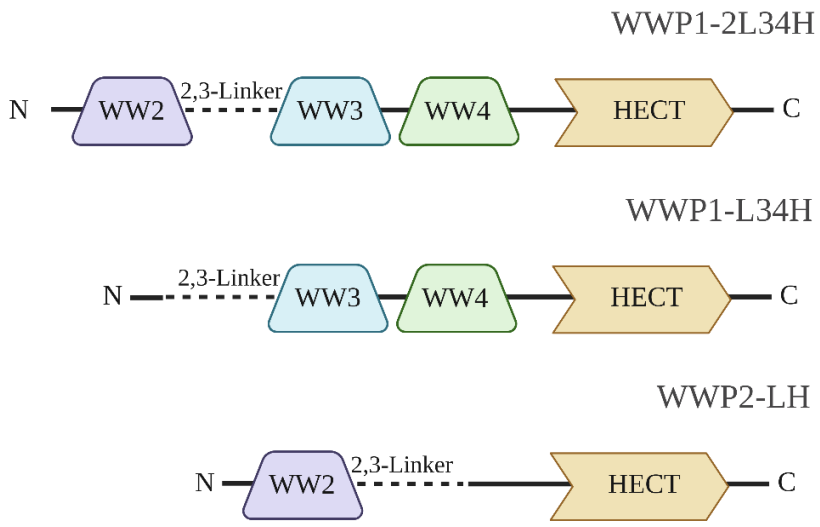


Figure 5-5: Domain architecture of the WWP ligase constructs used in Chapter 5. The WWP1-L34H and WWP1-2L34H construct, previously reported by Wang *et al.* (2019), consists of the WW3 (blue), WW4 (green) and HECT (orange) domains as well as the 2,3-linker (dashed line) located at the N-terminus, with WWP1-2L34H also containing an N-terminal WW2 (purple) domain. The WWP2-LH construct, previously reported by Chen *et al.* (2017), consists of the WW2 and HECT domains separated by the 2,3-linker region.

5.2 Results

5.2.1 WWP1 Crystallisation Efforts

The WWP1-L34H (PDB ID: 6J1Y) and WWP1-2L34H (PDB ID: 6J1X) constructs were chosen for X-ray crystallographic studies given their restrained flexibility and the reasonable resolutions of available apo crystal structures at 2.5 and 2.3 Å, respectively (Wang *et al*., 2019). With these constructs not only possessing the HECT domain but also various WW domains, this would provide the potential to observe other sites of interaction not possible with the HECT-only construct.

5.2.1.1 Protein purification optimised using buffer screen

DNA was kindly gifted by Wang *et al*. (2019) pre-cloned into pET32a plasmids into pET32a plasmids with Trx-His solubility/affinity purification tags. These were transformed into *E. coli* BL21(DE3)-codon plus RP cells for overexpression. Our initial efforts to purify these protein constructs following the published protocols resulted in solubility problems of the recombinant protein when using the buffer systems stated. With small adjustments to the incubation temperatures (10 °C) and the addition of glycerol (5%) to buffers, a standard problem-solving strategy (Christopher, Phipps and Gray, 1998; Vagenende, Yap and Trout, 2009), WWP1-L34H was successfully purified. For the complete protocol see Section 2.2.7.3, with SDS-PAGE analysis shown in Figure 3-5. WWP1-2L34H was the more desirable of the two constructs for the purposes of this study having both previously yielded crystals diffracting to a higher resolution and in addition possessing the additional WW2 domain. Its purification was achieved using the published protocol subject to the addition of 10% glycerol to buffers. However, significant problems with precipitation persisted and as such yielded very little soluble protein (~0.625 mg/L of culture).

To overcome these issues, experiments to identify a suitable buffer system were carried out exploiting DSF to screen various commercially available buffers and additives. These included the RUBIC buffer screen, RUBIC additive screen, and the Durham salt and Durham pH screens following the procedures as set out in Section 2.3.2. The most stabilising buffer systems and additives as well as various trends identified can be seen in Figure 5-6. In summary, sodium phosphate buffers demonstrated the highest ΔTm D of WWP1-2L34H compared to water controls, with higher buffer concentrations resulting in higher thermal stability. The pH screens showed WWP1-2L34H was most stable between pH 6.5 – 7, demonstrating some tolerance up until pH 8.5, whereas salt screens indicated the most stable NaCl concentrations were between 100 – 400 mM. As expected, glycerol was the most

stabilising additive, albeit at 20% v/v. Surprisingly, various sugars also showed to increase ΔTm D with Maltose (25 mM) having the greatest effect. More interestingly, imidazole, required at high concentration during His-NTA column elution, was observed to drastically reduce the stability of WWP1-2L34H. A custom buffer screen was generated based on these observations which also included Tris buffer as found in the original literature purification protocol (Wang *et al.*, 2019). The resulting buffer screen is shown in Table 2-5, with the relevant results shown in Figure 5-7A and B. As with the commercial screens, NaH_2PO_4 buffer pH 7.0 was found to have the highest ΔTm D (~ 5.5 °C), demonstrating a negative correlation to NaCl increasing concentrations. His-tagged WWP1-2L34H was also relatively stable in Tris pH 8.0, having ΔTm D (~ 5.0 °C) with 150 mM NaCl. In both cases, the increased buffer concentrations (150 mM and 100 mM respectively) also had a marked increase in ΔTm D. In further correlation with the commercial screen, glycerol was found to be the most stabilising additive in both sodium phosphate and Tris buffers, with maltose also displaying a stabilising effect. Although the addition of DTT and EDTA at 1 mM to the Tris buffers did show an increase in Tm D, the higher concentrations (5 mM) adversely affected stability. Following these optimised conditions, WWP1-2L34H purifications were attempted in two separate buffer systems as (A) 100 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT and 1 mM EDTA, 25 mM Maltose and (B) 150 mM NaH_2PO_4 pH 7.0, 1 mM DTT and 1 mM EDTA, 25 mM Maltose. Both purification attempts were initially successful, utilising gradient elution from Ni-NTA columns to keep imidazole concentrations as low as possible. However, after dialysing the purified proteins at 4 °C for 72 hrs, the samples in the NaH_2PO_4 buffer (B) showed substantial precipitation and so the optimised Tris-based buffer (A) was carried forward. After cleaving the His-Trx-3C tag the sample was much more soluble, and so was dialysed back into the original literature buffer ready for X-ray crystallographic analysis. The complete purification protocol can be found in Section 2.2.7.4, with the results of the SDS-PAGE analysis shown previously (Figure 4-6).

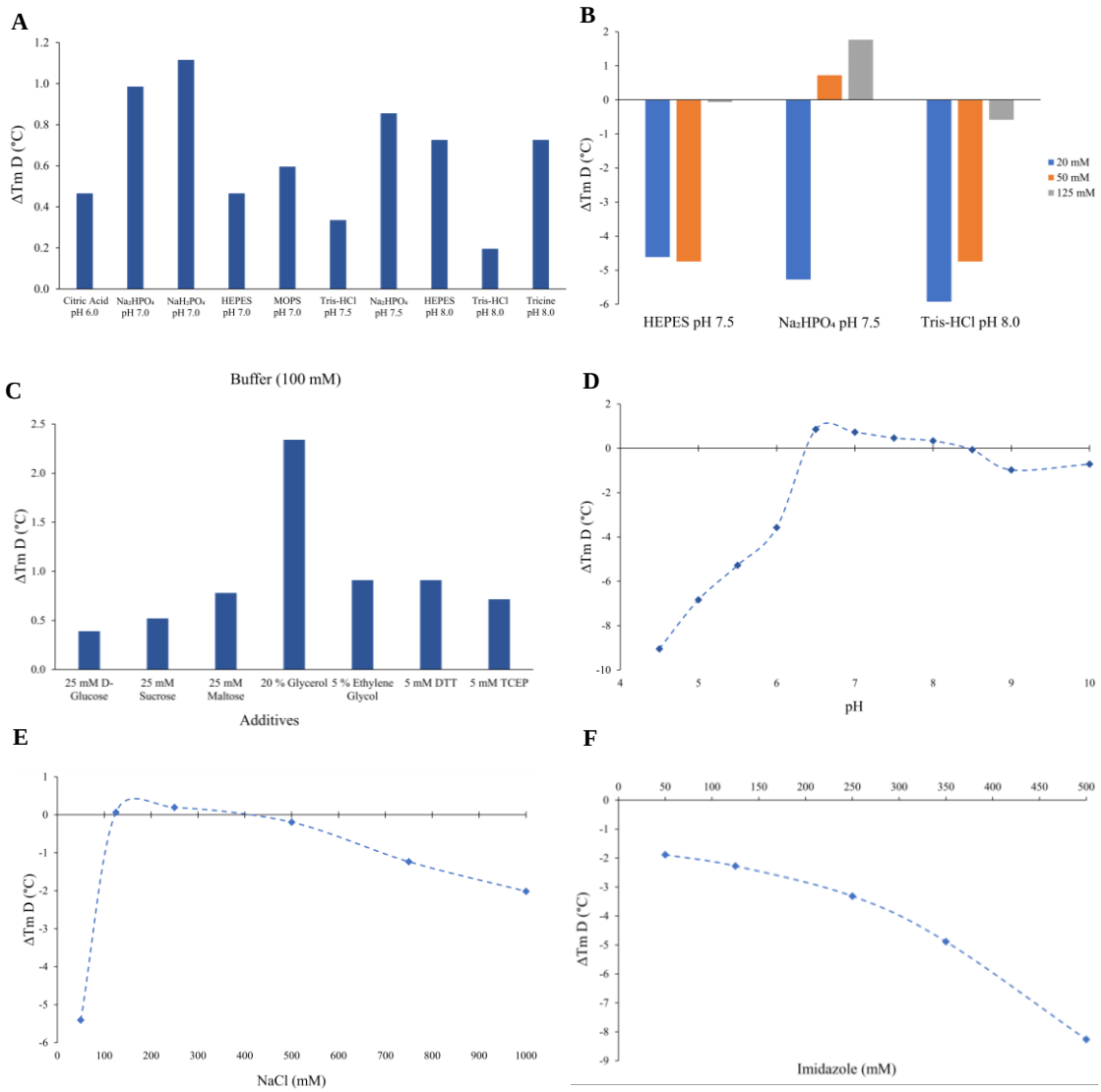


Figure 5-6: DSF commercial buffer screens against His-tagged WWP1-2L34H. All protocols were followed as stated by both the Rubic and Durham screens. The most stabilising buffers and additives, as well as various other relevant effects, are shown with ΔTm D calculated against water controls. (A) Effects of various buffer systems from pH 6 – 8.0. (B) Effects of HEPES, Na_2HPO4 and Tris.HCl at varying concentrations. (C) Effects of various additives and reducing agents. (D) Effect of ranging pH in SPG buffer (Succinic acid, NaH_2PO_4 and Glycine). (E) Effect of increasing NaCl concentrations in 50 mM Tris-HCl pH 8.0. (F) Effect of increasing imidazole concentrations at pH 7.5, containing 100 mM NaCl.

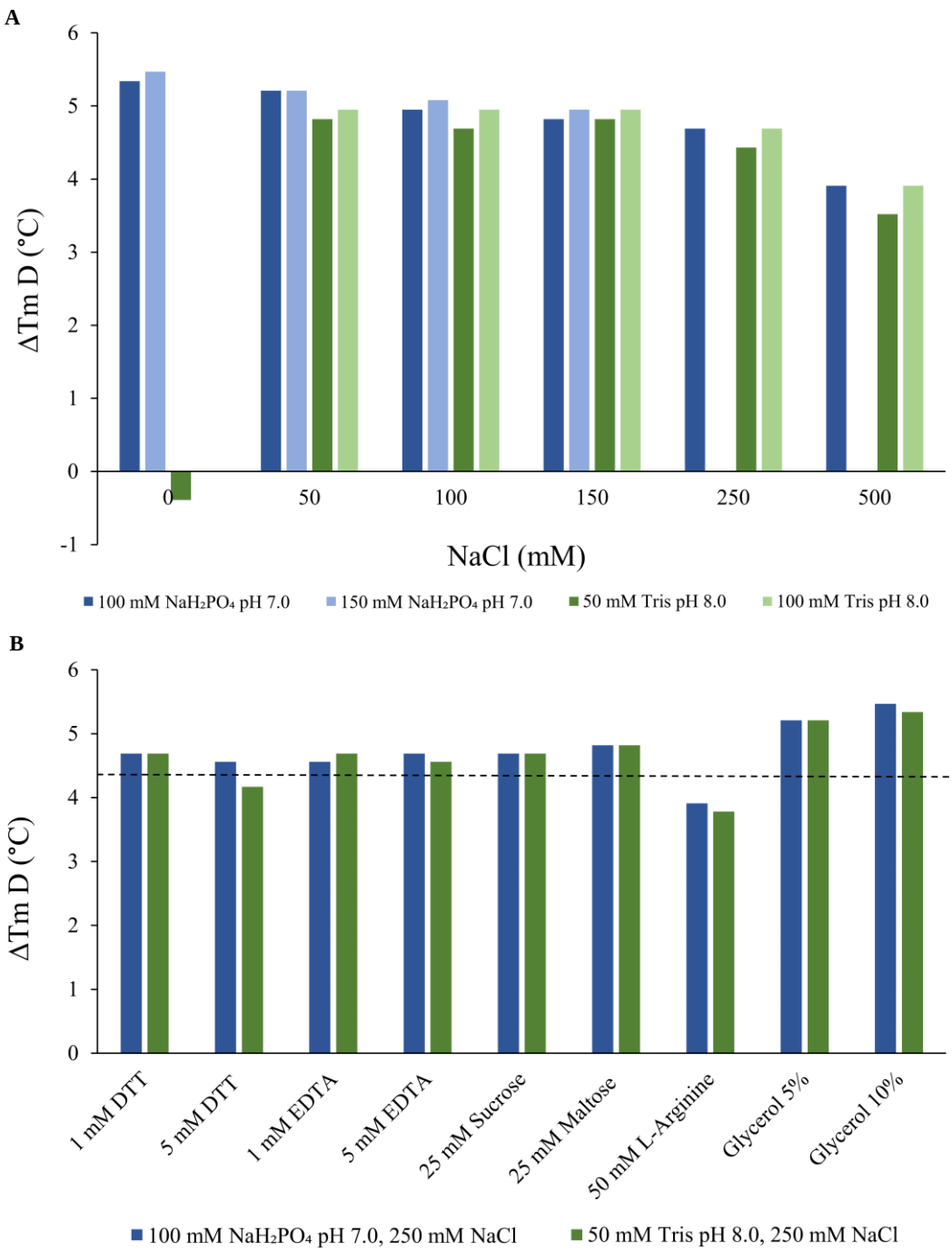


Figure 5-7: DSF custom buffer screen against His-tagged WWP1-2L34H. The most stabilising buffers and additives are shown with ΔTm D calculated against water controls. (A) Effect of increasing NaCl concentrations in NaH_2PO_4 (blue) and Tris buffers (green). Both buffer systems were screened at 100 or 50 mM (dark coloured) and 150 mM or 100 mM (light coloured), respectively. Data points at 0, 250 and 500 mM NaCl concentrations were omitted due to poor melting curves. (B) Effects of various additives and reducing agents. These were screened in either 100 mM NaH_2PO_4 pH 7.0 (blue) or 50 mM Tris pH 8.0 (green) with 250 mM NaCl. A black dashed line was included to indicate the ΔTm D control from 50 mM Tris pH 8.0, 250 mM containing no additives.

5.2.1.2 Crystals of WWP1-L34H failed to diffract at high resolution.

Crystallization of WWP1-L34H was initially attempted by sitting drop vapour diffusion at 16 °C and based on precipitant conditions reported in the literature (100 mM sodium malonate pH 5.0, 12% PEG3350) (Wang *et al.*, 2019), as described in Section 2.3.4.1. . Initial attempts used purified protein at 3.5 mg/mL, over a pH range of 4.5 – 5.5 and using 10 – 14% PEG3350. After 1 day, plate/tabular crystals began forming at pH 4.5, with all drops other than those at pH 5.5 containing numerous crystals within 7 days (Figure 5-8A). Unfortunately, the crystals showed severe defects suggesting possible distortions in the crystal lattice and it was decided that improvements were needed before data collection. To achieve this, a variety of crystallization screens were performed using a higher protein concentration (~9.5 mg/mL) closer to the 15 mg/mL used in literature. These were crystallised in either the presence of glycerol (5%), at a reduced temperature (4 °C) or seeded using the initially obtained crystals (see Section 2.3.4.2 for seeding protocols). The rationale behind this was to attempt to improve crystal nucleation either by aiding its formation during seeding, or by slowing it down to prevent overgrowth. However, the higher protein concentration simply produced more crystals with more defects (Figure 5-8B), with the other screens having no more success. Although crystals were grown when seeded, they typically formed as spherulite plate microcrystals and were deemed too small to harvest (Figure 5-8C).

It was assumed that possible impurities or protein misfolding during the WWP1-L34H isolation procedures may have contributed to the poor morphology. Various efforts to improve this included using a non-frozen sample (protein was isolated immediately from expression), extending the HisTrap wash volume and refreshing EDTA to remove potentially leached nickel. With lower concentrations appearing to give improved crystal morphology, WWP1-L34H was crystallized at 2 mg/mL resulting in the growth of significantly improved crystals with a cuboid habit (Figure 5-8D). The crystals were harvested following to protocols in Section 2.3.4.3, using 25% glycerol in mother liquor as a cryoprotectant. Data collection was performed on the I04 beamline at the Diamond Light Source (see Section 2.3.4.4). These crystals displayed weak Bragg diffraction. The resulting reflections when autoindexed in AutoProc gave a dataset at 3.8 Å resolution and the same P1 space group as found in the PDB (PDB ID: 6J1Y), summarised in Table 5-2.

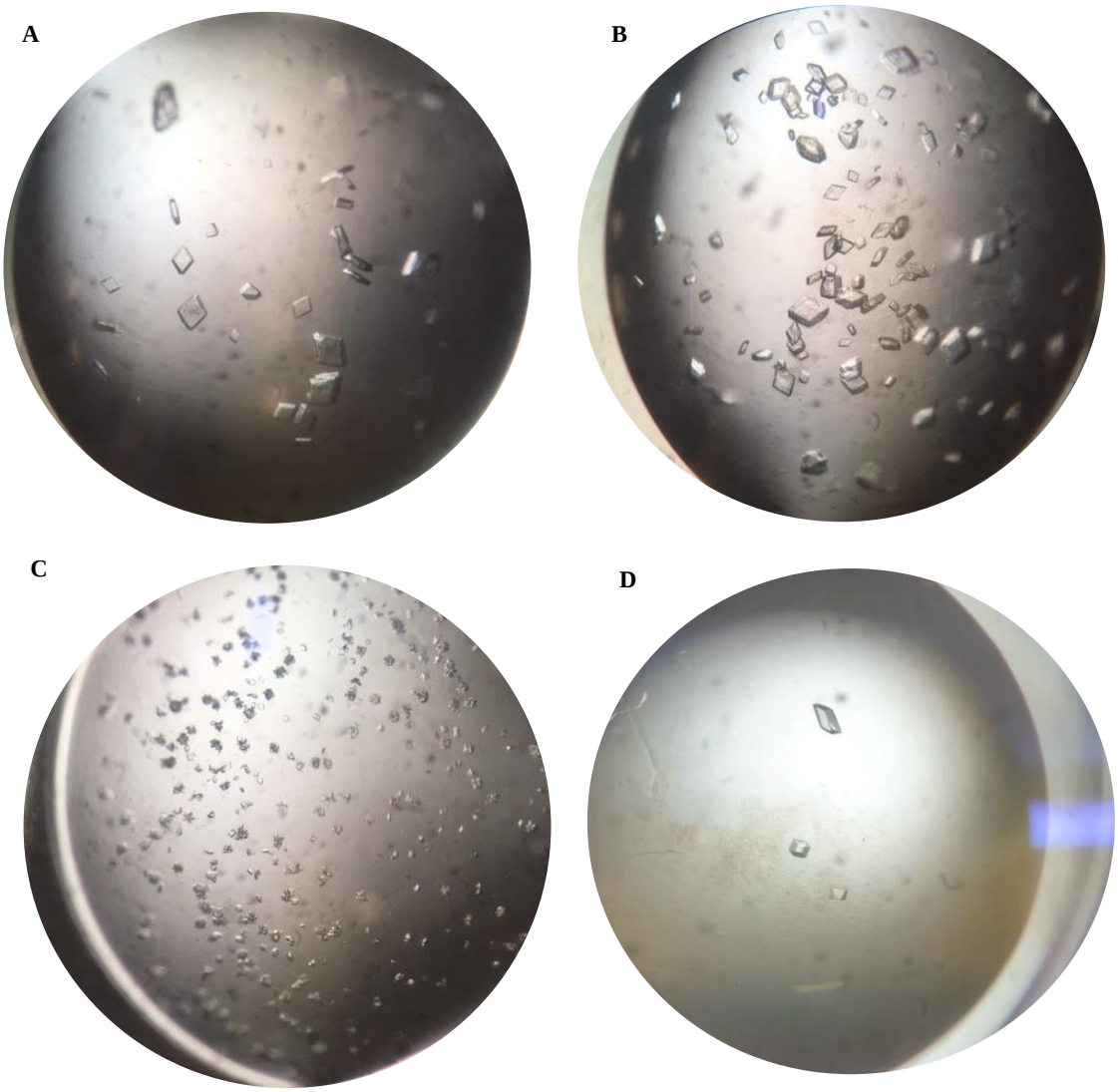


Figure 5-8: Crystal morphologies of WWP1-L34H. All samples in a solution of 50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA were crystallised by screening around 100 mM sodium malonate pH 5.0, 12% PEG3350. (A) Plate/tabular crystals from 3 mg/mL protein. (B) Plate/tabular crystals from 9.5 mg/mL protein. (C) Spherulite plate microcrystals from 8.5 mg/mL protein with seeding. (D) Cuboid crystals from 2 mg/mL protein with optimised WWP1-L34H isolation, crystallised at pH 5.0, 10% PEG3350.

Table 5-2: Summary of data collection statistics for WWP1-L34H.

	WWP1-L34H (WWP1_18_1)
Wavelength/ Å	0.9795
Resolution range	50.76 - 3.172 (3.286 - 3.172)
Space group	P 1
Unit cell	59.0994 59.7847 86.2992 91.4175 100.408 109.293
Total reflections	66506 (6714)
Unique reflections	18306 (1797)
Multiplicity	3.6 (3.7)
Completeness (%)	98.74 (96.30)
Mean I/sigma(I)	4.39 (0.48)
Wilson B-factor	73.48
R-merge	0.3117 (2.157)
R-meas	0.3664 (2.521)
R-pim	0.191 (1.298)
CC1/2	0.985 (0.421)
CC*	0.996 (0.77)

5.2.1.3 Novel C1 2 1 space group of WWP1-2L34H enables modelling of all WW domains.

The crystallisation of WWP1-2L34H was achieved following the conditions stated in literature (100 mM Tris.HCl pH 7.0, 15% reagent alcohol) utilising sitting drop vapour diffusion at 16 °C, as described in Section 2.3.4.1. WWP1-2L34H was plated at 4 mg/mL, screening between pH 6.5 – 7.5 and 12.5 – 17.5% reagent alcohol, with an alcohol ratio of 90% ethanol, 5% methanol and 5% isopropanol. Crystals were formed after 12 days in two distinct habits both at 17.5 % reagent alcohol, either as spherulite blades or rods (Figure 5-11).

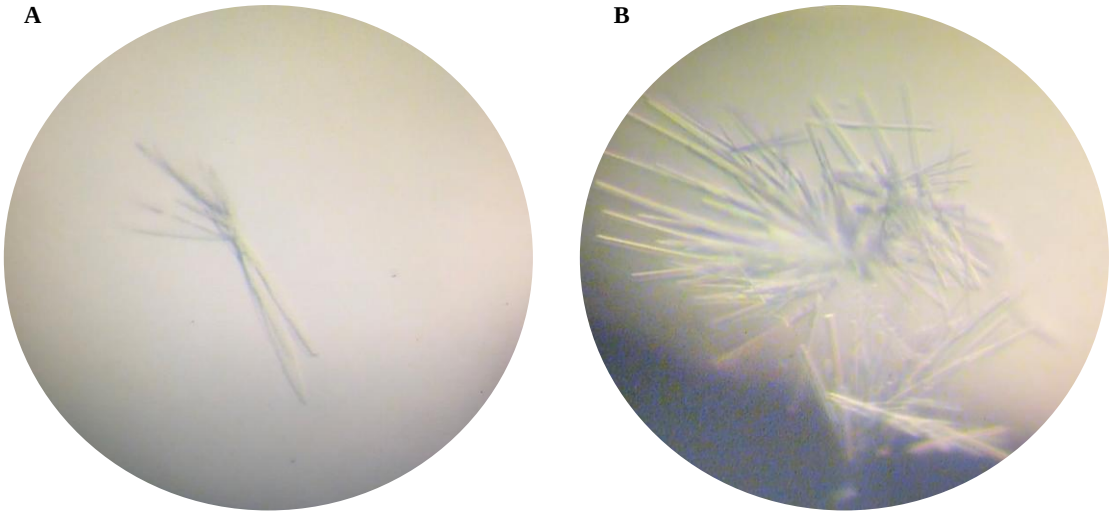


Figure 5-9: Crystal morphologies of WWP1-2L34H. All samples in a solution of 50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 1 mM EDTA were crystallised at 4 mg/mL protein screening around 100 mM Tris.HCl pH 7.0, 15% reagent alcohol. (A) Spherulite blade crystal formed at pH 6.5, 17.5% reagent alcohol. Autoindexed in the C1 2 1 space group. (B) Spherulite rod crystal formed at pH 7.0, 17.5% reagent alcohol. Autoindexed in the P 1 2_1 1 space group.

To prevent overgrowth, crystals were harvested immediately following protocols in Section 2.3.4.3. However, when using the stated cryoprotectant at 30% glycerol in the mother liquor, the crystals dissolved. After various efforts, the use of a 5% more concentrated crystallization buffer with 30% glycerol was found to be the most stable and so was carried forward. Data collection was run on the I24 beamline at the Diamond synchrotron, see Section 2.3.4.4. Interestingly, the two observed crystal morphologies were autoindexed in distinct space groups, with the rod crystals processed via Xia2 in the P1 2_1 1 space group found in literature at a resolution of 3.5 Å, and the blade crystals processed via DIALS into a new C1 2 1 space group

at a resolution of 3.0 Å. The data collected from the new C1 2 1 space group was solved using the structure PDB ID: 6J1X for molecular replacement in PHASER, as part of the CCP4 suite, before further refining in REFMAC and Phenix, as described in Section 2.3.4.4. The relevant data collection and model statistics are shown in Table 5-3, with the full structure submitted as PDB ID: 9EQK. Two molecules were found in the asymmetric unit both containing significant density required to model the WW3 domain found previously disordered in the alternative P1 2_1 1 space group. Interestingly, although both molecules displayed the WW3 domain in the same position, the WW4 domains were found in alternative conformations, resulting in an RMSD = 0.733. With the WW4 domain in Chain B located in the same position as the PDB ID: 6J1X, it was selected for in silico molecular modelling studies (see Section 5.2.3) and is shown in Figure 5-10.

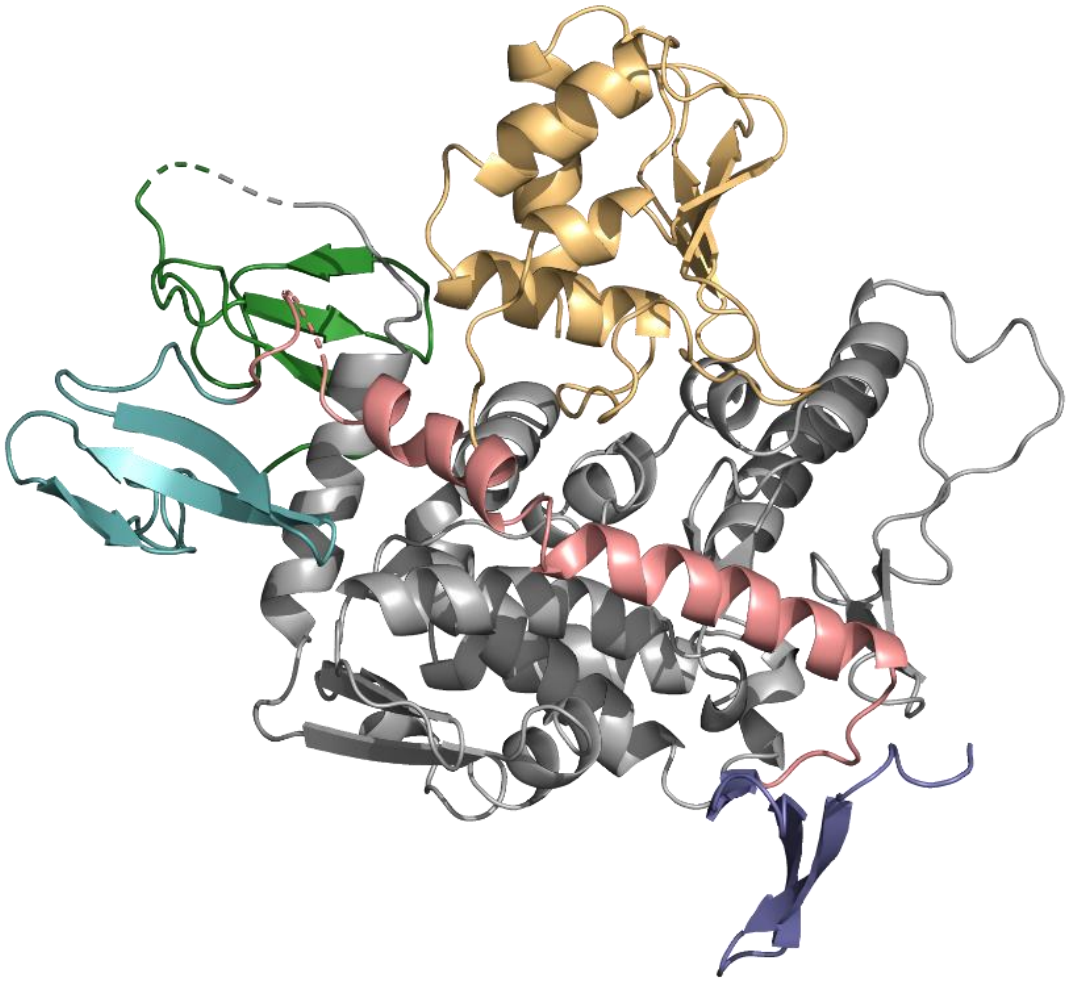


Figure 5-10: Crystal structure of WWP1-2L34H (PDB ID: 9EQK). WWP1-2L34H structure shown is chain B, having two molecules in the asymmetric unit, chain A WW3 and WW4 are more disordered and are conformationally different. The HECT domain contains the N lobe (grey) and C lobe (orange). The WW2 (purple) domain is binding adjacent the Ub exo site with the 2,3-linker (pink) binding across the hinge region, finally resulting in the WW3 (blue) and WW4 (green) domains. Image created in PyMOL.

Table 5-3: Summary of data collection and model statistics for WWP1-2L34H.

	WWP1-2L34H (WWP1_19_4)
Wavelength/ Å	0.9763
Resolution range	84.49 - 3.0 (3.11 - 3.00)
Space group	C 1 2 1
Unit cell	227.0 59.7 108.3 90 99.4 90
Total reflections	201784 (19744)
Unique reflections	29068 (2805)
Multiplicity	6.9 (7.0)
Completeness (%)	99.66 (97.87)
Mean I/sigma(I)	5.36 (0.74)
Wilson B-factor	66.38
R-merge	0.317 (1.895)
R-meas	0.3427 (2.047)
R-pim	0.1292 (0.7681)
CC1/2	0.981 (0.469)
CC*	0.995 (0.799)
Reflections used in refinement	29029 (2800)
Reflections used for R-free	1492 (138)
R-work	0.2038 (0.3358)
R-free	0.2692 (0.3974)
CC(work)	0.956 (0.677)
CC(free)	0.931 (0.517)
Number of non-hydrogen atoms	8686
macromolecules	8667
ligands	3
solvent	16
Protein residues	1056
RMS(bonds)	0.011
RMS(angles)	1.31
Ramachandran favored (%)	88.37
Ramachandran allowed (%)	9.71
Ramachandran outliers (%)	1.92
Rotamer outliers (%)	3.91
Clashscore	12.22
Average B-factor	78.36
macromolecules	78.40
ligands	56.04
solvent	63.89
Number of TLS groups	11

5.2.2 WWP2 Crystallisation Efforts

5.2.2.1 Hanging drop crystallisation of WWP2-LH improved yield and quality.

With WWP2 being the original target for our work, a significant proportion of time was spent on crystallising the WWP2-LH construct, chosen for its relatively high resolution of 2.3 Å observed in the structure PDB ID: 5TJ8. The WWP2-LH purification strategies have been previously discussed in Section 3.2.1.1, with specific conditions found in Section 2.2.7.5. Crystallization of WWP2-LH was achieved following the conditions as stated in literature (100 mM MMT pH 6.0, 25% PEG1500) utilising both sitting and hanging drop vapour diffusion at 16 °C, as described in Section 2.3.4.1 (Chen *et al*., 2017). With various optimisations, mainly the speed of isolation using non-frozen samples, crystallization reliably occurred at 1 – 2 mg/mL screening between pH 5.5 – 6.5, 24 – 26% PEG1500, forming crystals with plate or rod habits (Figure 5-11).

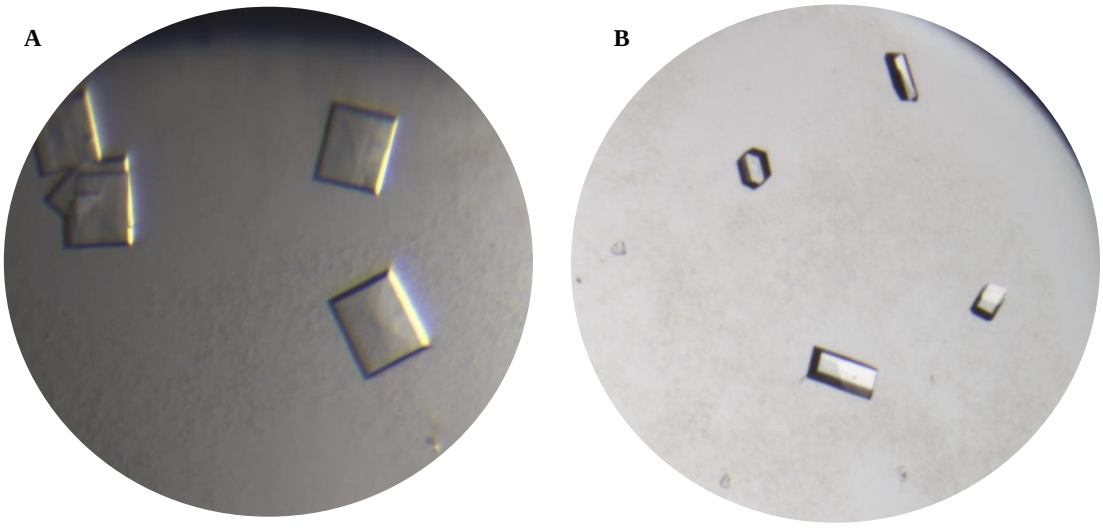


Figure 5-11: Crystal morphologies of WWP2-LH. All samples in a solution of 25 mM Tris.HCl pH 7.5, 150 mM NaCl, 5 mM DTT were crystallised screening around 100 mM MMT pH 6.0, 25% PEG1500, shown here using hanging drop vapour diffusion. (A) Plate crystals. (B) Rod crystals.

Interestingly, it was discovered that the hanging drop vapour diffusion method significantly improved both the speed crystal formation and drop count after the two methods were compared over 12 days, standardising to a 96-well dual drop plate as shown in Figure 5-12. The hanging drop method also resulted in each drop containing more crystals at a larger size ranging from 3 – 5 crystals at 50 – 100 μm^2 for sitting drop to 5 – 25 crystals at 150 – 200 μm^2 for hanging drop. Although not required, it was further noted that a short exposure/drying period of ~ 60 seconds during later stages resulted in increased nucleation, producing over 100 crystals per drop.

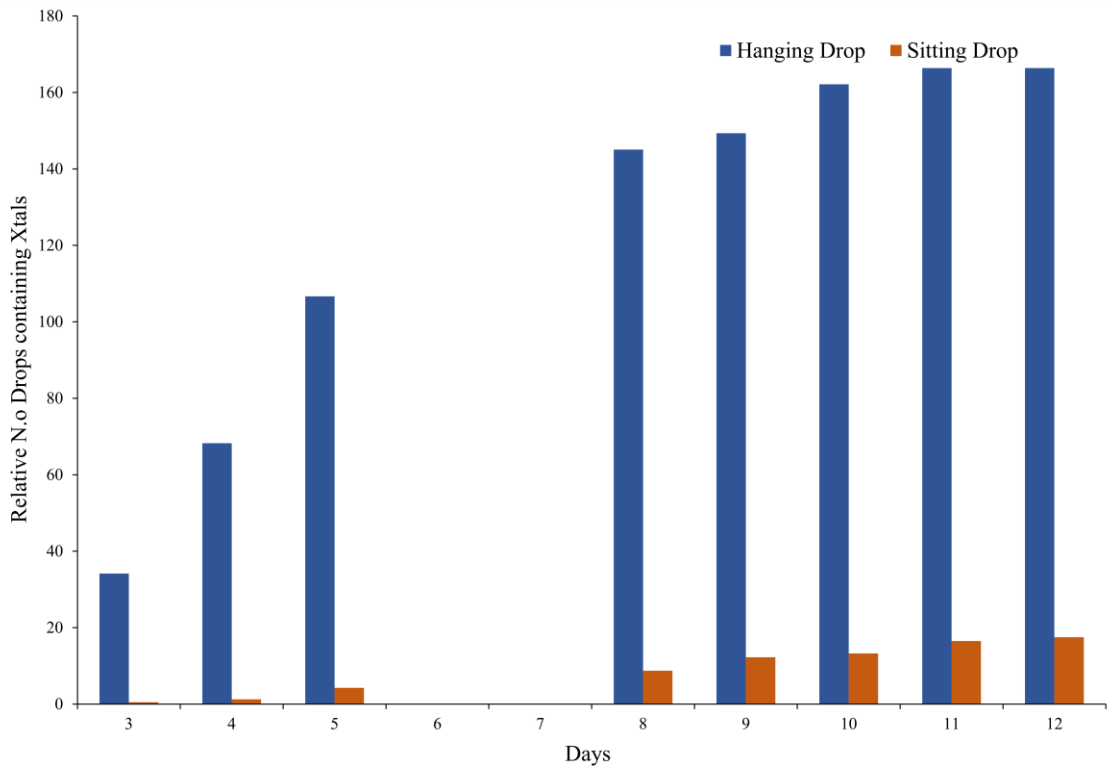


Figure 5-12: WWP2-LH crystal formation rates using hanging (blue) and sitting drop (orange) vapour diffusion. The two methods were compared over 12 days, excluding days 6 and 7, counting the number of drops containing crystals. This was standardised to a 96-well dual drop plate.

Apo crystals were harvested following protocols in Section 2.3.4.3, using 20% glycerol in mother liquor as a cryoprotectant. Data collection was run on either the I04 or I24 beamline at the Diamond Light Source, see Section 2.3.4.4. Unlike WWP1-2L34H, the two observed crystal morphologies autoindexed into the same $P2_1\,2_1\,2_1$ space group as found in literature. Our best structure was solved at a resolution of 2.06 Å using the structure PDB ID: 5TJ8 for molecular replacement in Phaser, as part of the CCP4 suite, before further refining in Refmac and Phenix. The relevant data collection statistics are shown in Table 5-4, with the full structure submitted as PDB ID: 9EQH, also shown in Figure 5-13. Although this model retained the same issue as PDB ID: 5TJ8 where the WW2 domain is disordered, the overall R values were reduced indicating an improved global phasing. As such, this model was later used in our in silico molecular modelling studies found in Section 5.2.3.

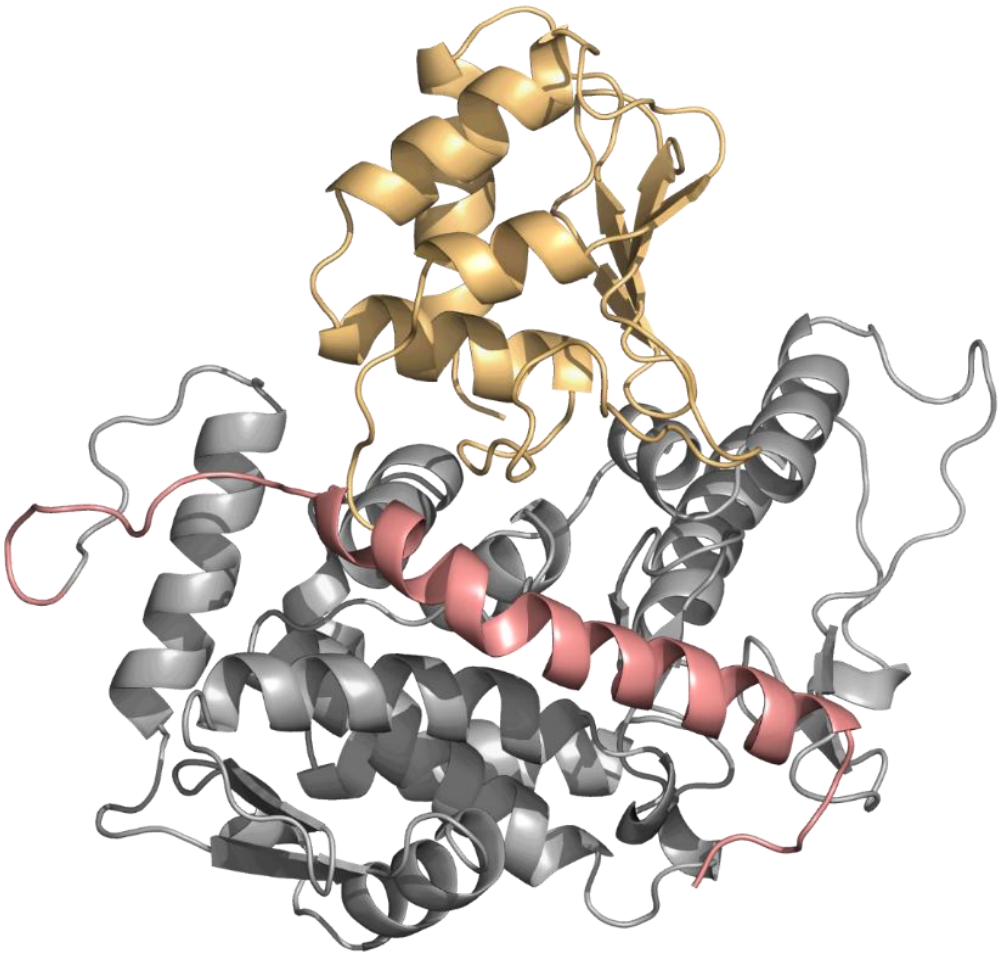


Figure 5-13: Crystal structure of WWP2-LH (PDB ID: 9EQH). The HECT domain contains the N lobe (grey) and C lobe (orange). The 2,3-linker (pink) is binding across the hinge region, with the WW2 domain missing due to being disordered. Image created in PyMOL.

Table 5-4: Summary of data collection and model statistics for WWP2-LH.

	WWP2-LH (WWP2_83_3)
Wavelength/ Å	0.9795
Resolution range	45.15 - 2.06 (2.134 - 2.06)
Space group	P 21 21 21
Unit cell	44.03 90.31 111.25 90 90 90
Total reflections	329940 (30499)
Unique reflections	28211 (2767)
Multiplicity	11.7 (11.0)
Completeness (%)	99.91 (99.96)
Mean I/sigma(I)	6.18 (0.98)
Wilson B-factor	36.38
R-merge	0.3231 (3.378)
R-meas	0.3386 (3.543)
R-pim	0.09963 (1.059)
CC1/2	0.992 (0.38)
CC*	0.998 (0.742)
Reflections used in refinement	28206 (2766)
Reflections used for R-free	1390 (130)
R-work	0.2165 (0.3183)
R-free	0.2655 (0.3489)
CC(work)	0.936 (0.699)
CC(free)	0.912 (0.640)
Number of non-hydrogen atoms	3642
macromolecules	3524
ligands	29
solvent	105
Protein residues	422
RMS(bonds)	0.015
RMS(angles)	1.36
Ramachandran favored (%)	91.43
Ramachandran allowed (%)	7.86
Ramachandran outliers (%)	0.71
Rotamer outliers (%)	2.65
Clashscore	10.37
Average B-factor	53.78
macromolecules	54.00
ligands	62.83
solvent	45.31
Number of TLS groups	1

5.2.2.2 Ligand-soaking efforts of WWP2-LH were unsuccessful

After the successful crystallization of WWP2-LH, we next wanted to soak the crystals in our hit inhibitors to visualise their binding interactions. Unfortunately, the WWP1-2L34H crystals were solved too late in the project to be used. WWP2-LH crystals were grown as previously described (Section 5.2.2), with soaking solutions generated from the mother liquor, see Section 2.3.4.3. Previous efforts undertaken by Dr Gregory Hughes with NSC-2805 and NSC-288387 had found that the maximum DMSO concentration these crystals could withstand without considerable loss in resolution was 2.5%, and so this was used as our limit. With the soaking experiments being undertaken throughout the project, the full characterisation of both binding and inhibition against WWP2 was not always known. With this in mind, a considerable amount of effort was exhausted attempting to observe ligand-bound states of the parent compounds, I3C and NSC-217913. This included varying both the soaking concentrations from 2.5 – 25 mM and soaking time from as short as 3 hours to 7 days. NSC-217913 was limited at 6.25 mM having solubility issues at high stock concentrations required for a 2.5% DMSO soak.

Despite our best efforts, neither I3C nor NSC-217913 soaked crystals showed significant positive electron density (Fo-Fc) that could be associated with a bound ligand state. There were noticeable differences between the soaked and apo structures, including a distinct change in the channel between the C and N lobe, found adjacent to the active site Cys838. The data collection and model statistics are shown in Table 5-5. This was later discovered to be the movement of bound glycerol close to a potential metal binding site (Figure 5-14A and B). This metal position is highlighted as a green sphere, and although we have shown it modelled as water due to having no significant data to model otherwise, another more recent study has shown this as a potential Zn(II) binding site (PDB ID: 8EI8). This soaked structure also demonstrates a preferred conformation for the active site Cys838 sulfur side chain to point towards the pocket, having been found to adopt two conformations (static disorder) in our apo structure. Interestingly, ligand-soaked WWP2-LH crystals also routinely demonstrated a significant shift in the last four amino acids of the C-terminal tail, resulting in this C/N lobe channel transitioning into a large ‘active site’ pocket (Figure 5-14C and D). This movement breaks various crystal lattice contacts and suggests a relaxing of the structure into a more native conformation with the addition of DMSO. The most intriguing discovery was observed in structures solved after the longest 7-day soaks, wherein a significant shift in the Cys588/Tyr587 ‘active site’ loop occurred. However, modelling of this alternate conformation is difficult given the extent of the disorder. This change was most likely associated with the oxidation of a nearby cysteine residue pair (Cys759, Cys802) forming a disulfide bridge, these residues were found reduced in the apo structure. Later on in the project, ligand soaking was revisited using the hit I3C analogues DIM, N-tosyl-I3C and even I19. It was quickly revealed that any attempt to soak

these particular hits led to a drastic loss of the resolution of diffraction (~6 Å), even at short 1 hour soaks. This crystal behaviour has previously been observed by Dr Gregory Hughes when looking at ligand soaking of NSC-288387 and led to the adoption of the alternative co-crystallization approach.

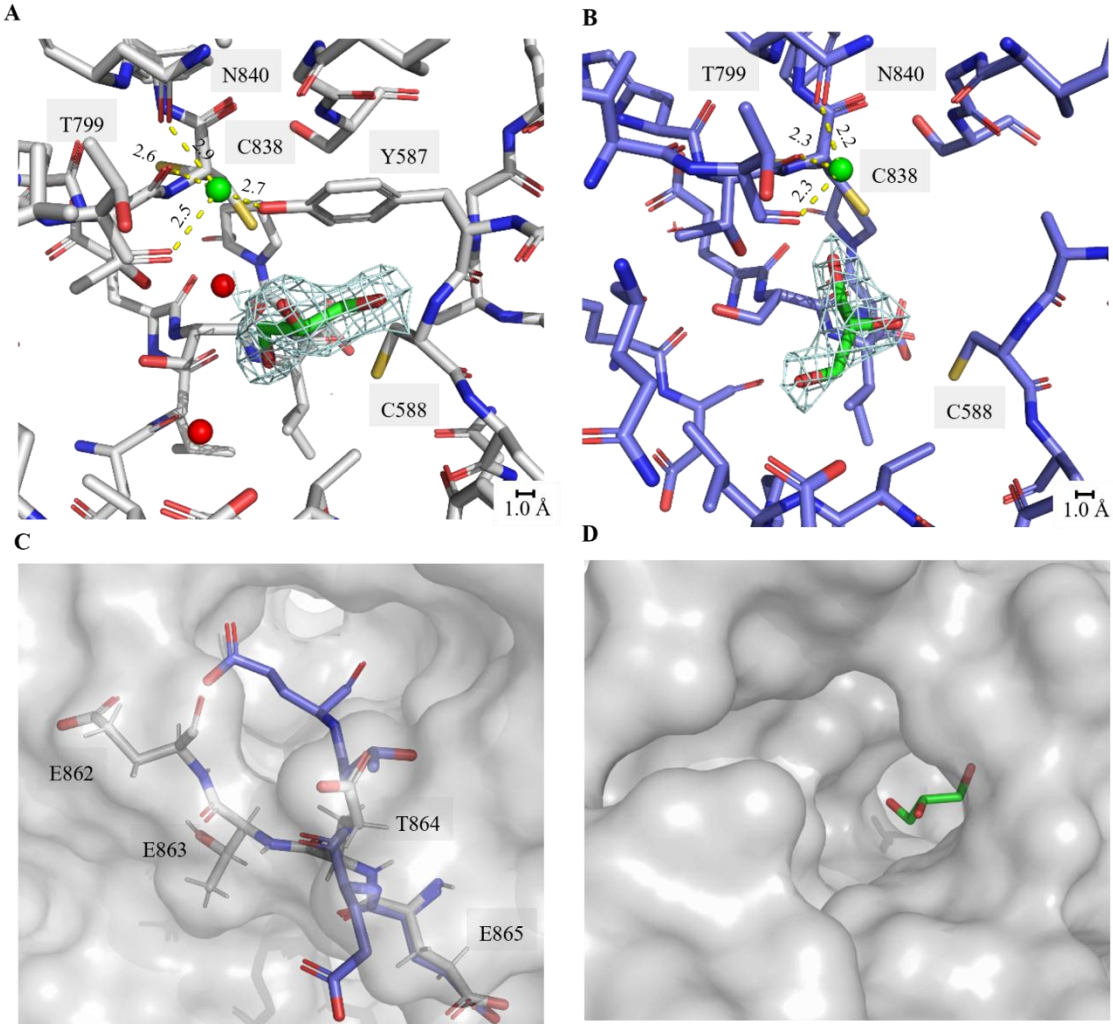


Figure 5-14: Structural changes between apo and ligand-soaked WWP2-LH crystals. Comparison between PDB ID: 9EQH (grey), and a 2.3 Å structure soaked with 2.5 mM I3C (2.5% DMSO) for 3 hours (purple). The 2Fo-FC electron density maps (light blue mesh) are contoured to 1 σ. The modelled water/potential metal of interest (green sphere) is shown with measured distances given in Angstroms (Å). (A) Glycerol (green) bound state in Apo structure. (B) Glycerol (green) bound state in soaked structure. (C) Comparison of last 4 amino acids (Glu862, Glu863, Thr864 and Glu865) position between apo (grey) and ligand soaked (purple) structures. (D) Surface representation of 'active site' pocket from ligand-soaked structure, with bound glycerol (green).

Table 5-5: Summary of data collection and model statistics for WWP2-LH (unbound) I3C soak.

	WWP2-LH (WWP2_68_3)
Wavelength/ Å	0.9795
Resolution range	70.7 - 2.3 (2.382 - 2.3)
Space group	P 21 21 21
Unit cell	44.3023 91.146 112.007 90 90 90
Total reflections	268763 (26411)
Unique reflections	20867 (2042)
Multiplicity	12.9 (12.9)
Completeness (%)	99.67 (99.12)
Mean I/sigma(I)	9.20 (0.51)
Wilson B-factor	53.54
R-merge	0.1945 (4.457)
R-meas	0.2026 (4.641)
R-pim	0.05632 (1.285)
CC1/2	0.998 (0.611)
CC*	0.999 (0.871)
Reflections used in refinement	20801 (2028)
Reflections used for R-free	1020 (104)
R-work	0.2185 (0.3490)
R-free	0.2731 (0.3838)
CC(work)	0.958 (0.651)
CC(free)	0.926 (0.646)
Number of non-hydrogen atoms	3487
macromolecules	3473
ligands	7
solvent	7
Protein residues	415
RMS(bonds)	0.009
RMS(angles)	1.14
Ramachandran favored (%)	92.94
Ramachandran allowed (%)	6.57
Ramachandran outliers (%)	0.49
Rotamer outliers (%)	3.48
Clashscore	9.66
Average B-factor	81.59
macromolecules	81.65
ligands	65.84
solvent	65.09
Number of TLS groups	1

5.2.2.3 Co-crystallisation efforts of WWP2-LH were unsuccessful

To account for possible conformational changes required for ligand binding, we attempted co-crystallization of the WWP2-LH construct with our hit inhibitors. Given the demanding nature of co-crystallization, in essence requiring a full protein crystallization screen for each ligand only I3C, NSC-217913 and I19 (believed at the time to be our best hit) were attempted. The sitting drop vapour diffusion method at 16 °C was adopted, using only 1 mM compound to maintain a low (1%) DMSO concentration. The purified WWP2-LH protein concentrated at 2 mg/mL was first incubated with each compound for 3 hours on ice before spinning at 20,000 g to remove any insoluble material and plated as per the described procedures in Section 2.3.4.1. Both I3C and I19 were screened exhaustively, against all 10 purchased commercial screens (see Section 2.3.4.2), with NSC-217913 only screened against Structure, PACT and PEG/ION crystallisation screens.

I3C co-crystallization yielded two repeatable conditions forming crystals with tabular habit after 2 weeks of incubation in either 100 mM bis-tris propane pH 7.5, 200 mM sodium acetate trihydrate, 20% w/v PEG 3350 or 100 mM sodium HEPES pH 7.5, 100 mM sodium chloride, 1.6 M ammonium sulfate. I19 co-crystallization yielded three repeatable conditions forming spherulite needle crystals after 6-weeks of incubation, growing to sufficient size for harvesting after a further 4 weeks in either 100 mM MES pH 6.0, containing either 200 mM $MgCl_2$, or 200 mM $CaCl_2$, or 100 mM HEPES pH 7.0 containing 100 mM $ZnCl_2$, with all solutions containing 20% PEG6000. Unfortunately, NSC-217913 co-crystallization efforts yielded no crystals. The crystals were cryoprotected with mother liquor containing 25% glycerol, with either 5 mM I3C or 2.5 mM I19 (2.5% DMSO) to prevent ligand displacement. Data collection was performed on either the I04 or I24 beamlines at the Diamond Light Source, see Section 2.3.4.4. Only the Bis-Tris-based I3C co-crystal demonstrated sufficient Braggs diffraction for indexing, achieved using DIALS to give a 2.9 Å resolution in the same $P2_1 2_1 2_1$ space group, with data collection statistics shown in Table 5-6. Unfortunately, as with our soaking experiments, no sufficient Fo-FC electron density was observed to enable I3C to be fitted. For the I19 co-crystallised samples, the MES/MgCl2-based co-crystals were found to be small molecule crystals of the I19 compound, and so were developed no further.

Table 5-6: Summary of data collection statistics for co-crystalised WWP2-LH with I3C (unbound).

	WWP2-LH – Co I3C (WWP2_75_1)
Wavelength/ Å	0.9795
Resolution range	47.62 - 2.904 (3.008 - 2.904)
Space group	P 21 21 21
Unit cell	44.1794 90.8219 111.843 90 90 90
Total reflections	134227 (13011)
Unique reflections	10429 (987)
Multiplicity	12.9 (13.2)
Completeness (%)	99.48 (95.73)
Mean I/sigma(I)	3.45 (0.35)
Wilson B-factor	63.38
R-merge	0.3929 (5.456)
R-meas	0.4091 (5.675)
R-pim	0.1129 (1.547)
CC1/2	0.992 (0.495)
CC*	0.998 (0.814)

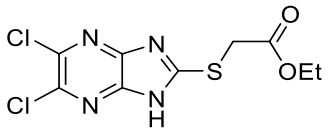
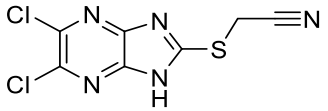
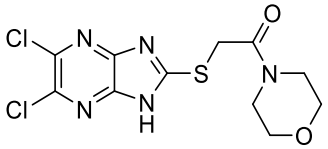
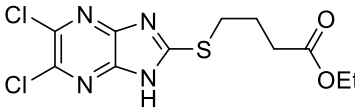
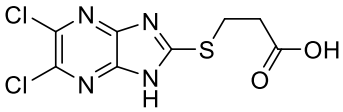
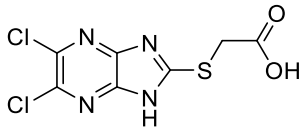
5.2.3 In Silico Molecular Docking against WWP1-2L34H and WWP2-LH

Despite our best efforts to solve enzyme-bound inhibitor structures, our findings were inconclusive. To gain insight into the possible SAR between our NSC-217913 and I3C analogues with both WWP1 and WWP2, in silico molecular docking was undertaken as an alternative approach to gain insight into their potential binding modes and location.

5.2.3.1 NSC-217913 analogues modelled into the active site pocket displayed a conserved heterocyclic orientation.

As NSC-217913 and its various hit analogues are novel WWP1/2 inhibitors, the 'true' binding position is unknown. As such, we selected a pseudo-binding location based on the assumption that this would be close to the active site cysteines 890 (WWP1) and 838 (WWP2), given the NSC-217913's ability to interfere with ubiquitination. Such a site has been previously highlighted in our WWP2 apo structures (Figure 5-14D) with a large cavity found close to the active site cysteine. This cavity has also been previously suggested to bind the WWP2 small molecule inhibitor NSC-288387, with the relevant contact points validated through STD-NMR epitope mapping (Watt *et al*., 2018; Hughes, 2019). Being an ambiguous site, a relatively non-intensive docking simulation was performed using the Lead Finder program on Cresset Flare as described in Section 2.3.5.1, with grid boxes set to surround this pocket. Our solved structures of WWP1-2L34H (PDB ID: 9EQK) and WWP2-LH (PDB ID: 9EQH) were selected for these docking simulations having improved Rfree values to those previously found in literature and/or more ordered regions. Simulations of the hit analogues of NSC-21713 including itself (Table 4-3) were run, excluding compound S17 given the purify issues. The cresset docking scores and estimated free energy of binding are shown in Table 5-7, with the best docking poses shown in Figure 5-15 and Figure 5-16, for WWP1-2L34H and WWP2-LH respectively.

Table 5-7: Cresset scoring functions and estimated binding free energies of hit NSC-217913 analogues against WWP1-2L34H and WWP2-LH.

Compound (S)	Structure	WWP1-2L34H		WWP2-LH	
		Docking score	Binding energy (kcal/mol)	Docking Score	Binding energy (kcal/mol)
6/NSC-217913		-8.259	-8.545	-8.686	-9.188
7		-7.732	-8.012	-8.147	-8.267
9		-8.801	-9.146	-8.876	-9.708
11		-8.221	-8.641	-8.893	-9.876
13		-7.504	-8.11	-7.943	-8.877
15		-7.566	-8.176	-7.654	-8.394

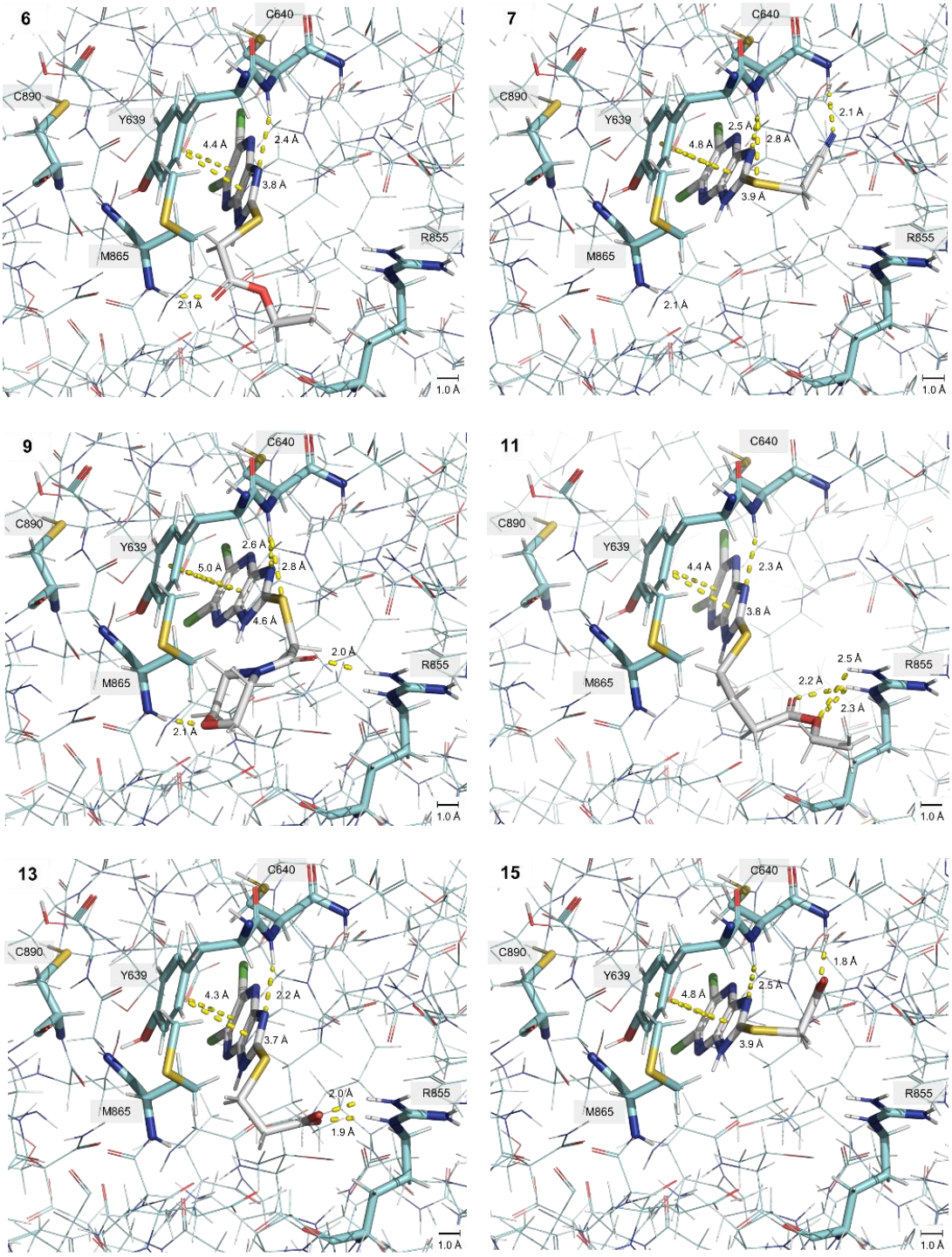


Figure 5-15: Ligand poses of NSC-217913 analogues 6, 7, 9, 11, 13 and 15 within WWP1-2L34H. Compounds were docked using Cresset Flare software into the Cys890 active site pocket, with key residues Tyr639, Cys640, Met865 and Arg855 represented as sticks. Both polar contacts and π-π interaction distances are given in angstroms (Å). The images were created in PyMOL.

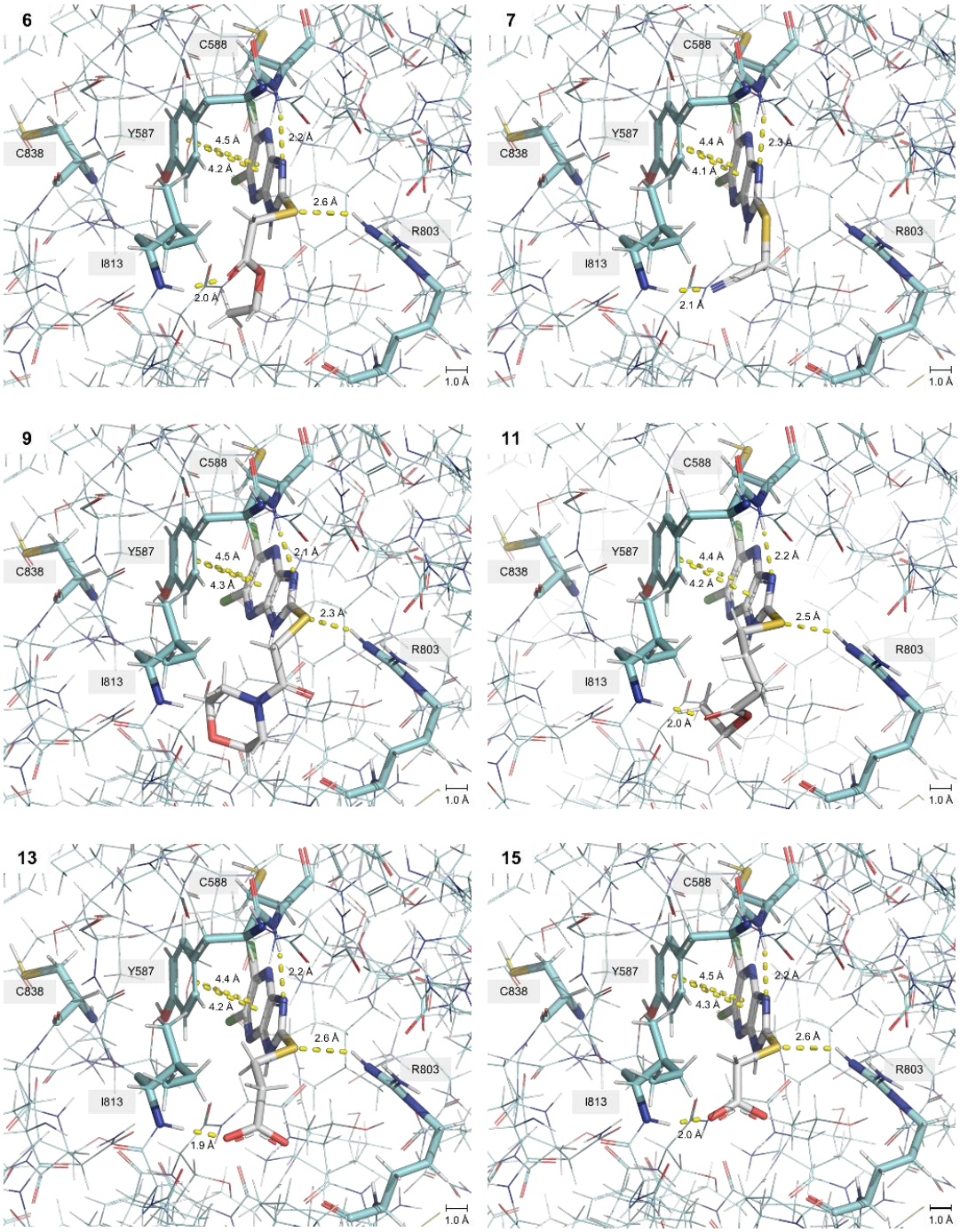


Figure 5-16: Ligand poses of NSC-217913 analogues 6, 7, 9, 11, 13 and 15 within WWP2-LH. Compounds were docked using Cresset Flare software into the Cys838 active site pocket, with key residues Tyr587, Cys588, Ile813 and Arg803 represented as sticks. Both polar contacts and π-π interaction distances are given in angstroms (Å). The images were created in PyMOL.

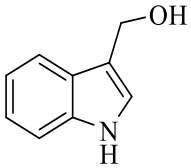
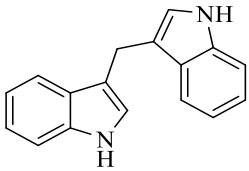
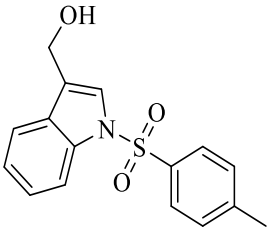
Again, given the ambiguous nature of the docking site, any observations should be carefully considered. That being said, all of the docked compounds demonstrated impressive Cresset scores with calculated free energy of binding no higher than – 8.0 kcal/mol for either WWP1-2L34H and WWP2-LH, most likely associated with the similarities of the proposed binding sites between these two enzymes. As shown, the heterocyclic portion of the NSC-217913 analogues all pick up at least the same π-π interactions in WWP1-2L34H/WWP2-LH with the phenolic side chains of Tyr639/Tyr587 and hydrogen bonding interactions with the backbone amides of non-catalytic Cys640/Cys588. Although not easily observed in Figure 5-15 and Figure 5-16, the chloride moieties may also be forming possible halogen (Cl) bonding interactions with Ser698/Ser646 at the back of the pocket. Depending on the thioether identity, interactions with either the Arg855/Arg803 and/or Met865/Ile813 residues also occur with only compound S7 shown to form a hydrogen bound with the amide backbone of Leu641 in WWP1-2L34H.

The most potent NSC-217913 analogue S11 provided a rank score of -8.221 (3rd in rank order) and -8.893 (1st in rank order), with estimated binding free energies of -8.641 kcal/mol and -9.876 kcal/mol for WWP1-2L34H and WWP2-LH, respectively. The estimated free energy of binding predicted compound 11 as the second best against WWP1-2L34H and best within the compounds under investigation against WWP2-LH. The ligand poses of compound S11 in both WWP1-2L34H and WWP2-LH are somewhat similar, again showing the imidazo[4,5-b]pyrazine moiety engaging in the same π-π interactions and hydrogen bonding interactions previously mentioned. However, the ethyl butyrate thioether chain interactions differ significantly between the two enzymes. When docked against WWP1-2L34H, the side chain of S11 orients itself to interact with the guanidium ion of Arg855. Against WWP2-LH, the sulfur interacts with Arg803, with the butyrate chain having multiple hydrophobic contacts with the aromatic ring of Tyr587 and the alkyl side chain of Ile813. Lastly, a backbone amide from Ile813 is hydrogen bonding with the carbonyl oxygen of the ester. As the heterocycle is oriented very similarly in the predicted complexes with both WWP1-2L34H and WWP2-LH, it may be that the ion-dipole interactions between the ester and guanidinium functionalities of S11 docked to WWP1-2L34H provide the additional strength of binding observed in the IC_{50} values. Against WWP2-LH, the ester is participating in weaker hydrogen bonding interactions against Ile813.

5.2.3.2 Indole-3-carbinol derivatives modelled into the Ub exo site displayed improved interactions compared to I3C.

Unlike NSC-217913, a possible binding location for I3C has been suggested, with various studies having modelled I3C against both NEDD4-1 HECT and WWP1 HECT, highlighting significant interactions of the indole ring in a hydrophobic cavity located close to the non-catalytic cysteines(Quirit *et al.*, 2017; Lee *et al.*, 2019). However, these studies have not considered the possible binding contributions of the WW2 domain (missing in PDB ID: 9EQH). To overcome this, we superposed WWP1 (PDB ID: 9EQK) and WWP2 (PDB ID: 6J1Z) structures containing WW2 on the previously solved NEDD4 I3C-derived covalently bound structure (PDB ID: 5C91), before aligning our target I3C derivatives of I3C, DIM and N-Tosyl-I3C(Kathman *et al.*, 2015). With the non-bound WWP1 and WWP2 structures having much tighter binding sites than observed in the covalently bound NEDD4-1 structure, we first minimised the surrounding pocket to generate a pseudo-bound state, before using the Schrödinger Glide Docking protocol to simulate ligand interactions, as described in Section 2.3.5.2. The corresponding GlideScore for both WWP1 and WWP2 are shown below in Table 5-8, with the representative poses of I3C, DIM and N-Tosyl-I3C against WWP1 and WWP2 shown in Figure 5-17 and Figure 5-18, respectively.

Table 5-8: Glide scoring functions of hit I3C derivatives against WWP1-2L34H (PDB ID: 9EQK) and WWP2-LH (PDB ID: 6J1Z).

Compound	Structure	WWP1	WWP2
		Glide Score (kcal/mol)	Glide Score (kcal/mol)
I3C		- 6.28	-6.45
DIM		-7.09	-7.79
N-Tosyl-I3C		-8.44	-7.28

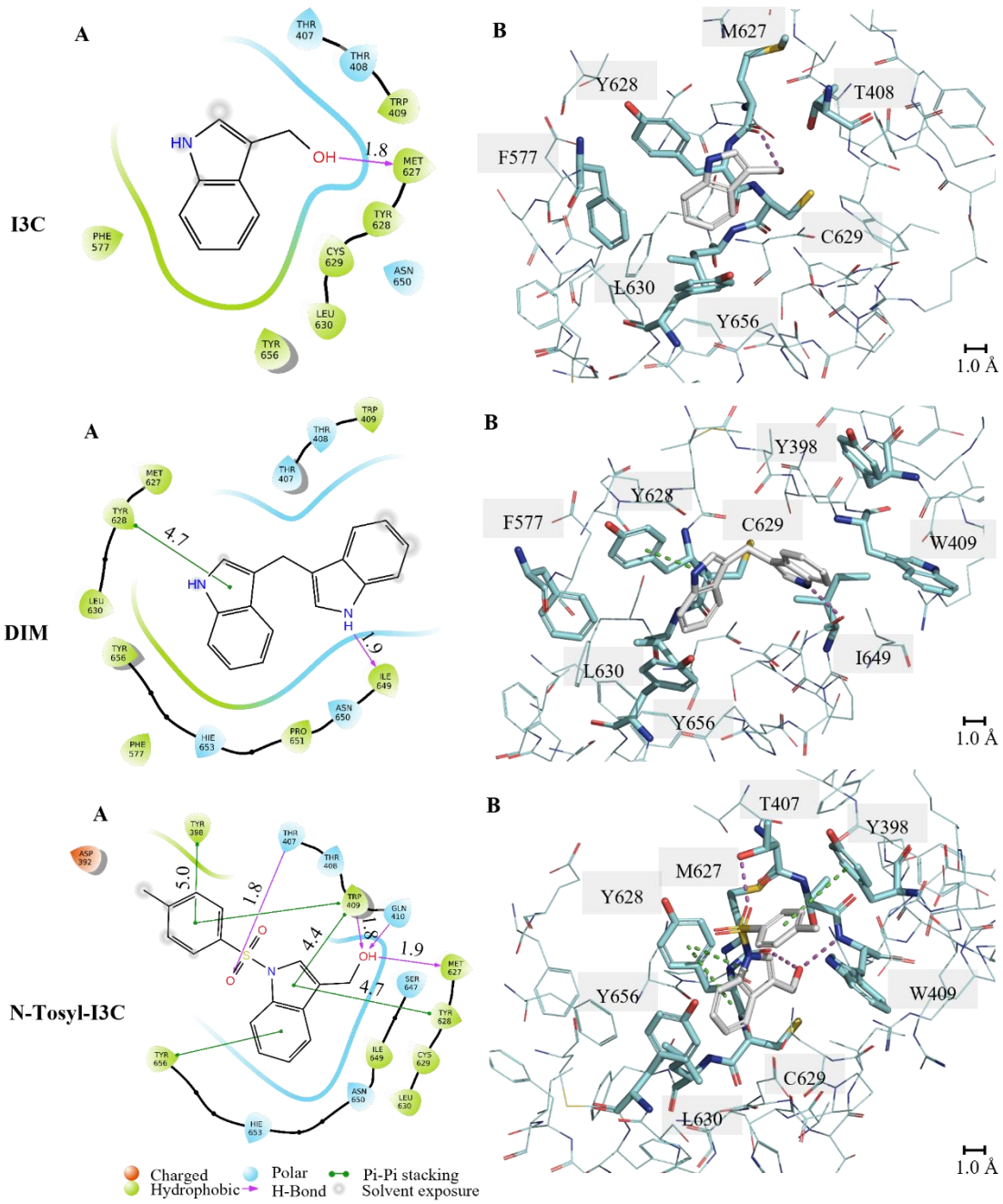


Figure 5-17: Ligand poses of I3C, DIM and N-Tosyl-I3C with WWP1. Compounds were minimised and re-docked into the exo Ub site of PDB entry 9EQK using Glide software. Hydrophobic (green), hydrophilic (blue) and charged (orange) residues are shown, with key polar (purple) and π-π (green) interactions given in Angstroms (Å). A. 2D interaction interface taken from Maestro. B. 3D interaction interface with hydrogens removed. Images were created in PyMOL with key residues represented as sticks.

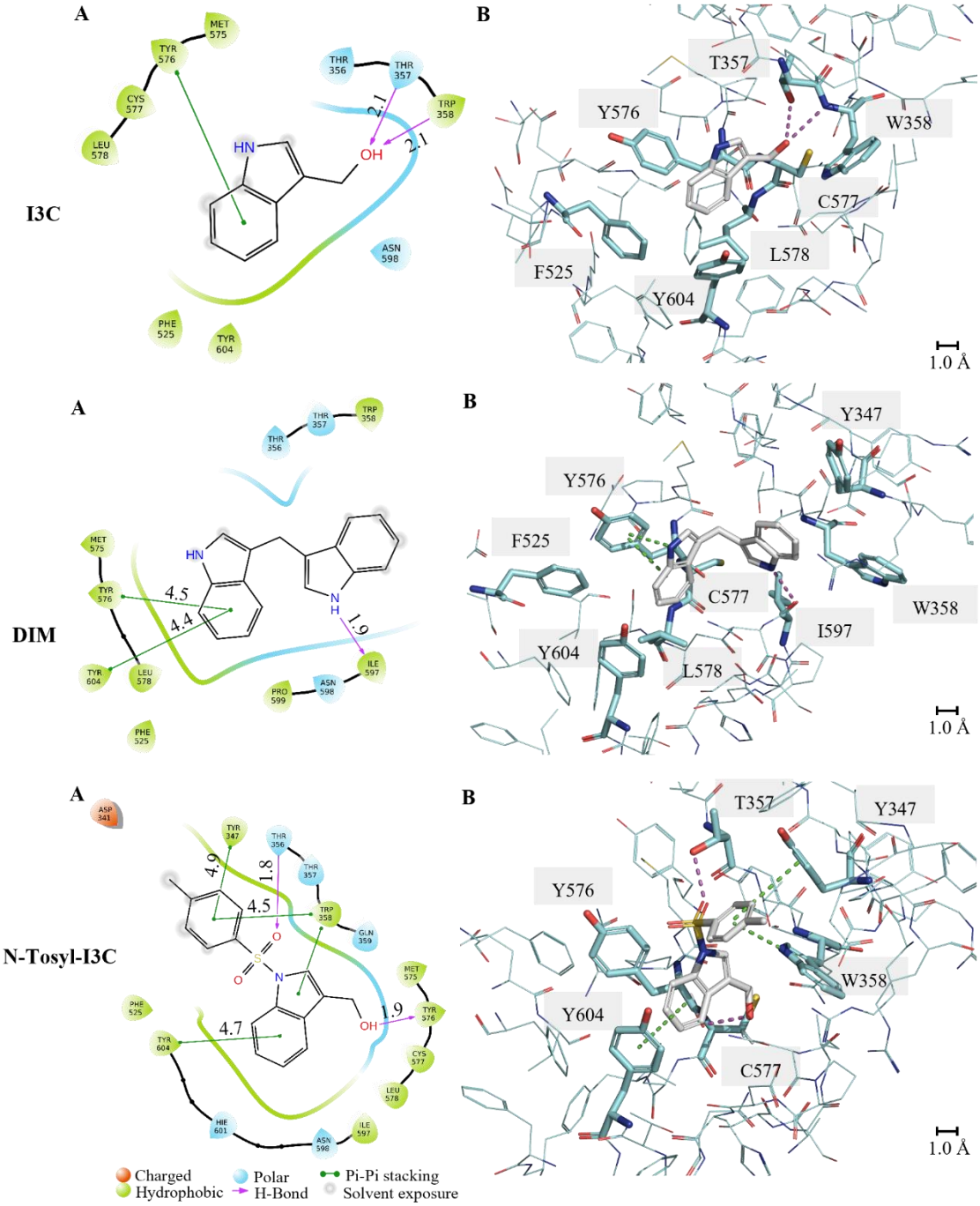


Figure 5-18: Ligand poses of I3C, DIM and N-Tosyl-I3C with WWP2. Compounds were minimised and re-docked into the exo Ub site of PDB entry 6J1Z using Glide software. Hydrophobic (green), hydrophilic (blue) and charged (orange) residues are shown, with key polar (purple) and π-π (green) interactions given in Angstroms (Å). A. 2D interaction interface taken from Maestro. B. 3D interaction interface with hydrogens removed. Images were created in PyMOL with key residues represented as sticks.

Correlating with a weaker binding and higher IC_{50}, I3C's GlideScore was the lowest of all derivatives at -6.28 and -6.45 kcal/mol for WWP1 and WWP2 respectively. The second pose of WWP1-I3C was chosen due to its agreement with STD NMR previously collected by the Chantry group (Appendix Figure 7-1) having the indole located deeper into the hydrophobic cavity appearing to be further supported by the aromatic Phe577, interestingly mutated from Leu553 in NEDD4-1. Although having a better GlideScore, WWP2-I3C generated various differentiated poses suggesting a weaker overall preference for binding and aligned well with the STD NMR. As expected DIM produced a higher GlideScore of -7.09 and -7.79 kcal/mol for WWP1 and WWP2 respectively. DIM makes very similar contacts between WWP1 and WWP2 not only fitting into the hydrophobic pocket but further making π-π stacking interactions with Tyr628/Tyr576 as well as the hydroxyl position hydrogen bonding with backbone nitrogen of Ile649/Ile597, the mutation position to NEDD4's Cys627. Although not clear from the 2D interface, DIM also makes hydrophobic interactions with Y398/Y347 and W409/W358 residues located on the WW2 domain. N-Tosyl-I3C displayed the highest GlideScore of -8.44 and -7.28 kcal/mol for WWP1 and WWP2 respectively, making significant interactions with both the HECT and WW2 domain. Although orientated further out of the hydrophobic cavity, the indole still participates in π-π stacking with Tyr656/Tyr604 and forms a hydrogen bond to the Tyr628/Tyr576 backbone from the hydroxyl position. The N-tosyl moiety makes closer contacts to Y398/Y347 and W409/W358 enabling π-π stacking, with the sulfonyl oxygen forming a hydrogen bond to Thr407/Thr357 also located on the WW2 domain.

5.3 Discussion

This chapter aimed to structurally characterise NSC-217913, I3C and their derivatives against the targets WWP1 and WWP2 E3 ligases using X-ray crystallography and in silico modelling, with the hope of aiding a future structure-informed approach. WWP1 and WWP2 were first re-crystallised by screening around literature conditions. Initially, WWP1-L34H crystallisation was attempted but later abandoned due to diminishing returns. After a more stable buffer was introduced using DSF, the alternative WWP1-2L34H crystals were solved both in the original literature and a new C 1 2 1 space group, at a resolution of 3.0 Å, resulting in the most complete WWP1 structure found to date (PDB ID: 9EQK). The WWP2-LH crystallisation was more straightforward, with optimisations able to increase the resolution to 2.06 Å, with better overall R values (PDB ID: 9EQH). Ligand soaking experiments were attempted using the WWP2 crystals, however, no significant density could be associated with I3C and NSC-217913, with the more potent inhibitors causing a loss in resolution. Although the co-crystallisation efforts of I3C did discover a new condition for WWP2 crystallisation at 2.9 Å, the lattice was in the same P 2_1 2_1 2_1 space group, and unfortunately, again there was additional density associated with I3C. In silico molecular docking efforts were more successful, using the two solved WWP1 and WWP2 structures. NSC-217913 and its most potent analogues were docked using Cresset into a cavity close to the active site cysteines, with all compounds demonstrating reasonable free energy of binding (- 8.0 kcal/mol or less). The subsequent poses also highlighted a potentially important π-π interaction with Tyr639/587 as well as various other maintained interactions within the site. Glide docking of I3C, DIM and N-Tosyl-I3C into a pseudo bound Ub exo site followed STD-NMR determined characteristics, with the indole modelling within the well-defined hydrophobic pocket (Tyr628/576, Phe577/525, Tyr656/604 and Leu630/578). Interestingly, the inclusion of the WW2 domain demonstrated its possible importance for binding in this site for DIM and N-Tosyl-I3C. Here both the crystallisation efforts of WWP1 and WWP2, as well as the approach to solving ligand-protein complexes will be discussed in more detail before finally discussing the in-silico modelling results to highlight possible SARs of the docked analogues.

5.3.1 Crystallisation efforts

5.3.1.1 Justification for WWP1 and WWP2 construct choice and crystallisation strategy.

Many variables contribute to the formation of protein crystals, ranging from the more obvious (protein purity, concentration, buffers, pH, salts, precipitants, etc) to more target-specific (size,

hydrophilicity, flexibility, etc) (Rhodes, 2006). In truth, this stage of X-ray crystallography is the least understood and can be an exhaustive process of screening these parameters to obtain the perfect conditions for crystallisation. The introduction of machine learning is beginning to provide some level of greater understanding, having access to a plethora of training data in the various depository data banks, yet is still at its earliest stages (Matinyan, Filipcik and Abrahams, 2024). Despite our best efforts, numerous proteins are simply unable to undergo crystallisation in their native forms, with membrane-bound proteins being well characterised in this predicament, whether it's their sheer size and/or hydrophobic surfaces interfering with crystal packing. This problem is typically translated to multidomain proteins, such as HECT E3 ligases, being also large (~100 kDa) and undergoing dynamic movements as part of their function. To overcome these issues, many of these proteins are segmented into smaller sections, either expressed as such or proteolytically cleaved, with at minimum augmented to retain stability. In rare cases, site-specific mutations are incorporated to improve their capabilities of crystal packing, but this is usually employed where a preliminary structure has already been solved (Derewenda, 2004). In the case of WWP1 and WWP2, significant efforts had already been employed to obtain crystal structures, originally as HECT-only constructs, as PDB ID: 1ND7 and PDB ID: 4Y07, respectively, to the more complex variants previously mentioned as WWP1-L34H, WWP1-2L34H and WWP2-LH (Verdecia *et al.*, 2003; Gong *et al.*, 2015; Chen *et al.*, 2017; Wang *et al.*, 2019). Although solved, these complex structures displayed disordered domains, suggesting either misfolding or at least domain movements given that electron density maps are essentially an average of the entire lattice. Even with this issue, the inclusion of the surrounding domains observed stability of these constructs and the apparent relative ease of their purifications, made these complex constructs the best choice for the later protein-ligand complex efforts. This was further supported by the purification struggles of WWP2-HECT noted in the previous chapter (Section 4.2.4.1). In the recrystallization efforts, both structures were overall improved. As highlighted, the new C 1 2 1 space group discovered for WWP1-2L34H enabled the modelling of the previously disordered WW3 domain (Figure 5-19).

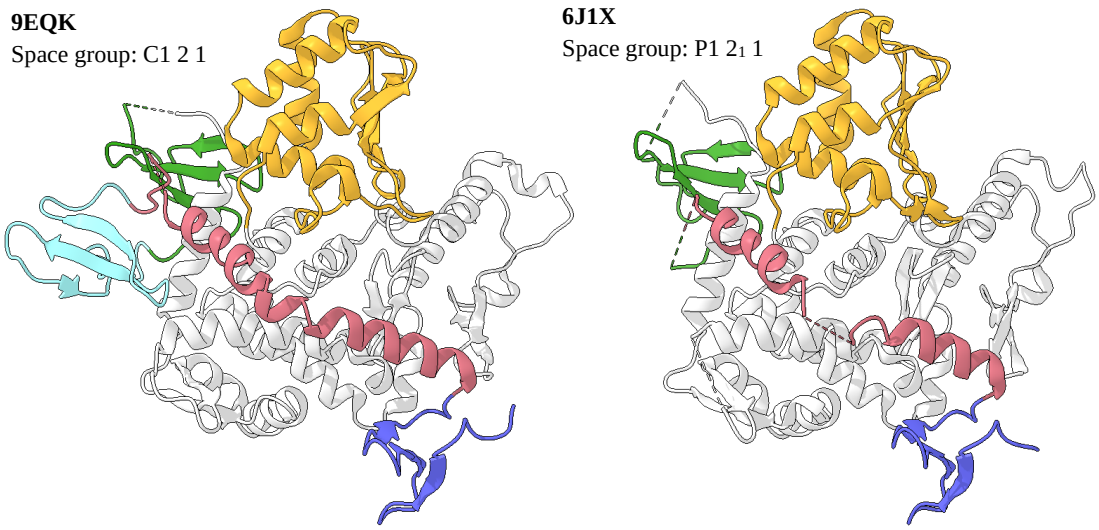


Figure 5-19: WWP1-2L34H model comparison between PDB ID: 9EQK (left) and PDB ID: 6J1X (right). The PDB entry 9EQK was solved here, in the new space group C1 2 1, alternate to the P1 2_1 1 previously reported in literature (Wang *et al*., 2019). For both structures, the HECT domain is shown to contain an N lobe (grey) and C lobe (orange). The WW2 (purple) domain is shown binding adjacent to the Ub exo site with the 2,3-linker (pink) binding across the hinge region, finally resulting in the WW3 (blue) (disordered in 6J1X) and WW4 (green) domains. Image created in ChimeraX.

For WWP2-LH both the resolution and validation statistics were improved including the R values, a measure of correlation between the model and the experimental data. Interestingly, the crystallisation of this construct via the hanging drop over sitting drop vapour diffusion significantly improved crystal formation speed, crystal size and morphology (Figure 5-12). These observations have been noted previously by Hou and colleagues (2019), highlighting the effect of gravity on concentration distribution within the drops, with hanging drops having a larger protein concentration at the air/drop interface compared to the solid/drop interface observed in sitting drop plates (Figure 5-20). This translates into their point of nucleation, with hanging drops crystals forming away from the plate, preventing lattice mismatch from occurring from surface contacts. In a more practical sense, these crystals were also easier to harvest, meaning less damage occurred during handling. Some studies have taken this to the extreme, looking at the benefits of a contactless crystallisation system (Maki and Hagiwara, 2022).

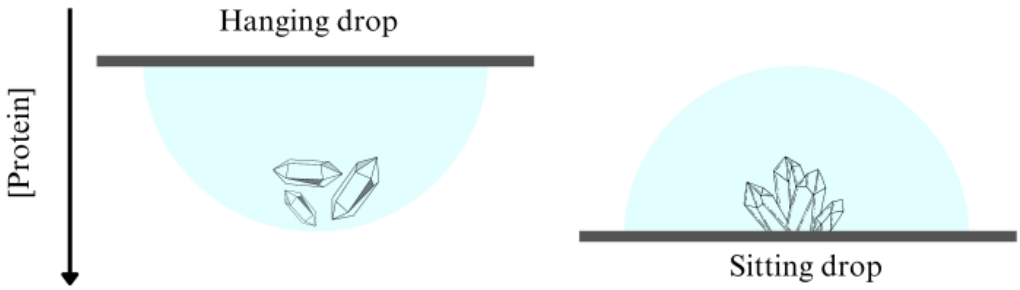


Figure 5-20: Protein distribution within hanging and sitting drop vapour diffusion. Hanging drops: Protein most concentrated at air/drop interface, leading to nucleation and unimpeded lattice formation. Sitting drops: Protein most concentrated at solid/drop interface, leading to nucleation close or direct at plate contact, potentially causing lattice mismatch. During harvesting, crystals commonly require breaking off the plastic, resulting in damage.

5.3.1.2 Low-resolution WWP1-2L34H crystals may have been due to poor cryoprotection.

Unfortunately, the WWP1-2L34H construct was crystallised at a low resolution of 3.0 Å, late into the project. Although there is no defined resolution limit for ligand-protein complexes, in the case of small molecules, higher resolutions (~ 2.0 Å) are desired to be confident in the ligand model itself. This is more important given that weaker ligands are found at lower occupancy in crystal structures, meaning they are not bound to every protein throughout the crystal lattice. As mentioned, this translates into a loss of observed electron density from the averaging effect. Such low resolution of WWP1-2L34H was most likely due to lattice 'melting' during the cryoprotection stage. Cryoprotection is a vital step to prevent damage during snap-freezing from ice formation. However, the choice of cryoprotectant is important and needs to be able to enable crystal vitrification whilst not degrading the lattice. There are many choices available, ranging from the most common small polyols (glycerol and ethylene glycol) and low molecular weight PEGs (200, 400, 600) to sugars and even oils (Pflugrath, 2015). Various studies tend to add the cryoprotectant directly into the 'mother liquor', however, this can result in the crystals destabilising through the dilution of the crystallisation buffer, as observed with WWP1-2L34H. We did look to use both glycerol and ethylene glycol, however, were only able to perform minimum optimisations given the late stage of the project. The final cryoprotectant (5% concentrated crystallisation buffer and 30% glycerol) still showed signs of crystal damage after soaking for five minutes. A short-handing period may have aided this, however given the crystals were grown in a reagent alcohol precipitant, harvesting was particularly difficult. Isopropanol is a very volatile alcohol, and boils at the low volumes found in drops causing the

crystals to ‘dance’. Once the isopropanol vapours had boiled off, the low ‘water’ retention caused the crystal to sink to the bottom of the well, requiring physical manipulation to harvest. Future work with these crystals may find more success in screening for a stable cryoprotectant buffer absent of an alcohol precipitant or simply undergoing a full rescreen of crystallisation conditions.

5.3.1.3 Ambiguous unoccupied density led to over-interpretation of the WWP2-LH ligand-soaked crystals.

The WWP2-LH crystals did provide more hope towards discovering a ligand-protein complex, given much of the optimisation was performed previously by Dr Gregory Hughes. However, despite a significant amount of effort (~800 crystals undergoing data collection), no positive Fo-FC density could be associated with the soaked or co-crystalised ligand. Various soaking conditions were looked at including varying the soaking times. Short soaks are known to limit potential lattice damage whereas longer soaks enable the ligand to diffuse into the centre of the crystals, resulting in improved occupancies (Geremia *et al.*, 2006). In the case of WWP2-LH, longer soaks reduced the resolution of the crystals, and even in some cases caused conformational transitions through oxidation. Although outside this project’s reach, the use of an acoustic injection of small molecules can overcome both these limitations and is used as part of the XChem strategy (Collins *et al.*, 2017). Rather frustratingly, at two points in the project, there were signs of potential binding. Initially, soaked crystals with I3C demonstrated a large unoccupied Fo-FC density close to the active site Cys838. However, this was difficult to model, at the time believed to be due to low occupancy and so poldar omit maps were employed. Omit maps remove the bias phase contributions of a modelled ligand before comparing the electron density maps. Poldar maps take this one further removing bulk solvent contributions within the omitted region to aid low occupancy ligands (Liebschner *et al.*, 2017). These poldar maps were still inconclusive. Such uncertainties were further caused by the poor R values in the WWP2-LH structure and its disordered WW2 domain, associated with poor global phasing. The effect of this was that the electron density maps from the perceived high-resolution 2.0 Å structure, looked more like 3.0 – 3.5 Å maps, making modelling a small molecule even more difficult. As observed, it was later discovered that the unoccupied electron density was an alternative glycerol conformation, potentially associated with coordination with an unknown metal (Figure 5-14B). In hindsight, a considerable amount of labour was spent looking to find NSC-217913 and I3C bound structures, now known to either not interact with WWP2, or at an insufficient strength to be observed by our bioassays. Despite the outcome of this study, co-crystallisation would be the best approach moving forward with the most potent compounds such as S11 and DIM, given this has already been shown to work using NEDD4-1 and its covalently bound I3C-

derivative and also would enable conformational changes without the apparent loss of resolution (Kathman *et al.*, 2015).

5.3.2 In-silico Modelling Efforts

5.3.2.1 Justification for modelling programs and strategies used.

Being unable to gain any structural information experimentally, the use of in silico molecular modelling was employed. These techniques are computationally expensive, requiring simulated exploration of the vast chemical space around a potential binding site. Luckily, I3C's binding to the Ub exo site is already well characterised, not only structurally against NEDD4 but also through mutational activity studies against WWP1, discussed previously (Lee *et al.*, 2019). Unfortunately, we were unable to characterise the NSC-217913 binding location. This led to some ambiguity in a choice between four possible locations (Ub exo site, E2 binding site, active site Cys, and flexible linker) discussed previously. Given similarities between the indole scaffold of NSC-217913 and I3C, the Ub exo site appeared to be an attractive site. However, after careful consideration, the highly hydrophilic nature of NSC-217913's indole scaffold made it unlikely to bind at the same hydrophobic cavity found vital for I3C interaction. With NSC-217913's small size, it was also unlikely that it would be able to interfere with the protein-protein interactions found at the E2 binding site and flexible linker and thus the cavity found close to the active site cysteine was chosen. With this in mind, two different approaches were taken for the two 'parent' scaffolds, with the faster and less computationally expensive Cresset Flare docking software used for the more ambiguous NSC-217913 site, and the exhaustive, gold-standard GLIDE Schrödinger software used for the I3C docking. Both these programs use flexible ligand docking, enabling the ligand to undergo conformation changes 'on the fly', whilst maintaining a rigid receptor. Cresset Flare utilises Lead Finder, a stochastic search algorithm originating from the classical genetic algorithm (Stroganov *et al*., 2008). In essence, this algorithm randomly generates a library of initial poses it terms 'chromosomes' before pairing high-scoring chromosomes to generate 'children'. These children are themselves randomly generated based on the features of their 'parents' (torsion angles, orientations etc) (Judson, Jaeger and Treasurywala, 1994). Similar to Darwinian evolution, the highest-scoring children poses are then selected for the next generation until no further improvements are observed. In this staged approach, both empirical and force-field based scoring functions are incorporated to rank the produced poses. Although user-friendly, this docking approach can result in the final poses not being optimal. The more exhaustive GLIDE docking software uses a complete systematic search algorithm of conformational, orientational, and positional space of the docked ligand, reducing the computational stress through a funnelling approach (Friesner *et*

al., 2004). This funnelling uses a rough scoring period, again through both empirical and force-field based functions, including a small receptor energy minimisation period allowing some receptor flexibility before continuing a more refined iterative pose selection. This software also contains induced-fit docking, further increasing the receptor flexibility to mimic conformational changes that would be observed under native conditions. However, given that an I3C-derivative NEDD4 bound structure was already available we were able to directly align both WWP1 and WWP2 as well as the I3C derivatives to undergo a more selective 'induced-fit' using energy minimisations before redocking into the new pseudo-bound site.

5.3.2.2 The active site pocket may indeed be NSC-217913's binding site.

If taken at face value, the NSC-217913 docking efforts highlighted some interesting features. Initial interpretations demonstrate a clear orientation and positioning of the heterocyclic portions of all the modelled poses against WWP1 and WWP2 are undergoing at least the same π-π interactions with the Tyr636/587 phenolic side chains and hydrogen bonding interactions with the Cys640/588 residues as seen in the most potent compound S11. However, depending on the thioether identity, interactions with either the Arg855/803 residue (right of the compounds), Met865/Ile813 (left of the compounds) or Leu641 (in WWP1 above the heterocyclic core) residues occur (Figure 5-15 Figure 5-16). This is intriguing and to some degree supports the selected position, especially given that a large receptor grid (8.0 Å) was used for the docking simulations that encompassed the entire binding cavity. Concerning the observed SARs from the bioassays, the dichlorinated positions were highlighted to demonstrate increased potency for both WWP1 and WWP2. Although unclear in the figures, the lower chlorines have the potential to participate in halogen bonding to Ser698/646 and Thr889/837, when considering the inflexibility of the docked receptor. The major contributions are still most likely through the augmenting dipole-dipole interactions between the indole and tyrosine aromatic rings as suspected from the earlier bioassays and now highlighted to be the most important binding feature (Wheeler *et al.*, 2010). The Arg855/803 is modelled to participate in N-S hydrogen bonding for most compounds, however, the most potent S11 appears to make significant interactions through its extended thioether chain that could further explain its increased potency. The increased potency towards WWP1 over WWP2 may indeed be due to the Met865/Ile813 mutated position given its potential involvement with the thioether group. Earlier bioassay screens highlighted that both more liner and hydrophilic thioether groups had a tentative trend towards NSC-217913 analogue potency. This mutated position could therefore potentially be of importance, with the WWP2 isoleucine significantly increasing the hydrophobic contacts in this region. This could further relate to compound **S9**, containing a cyclic ether increasing these hydrophobic interactions and shown to be more potent towards

WWP2. Given the ambiguity of this position, any SAR analysis of these proposed docking poses should be considered even more carefully than with any traditional simulated models. For instance, the apparent similarity in the heterocyclic position may simply be an artefact of the modelling strategy due to the minimal flexibility in the receptor site. Future studies would benefit from first confirming the binding site of NSC-217913 analogues, primarily S11, either through mutational studies such as Tyr636/587 to alanine or through other biophysical techniques such as STD-NMR or HDX-MS.

5.3.2.3 The WW2 domain appears to be important for DIM and N-Tosyl-I3C interactions at the Ub exo site.

In contrast to the NSC-217913 WWP2 docking, the I3C GLIDE docking efforts were achieved using an alternative WWP2 structure (PDB ID: 6J1Z), found to have an ordered WW2 domain. This was important given that WW2 is located at the Ub exo site and plays a vital role in the autoinhibitory actions of WWP2 (Chen *et al.*, 2017). Interestingly, such interactions do not occur in NEDD4-1, found to bind the C2 domain in an alternative mechanism (Wang *et al.*, 2019). Given the I3C, DIM and N-Tosyl-I3C are more chemically distinct than the NSC-217913 hit analogues, no major SAR trends could be observed. Despite this, some interesting features may provide some insight into their possible inhibitory action. Previous work by Lee *et al.* (2019), highlighted the importance of the hydrophobic cavity for I3C binding to WWP1 after alanine mutations of Phe577 and Tyr656 resulted in desensitization and loss of binding observed microscale thermophoresis assays. The docking of I3C here further supported this hydrophobic cavity, with not only the WWP1-I3C poses found to agree with previous STD-NMR observations, but the various WWP2 poses were so ambiguous they indicated no significant interactions, similar to the DSF and autoubiquitination assay results (Figure 5-17). Docking of both DIM and N-Tosyl-I3C, also demonstrated hydrophobic as well as dipole-dipole interactions with Tyr628/Tyr576 and Tyr656/Tyr604, respectively. Interestingly, DIM makes further hydrogen bonding interactions with the Ile649/597 of WWP1 and WWP2, a mutated position of the non-catalytic NEDD4 Cys627, found adjacently reintroduced at Cys629/577. The docked poses also highlighted DIM undergoing hydrophobic interactions with the WW2 domains Y398/Y347 and W409/W358 residues. Such interactions close to the Ub binding site could be of importance for DIM inhibitory action, however, does not explain the difference observed between DIM's inhibition of WWP1 and WWP2. This is made more complex given our autoubiquitination assays exploited a variation of WWP1 (WWP1-L34H) without the WW2 domain due to a problem associated with a strong autoinhibition mechanism observed in full-length WWP1 constructs making them mostly inactive (Wang *et al.*, 2019). It may be that the absence of the WW2 domain enabled the DIM inhibition in WWP1, having the opposite effect

on WWP2. This prevention of DIM inhibition could be through steric hindrance; however, is unlikely given that DIM demonstrated strong binding to WWP2-LH in the earlier DSF assays. In either case, this does demonstrate that the WW2 interactions are not vital for the inhibition of these I3C derivatives, even if they still may contribute to their binding interactions. Such contributions may be most important in N-Tosyl-I3C, undergoing π-π stacking with the WW2 domain's Tyr398/347 and Trp409/358, and its sulfonyl oxygen forming a hydrogen bond to Thr407/357. Although more work is required to gain a more detailed understanding of their inhibitory action, the differences in I3C and 1-Benzyl-I3C inhibition discussed previously (Section 4.3) alongside various mutations in the Ub exo site pocket and possible WW2/C2 interactions, suggest significant differences in the binding environment and therefore the possibility of a selective therapeutic approach between NEDD4-1 and WWP1/WWP2 using an I3C-derived scaffold.

Chapter 6: Discussion

In this final discussion, the overall strategies taken to discover NSC-217913, as well as the proposed SARs of the most potent NSC-217913 and I3C analogues are summarised and discussed, including suggestions for future studies. This project provides the basis for lead compound development in the next Lead-optimisation phase of the drug discovery pipeline.

6.1 The Hit-identification Phase

6.1.1 The High-throughput Screening Strategy

At the beginning of this project, the Chantry group had already highlighted several hit compound inhibitors against WWP2 as part of their previous screening efforts, most notably NSC-288387 and NSC-2805 (Watt *et al.*, 2018). The Hit-to-Lead SAR of NSC-288387 had already been undertaken by Dr Gregory Hughes, highlighting its potential binding site in the large cavity close to the active site Cys838 residue of WWP2. However, this compound had issues around its chemical tractability and demonstrated a potential PAINS redox property (Hughes, 2019). NSC-2805 was much more synthetically accessible, and so a substantial SAR was undertaken. Unfortunately, this compound demonstrated significant PAINS properties, associated with its di-hydroquinone moiety, as highlighted in Section 3.2.3.2.

Given the development of the most promising WWP2 hits had failed, it was necessary to return to the Hit-identification phase to address the shortage of available small molecule inhibitors, achieved here through screening the 'next-in-series' NCI Div VI compound library, as described in Chapter 3. The project was also expanded to investigate WWP1, an arguably more promising HECT E3 target being better characterised in its oncogenic roles with two small molecule inhibitors (Heclin and I3C) already shown to interact (Mund *et al.*, 2014; Lee *et al.*, 2019). The previous screening strategy was also improved, including the introduction of a DSF assay, reducing the compound library (~1600 compounds) to a more manageable size (24 compounds for both WWP1 and WWP2). The DSF assay was optimised by Dr Gregory Hughes, moving from the HECT-only constructs to the more thermally stable WWP1-L34H and WWP2-LH (Hughes, 2019). Although other in vitro activity assays were available, significant optimisation work had already been completed by Dr Jessica Watt on the ELISA autoubiquitination assay, and so this was carried forward alongside a thorough PAINS filtering and counter-assay process. These combined strategies led to the discovery of NSC-217913 as the sole hit inhibitor demonstrating an IC_{50} of 158.3 µM against WWP1, once re-synthesized.

6.1.1.1 Future high-throughput strategy improvements

In retrospect, there were potentially many oversights that could have improved the hit-ratio

without significant changes to the screening strategy. Firstly, although moving away from the HECT-only constructs seemed to improve the background noise of the DSF assay, we can assume this may have been the result of the IMAC purification strategy utilised at the time leading to reduced purity (See Section 4.2.4.1). The use of the HECT-only constructs may have also improved the ΔTm D signal, as the stabilising effect of a binding partner could potentially be more significant when the apo protein is more thermally unstable. Secondly, although the DSF assay was particularly useful in deducing the library size, this could have been used to greater effect. For instance, more than a single screening library could have used with a higher cut-off, such as 2 °C, to reduce the number of false positives and increase the chances of identifying stronger interactions, particularly as the assay was quick and cheap in comparison to the downstream activity assays. Although compounds that increased the ΔTm D of WWP1 and WWP2 were selected, any destabilising compounds were omitted. This was reasonable if searching for ligands, however, these compounds could have demonstrated inhibition by interfering with key intramolecular interactions, and as such in future these compounds should also be considered for activity assays. Finally, when considering the activity assay used, a major concern was that the HRP-TMB reaction was not only being interfered with by the small molecules but also was a time-specific reaction, resulting in discrepancies. As discussed in Section 3.3.1, future studies could consider removing the ELISA-based steps, and replacing the flag-tagged Ub with Fluorescein-Ubiquitin therefore able to measure ubiquitination directly. This could further be modified and combined with other Ub-tags to measure Ub chain formation in a TR-FRET based assay (Madiraju *et al*., 2012).

6.1.2 Other Potential Small-molecule Inhibitors

Given that NSC-217913 was the only compound selected from the screening strategy, a closer inspection of the earlier omitted small molecules may have provided alternative development opportunities as therapeutic hits. In particular, NSC-13151 and NSC-57103 were shown to inhibit WWP1 (IC_{50} of 396.9 µM) and WWP2 (IC_{50} of 513.0 µM), respectively (Figure 6-1, Figure 3-10). These IC_{50} values were calculated using a non-logarithmic concentration range resulting in potential accuracy issues, and as such these compounds may indeed be more or less potent. These accuracy issues can be further extended in the case of NSC-13151, provided as a mixed stereoisomer, meaning if only one stereoisomer was active, the observed potency would be lower. Furthermore, in line with the oxidation issues of NSC-217913, once freshly synthesised these compounds may demonstrate altered potency. In either case, both compounds will still require the later cross- and counter-assay stages to ensure there are no off-target interactions.

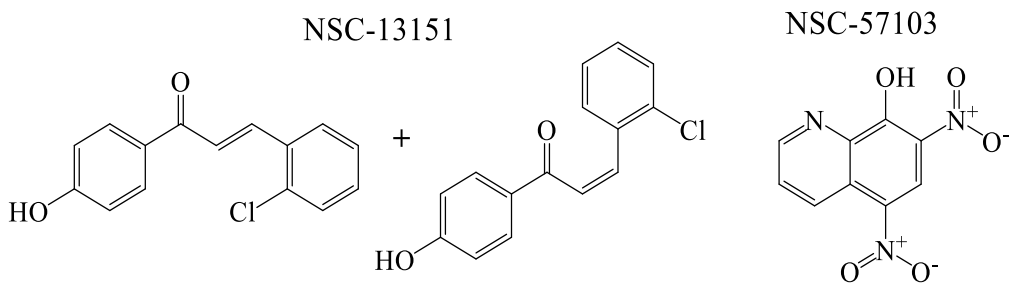


Figure 6-1: Chemical structures of NSC-13151 and NSC-57103. NSC-13151 was provided as a mixture of cis- and trans-isomers. Structures drawn in ChemDraw.

6.1.3 Consideration of Proteolysis Targeting Chimeras

As WWP1 and WWP2 are E3 ligases, the novel PROTAC technology provides an attractive option for future drug development, designed to selectively target a protein of choice for destruction. This approach was first suggested in 2001 and is similar to the 'Molecular glue' approach whereby two substrates are brought into close proximity (Sakamoto *et al.*, 2001; Che *et al.*, 2018). In the case of molecular glue's, a large ligand is utilised that interacts with both the substrates co-dependently. In contrast, PROTACs consist of two independent ligands joined via a flexible linker region. This hetero-bifunctional molecule is able to bind to a chosen E3 ligase and targeted protein, which upon recruitment using an appropriate linker size, enables the Ub-transfer to take place, resulting in downstream degradation of the targeted protein (Pei *et al.*, 2019). The PROTAC technology has many benefits being able to not only target previously 'undruggable' targets but further allow selective alteration of dysfunctional cellular signalling pathways by downregulation malignant signalling molecules, thus hopefully leading towards treatments with fewer side effects. Currently, only the RING E3 ligases, mainly Cereblon or von Hippel-Lindau (VHL), a receptor for Cullin, have been targeted by the PROTAC approach (Lee *et al.*, 2022). HECT-E3 ligases have to date not be used in the PROTAC field, possibly due to their increased size in comparison to RING-E3's, meaning the ligand position and linker is much more sensitive. In either case, they do provide an intriguing opportunity given their independent involvement in substrate ubiquitination, limiting possible bottlenecks involved in accessary protein recruitment in tissues with lower expression. Beyond hijacking the E3 ligase function, the PROTAC strategy can cause self-degradation as HomoPROTACs, or enable more novel ideas such as light-sensitive conformations switches within the flexible linker itself meaning future treatment could be light-activated (Hughes *et al.*, 2021). The application of this strategy has made small molecule E3 ligands almost as valuable as classical inhibitors.

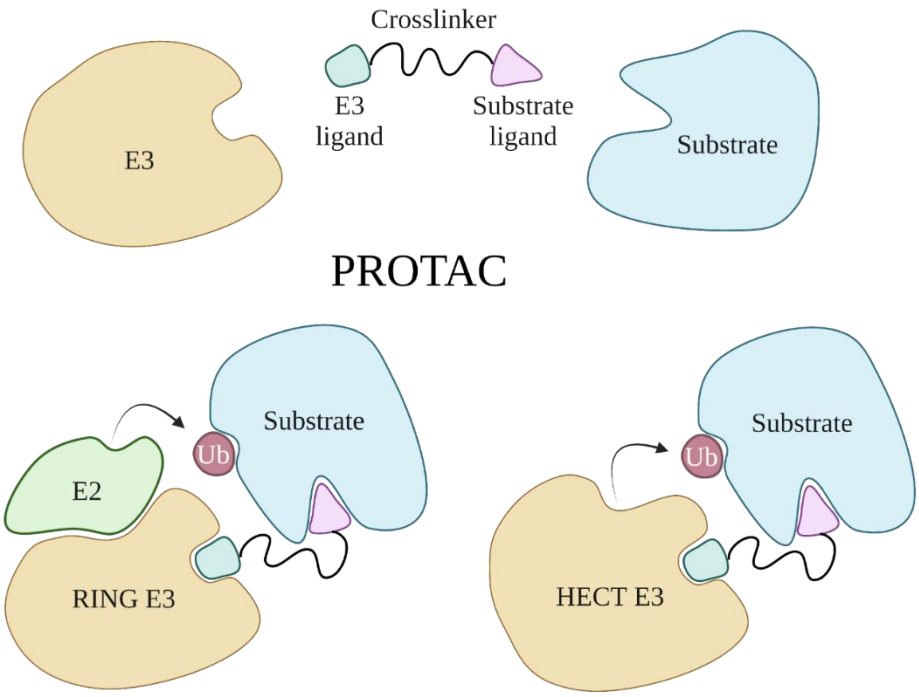


Figure 6-2: Schematic diagram of the small-molecule PROTAC approach. This drug-design approach utilises a heterobifunctional molecule consisting of three parts; an E3 ligand (blue square) and a substrate ligand (pink triangle) connected via a flexible crosslinker region (black line). This PROTAC can therefore mediate the interactions between the substrate (blue) and selected E3 ligase (orange). Once in proximity, the E3 ligase can either transfer the bound ubiquitin (red) onto the substrate directly (HECT E3) or indirectly through an E2 enzyme (green) (RING E3), building up sequential Ub chains resulting in downstream degradation. The image was created using BioRender.

6.1.3.1 Possible ligands are available for PROTAC development

The introduction of the DSF assay to the high-throughput screening strategy led to the added benefit of highlighting a variety of possible compounds that could provide promise as PROTAC ligands. After removing potential PAINS compounds and those showing inhibition, 16 and 14 compounds were found to have a significant ΔTm D shift against WWP1 and WWP2, respectively. The strongest interactions, that being compounds displaying a ΔTm D shift over 2 °C, would most likely be the best targets to take forward for characterisation. For WWP1, this included four compounds, NSC-30260, NSC-107582, NSC-332670 and NSC-522131, with only NSC-136513 displaying such a shift for WWP2 (Figure 6-3).

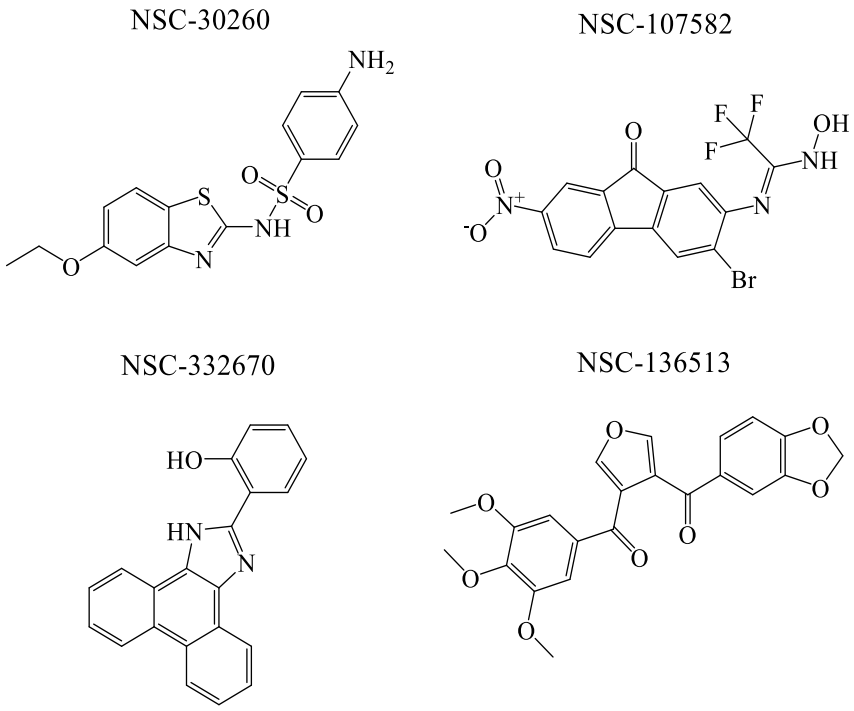


Figure 6-3: Chemical structures of NSC-30260, NSC-107582, NSC-332670 and NSC-136513. These compounds were found to give a ΔTm D over 2 °C against either WWP1 or WWP2 but showed no signs of inhibition. Structures drawn in ChemDraw.

Out of these compounds, only NSC-332670 has previously shown off-target activity, found to inhibit both foot-and-mouth disease virus (FMDV) 3C protease and Anopheles minimus Carboxypeptidase B in malaria (Mongkol *et al.*, 2015; Theerawatanasirikul *et al.*, 2023). A closer inspection into the distinct characteristics of the hit compounds identified several containing structural similarities to known NEDD4 family inhibitors. Interestingly, NSC-30260 contains an indole core scaffold, also observed in NSC-217913 and indole-3-carbinol, both known WWP1 interactors, thus further supporting its possible ligandability. Although not a substantial DSF hit (ΔTm ~ 1 °C), NSC-375981 exhibits structural similarities to the known NEDD4 family inhibitor Heclin, containing a benzene and furan ring, separated by a peptide bond (Mund *et al*., 2014). Previous work in the Chantry lab has already determined that Heclin does not inhibit WWP2 (Hughes, 2019). Despite this, some concern should be warranted regarding NSC-375981's furan ring, a moiety found to be metabolically unstable in Heclin and as such could limit its success (Chen, Gehringer and Lorenz, 2018). NSC-522131 is arguably the most interesting DSF hit, demonstrating an impressive ΔTm ~ 4 °C for WWP1 and 1.5 °C for WWP2. This compound also displays similar structural features to the SMURF1 inhibitor, HS-152 consisting of a urea moiety located between two electron-rich aromatic groups (Tian *et al.*, 2019). Not only have these structural features been noted in previous screening efforts of WWP2 but are observed in four additional DSF hits in our WWP2 screen (NSC-46492, NSC-

319994, NSC-320218 and NSC-329249). This common structural feature of the WWP2-screened compounds could suggest that a similar chemical space is being explored (Figure 6-4). Unfortunately, NSC-522131 was removed before the activity assays due to an apparent free radical quenching mechanism, a problematic feature when using a TMB/HRP reaction (Berggren *et al.*, 2019). Whether this compound could be applicable as a PROTAC ligand, a small molecule inhibitor or simply a PAINS compound is still unknown.

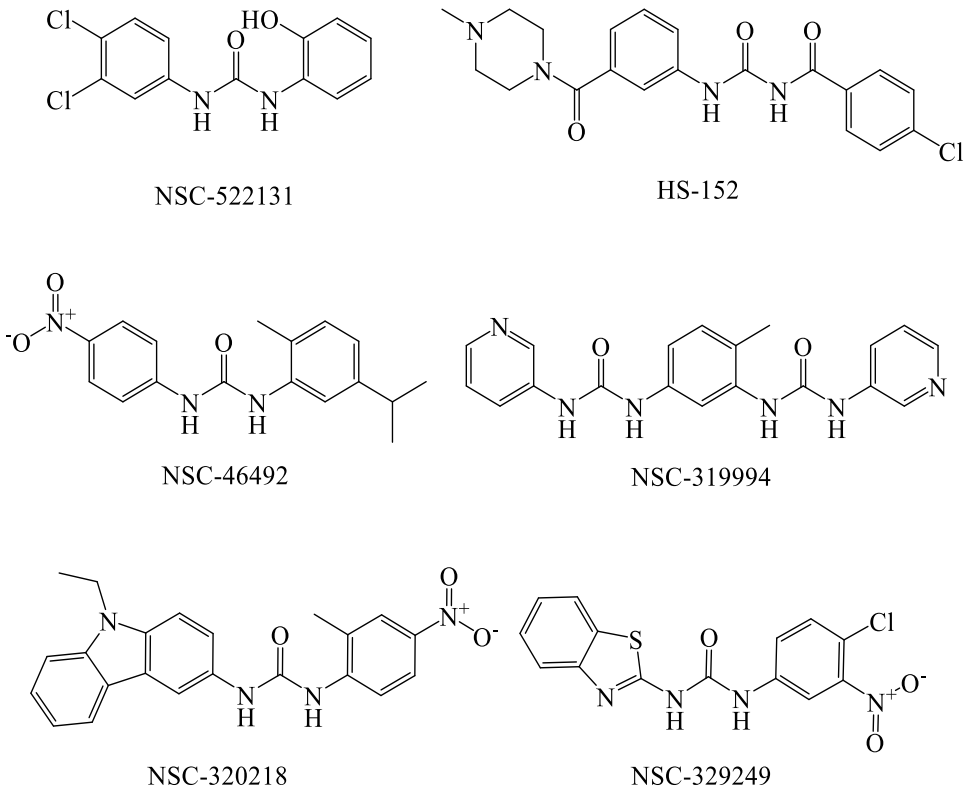


Figure 6-4: Structures of NSC-522131, HS-152 and other structurally similar DSF hits. These non-inhibitory DSF hit structures were found to be structurally similar to SMURF1 inhibitor HS-152, containing a urea moiety between two electron-rich aromatic groups. Structures are drawn in ChemDraw.

These potential ligands will require further validation using a gold-standard technique such as ITC. Such efforts were attempted as part of this project, however, optimisations failed, most likely associated with the choice of WWP2 construct. ITC is renowned for its high sample demands, which led to the relatively higher expressing WWP2-H construct being utilised. Previous work on crystalising WWP2-H demonstrated many issues with finding a stable construct (Gong *et al.*, 2015). Unfortunately, the final WWP2-H construct still had many purification issues despite significant efforts, including changing the affinity tag (see Section 4.2.4.2). In future, ITC experiments would benefit from using an alternative stable construct such as WWP2-LH despite its lower yield; This problem was somewhat overcome by the end of

the project with the final protocol given in Section 2.2.7.5 (~2.5 mg to 10 mg/L of culture). WWP1 may also be more fruitful with various stable constructs available, including the WWP1-H construct previously crystalised (Verdecia *et al.*, 2003).

6.2 The Hit-to-Lead Phase

At the Hit-to-Lead phase of the drug discovery pipeline, many hit compounds tend to fail due to unforeseen reasons. Sometimes these can be unaware PAINS issues as with NSC-2805, poor chemical tractability as with NSC-288387 or even chemically unstable moieties as with Heclin (Hughes, 2019). For this reason, the project was expanded beyond NSC-217913 SAR studies and introduced a smaller SAR of NEDD4-1 inhibitor I3C. Interestingly, although work had been completed to show that I3C interacts with WWP1, particularly at the Ub exosite through mutational studies, as well as seemingly having a WWP1-specific anti-proliferation role, no in vitro enzymatic inhibition had been recorded to date (Lee *et al.*, 2019). Both the SARs of NSC-217913 and I3C are summarised, supported by molecular docking.

6.2.1 In Vitro and Silico SAR of NSC-2171913

A total of 27 NSC-217913 analogues were screened using both the DSF and autoubiquitination assays. A detailed breakdown of the SAR is already discussed in Section 4.3, further supported through in silico molecular modelling in Section 5.3. However, in summary, the SAR studies identified the imidazo[4,5-b]pyrazine heterocyclic core of NSC-217913 as an important scaffold, docked consistently to undergo vital π-π interactions with Tyr636/587 aromatic side chains and hydrogen bonding interactions with the backbone of Cys640/588, in a large cavity close to active site cysteine residues of WWP1/ WWP2 (Figure 5-15 and Figure 5-16). Previous in silico docking with NSC-288387 has also highlighted this pocket and the WWP2 Tyr587 π-π stacking interaction, correlating well with STD-NMR data (Hughes, 2019). Functionalisation of the heterocyclic scaffold was shown to favour dichlorination, with hydrophilic interactions seemingly being preferred on linear thioether groups deeper within the binding site, as found with the most potent S11 analogue with an IC_{50} of 32.74 µM and 269.2 µM, for WWP1 and WWP2 respectively (Figure 6-5). Depending on the thioether identity, these deeper interactions potentially occur with Arg855/803, Met865/Ile813 or Leu641 (in WWP1) residues. In an effort to improve potency and selectivity between WWP1 and WWP2, future work should focus on the ethyl butyrate side chain and imidazo[4,5-b]pyrazine core looking to alter its functionalisation including the type of halogen present.

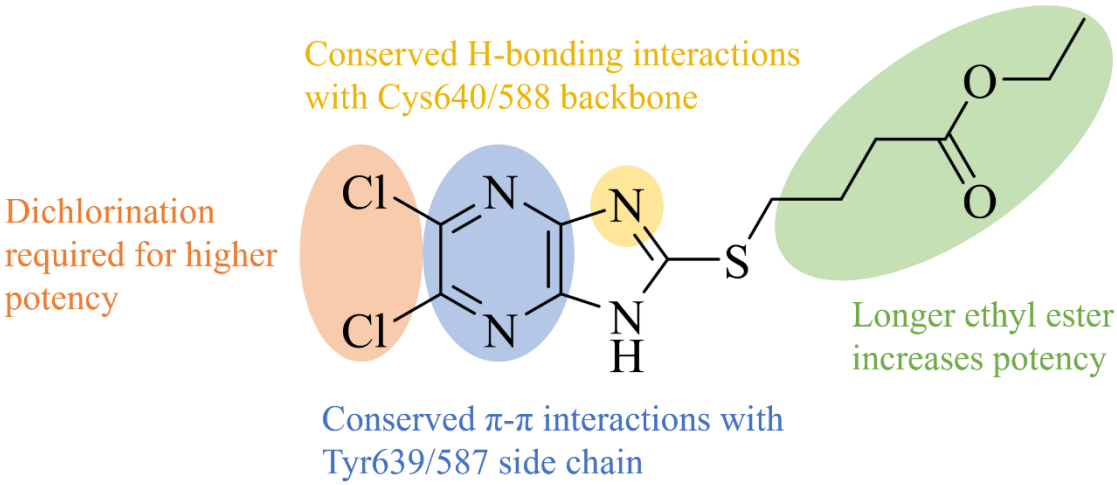


Figure 6-5: SAR summary of NSC-217913 analogue S11. The structure of the most potent NSC-217913 analogue with an IC_{50} of 32.74 and 269.2 µM, for WWP1 and WWP2 respectively. Dichlorination (orange) was found to be important to maintain potency, with the length and hydrophilicity of the thioether chain (green) shown to increase it. Molecular modelling to a cavity close to the active site cysteine highlighted a conserved π-π stacking interaction with Tyr636/587 of the aromatic core, and the Nitrogen undergoing hydrogen bonding with Cys640/588 backbone. Structure drawn in ChemDraw.

6.2.1.1 Substrate and toxicity assays should be considered

The use of an autoubiquitination assay to measure E3 activity is relatively standard, with many commercial assays available. However, there is a chance that the inhibition observed is specific to the autoubiquitination process and not the overall enzymatic activity. At this stage, other assays should be introduced to further validate the NSC-217913 inhibitor scaffold. A western blot assay may provide a good starting point as another common assay used to compare E3-directed substrate ubiquitination and has already proven suitable when looking at small molecule modulation of both WWP1 and WWP2-mediated PTEN ubiquitination (Watt *et al*., 2018; Lee *et al*., 2019). Following the drug discovery pipeline, in vivo cytotoxicity assays will also be required to validate the potential of the NSC-217913 scaffold for future therapeutic application, ensuring its presence at the concentrations required for inhibition is not fatal to healthy cells (Hughes *et al*., 2011).

6.2.2 In Vitro and Silico SAR of I3C and Stable Derivatives

A total of 21 I3C derivatives were screened using both DSF and autoubiquitination assays against WWP1 and WWP2. This SAR by catalogue approach was not successful, however, it did demonstrate an increased potency using the I3C acid-condensation product DIM as well as the stable I3C-derivative N-Tosyl-I3C, as discussed in Section 4.3. To summarise, although I3C

was validated to interact with WWP1 using a DSF assay and STD-NMR epitope mapping, the in vitro I3C inhibition recorded was too weak to measure an IC_{50} curve accurately. Conversely, DIM was significantly more potent, recording an IC_{50} of 111.2 µM against WWP1 and was shown to interact with both WWP1 and WWP2 using DSF (Table 4-4). Molecular modelling further suggested DIM had stronger interactions with the Ub exosite, with a higher GlideScore (-7.09 and -7.79 kcal/mol) compared to I3C (-6.28 and -6.45 kcal/mol), shown to maintain interactions with the hydrophobic pocket found vital for I3C interaction (Lee *et al*., 2019) (Figure 5-17 and Figure 5-18). DIM is formed as a result of I3C's acid instability as well as other minor acid-condensation products (Bradlow and Zeligs, 2010; Grose and Bjeldanes, 1992). As such the I3C anti-proliferation properties associated with WWP1 may in part be due to I3C's in vivo conversion to the more potent DIM. This may also answer why the I3C SAR failed, given the relatively similar I3C derivatives used. Future SARs may benefit from using DIM as the starting scaffold, with various DIM derivatives already having demonstrated improved anti-proliferation properties (Omar *et al*., 2009; Vanderlaag *et al*., 2006). In contrast, the acid-stable I3C derivative N-Tosyl-I3C/Compound I15 does provide some promise towards using I3C as a scaffold, intriguingly shown to inhibit both WWP1 and WWP2 at a similar potency with an IC_{50} of 218.3 and 223.7 µM, respectively. Based on the highly potent OSU-A9, this non-selectivity correlates well with OSU-19's wide response profile shown to interfere with many anti-proliferative pathways (Weng *et al*., 2007).

6.2.2.1 Detergent sensitivity assays are required

The N-Tosyl-I3C derivative alongside its tri-fluorinated analogue compound I16 demonstrated an unusual DSF trace, with a high fluorescence observed at low temperatures (~30 °C) and therefore no Tm could be accurately measured (Appendix Figure 7-3). Molecular modelling of N-Tosyl-I3C into the Ub exosite of WWP1 and WWP2 showed significant interactions with both the exosite and WW2 domains, in particular undergoing π-π stacking with the WW2 domain's Tyr398/347 and Trp409/358 (Figure 5-17 and Figure 5-18). However, given that this is a simulated model, there is a possibility that these derivatives interfere with the WW2 interactions at the Ub exosite. This could explain both the inhibition observed (preventing Ub binding) as well as the possible destabilisation effects on the T-shape conformation (elevating the Headband conformation) as seen in the WWP1-2L34H and WWP2-LH constructs used in the DSF assays. Other less elaborate explanations could be associated with compound aggregation, a feature unfortunately not taken into account in the assay design. Future studies should look to include a detergent sensitivity assay to rule out possible assay interference of both the NSC-217913 and I3C scaffolds (Jadhav *et al*., 2010).

6.2.3 The Structural Studies of Small Molecule Binding Modes

A structure-informed approach to small molecule SAR is a vital step in the drug discovery pipeline. Throughout the project, significant effort was placed on discovering the binding mode of the parent NSC-217913 and I3C compounds to WWP1 and WWP2 using X-ray crystallographic studies, discussed in Section 5.3.

6.2.3.1 X-ray crystallographic studies were conducted too early

Despite solving both apo WWP1-2L34H and WWP2-LH structures, efforts to solve ligand-bound structures were unsuccessful. Unfortunately, the crystallisation of WWP1 was achieved too late into the project to be useful. Ligand soaking studies of WWP2-LH were exhausted with both I3C and NSC-217913, as well as a handful of analogues. A smaller subset of co-crystallisation studies was also attempted to no avail. Unfortunately, a significant portion of the work was spent on this, and in retrospect was wasted labour given many of these compounds were later confirmed to show minimal, if no interaction with WWP2. This highlighted a problem with the strategy used, whereby X-ray structural studies were attempted during the screening and SAR stage, mainly due to the time required for crystallisation, meaning other experiments could be used between the plating and harvesting period. A more linear approach would most likely have shown improved results, by either fully characterising the parent compounds and their interactions with WWP1 and WWP2 or completing the small SAR and utilising the most potent compounds such as S11 for ligand soaking. In this regard, future X-ray crystallographic studies should focus on WWP1 having demonstrated increased potency towards I3C and NSC-217913 scaffolds. In the case of the WWP1-2L34H structure, co-crystallisation studies or at least rescreening for a better crystallisation condition may be beneficial, with harvesting and cryoprotecting found to be difficult when using a reagent alcohol precipitant. Other success may come from using the WWP1-HECT construct, having been successfully crystalised elsewhere (Verdecia *et al*., 2003). Although the WWP2-LH crystals failed to show a ligand-bound state, most likely as the inhibitor interactions were too weak, these crystals may benefit from the XChem program at the Diamond Light Source. A main requirement of this program is that the target protein forms a large number (~50-100) of high-quality crystals per drop, a parameter reached for WWP2-LH due to the various crystal optimisation efforts completed as part of this project (Collins *et al*., 2017).

6.2.3.2 Mutational studies of the proposed binding sites for validation

Beyond X-ray crystallisation, molecular docking was used as an alternative method to gauge possible binding poses. As the I3C binding location has already been confirmed, the resulting

poses in this site were considered with some confidence. However, in the case of NSC-217913, a more ambiguous docking location was chosen at a cavity previously validated with NSC-288387, close to the active site cysteine (Hughes, 2019). Although the docked NSC-217913 analogue did show a conserved pose, they may indeed bind elsewhere such as the Ub exosite. Following previous studies of I3C WWP1 interactions, future studies should look to include alanine site mutations in both the Ub exosite site (F577/525 and Y656/604 residues) and the potential binding sites in the active site cavity (Tyr 639/587, Arg855/803 and M865/Ile813 residues) to indicate the WWP1 and WWP2 binding sites highlighted through a change in NSC-217913 potency, as long as in vitro ligase activity is unaffected by the mutation (Lee *et al*., 2019).

6.2.3.3 Is the similarity of E3 ligase HECT domains a cause for concern?

Despite significant similarities between E3 ligase HECT domains, there seem to be opportunities for selective targeting, particularly within the NEDD4 family. Even within this study, there was a clear selection for WWP1 over WWP2, despite these enzymes sharing a 70% sequence identity. For example, the suggested active site cavity for NSC-217913, as well as its more potent analogues, pointed to residue mutations such as M865/Ile813 that resulted in distinctive binding modes relating to the change in compound activity. This contrasts with Heclin having broad activity towards many HECT E3 ligases, also known to target the active site (Mund *et al*., 2014). However, it is important to highlight that this small molecule does act through a unique mechanism, exposing the active site cysteine causing its oxidation rather than blocking the active site or reacting directly with the cysteine and as such is not representative of selectivity issues for HECT ligase inhibitors. Beyond the active site, the Ub exosite also displays various residue changes between NEDD4 and WWP1/2 that can clearly be targeted independently, at least when considering 1-Benzyl-I3C, having an IC_{50} of 12.3 µM against NEDD4-1, yet displaying no inhibition towards WWP1 or WWP2 (Quirit *et al*., 2017). The non-catalytic cysteine in the Ub exosite has also been targeted by covalent inhibitors, for example, compound 1 and compound 3 as shown by Kathman and colleagues (2015). This study counter-screened these NEDD4 inhibitors against a variety of HECT-containing E3 ligases, including E6AP, WWP1 and NEDD4-L as well as other enzymes with active site cysteines, using whole protein electrospray ionization (ESI)-MS to identify interaction. The most potent inhibitor, compound 3, was found to only interact with NEDD4 and its homolog NEDD4-L due to the high similarity of their Ub exosite and positioning of the non-catalytic cysteine. Arguably, this is further a positive indication of HECT selectivity, as selecting between direct homologues has always been a challenge in small molecule drug discovery and is not specific to the HECT domain (Tanramluk *et al*., 2009).

6.3 Future Considerations

6.3.1 Are Covalent HECT Inhibitors a Better Strategy?

Covalent inhibitors offer several advantages over traditional non-covalent inhibitors, including extended duration of action, enhanced ligand efficiency, and reduced risk of drug resistance when targeting residues essential for enzyme catalysis (Sutanto, Konstantinidou and Dömling, 2020). Various residues can be targeted by covalent reactions, including cysteine, lysine, glutamate, serine, and threonine (Powers *et al*., 2002; Martín-Gago *et al*., 2017; Quach *et al*., 2021). However, both cysteine and lysine residues are the most prevalent targets, as shown in the case of protein kinases, with both residues shown to have vital roles in the ATP active sites. For example, the natural fungal product Hypothemycin has been shown to target Cys166 of Mitogen-activated protein kinase 1 (MAPK), alongside other kinases containing the Cys-Asp-Xaa-Gly motif (Ohori *et al*., 2007). More importantly, irreversible lysine inhibitors have had success in clinical trials, with the anticancer drug ibrutinib used to target Bruton's tyrosine kinase (BTK) for the treatment of chronic lymphocytic leukaemia (Tucker and Rule, 2015). To no surprise, the use of electrophilic compounds has been a common strategy in the development of HECT E3 ligase inhibitors to directly target the active site cysteine and was the original strategy employed by Kathman and colleagues (2015) before discovering the non-catalytic Ub exosite cysteine, based on an irreversible covalent fragment screen they developed for cysteine proteases (Kathman, Xu and Statsyuk, 2014). Although showing successful selectivity, only targeting NEDD4-1 and its close homolog NEDD4-2, this study has only been used in vitro with a small sample of cysteine-containing enzymes and thus still risks future off-targeting, a significant challenge for covalent inhibitors.

More recent ligand discovery strategies have taken advantage of this ambiguity, undertaking screening of electrophile 'reporter molecules' to search for ligandable targets in their native biological setting; advantageous when purified proteins are not easily obtained. Such activity-based protein profiling (ABPP) approaches measure the binding of a small molecular probe containing a reactive group, reporter tag and binding group to interact with proteins in vivo before using untargeted MS-based proteomics for detection (Niphakis and Cravatt, 2024). This strategy is known to give high-content but low throughput, problematic as a drug-discovery approach. In this regard, limited screening libraries have been generated to contain molecules with preferred reactivities for types of proteins or residue environments, such as cysteine-reactive chemotypes for targeting deubiquitinates (Conole *et al.*, 2023). Other more recent studies have looked to go one step further in screening against preferred reactivity centres that are also stereoselective. Diversity-oriented synthesis (DOS)-based photoreactive stereoprobes are one such strategy that has recently been added to this toolbox, described as

stereochemically defined compounds bearing diazirine and alkyne units for UV light-induced covalent modification (Ogasawara *et al.*, 2024). This study further demonstrated these probes could be used in a high throughput nano bioluminescence resonance energy transfer (BRET) assay through competitive profiling, i.e. using the probe to bind competitively with a screened non-covalent ligand/inhibitor at a target active site.

Here in this study, the risk vs reward for a covalent inhibition strategy for HECT E3 ligases was deemed too high, despite the concept being proven by the Kathman study. However, it may be worth consideration in the future, or adapted through a similar strategy to the nanoBRET competitive profiling approach, using a covalent probe to compete with possible targets at the active site or the Ub exosite.

6.3.2 Future Therapeutic Applications

With the results of this study pointing towards WWP1 selectivity, associated diseases would be the most promising targets for future application of DIM and NSC-217913-based inhibitors. In particular, hepatocellular carcinoma, otherwise known as liver cancer, is an ideal disease model given the poor survival rate of patients when WWP1 is highly expressed (X. F. Zhang *et al*., 2015). Various cell lines such as HepG2 are also available for in vitro 2D and 3D modelling and would enable WWP1 inhibition to be analysed in a disease-context environment (Blidisel *et al*., 2021). Future work could also consider if WWP1 mutations such as those found in patients demonstrating PTEN-mutation-free Cowden-like syndrome (R86H, K740N, and N745S) do not affect the action of these small molecule scaffolds (Lee *et al*., 2020). Of course, this could be achieved in vitro, with a similar activity-based assay as seen in this study using the mutate WWP1 and would most likely enable the use of full-length enzyme given these mutations to interfere with autoinhibition. After such as study, it could be informative to also use an in vivo model, first to enable these mutations to cause a disease-like state, before treating with the DIM and NSC-217913 scaffolds and assessing the downregulation of PTEN using western blotting. In either case, the inhibitor scaffolds highlighted in this study are still in the early stages and will require more hit-to-lead development before such action can be taken forward.

Appendix

The data in this appendix has been provided to support the thesis. The author has not carried out this work. Only the final synthesis step and characterisation of the hit-compound analogues is provided with all other analogues and relevant synthesis strategies found in the Thesis of Jake Rigby (2024).

7.1 STD-NMR of I3C against WWP1 and WWP2

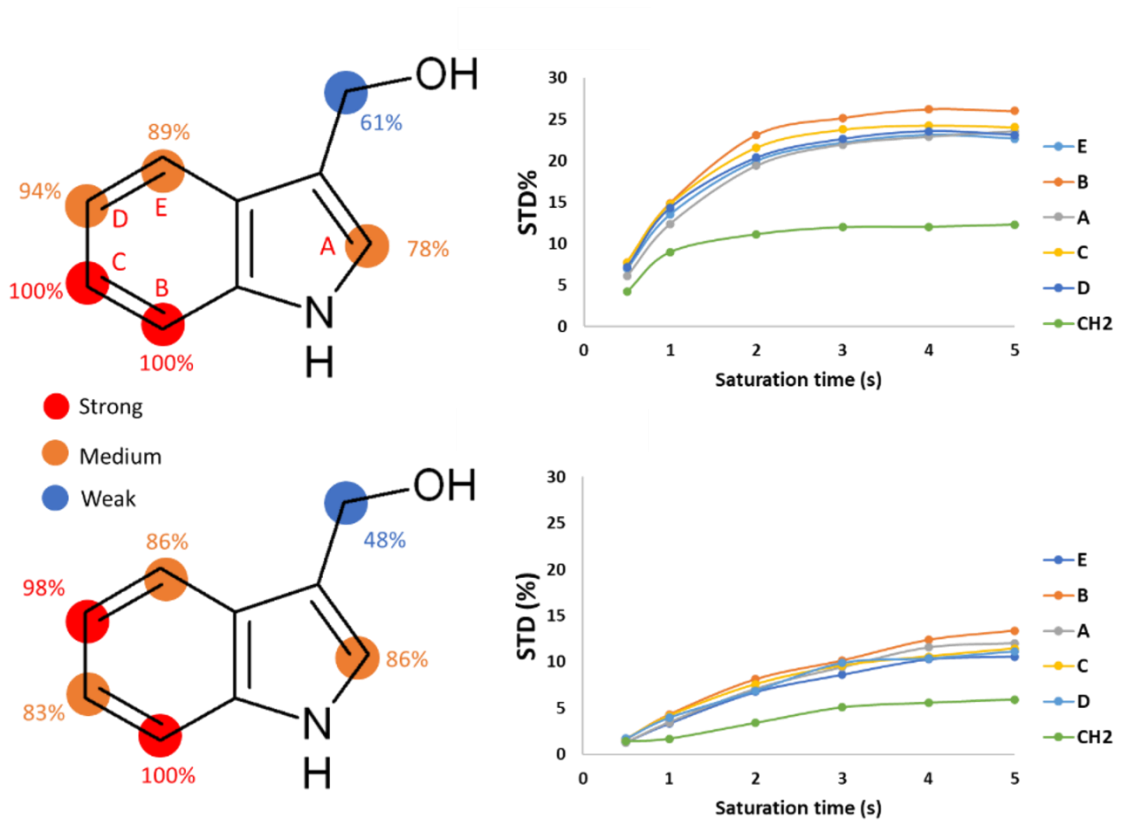


Figure 7-1. STD NMR binding epitope maps of I3C against WWP1 (top) and WWP2 (bottom) based on the normalized saturation transfer intensities (0-100%). Binding epitope maps were obtained from the initial slope of the relative build-up curves for each proton (right), by selective protein irradiation at 0.0 ppm. Legend indicates weak (blue), medium (orange) and strong (red) intensities. Raw STD spectra are reported in Figure 7-2.

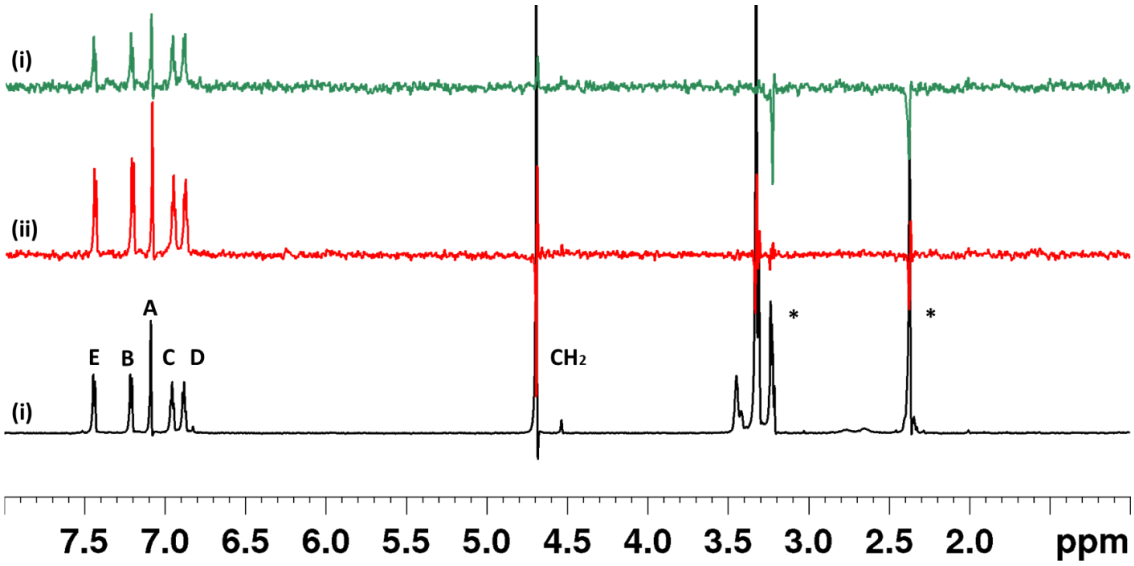


Figure 7-2. STD NMR binding experiments for Indole-3-carbinol (I3C) in complex with WWP1 and WWP2. Data was obtained at 2 s saturation time, with selective protein irradiation at 0.0 ppm. (i) Reference spectrum for the WWP1-H I3C complex. (ii) STD difference spectrum for the I3C WWP1-H complex. (iii) STD difference spectrum for the WWP1-H I3C complex. The ligand assignment is given as labels on the reference spectrum, while the residual buffer and d^6-DMSO are labelled with a *.

7.2 DSF Trace of N-Tosyl-I3C against WWP2

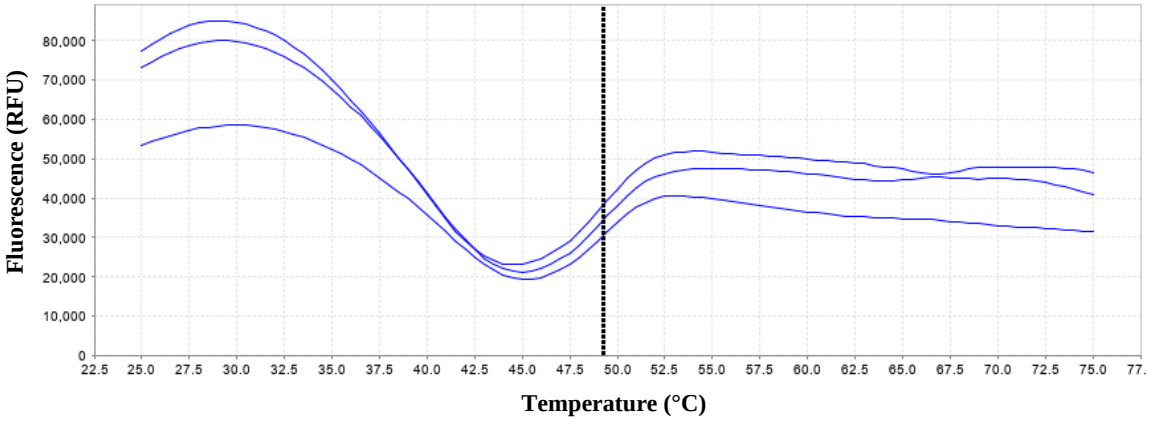


Figure 7-3. DSF trace of N-Tosyl-I3C against WWP2-LH. The assay was carried out using 100 µM of the compound, containing 0.1% DMSO, with all other parameters described in Section 2.3.2.1. Graph taken directly from the Protein Thermal Shift Software v1.4.

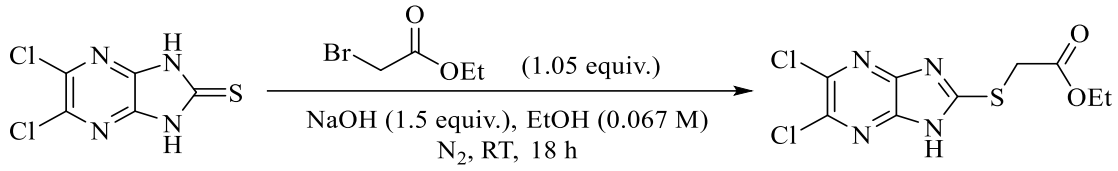
7.3 Chemical Characterisation

Unless specified, all reagents and starting materials were purchased from commercial sources (Sigma-Aldrich (Merck Life Sciences), Fluorochem (Doug Discovery), Fischer Scientific, Alfa Aesar) and used as supplied. Thin-layer chromatography was performed on Merck silica gel 60 F254 plates and visualised by UV absorption, purchased from VWR International. Flash column chromatography was carried out using Silica Gel 60 purchased from Material Harvest. 'Concentrated' refers to the removal of volatile organic solvents *via* distillation using a rotary evaporator. 'Dried' refers to pouring onto or adding anhydrous $MgSO_4$ or Na_2SO_4 to (as specified), followed by filtration. Water refers to deionised water.

NMR spectra were recorded on 400 or 500 MHz Bruker NMR spectrometer using the deuterated solvent stated in the reported data. ^{1}H, ^{13}C and ^{19}F NMR samples were prepared by dissolving a sample in 0.4 mL – 0.7 mL deuterated solvent. All deuterated solvents were purchased from Cambridge Isotopes and used as received, solvents were stored under 4 Å molecular sieves after opening. All spectra were referenced to the residual solvent peaks of the solvent used. NMR spectra chemical shifts (δ) are reported in ppm and coupling constants (J) are reported in hertz (Hz). Abbreviations for NMR splitting are s (singlet), d (doublet), t (triplet), q (quartet), and m (multiplet). Infrared spectra were recorded using a Perkin Elmer Spectrum Two LITA. High-resolution mass spectrometry was performed at the University of East Anglia using a UPLC-HRMS (ACQUITY H-Class PLUS UPLC and Waters SYNAPT XS High-Resolution Mass Spectrometer) setup with electrospray ionisation using ca. 1 μgmL^{-1} solution in acetonitrile or methanol. Melting points (not corrected) were recorded on a Büchi Melting Point B-545 using capillary melting point tubes made in-house.

7.3.1 NSC-217913 Hit Analogues

7.3.1.1 ***Compound S6/NSC-217913*** *– Ethyl 2-({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)-acetate*



To an 8 mL vial was added 5,6-dichloroimidazo[4,5-*b*]pyrazine-2(1,3*H*)-thione (100 mg, 0.452 mmol) and sodium hydroxide (28 mg, 1.5 equiv.). The vial was then sealed with a suba seal and wrapped with parafilm. Ethanol (6.75 mL, 0.067 M) was added *via* syringe and the mixture stirred until all solids had dissolved. Ethyl bromoacetate (53 µL, 1.05 equiv.) was added *via* microsyringe and the reaction was left to stir for 18 h. The solvent was removed under reduced pressure and the solid redissolved in NH_4Cl solution and EtOAc (*ca.* 5 mL each) and the mixture transferred to a separatory funnel. The organic layer was removed, and the aqueous layer extracted with EtOAc (3x 10 mL). The organic layers were collected and dried ($MgSO_4$) and solvent removed *in vacuo*. The crude material was purified by column chromatography eluting with Pet. E. : EtOAc (7:3) followed by triturating with cold *n*-hexane (*ca.* 4 mL) and recrystallising from ethanol to provide a white solid of ethyl 2-({5,6-dichloro-1*H*-imidazo[4,5-*b*]pyrazine-2-yl}sulfanyl)-acetate (106 mg, 0.344 mmol, 76%).

^{1}H NMR (400 MHz, C*D*$_3$CN) δ 4.19 (q, *J* = 7.1 Hz, 2H), 4.17 (s, 2H), 1.24 (t, *J* = 7.1 Hz, 3H).

^{13}C NMR (126 MHz, C*D*$_3$CN) δ 168.9, 160.7, 139.5, 62.9, 34.3, 14.4, (6 out of 7 carbon resonances observed).

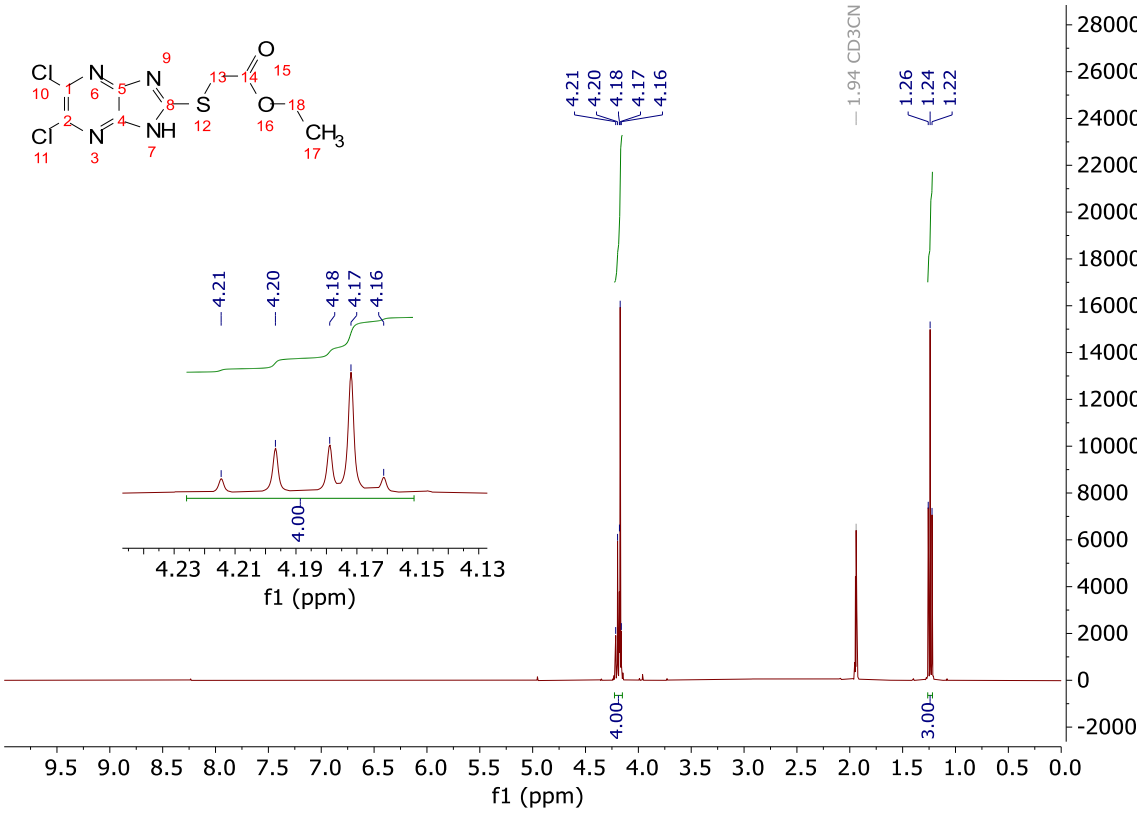
IR (cm^{-1}) 3084 (NH), 1731 (C=O), 1595 (NH bend).

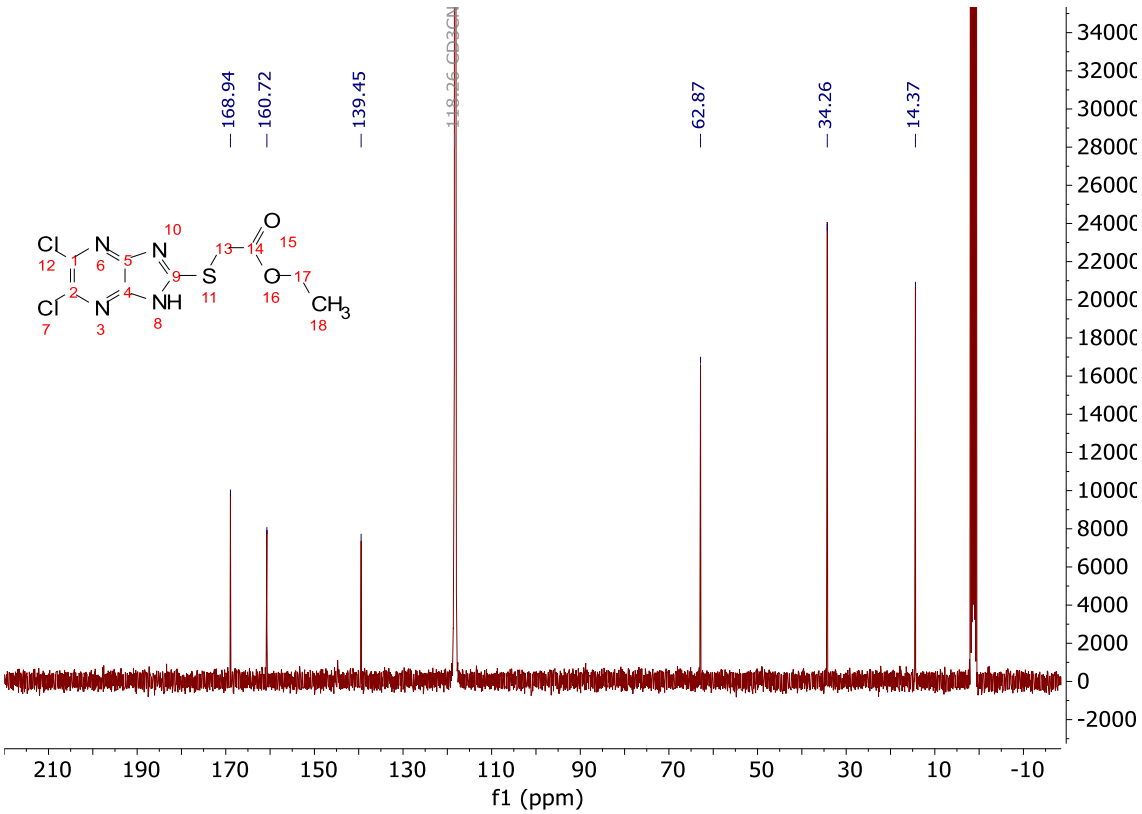
M.P. 167.6 – 168.1 ºC.

MS ES+ m/z calcd for $C_9H_8Cl_2N_4O_2S$ $(M+H)^+$ 308.9794, found: 308.9792.

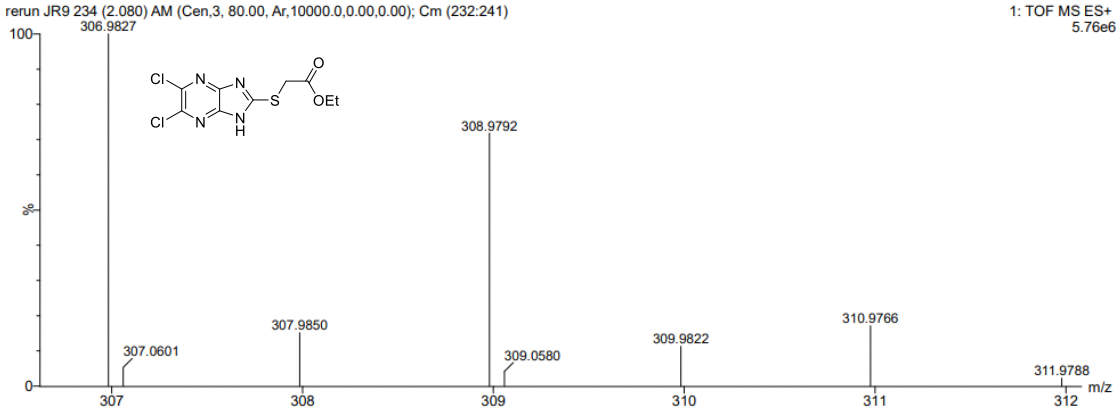
Adapted from the literature procedure (Tong, 1981).

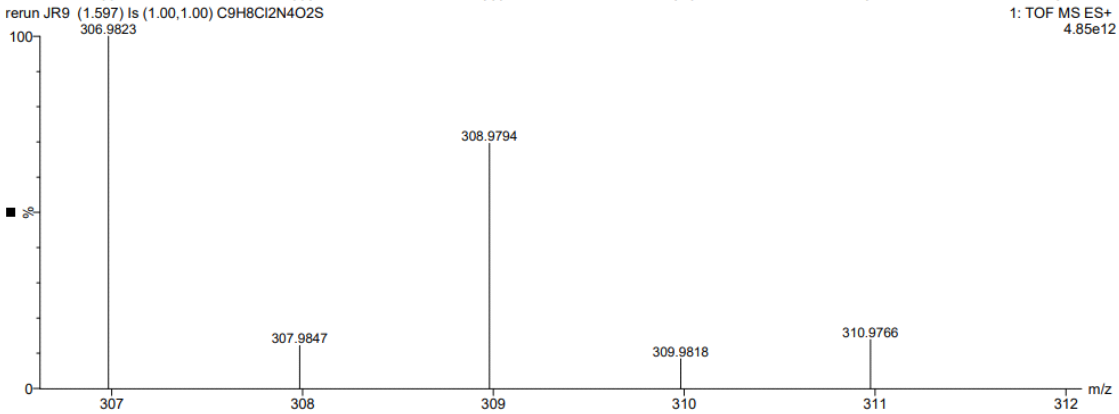
NMR Characterisation:



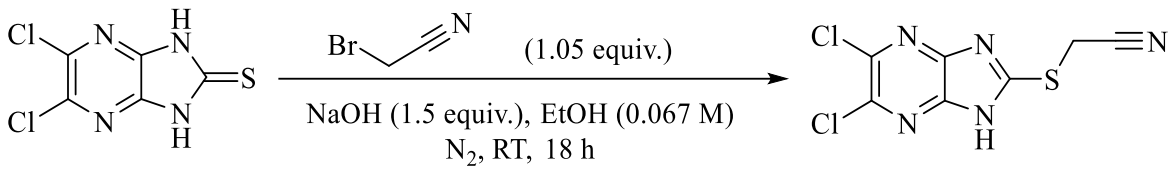


HRMS Characterisation:





7.3.1.2 ***Compound S7*** *– 2-({5,6-Dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)acetonitrile*



To an 8 mL vial was added 5,6-dichloroimidazo[4,5-*b*]pyrazine-2(1,3*H*)-thione (100 mg, 0.452 mmol) and sodium hydroxide (28 mg, 1.5 equiv.). The vial was then sealed with a suba seal and wrapped with parafilm. Ethanol (6.75 mL, 0.067 M) was added *via* syringe and the mixture stirred until all solids had dissolved. Bromoacetonitrile (33 µL, 1.05 equiv.) was added *via* microsyringe, and the reaction was left to stir for 18 h. The solvent was removed under reduced pressure and the solid redissolved in ammonium chloride solution and ethanol (*ca.* 5 mL each) and the mixture transferred to a separatory funnel. Separation of the single-phase mixture was achieved by adding a small amount of brine solution. The organic layer was removed, and the aqueous layer extracted with EtOAc (3x 10 mL). The organic layers were collected and dried ($MgSO_4$) and solvent removed *in vacuo.* The crude material was purified by column chromatography eluting with Pet. E. : EtOAc (7:3) followed by triturating with cold *n*-hexane (*ca.* 4 mL) and recrystallising from ethanol to provide a light-yellow solid of 2-({5,6-dichloro-1*H*-imidazo[4,5-*b*]pyrazine-2-yl}sulfanyl)acetonitrile (63.3 mg, 0.243 mmol, 54%).

^{1}H NMR (400 MHz, C*D*$_3$CN) δ 4.25 (s, 2H).

^{13}C NMR (101 MHz, C*D*$_3$CN) δ 158.3, 140.0, 117.5, 18.0 (4 out of 5 carbon resonances observed).

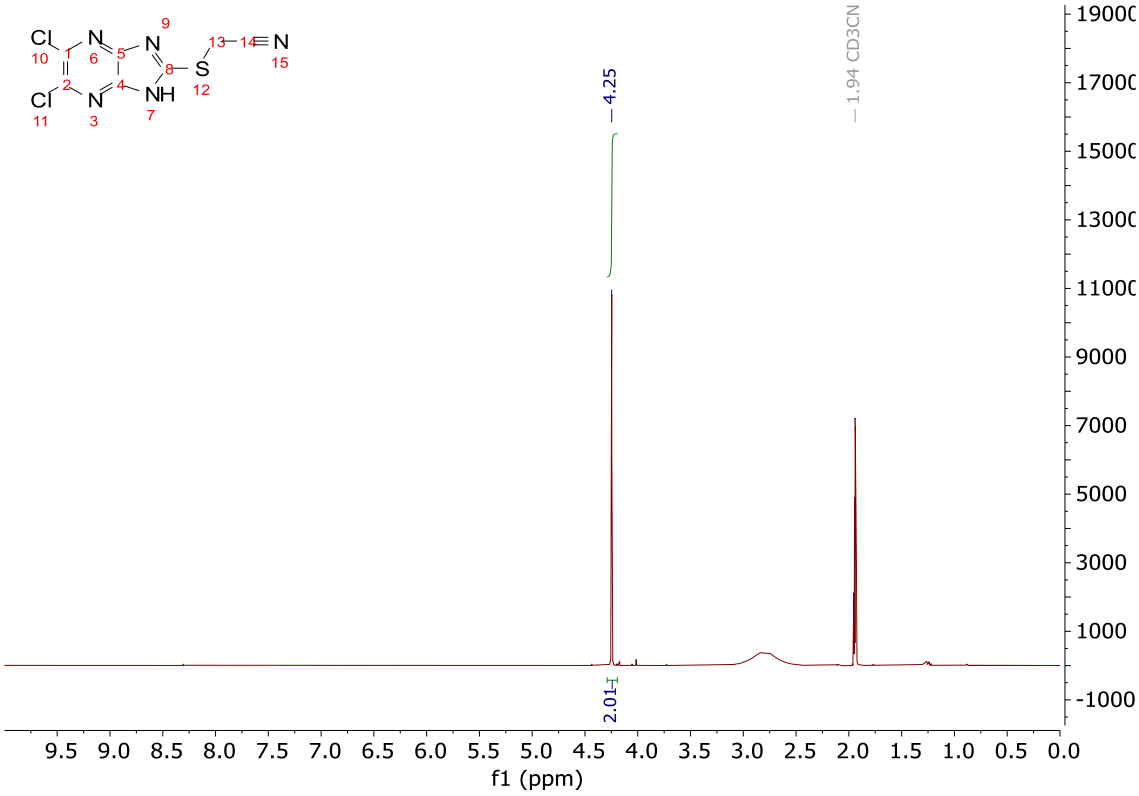
IR (cm^{-1}) 3100 (NH), 2248 (C≡N).

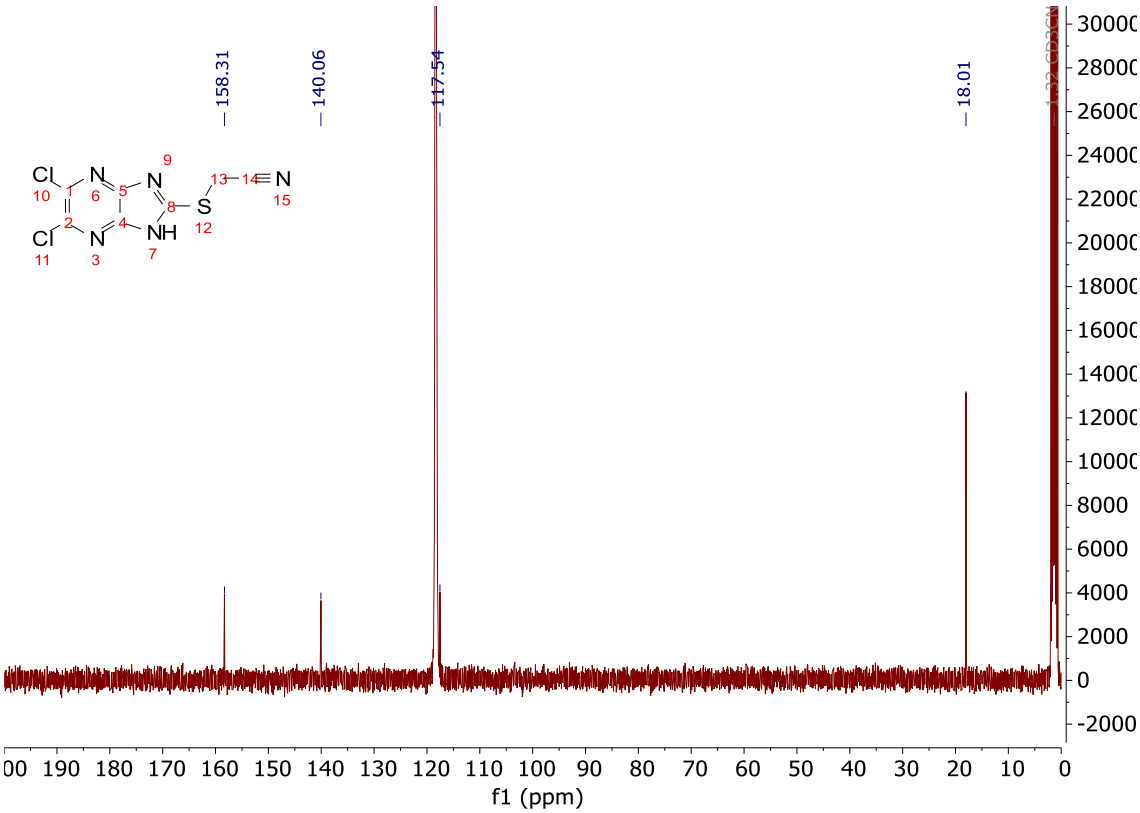
M.P. 213.5 – 216 ºC (deg.).

MS ES+ m/z calcd for $C_7H_3Cl_2N_5S$ $(M+H)^+$: 260.9584, found: 260.9583.

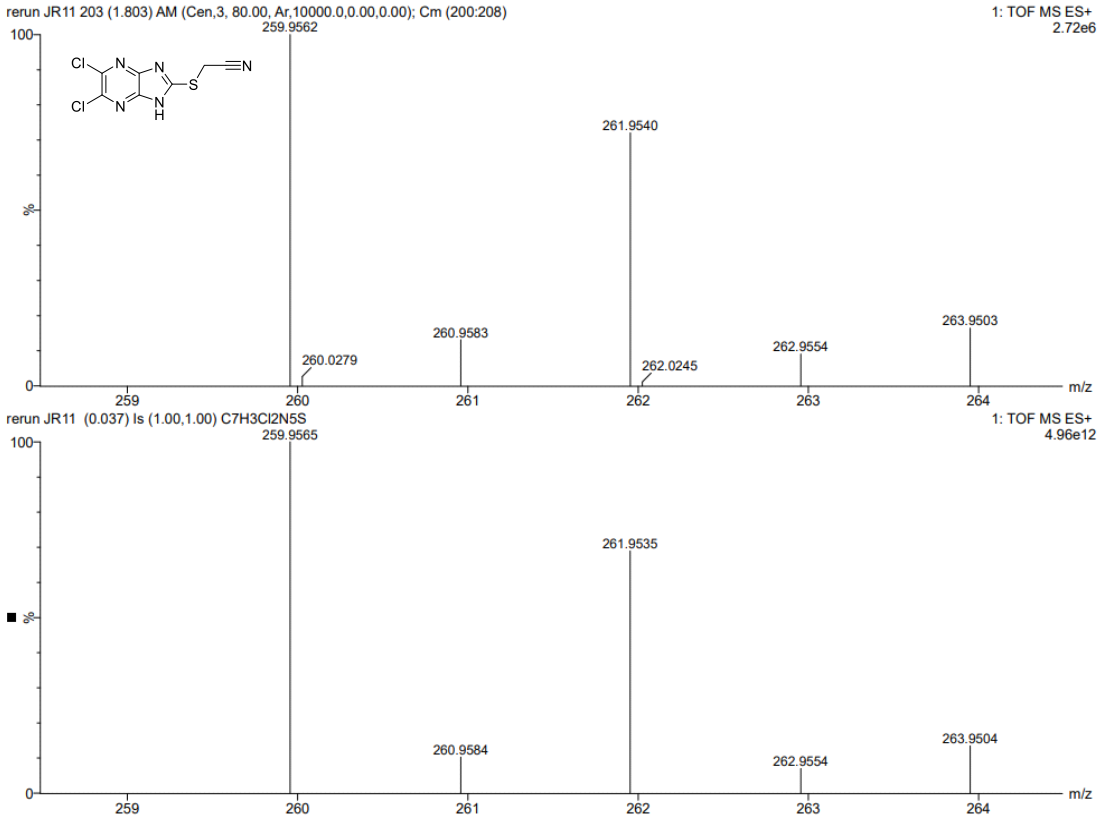
Adapted from the literature procedure (Tong, 1981).

NMR Characterisation:

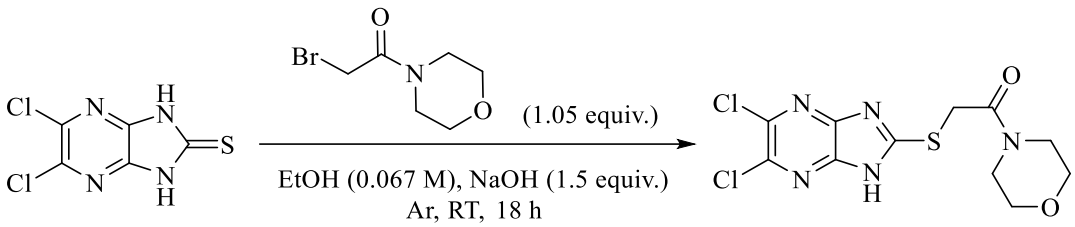




HRMS Characterisation:



7.3.1.3 ***Compound S9*** *– 2-({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)acetylmorpholine*



To an 8 mL vial was added a stirrer bar, 5,6-dichloroimidazo[4,5-*b*]pyrazine-2(1,3*H*)-thione (75 mg, 0.34 mmol), sodium hydroxide (20.3 mg, 1.5 equiv.). Ethanol (over mol. sieves, 5.0 mL) was added *via* syringe. *N*-(bromoacetyl)morpholine (74.1 mg, 1.05 equiv.) was added in one batch and the reaction placed under argon atmosphere. Left to stir at RT for 18 h. Afterwards, the solvent is removed *in vacuo.* and the residue redissolved in sat. soln. NH_4Cl (10 mL), and EtOAc (10 mL), transferred to a separatory funnel and the organic layer removed. Aqueous layer extracted with EtOAc (3x 10 mL), organic layers collected and dried ($MgSO_4$) and solvent removed *in vacuo.* The residue purified by column chromatography (6:4 Pet. E : EtOAc -> EtOAc) to provide an off-white solid of 2-({5,6-dichloro-1*H*-imidazo[4,5-*b*]pyrazine-2-yl}sulfanyl)acetylmorpholine (53.4 mg, 0.153 mmol, 45%).

^{1}H NMR (400 MHz, C*D*$_3$CN) δ 4.37 (s, 2H), 3.73 – 3.66 (m, 2H), 3.65 – 3.61 (m, 2H), 3.61 – 3.54 (m, 4H).

^{13}C NMR (101 MHz, C*D*$_3$CN) δ 166.8, 161.7, 139.3, 67.1, 67.1, 47.3, 43.4, 35.6 (8 out of 9 carbon resonances observed).

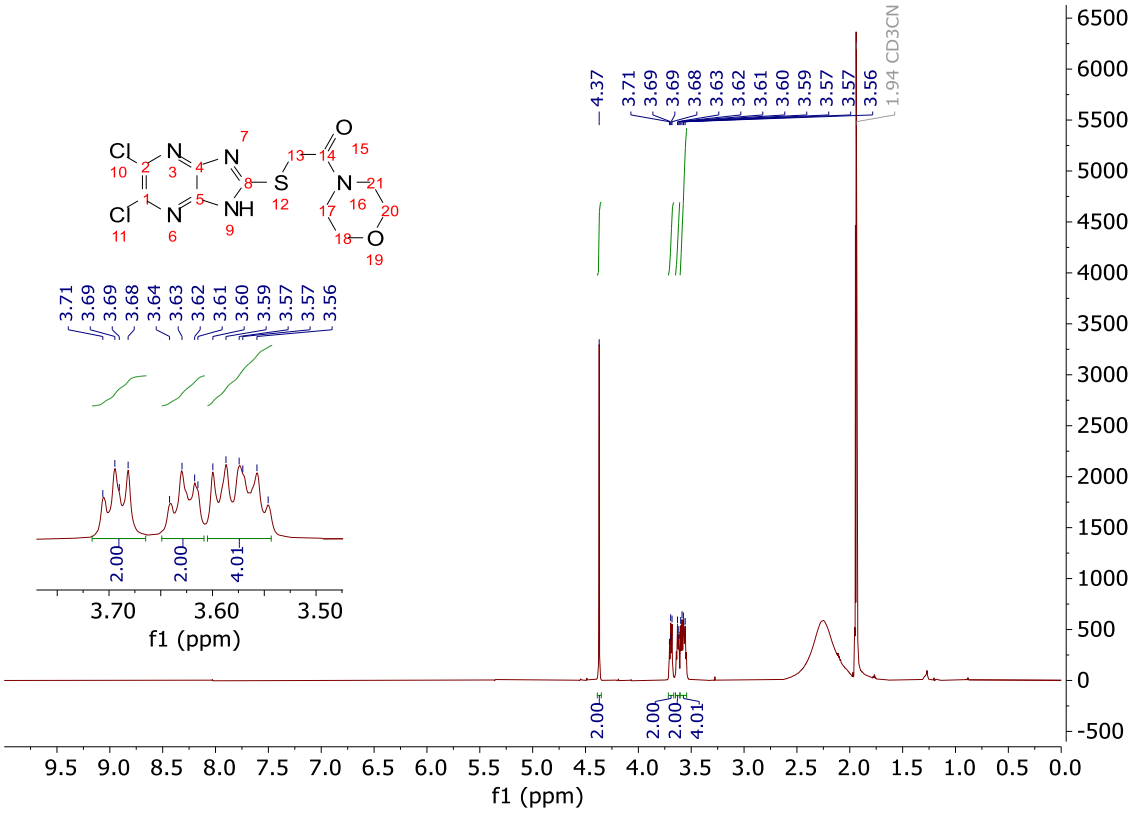
IR (cm^{-1}) 3143 (NH), 2771, 1623 (C=O).

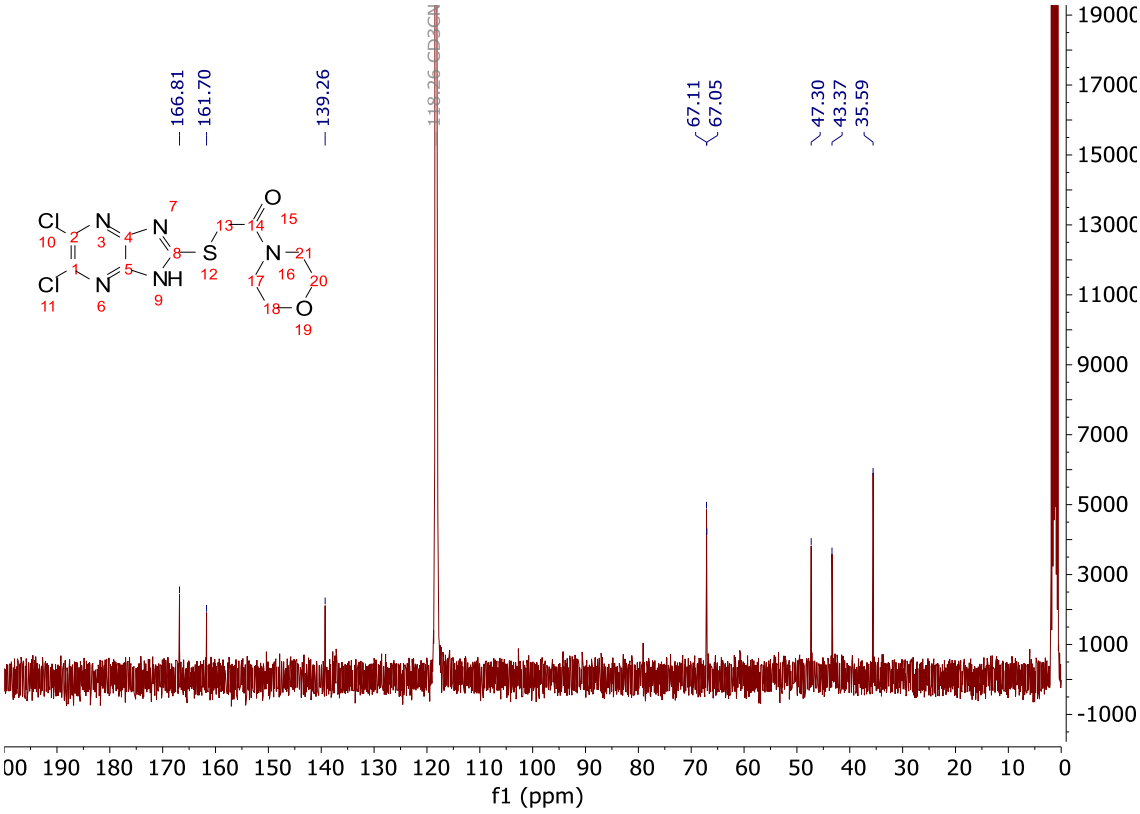
M.P. 214.4 – 215.1 ºC.

MS ES+ m/z calcd for $C_{11}H_{11}Cl_2N_5O_2S$ $(M+H)^+$: 349.0113, found: 349.0112.

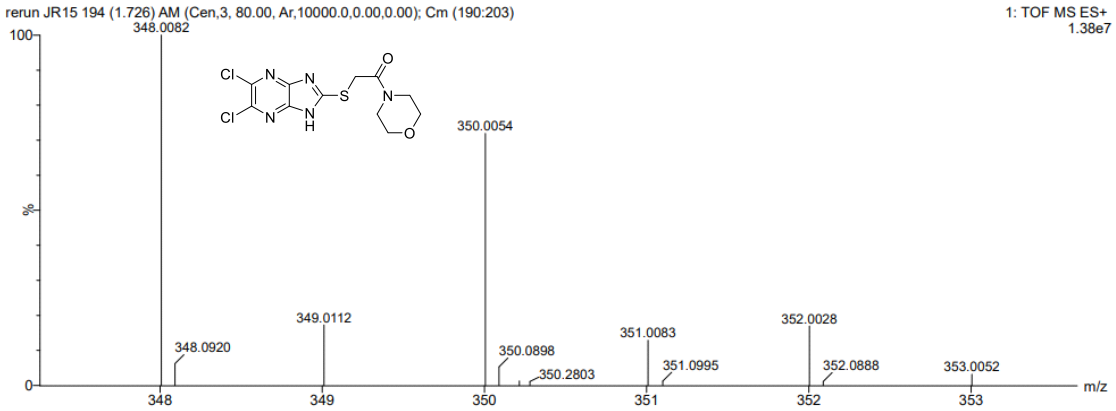
Adapted from the literature procedure (Tong, 1981).

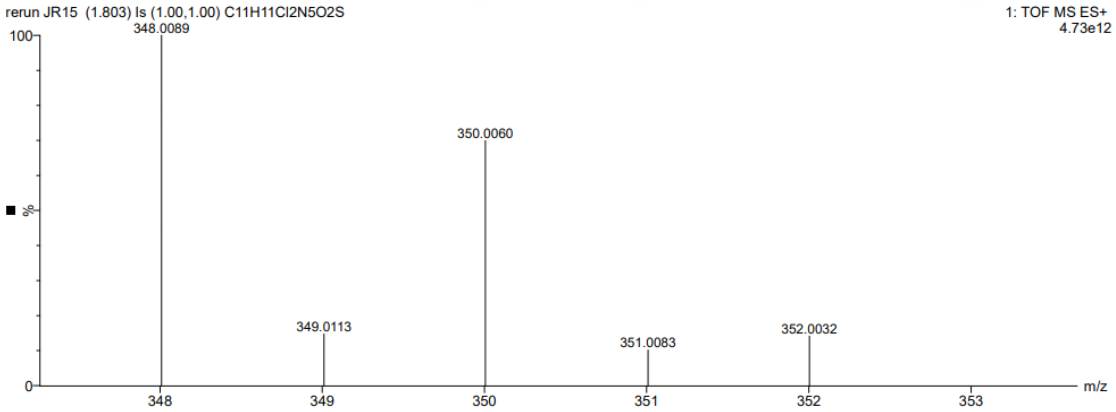
NMR Characterisation:



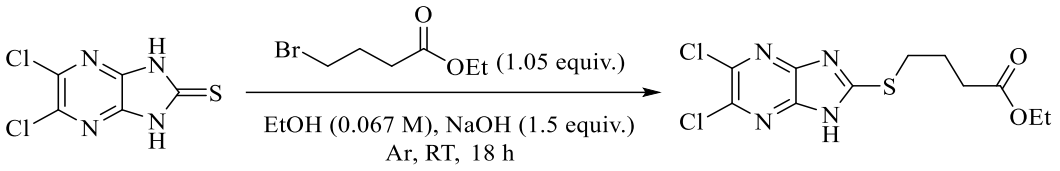


HRMS Characterisation:





7.3.1.4 ***Compound S11*** – *Ethyl 2-({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)butyrate*



To a 25 mL RBF was added a stirrer bar, 5,6-dichloroimidazo[4,5-*b*]pyrazine-2(1,3*H*)-thione (100 mg, 0.452 mmol) and sodium hydroxide (27.4 mg, 1.5 equiv.) and placed under argon atmosphere. Ethanol (over mol. sieves, 6.75 mL) was added with stirring. Ethyl 4-bromobutyrate (67.9 µL, 1.05 equiv.) was added via syringe with stirring. Allowed to stir at RT for 18 h. The solvent was removed under reduced pressure and the solid redissolved in sat. soln. NH_4Cl (10 mL) and EtOAc (10 mL). Transferred to a separatory funnel and the organic layer separated. The aqueous layer was extracted with EtOAc (3x 10 mL), the organic layers were collected and dried ($MgSO_4$). EtOAc was removed *in vacuo.* to provide a beige solid which was purified by column chromatography, eluting with 8:2 Pet. E : EtOAc to provide a colourless oil which solidifies over time into a light-orange solid. Further purified by triturating with *n*-hexane (5 mL) and ice-cold ethanol (5 mL) to provide a white solid of ethyl 2-({5,6-dichloro-1*H*-imidazo[4,5-*b*]pyrazine-2-yl}sulfanyl)butyrate (70.0 mg, 0.20 mmol, 46%).

^{1}H NMR (400 MHz, C*D*$_3$CN) δ 11.19 (br s, 1H), 4.11 (q, *J* = 7.2 Hz, 1H), 3.39 (t, *J* = 7.3 Hz, 1H), 2.47 (t, *J* = 7.3 Hz, 1H), 2.09 (p, *J* = 7.3 Hz, 1H), 1.21 (t, *J* = 7.2 Hz, 1H).

^{13}C NMR (101 MHz, C*D*$_3$CN) *δ* 173.7, 162.1, 139.1, 61.3, 33.3, 31.3, 25.7, 14.5 (8 out of 9 carbon resonances observed).

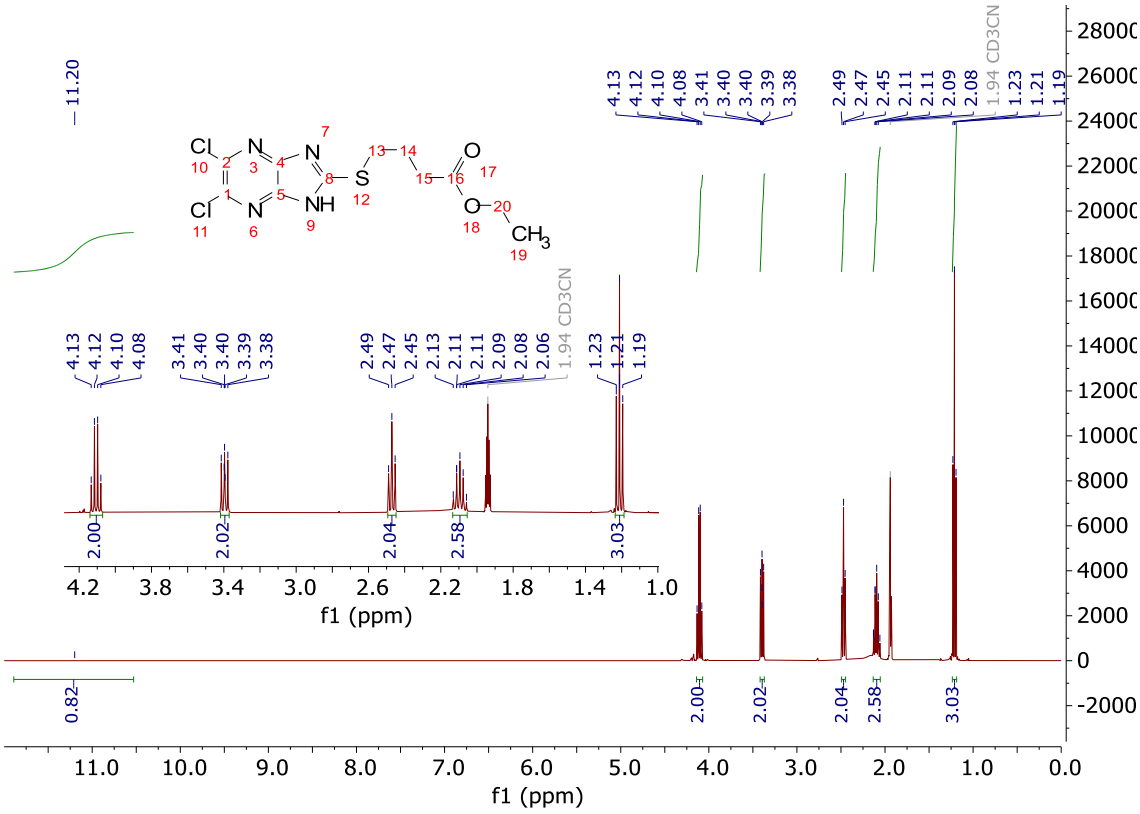
IR (cm^{-1}): 3151 (NH), 1691 (C=O).

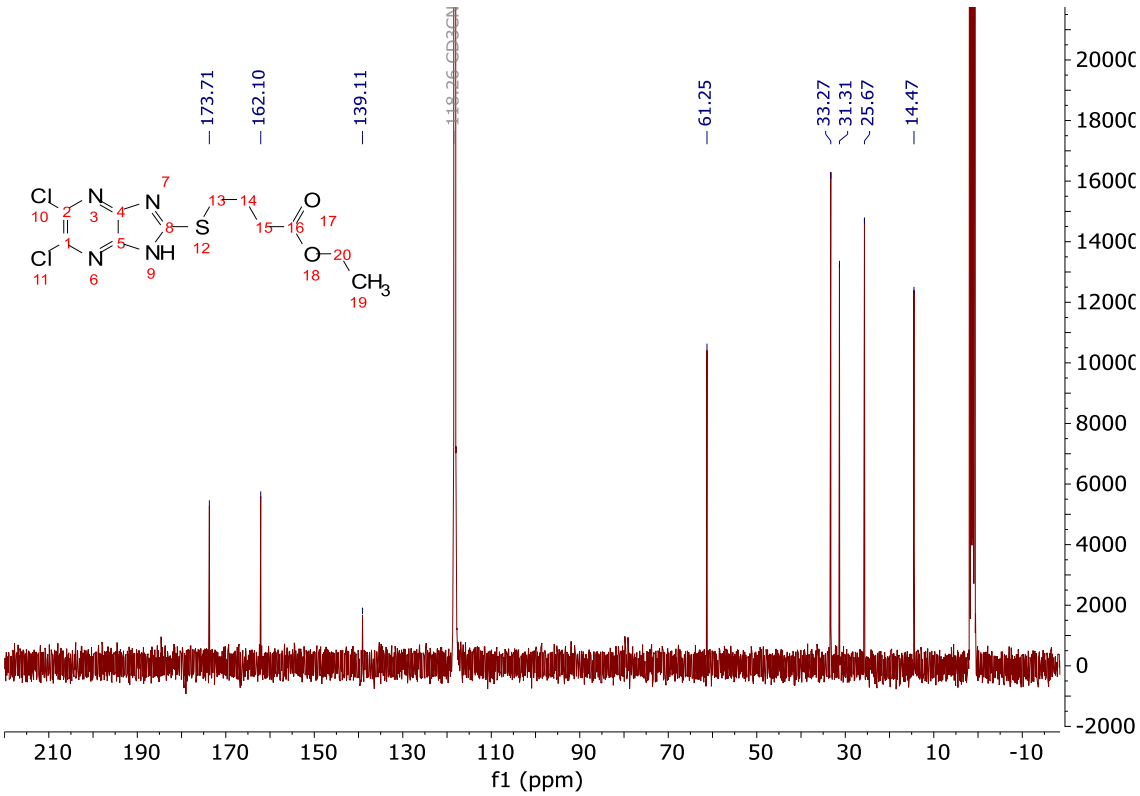
M.P. 120.3 – 121.3 ºC.

MS ES+ m/z calcd for $C_{11}H_{12}Cl_2N_4O_2S$ $(M+H)^+$: 335.0136, found: 335.0135.

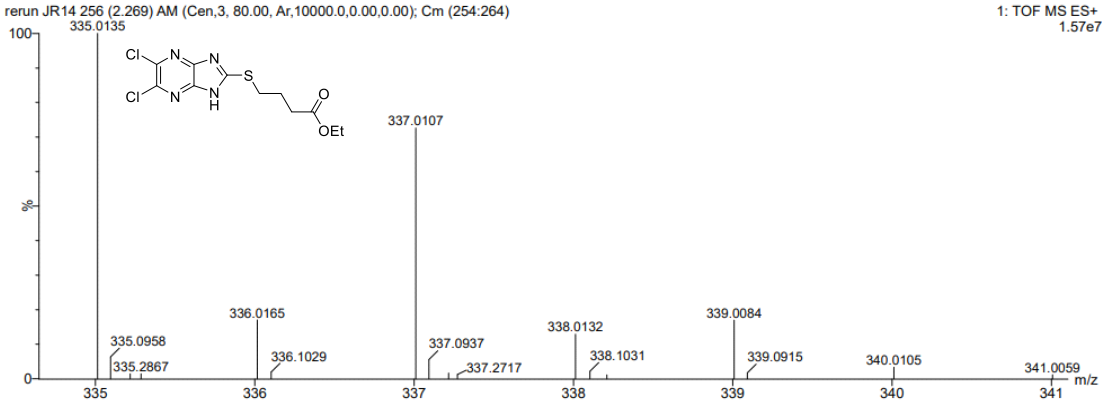
Adapted from the literature procedure (Tong, 1981).

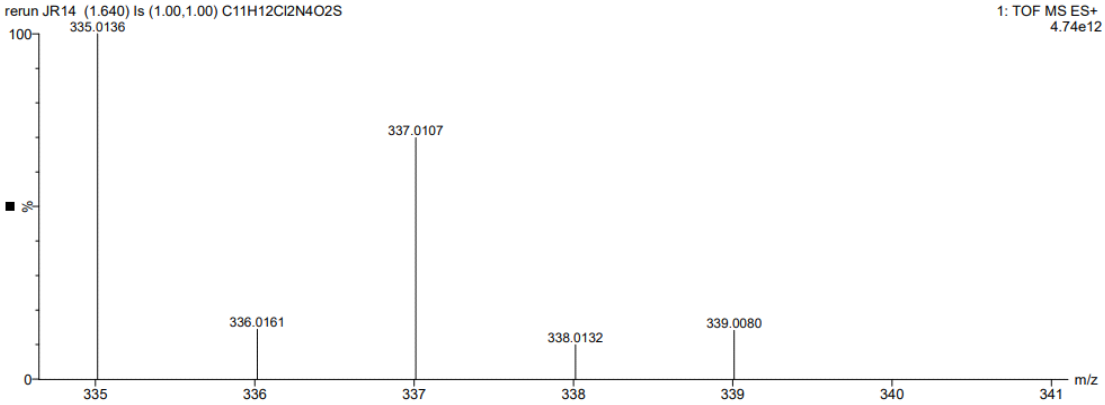
NMR Characterisation:



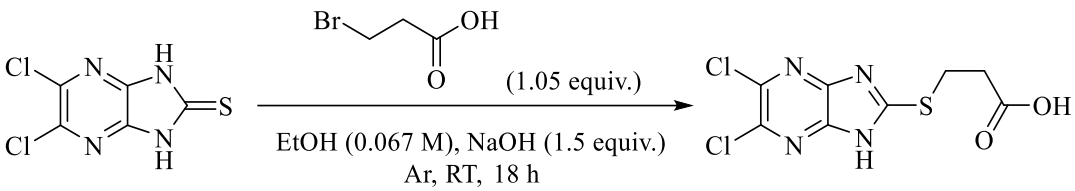


HRMS Characterisation:





7.3.1.5 ***Compound S13*** *– 3-({5,6-Dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)propionic acid*



To an 8 mL vial was added a stirrer bar, 5,6-dichloroimidaazo[4,5-*b*]pyrazine-2(1,3*H*)-thione (75 mg, 0.34 mmol), sodium hydroxide (20. 3 mg, 1.5 equiv.) and ethanol (over mol. sieves, 5.0 mL), stirred to dissolve. 3-bromopropionic acid (54.5 mg, 1.05 equiv.) was added and the vial placed under argon, left to stir for 18 h at RT. The solvent was removed *in vacuo.* and the residue dissolved in water (5 mL), extracted twice with diethyl ether (2x 10 mL), the aqueous layer was then acidified to pH approx. 2 and re-extracted with EtOAc (3x 10 mL), the EtOAc layers were then dried ($MgSO_4$) and solvent removed *in vacuo.* Further purified by column chromatography, eluting with Pet. E : EtOAc (6:4 with 0.4% acetic acid) to provide 2-({5,6-dichloro-1*H*-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)propionic acid (23.8 mg, 0.08 mmol, 24%) as a yellow solid.

^{1}H NMR (400 MHz, C*D*$_3$CN) δ 3.55 (t, *J* = 6.8 Hz, 2H), 2.87 (t, *J* = 6.8 Hz, 2H).

^{13}C NMR (101 MHz, C*D*$_3$CN) δ 173.0, 161.9, 139.2, 34.4, 27.4 (5 out of 6 carbon resonances observed).

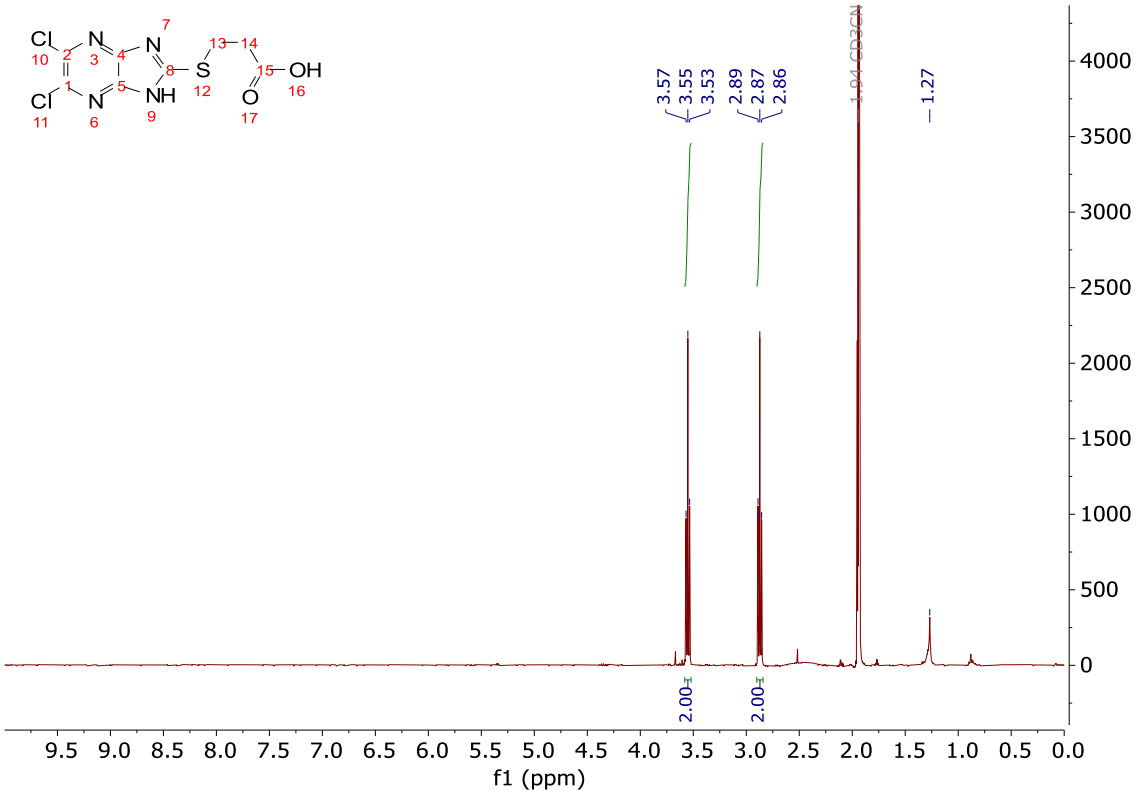
IR (cm^{-1}): 3111 (NH), 3023 (OH), 1721 (C=O).

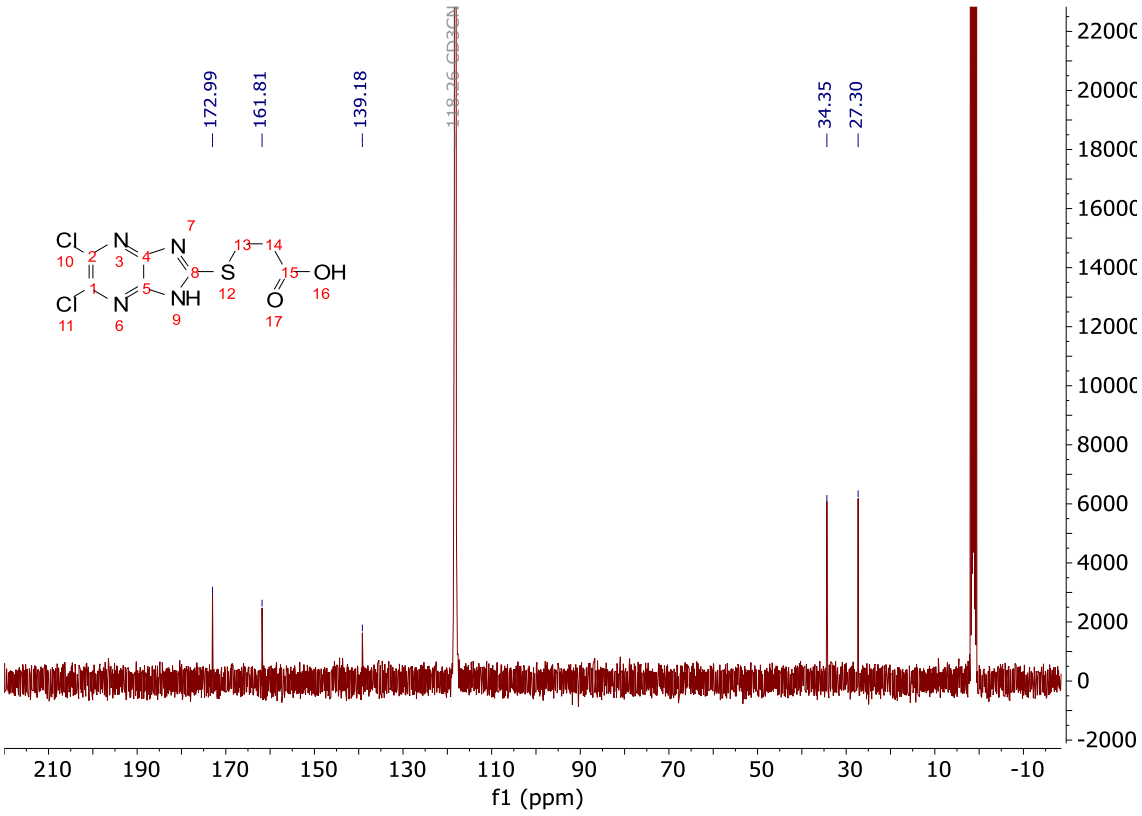
M.P. 205.1 – 206.2 ^{o}C (deg.).

MS ES+ m/z calcd for $C_8H_6Cl_2N_4O_2S$ $(M+H)^+$: 293.9689, found: 293.9688.

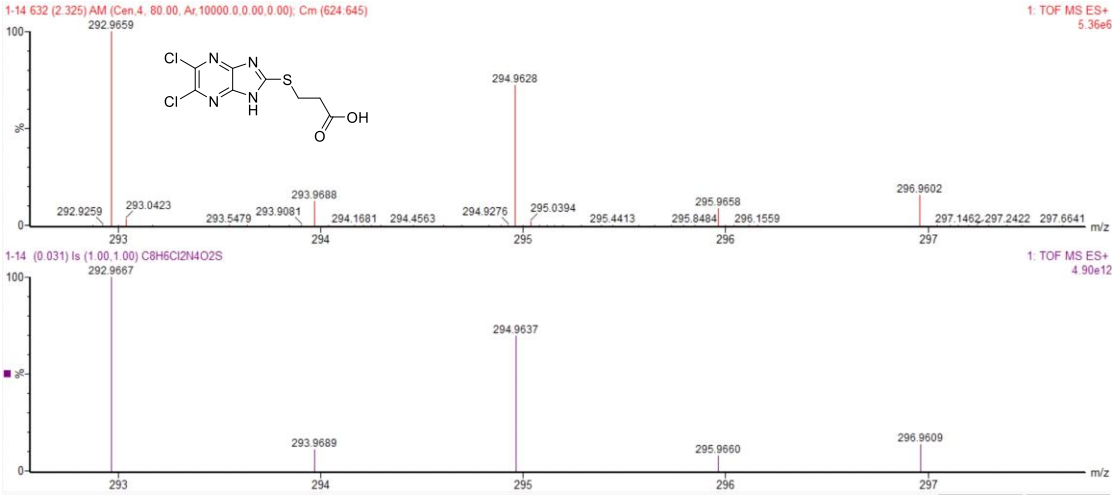
Adapted from the literature procedure (Tong, 1981).

NMR Characterisation:

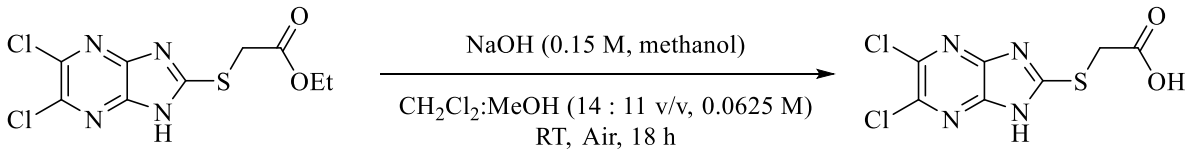




HRMS Characterisation:



7.3.1.6 ***Compound S15*** – *({5,6-dichloro-1H-imidazo[4,5-b]pyrazine-2-yl}sulfanyl)acetic acid*



Followed general procedure 1. Ethyl 2-({5,6-dichloro-1*H*-imidazo[4,5-*b*]pyrazine-2-yl}sulfanyl)-acetate (50 mg, 0.162 mmol), DCM : MeOH solution (1.62 mL). NaOH (0.97 mL). Provided ({5,6-dichloro-1*H*-imidazo[4,5-*b*]pyrazine-2-yl}sulfanyl)acetic acid as a light-yellow solid (31.4 mg, 0.11 mmol, 69%).

^{1}H NMR (400 MHz, C*D*$_3$CN) δ 4.14 (s, 2H).

^{13}C NMR (101 MHz, C*D*$_3$CN) δ 169.6, 161.2, 144.8, 139.5, 34.2.

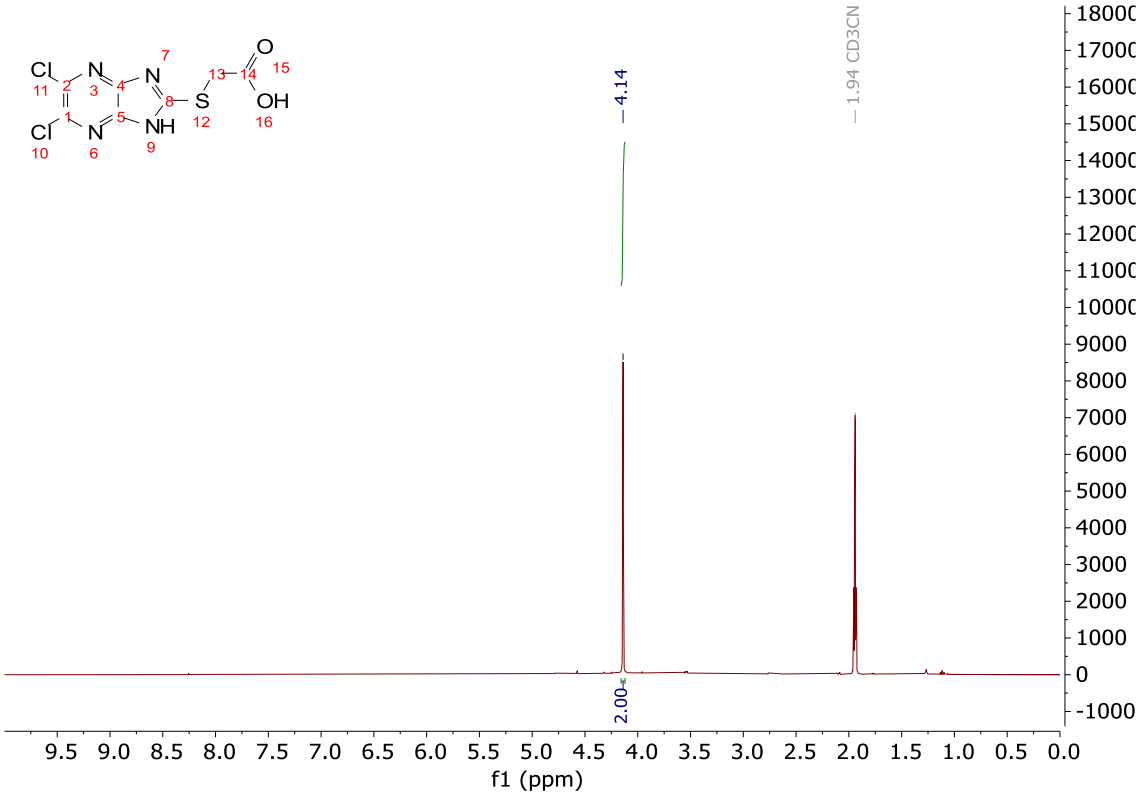
IR (cm^{-1}) 3111 (NH), 3027 (OH), 1740 (C=O).

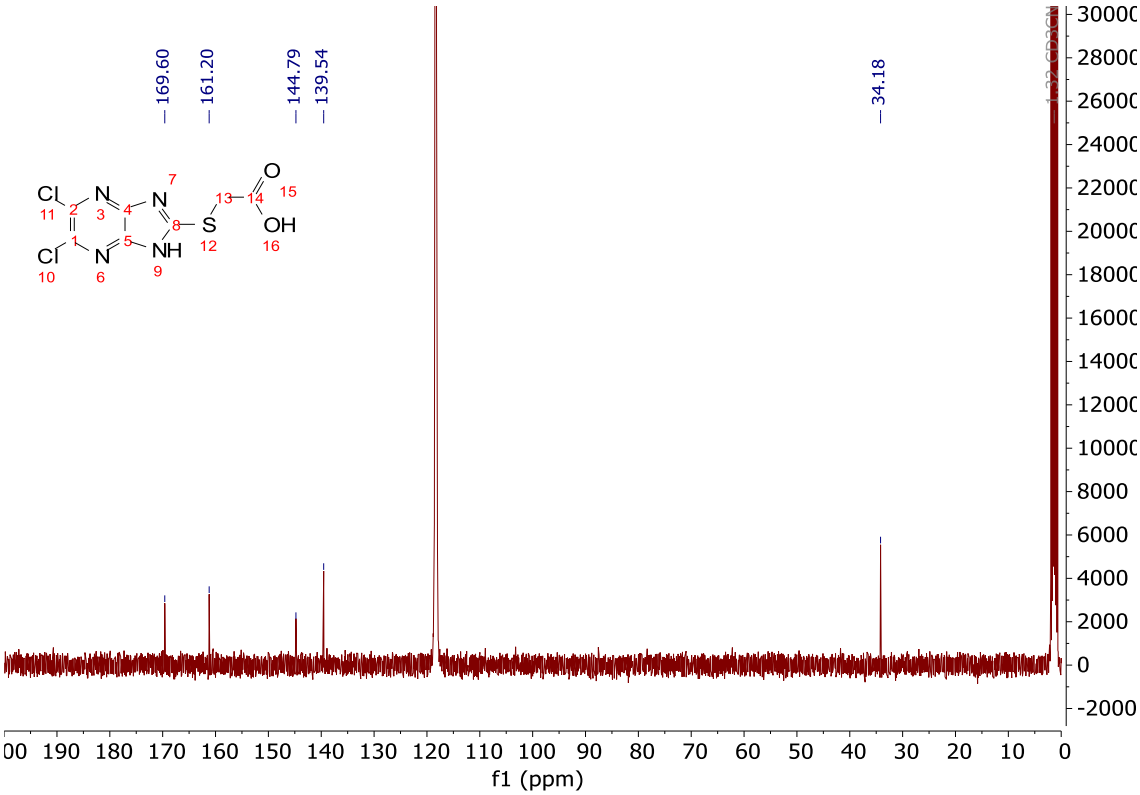
M.P. 219.1 – 220.7 ^{o}C.

MS ES+ m/z calcd for $C_7H_4Cl_2N_4O_2S$ $(M+H)^+$: 278.9510, found (-1.979 average neutral loss): 276.9719.

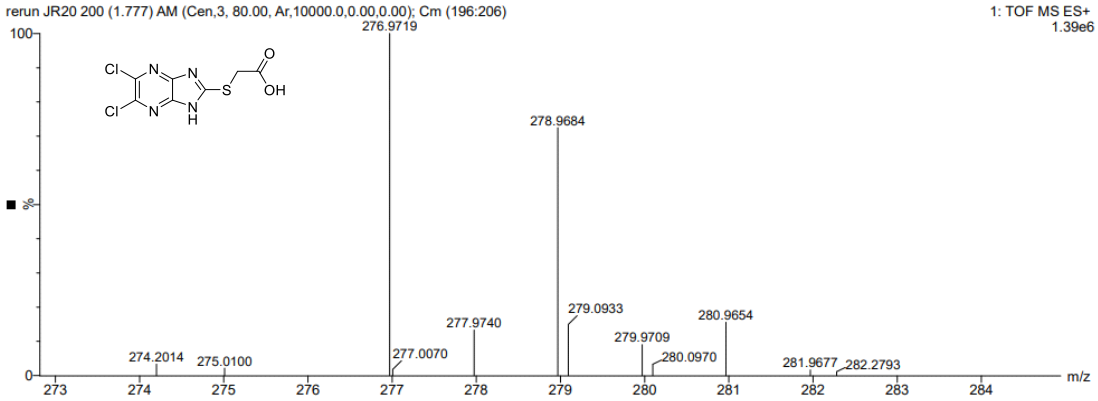
Adapted from the literature procedure (Theodorou *et al.*, 2007).

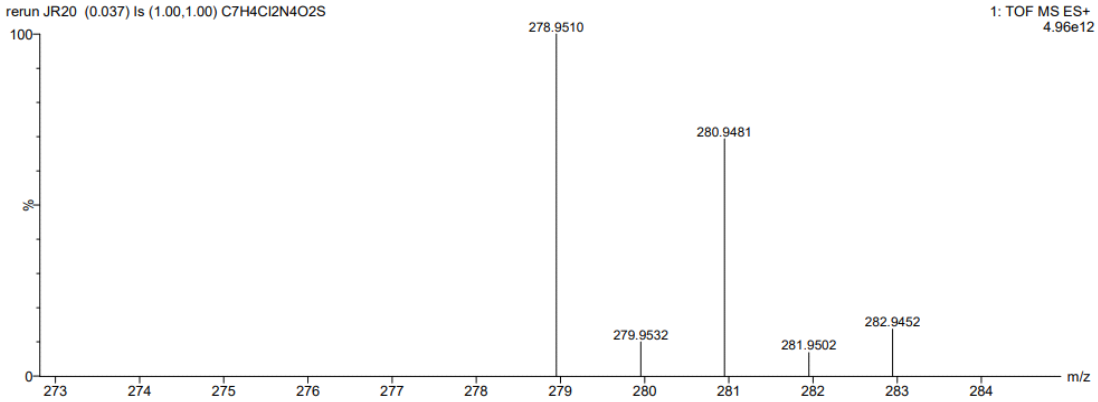
NMR Characterisation:





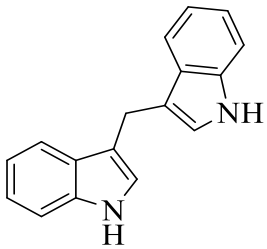
HRMS Characterisation:





7.3.2 I3C Hit Derivatives

7.3.2.1 ***DIM*** *– 3,3'-diindolylmethane*



To a 250 mL RBF was added indole-3-carbinol (1.0 g, 6.78 mmol) and 2.0 M NaOH solution (100 mL, 0.068 M). The mixture was heated to reflux for 1 h. The mixture was cooled and neutralised using CO_2 gas (dry ice). The precipitate was filtered and dried, then subjected to column chromatography (8:2 Hex. : EtOAc) to provide 3,3'-diindolylmethane as an off-white solid (341 mg, 1.38 mmol, 20%).

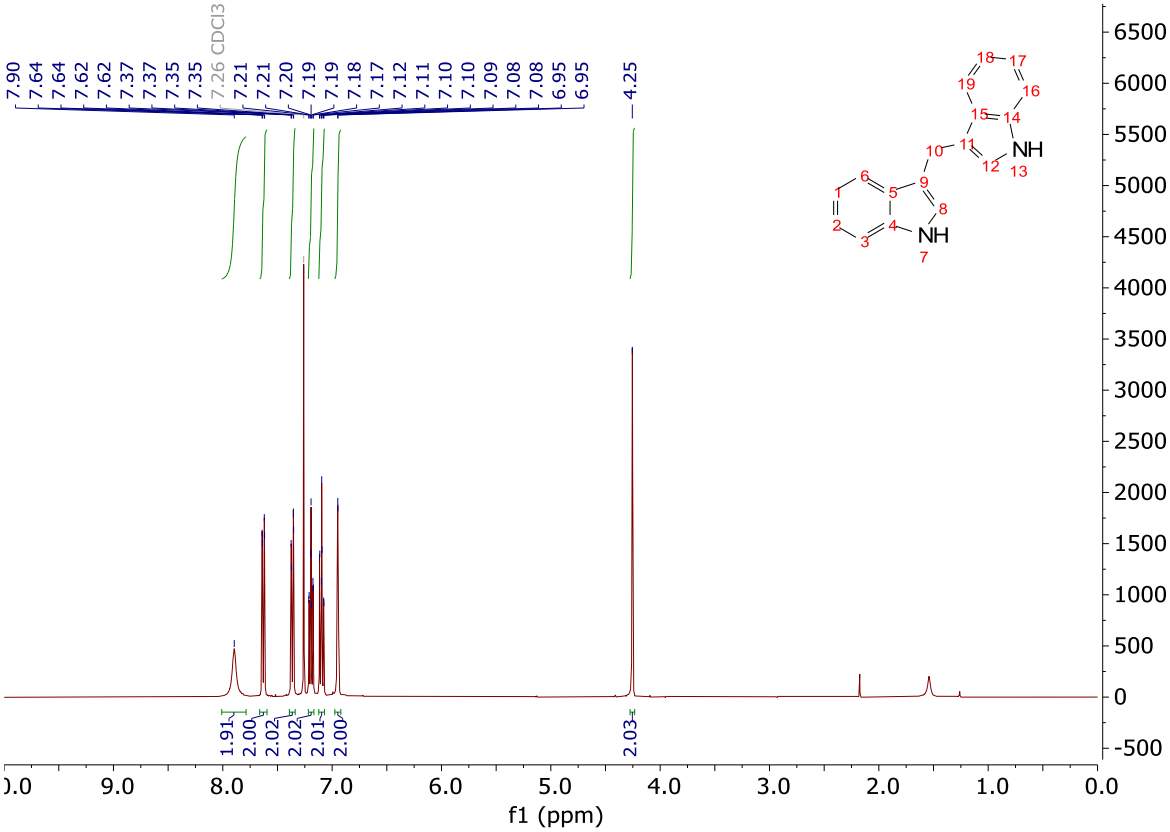
^{1}H NMR (400 MHz, C*D*Cl$_3$) δ 7.90 (br s, 2H), 7.63 (dd, J = 8.0, 1.1 Hz, 2H), 7.36 (dd, J = 8.0, 1.1 Hz, 2H), 7.19 (ddd, J = 8.2, 7.0, 1.2 Hz, 2H), 7.10 (ddd, J = 8.0, 7.0, 1.0 Hz, 2H), 6.97 – 6.92 (m, 2H), 4.25 (s, 2H).

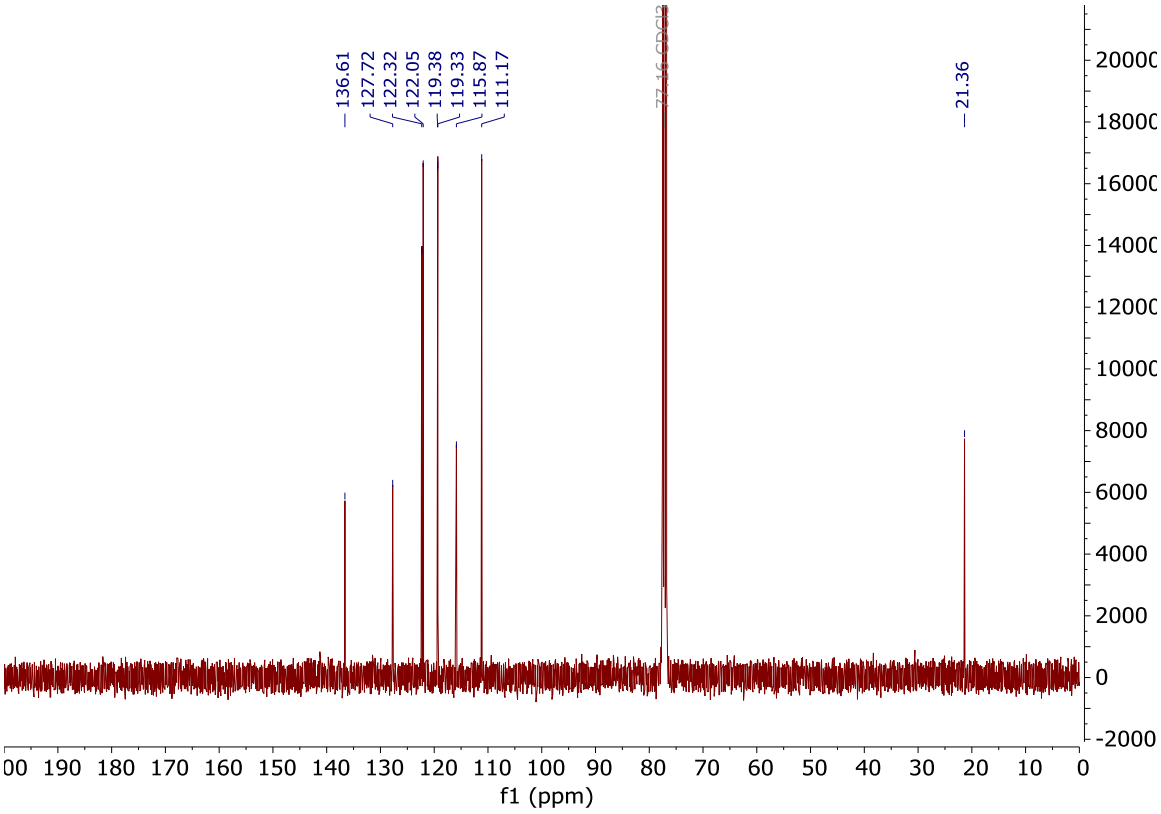
^{13}C NMR (101 MHz, C*D*Cl$_3$) δ 136.6, 127.7, 122.3, 122.1, 119.4, 119.3, 115.9, 111.2, 21.4.

IR (cm^{-1}): 3391 (NH).

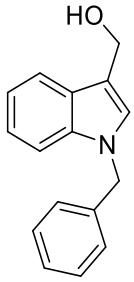
Followed literature procedure, data matches literature data (Chao *et al.*, 2007).

NMR Characterisation:





7.3.2.2 ***Compound I13/1-Benzyl-I3C*** *– N-benzylindole-3-carbinol*



To a 25 mL RBF was added *N*-benzylindole-3-carboxaldehyde (300 mg, 1.275 mmol), distilled THF (6.4 mL) and EtOH (6.4 mL) to reach a total concentration of 0.1 M. $NaBH_4$ (62.7 mg, 1.3 equiv.) was added in one batch at 0 ºC and the mixture allowed to warm to RT and stirred for 18 h. Afterwards, water (20 mL) was added to the suspension and the solvents were removed *in vacuo.* The aqueous suspension was then transferred to a separatory funnel and extracted with EtOAc (3x 20 mL). The organic layers were collected and washed with brine (20 mL), dried ($MgSO_4$) and solvent removed under reduced pressure to provide a white solid. Purified by column chromatography (8:2 -> 7:3 Hex. : EtOAc) to provide *N*-benzylindole-3-carbinol as a white solid (197 mg, 0.83 mmol, 65%).

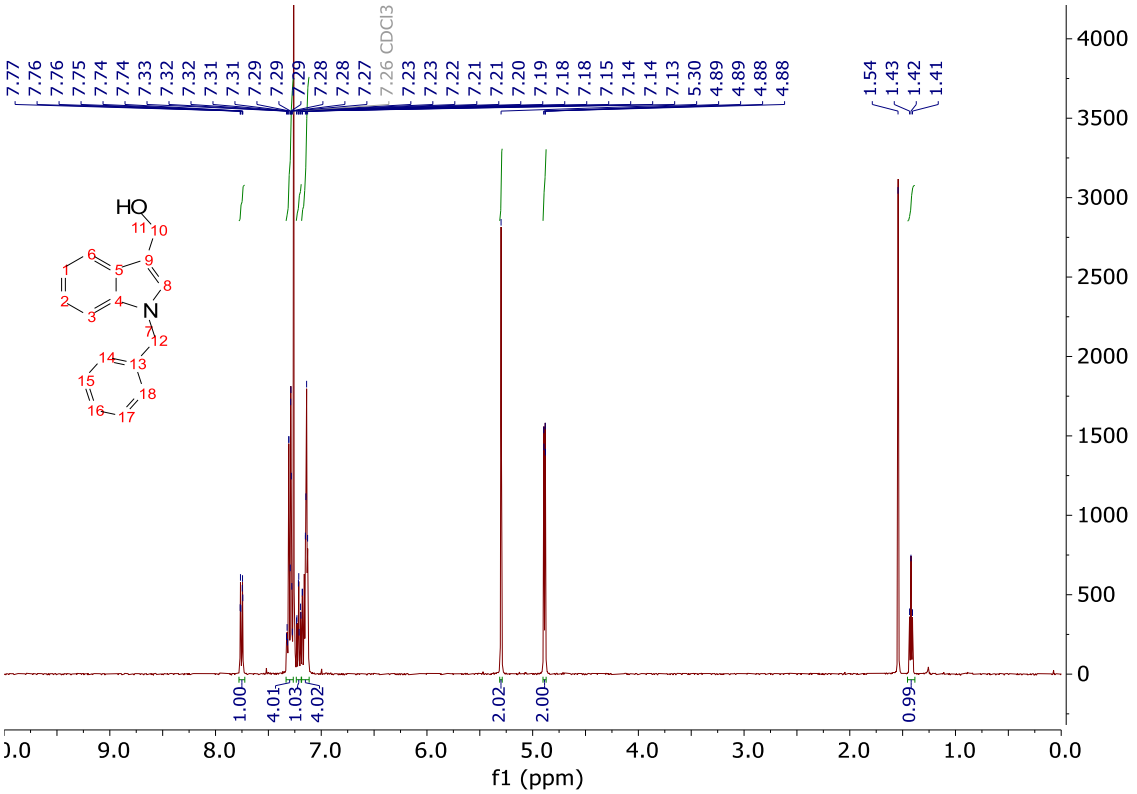
^{1}H NMR (400 MHz, C*D*Cl_3) δ 7.75 (dt, *J* = 7.5, 1.0 Hz, 1H), 7.33 – 7.26 (m, 4H), 7.21 (ddd, *J* = 8.2, 7.0, 1.4 Hz, 1H), 7.19 – 7.15 (m, 1H), 7.15 – 7.11 (m, 3H), 5.30 (s, 2H), 4.89 (d, *J* = 5.5 Hz, 2H), 1.42 (d, *J* = 5.5 Hz, 1H).

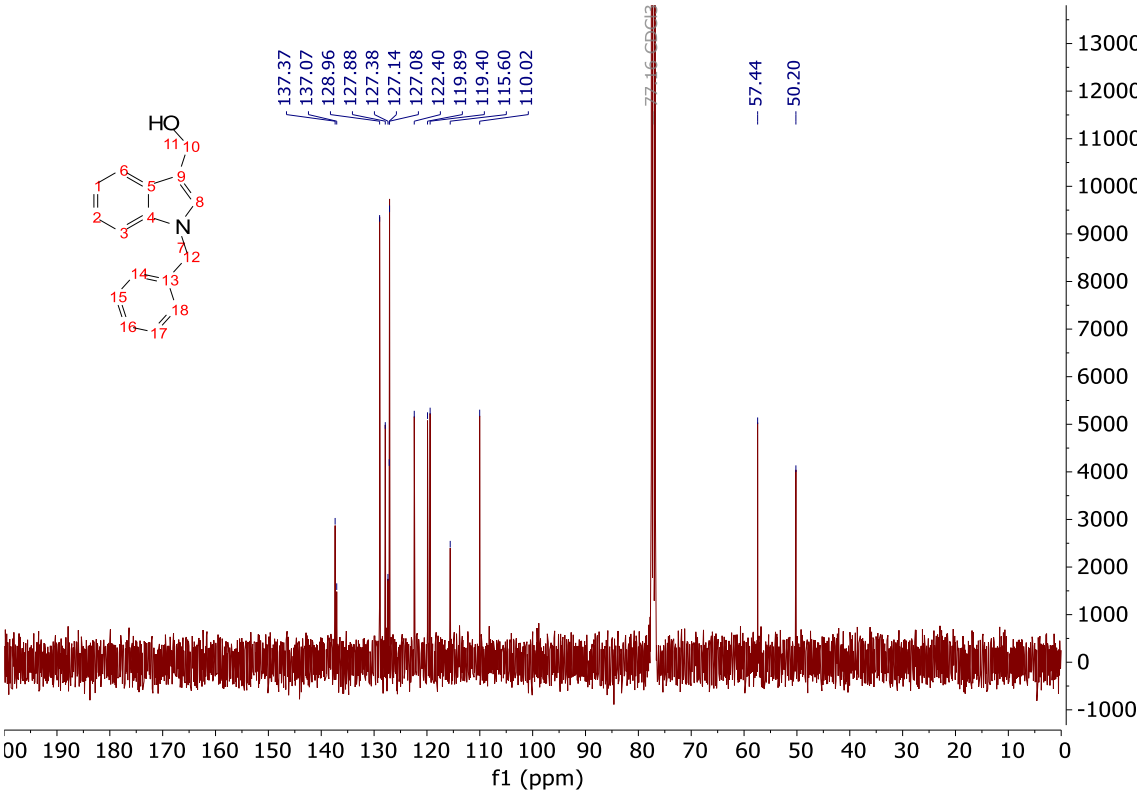
^{13}C NMR (101 MHz, C*D*Cl_3) δ 137.4, 137.0, 129.0, 127.9, 127.4, 127.14, 127.08, 122.4, 119.9, 119.4, 115.6, 110.0, 57.4, 50.2.

IR (cm^{-1}): 3367 (OH).

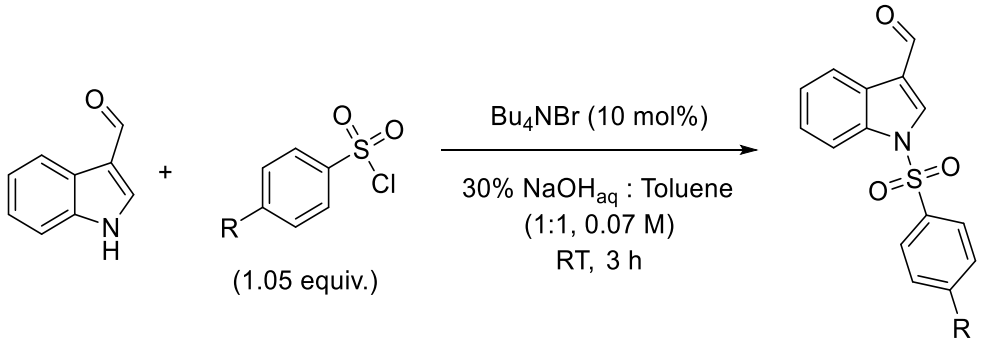
Data in line with literature (Arcadi *et al*., 2021).

NMR Characterisation:



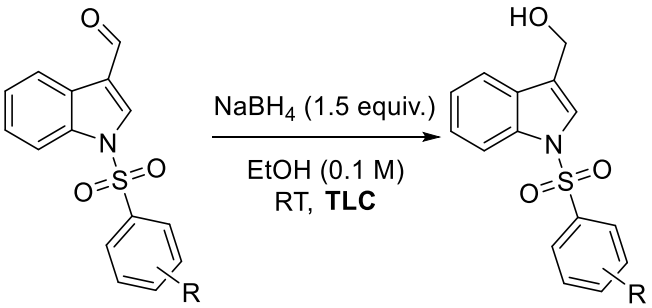


7.3.2.3 ***Compound I15/16*** *General Procedure 1: PTC Sulfonamide Formation*



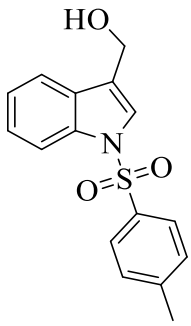
To an RBF was added indole-3-carboxaldehyde and toluene (0.14 M). To the stirred suspension was added 30% NaOH solution (0.14 M) and Bu_4NBr (10 mol%). Afterwards, the respective sulfonyl chloride (1.05 equiv.) was added, and the mixture vigorously stirred for 3 h. The mixture was transferred to a sep. funnel and separated, the aqueous phase washed with toluene (20 mL). The organic layers were collected and dried ($MgSO_4$) and solvent removed. Purified by column chromatography to provide the respective *N*-(arylsuflonyl)jndole-3-carboxaldehyde. Adapted from literature conditions(Kutschy *et al.*, 1998).

7.3.2.4 ***Compound I15/16*** *General Procedure 2: N-(aryl)indole-3-carboxaldehyde Reduction*



To a vial was added the respective *N*-functionalised indole-3-carboxaldehyde and ethanol (0.1 M). To the stirred solution was added $NaBH_4$ (1.5 equiv.) and allowed to stir at RT, monitored by TLC. H_2O (3 mL) was added to quench the reaction after completion, the mixture was transferred to a separatory funnel and brine was added (10 mL), the mixture extracted with EtOAc (3x 20 mL), the organic layers collected and washed with brine (10 mL), dried ($MgSO_4$) and solvent removed under reduced pressure at 27 - 28 ^{o}C. Purified by column chromatography (7:3 Hexane : EtOAc), followed by trituration with hexane.

7.3.2.5 ***Compound I15/N-Tosyl-I3C*** *– N-(4-methylbenzenesulfonyl)indole-3-carbinol*



Followed general procedure 2: *N*-(4-methylbenzenesulfonyl)indole-3-carboxaldehyde (200 mg, 0.668 mmol), EtOH (6.68 mL, 0.1 M), sodium borohydride (38 mg, 1.5 equiv.). 5 h reaction time. Routine workup provided *N*-(4-methylbenzenesulfonyl)indole-3-carbinol as a colourless oil (168 mg, 0.56 mmol, 83%).

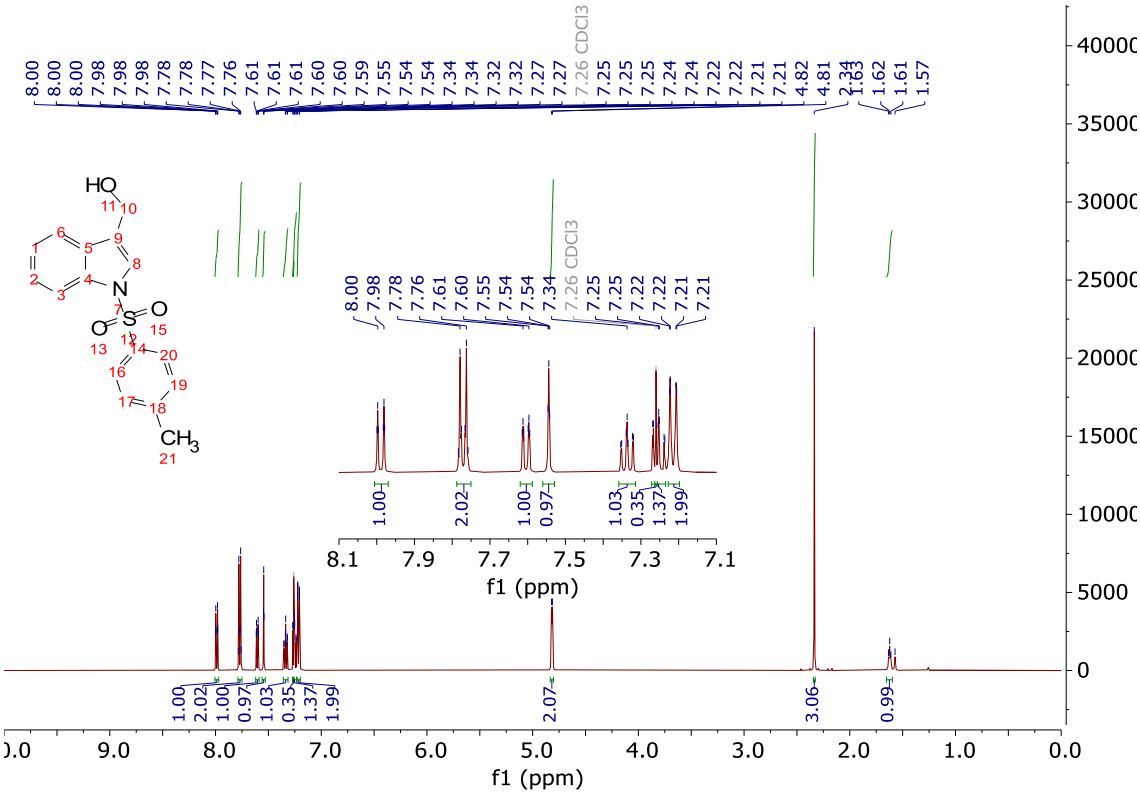
^{1}H NMR (400 MHz, C*D*Cl$_3$) δ 7.99 (dt, *J* = 8.4, 0.9 Hz, 1H), 7.81 – 7.73 (m, 2H), 7.61 (dt, *J* = 7.8, 1.1 Hz, 1H), 7.55 (d, *J* = 1.1 Hz, 1H), 7.34 (ddd, *J* = 8.4, 7.2, 1.3 Hz, 1H), 7.28 – 7.20 (m, 3H), 4.82 (dd, *J* = 5.7, 1.0 Hz, 2H), 2.34 (s, 3H), 1.59 (br t, *J* = 5.7 Hz, 1H).

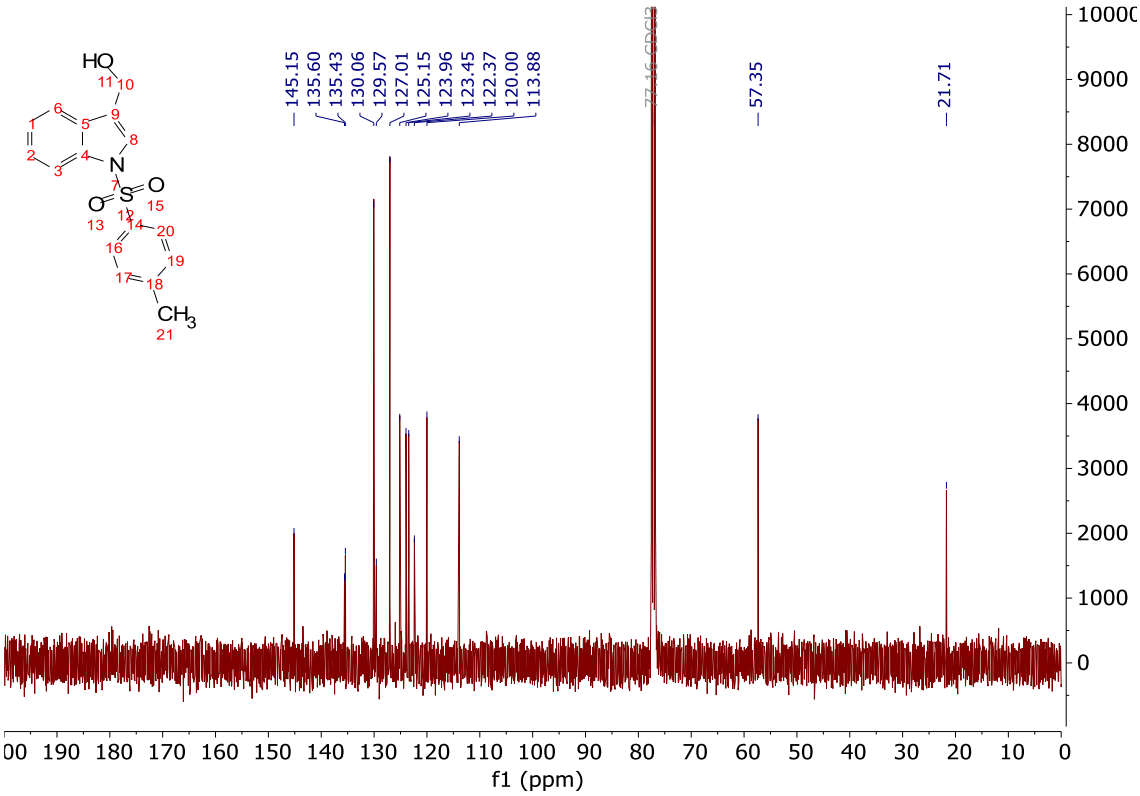
^{13}C NMR (101 MHz, C*D*Cl$_3$) δ 145.2, 135.6, 135.4, 130.1, 129.6, 127.0, 125.2, 124.0, 123.5, 122.4, 120.0, 113.9, 57.4, 21.7.

IR (cm^{-1}): 3384 (OH).

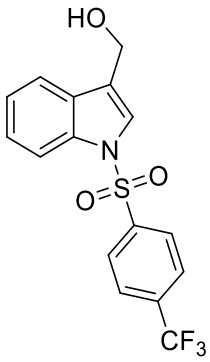
Data in line with the literature (Quirit *et al.*, 2017).

NMR Characterisation:





7.3.2.6 ***Compound I16*** *– N-(4-trifluoromethylbenzenesulfonyl)indole-3-carbinol*



Followed general procedure 2: *N*-(4-trifluoromethylbenzenesulfonyl)indole-3-carboxaldehyde (30 mg, 0.085 mmol), $NaBH_4$ (5 mg, 1.5 equiv.)n EtOH (1 mL). Reaction time 2 h. Provided an off-white solid of *N*-(4-trifluoromethylbenzenesulfonyl)indole-3-carbinol (30 mg, quant.).

^{1}H NMR (400 MHz, C*D*Cl_3) δ 8.04 – 7.97 (m, 3H), 7.73 – 7.67 (m, 2H), 7.65 – 7.59 (m, 1H), 7.54 (d, *J* = 1.1 Hz, 1H), 7.38 (ddd, *J* = 8.5, 7.3, 1.3 Hz, 1H), 7.29 (ddd, *J* = 8.3, 7.3, 1.1 Hz, 1H), 4.84 (dd, *J* = 5.7, 1.1 Hz, 2H), 1.62 (t, *J* = 5.7 Hz, 1H).

^{13}C NMR (101 MHz, C*D*Cl_3) δ 141.63 135.6, 135.4, 129.7, 127.5, 126.7 (q, J = 3.9 Hz), 125.6, 124.0, 123.6, 123.5, 120.3, 113.8, 57.3 (13 out of 14 carbon resonances observed).

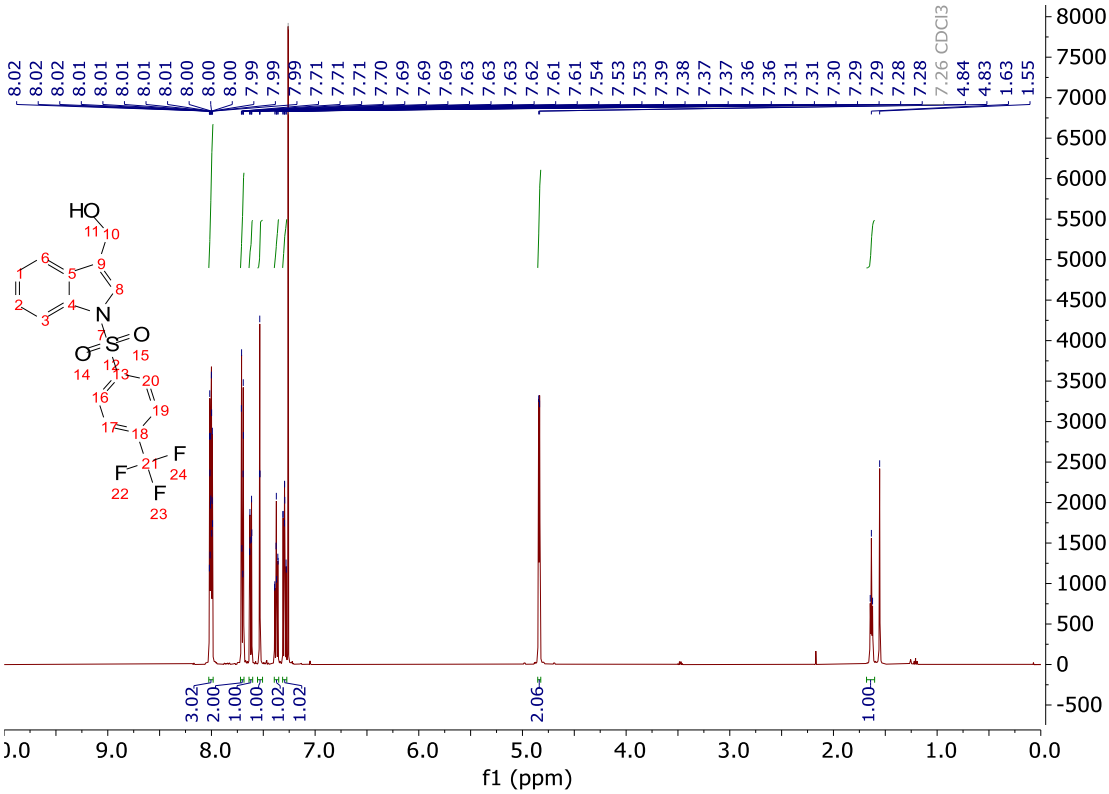
^{19}F NMR (376 MHz, C*D*Cl_3) δ -63.37.

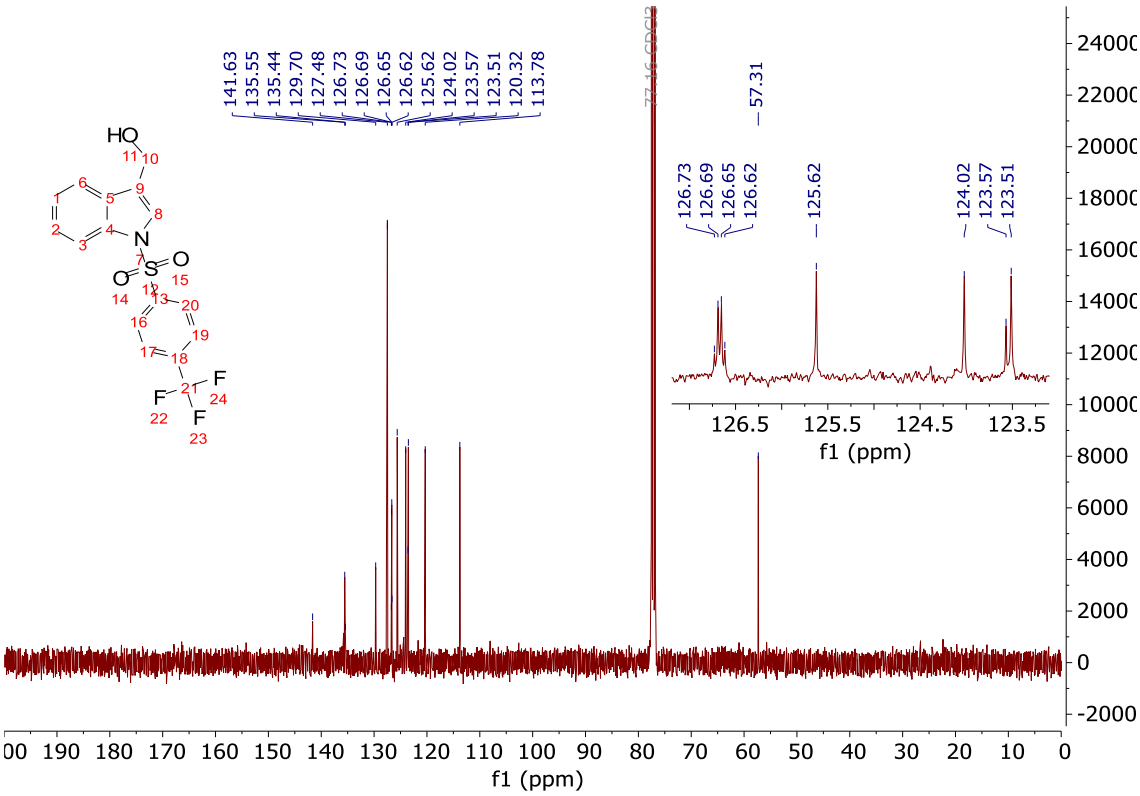
IR (cm^{-1}): 3246 (OH).

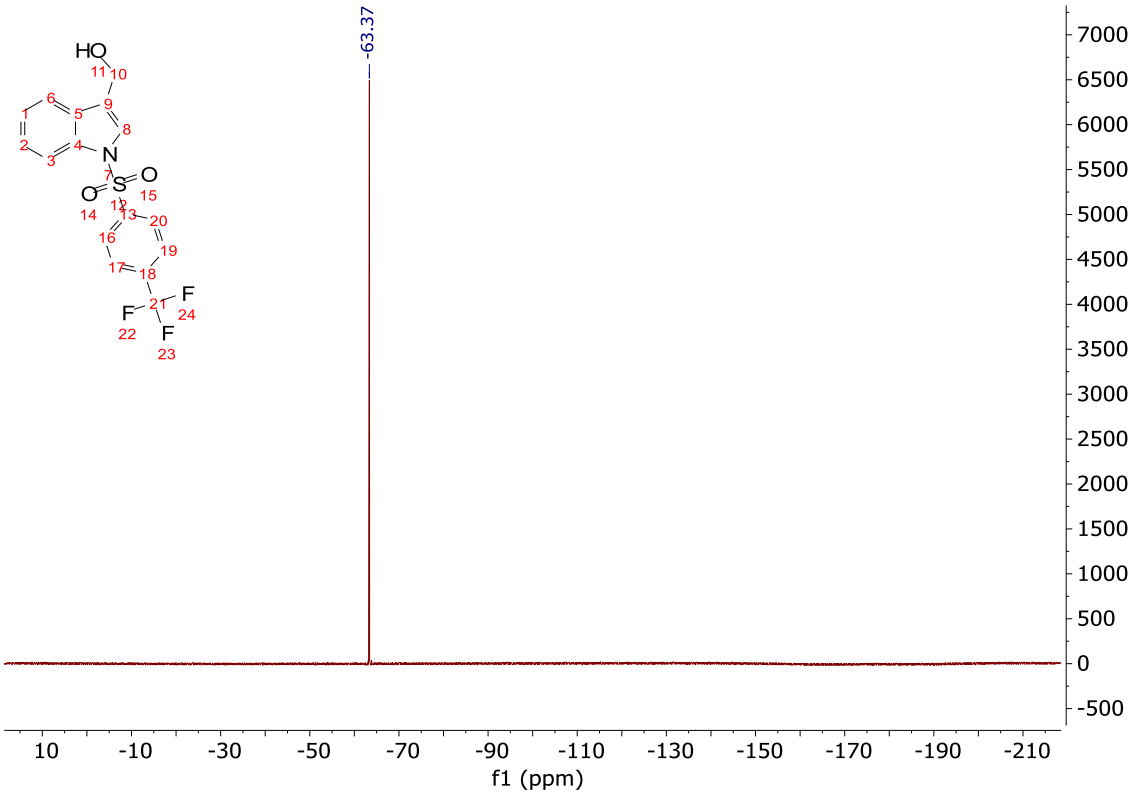
M.P. 164.1 – 165.5 ºC.

MS ESI+ Calc. for $C_{16}H_{11}F_3NO_2S$ (-OH): 339.0493, found: 339.0493.

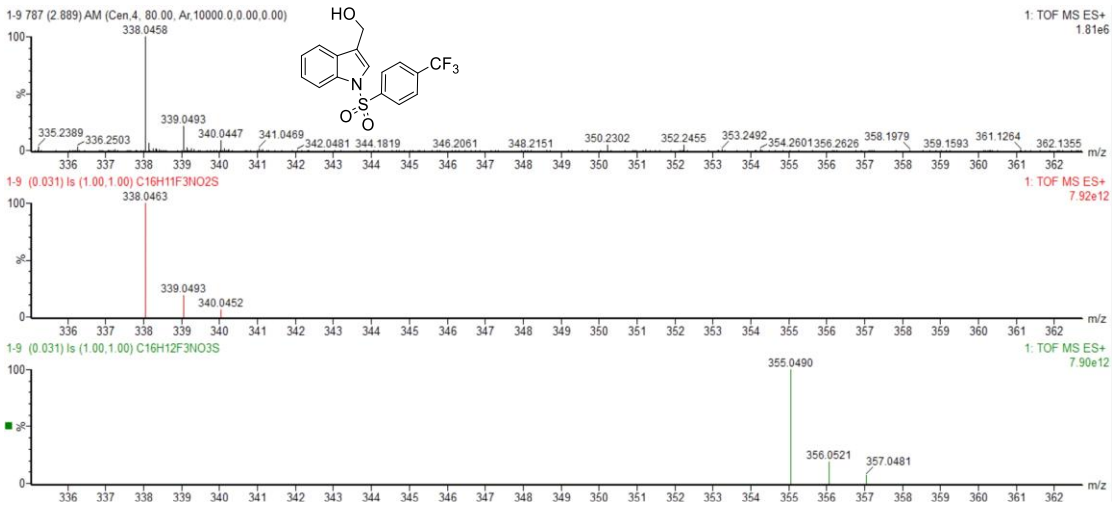
NMR Characterisation:







HRMS Characterisation:



References

Abramson, J. et al. (2024) 'Accurate structure prediction of biomolecular interactions with AlphaFold 3', *Nature*, 630(8016), pp. 493-500. Available at: https://doi.org/10.1038/s41586-024-07487-w.

Adams, M.J. et al. (1969) 'Structure of rhombohedral 2 zinc insulin crystals', *Nature*, 224(5218), pp. 491-495. Available at: https://doi.org/10.1038/224491a0.

Ağagündüz, D. et al. (2022) 'Cruciferous Vegetables and Their Bioactive Metabolites: From Prevention to Novel Therapies of Colorectal Cancer', *Evidence-based Complementary and Alternative Medicine* [Internet], e1534083. Available at: https://doi.org/10.1155/2022/1534083.

Agirre, J. et al. (2023) 'The CCP4 suite: integrative software for macromolecular crystallography', *Acta Crystallographica. Section D, Structural biology*, 79(6), pp. 449-461. Available at: https://doi.org/10.1107/S2059798323003595.

Alhajj, M., Zubair, M. and Farhana, A. (2023) 'Enzyme Linked Immunosorbent Assay' in *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK555922/ (Accessed: 23 June 2024)

Alpi, A.F., Chaugule, V. and Walden, H. (2016) 'Mechanism and disease association of E2-conjugating enzymes: Lessons from UBE2T and UBE2L3', *Biochemical Journal*, 473(20), pp. 3401-3419. Available at: https://doi.org/10.1042/BCJ20160028.

Amare, D.E. et al. (2020) 'Acid condensation products of indole-3-carbinol and their in-vitro (anti)estrogenic, (anti)androgenic and aryl hydrocarbon receptor activities', *Arabian Journal of Chemistry*, 13(9), pp. 7199-7211. Available at: https://doi.org/10.1016/j.arabjc.2020.08.002.

Amerik, A.Y. and Hochstrasser, M. (2004) 'Mechanism and function of deubiquitinating enzymes', *Biochimica et Biophysica Acta*, 1695(1-3), pp. 189-207. Available at: https://doi.org/10.1016/j.bbamcr.2004.10.003.

Anderton, M.J. et al. (2004) 'Pharmacokinetics and tissue disposition of indole-3-carbinol and its acid condensation products after oral administration to mice', *Clinical Cancer Research*, 10(15), pp. 5233-5241. Available at: https://doi.org/10.1158/1078-0432.CCR-04-0163.

Arcadi, A. et al. (2021) 'Synthesis of Indole/Benzofuran-Containing Diarylmethanes through Palladium-Catalyzed Reaction of Indolylmethyl or Benzofuranylmethyl Acetates with Boronic Acids', *Synthesis*, 54(3), pp. 741-753. Available at: https://doi.org/10.1055/s-0041-1737275.

Aronchik, I. et al. (2012) ‘Target protein interactions of indole-3-carbinol and the highly potent derivative 1-benzyl-I3C with the C-terminal domain of human elastase uncouples cell cycle arrest from apoptotic signalling’, *Molecular Carcinogenesis*, 51(11), pp. 881-894. Available at: https://doi.org/10.1002/mc.20857.

Aronchik, I. et al. (2014) ‘The antiproliferative response of indole-3-carbinol in human melanoma cells is triggered by an interaction with NEDD4-1 and disruption of wild-type PTEN degradation’, *Molecular Cancer Research*, 12(11), pp. 1621-1634. Available at: https://doi.org/10.1158/1541-7786.MCR-14-0018.

Aronchik, I., Bjeldanes, L.F. and Firestone, G.L. (2010) ‘Direct inhibition of elastase activity by indole-3-carbinol triggers a CD40-TRAF regulatory cascade that disrupts NF-κB transcriptional activity in human breast cancer cells’, *Cancer Research*, 70(12), 4961-4971. Available at: https://doi.org/10.1158/0008-5472.CAN-09-3349.

Auld, D.S., Inglese, J. and Dahlin, J.L. (2017) ‘Assay Interference by Aggregation’ in Markossian, S. et al (ed). *Assay Guidance Manual* [Internet], Bethesda (MD): Eli Lilly and Company and the National Center for Advancing Translational Sciences; 2004. Available at: https://www.ncbi.nlm.nih.gov/books/NBK442297/ (Accessed: 22 June 2024).

Baell, J.B. and Holloway, G.A. (2010) ‘New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays’, *Journal of Medicinal Chemistry*, 53(7), pp. 2719-2740. Available at: https://doi.org/10.1021/jm901137j.

Baell, J.B. and Nissink, J.W.M. (2018) ‘Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017 - Utility and Limitations’, *ACS Chemical Biology*, 13(1), pp 36-44. Available at: https://doi.org/10.1021/acschembio.7b00903.

Bai, L.Y. et al. (2013) ‘OSU-A9, an indole-3-carbinol derivative, induces cytotoxicity in acute myeloid leukaemia through reactive oxygen species-mediated apoptosis’, *Biochemical Pharmacology*, 86(10), pp 1430-1440. Available at: https://doi.org/10.1016/j.bcp.2013.09.002.

Banerjee, S. et al. (2011) ‘Attenuation of multi-targeted proliferation-linked signalling by 3,3’-diindolylmethane (DIM): From bench to clinic’, *Mutation Research - Reviews in Mutation* Research, 728(1-2), pp 47-66. Available at: https://doi.org/10.1016/j.mrrev.2011.06.001.

Barghout, S.H. and Schimmer, A.D. (2021) ‘E1 enzymes as therapeutic targets in cancers’, *Pharmacological Reviews*, 73(1), pp 1-58. Available at: https://doi.org/10.1124/pharmrev.120.000053.

Bastos, M. et al. (2023) ‘Isothermal titration calorimetry’, *Nature Reviews Methods Primers*, 3(1), e17. Available at: https://doi.org/10.1038/s43586-023-00199-x.

Bayer, P. et al. (1998) ‘Structure determination of the small ubiquitin-related modifier SUMO-1’, *Journal of Molecular Biology*, 280(2), pp 275-286. Available at: https://doi.org/10.1006/jmbi.1998.1839.

Behera, A. and Reddy, A.B.M. (2023) 'WWP1 E3 ligase at the crossroads of health and disease', *Cell Death and Disease*, 14(1), e853. Available at: https://doi.org/10.1038/s41419-023-06380-0.

Behrends, C. and Harper, J.W. (2011) 'Constructing and decoding unconventional ubiquitin chains', *Nature Structural and Molecular Biology*, 18(5), pp 520-528. Available at: https://doi.org/10.1038/nsmb.2066.

Berggren, G. et al. (2019) 'Compounds with capacity to quench the tyrosyl radical in Pseudomonas aeruginosa ribonucleotide reductase', *Journal of Biological Inorganic Chemistry*, 24(6), pp 841-848. Available at: https://doi.org/10.1007/s00775-019-01679-w.

Bie, P. and Ciechanover, A. (2011) 'Ubiquitination of E3 ligases: Self-regulation of the ubiquitin system via proteolytic and non-proteolytic mechanisms', *Cell Death and Differentiation*, 18(1), pp 1393-1402. Available at: https://doi.org/10.1038/cdd.2011.16.

Blidisel, A. et al. (2021) 'Experimental models of hepatocellular carcinoma—a preclinical perspective', *Cancers*, 13(5), p 3651. Available at: https://doi.org/10.3390/cancers13153651.

Błoch-Mechkour, A. et al. (2010) 'Radicals and radical ions derived from indole, indole-3-carbinol and diindolylmethane', *Journal of Physical Chemistry A*, 114(25), pp 6787-6794. Available at: https://doi.org/10.1021/jp912121y.

Blundell, T.L. et al. (1972) 'The crystal structure of rhombohedral 2 zinc insulin.', *Cold Spring Harbor Symposia on Quantitative Biology*, 36(1), pp 233-241. Available at: https://doi.org/10.1101/SQB.1972.036.01.031.

Bradlow, H.L. and Zeligs, M.A. (2010) 'Diindolylmethane (DIM) spontaneously forms from indole-3-carbinol (I3C) during cell culture experiments', *In Vivo*, 24(4), pp 387-391. Available at: https://pubmed.ncbi.nlm.nih.gov/20668304/ (Accessed: 26 June 2024).

Bragg, W.L. and Thomson, J.J. (1913) 'The Diffraction of Short Electromagnetic Waves by a Crystal', *Proceedings of the Cambridge Philosophical Society*, 17(1), pp 43-57.

Brooijmans, N. and Kuntz, I.D. (2003) 'Molecular recognition and docking algorithms', *Annual Review of Biophysics and Biomolecular Structure*, 32(1), pp 335-373. Available at: https://doi.org/10.1146/annurev.biophys.32.110601.142532.

Broomfield, S., Chow, B.L. and Xiao, W. (1998) 'MMS 2, encoding a ubiquitin-conjugating-enzyme-like protein, is a member of the yeast error-free postreplication repair pathway', *Proceedings of the National Academy of Sciences of the USA*, 95(10), pp 5678-5783. Available at: https://doi.org/10.1073/pnas.95.10.5678.

Callis, J. (2014) 'The Ubiquitination Machinery of the Ubiquitin System', *Arabidopsis Book*, 12, e0174. Available at: https://doi.org/10.1199/tab.0174.

Cao, Y. et al. (2014) 'Selective small molecule compounds increase BMP-2 responsiveness by inhibiting Smurf1-mediated Smad1/5 degradation', *Scientific Reports*, 4(1), p 4965. Available at: https://doi.org/10.1038/srep04965.

Cappadocia, L. and Lima, C.D. (2018) 'Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism', *Chemical Reviews*, 118(3), pp 890-911. Available at: https://doi.org/10.1021/acs.chemrev.6b00737.

Centofanti, F. et al. (2023) 'Synthetic Methodologies and Therapeutic Potential of Indole-3-Carbinol (I3C) and Its Derivatives', *Pharmaceuticals*, 16(2), p 240. Available at: https://doi.org/10.3390/ph16020240.

Chan, W.C. et al. (2023) 'Accelerating inhibitor discovery for deubiquitinating enzymes', *Nature Communications*, 14(1), e686. Available at: https://doi.org/10.1038/s41467-023-36246-0.

Chang, Y. et al. (2023) 'A Guide to In Silico Drug Design', *Pharmaceutics*, 15(1), p 49. Available at: https://doi.org/10.3390/pharmaceutics15010049.

Chang, Y.C. et al. (1999) 'Cytostatic and antiestrogenic effects of 2-(indol-3-ylmethyl)-3,3'-diindolylmethane, a major in vivo product of dietary indole-3-carbinol', *Biochemical Pharmacology*, 58(5), pp 825-834. Available at: https://doi.org/10.1016/S0006-2952(99)00165-3.

Chantry, A. (2011) 'WWP2 ubiquitin ligase and its isoforms: New biological insight and promising disease targets', *Cell Cycle*, 10(15), pp 2437-2439. Available at: https://doi.org/10.4161/cc.10.15.16874.

Chao, W.R. et al. (2007) 'Computer-aided rational drug design: A novel agent (SR13668) designed to mimic the unique anticancer mechanisms of dietary indole-3-carbinol to block Akt signalling', *Journal of Medicinal Chemistry*, 50(15), pp 3412-3415. Available at: https://doi.org/10.1021/jm070040e.

Che, Y. et al. (2018) 'Inducing protein-protein interactions with molecular glues', *Bioorganic and Medicinal Chemistry Letters*, 28(15), pp 2585-2592. Available at: https://doi.org/10.1016/j.bmcl.2018.04.046.

Cheeseright, T. et al. (2006) 'Molecular field extreme as descriptors of biological activity: Definition and validation', *Journal of Chemical Information and Modelling*, 46(2), pp 665-676. Available at: https://doi.org/10.1021/ci050357s.

Chen, C. and Matesic, L.E. (2007) 'The Nedd4-like family of E3 ubiquitin ligases and cancer', *Cancer and Metastasis Reviews*, 26(3-4), pp 587-604. Available at: https://doi.org/10.1007/s10555-007-9091-x.

Chen, C. et al. (2007) 'The amplified WWP1 gene is a potential molecular target in breast cancer', International Journal of Cancer, 121(1), pp 80-87. Available at: https://doi.org/10.1002/ijc.22653.

Chen, C. et al. (2007) 'Ubiquitin E3 ligase WWP1 as an oncogenic factor in human prostate cancer', *Oncogene*, 26(16), pp 2386-2394. Available at: https://doi.org/10.1038/sj.onc.1210021.

Chen, D., Gehringer, M. and Lorenz, S. (2018) 'Developing Small-Molecule Inhibitors of HECT-Type Ubiquitin Ligases for Therapeutic Applications: Challenges and Opportunities', *ChemBioChem*, 19(20), pp 2123-2135. Available at: https://doi.org/10.1002/cbic.201800321.

Chen, J.-J. and Zhang, W. (2018) 'High expression of WWP1 predicts poor prognosis and associates with tumor progression in human colorectal cancer', *American Journal of Cancer Research*, 8(2), pp 256-265.

Chen, Z. et al. (2017) 'A Tunable Brake for HECT Ubiquitin Ligases', *Molecular Cell*, 66(3), pp 345-357. Available at: https://doi.org/10.1016/j.molcel.2017.03.020.

Cheng, F. et al. (2021) 'NEDD4 E3 ubiquitin protein ligase serves an important role in cutaneous melanoma occurrence and development', *Experimental and Therapeutic Medicine*, 22(6), p 1382. Available at: https://doi.org/10.3892/etm.2021.10818.

Choi, J.H. et al. (2011) 'Antidiabetic actions of a non-agonist PPARγ ligand blocking Cdk5-mediated phosphorylation', *Nature*, 477(7365), pp 477-481. Available at: https://doi.org/10.1038/nature10383.

Christopher, G.K., Phipps, A.G. and Gray, R.J. (1998) 'Temperature-dependent solubility of selected proteins', *Journal of Crystal Growth*, 191(4), pp 820-826. Available at: https://doi.org/10.1016/S0022-0248(98)00355-8.

Ciechanover, A. (2015) 'The unravelling of the ubiquitin system', *Nature Reviews Molecular Cell Biology*, 16(5), pp 322-324. Available at: https://doi.org/10.1038/nrm3982.

Ciechanover, A. and Ben-Saadon, R. (2004) 'N-terminal ubiquitination: More protein substrates join in', *Trends in Cell Biology*, 14(3), pp 103-106. Available at: https://doi.org/10.1016/j.tcb.2004.01.004.

Ciechanover, A., Hod, Y. and Hershko, A. (1978) 'A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes', *Biochemical and Biophysical Research Communications*, 81(4), pp 1100-1105. Available at: https://doi.org/10.1016/0006-291X(78)91249-4.

Clague, M.J. and Urbé, S. (2010) 'Ubiquitin: Same molecule, different degradation pathways', *Cell*, 143(5), pp 682-685. Available at: https://doi.org/10.1016/j.cell.2010.11.012.

Clague, M.J., Urbé, S. and Komander, D. (2019) 'Breaking the chains: deubiquitylating enzyme specificity begets function', *Nature Reviews Molecular Cell Biology*, 20(6), pp 338-352. Available at: https://doi.org/10.1038/s41580-019-0099-1.

Clements, A.E. et al. (2015) 'WWP2 and its association with PTEN in endometrial cancer', *Gynecologic Oncology Reports*, 13(1), pp 26-29. Available at: https://doi.org/10.1016/j.gore.2015.05.004.

Cohen, P. and Tcherpakov, M. (2010) 'Will the ubiquitin system furnish as many drug targets as protein kinases?', *Cell*, 143(5), pp 686-693. Available at: https://doi.org/10.1016/j.cell.2010.11.016.

Collins, P.M. et al. (2017) 'Gentle, fast and effective crystal soaking by acoustic dispensing', *Acta Crystallographica Section D: Structural Biology*, 73(3), pp 246-255. Available at: https://doi.org/10.1107/S205979831700331X.

Conole, D. et al. (2023) 'Discovery of a Potent Deubiquitinase (DUB) Small-Molecule Activity-Based Probe Enables Broad Spectrum DUB Activity Profiling in Living Cells**', *Angewandte Chemie - International Edition*, 62(47), p e202311190. Available at: https://doi.org/10.1002/anie.202311190.

Conway, J.A., Kinsman, G. and Kramer, E.R. (2022) 'The Role of NEDD4 E3 Ubiquitin–Protein Ligases in Parkinson's Disease', *Genes*, 13(3), p 513. Available at: https://doi.org/10.3390/genes13030513.

Damgaard, R.B. (2021) 'The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities', *Cell Death and Differentiation*, 28(2), pp 423-426. Available at: https://doi.org/10.1038/s41418-020-00703-w.

Daviet, L. and Colland, F. (2008) 'Targeting ubiquitin specific proteases for drug discovery', *Biochimie*, 90(2), pp 270-283. Available at: https://doi.org/10.1016/j.biochi.2007.09.013.

Delano, W.L. (2002) 'The PyMOL Molecular Graphics System', *CCP4 Newsletter on protein crystallography*, 40(1), pp 44-53. Available at: https://legacy.ccp4.ac.uk/newsletters/newsletter40.pdf

Derewenda, Z.S. (2004) 'Rational protein crystallization by mutational surface engineering', *Structure*, 12(4), pp 529-535. Available at: https://doi.org/10.1016/j.str.2004.03.008.

Deshaies, R.J. and Joazeiro, C.A.P. (2009) 'RING domain E3 ubiquitin ligases', *Annual Review of Biochemistry*, 78(1), pp 399-434. Available at: https://doi.org/10.1146/annurev.biochem.78.101807.093809.

Dikic, I., Wakatsuki, S. and Walters, K.J. (2009) 'Ubiquitin-binding domains from structures to functions', *Nature Reviews Molecular Cell Biology*, 10(10), pp 659-671. Available at: https://doi.org/10.1038/nrm2767.

Dou, H. *et al.* (2012) 'BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer', *Nature Structural and Molecular Biology*, 19(9), pp 876-883. Available at: https://doi.org/10.1038/nsmb.2379.

Douangamath, A. et al. (2021) 'Achieving Efficient Fragment Screening at XChem Facility at Diamond Light Source', *Journal of Visualized Experiments* [Internet], 171, e62414. Available at: https://doi.org/10.3791/62414.

Dove, K.K. and Klevit, R.E. (2017) 'RING-Between-RING E3 Ligases: Emerging Themes amid the Variations', *Journal of Molecular Biology*, 429(22), pp 3363-3375. Available at: https://doi.org/10.1016/j.jmb.2017.08.008.

Duda, D.M. et al. (2008) 'Structural Insights into NEDD8 Activation of Cullin-RING Ligases: Conformational Control of Conjugation', *Cell*, 134(6), pp 995-1006. Available at: https://doi.org/10.1016/j.cell.2008.07.022.

Emsley, P. et al. (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp 486-501. Available at: https://doi.org/10.1107/S0907444910007493.

Espinosa, J.R. et al. (2016) 'Seeding approach to crystal nucleation', *Journal of Chemical Physics* [Internet], 144(3), e034501. Available at: https://doi.org/10.1063/1.4939641.

Eyal, E. et al. (2005) 'The limit of accuracy of protein modelling: Influence of crystal packing on protein structure', *Journal of Molecular Biology*, 351(2), pp 431-442. Available at: https://doi.org/10.1016/j.jmb.2005.05.066.

Fischer, E.S. et al. (2014) 'Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide', *Nature*, 512(7512), pp 49-53. Available at: https://doi.org/10.1038/nature13527.

Flasza, M. et al. (2002) 'Alternative splicing determines the domain structure of WWP1, a Nedd4 family protein', *Biochemical and Biophysical Research Communications*, 290(1), pp 431-437. Available at: https://doi.org/10.1006/bbrc.2001.6206.

French, M.E., Kretzmann, B.R. and Hicke, L. (2009) 'Regulation of the RSP5 ubiquitin ligase by an intrinsic ubiquitin-binding site', *Journal of Biological Chemistry*, 284(18), pp 12071-12079. Available at: https://doi.org/10.1074/jbc.M901106200.

Friesner, R.A. et al. (2004) 'Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy', *Journal of Medicinal Chemistry*, 47(7), pp 1739-1749. Available at: https://doi.org/10.1021/jm0306430.

Fukumoto, C. et al. (2014) 'WWP2 is overexpressed in human oral cancer, determining tumor size and poor prognosis in patients: Downregulation of WWP2 inhibits the AKT signaling and tumor growth in mice', *Oncoscience*, 1(12), pp 807-820. Available at: https://doi.org/10.18632/oncoscience.101.

García-Ruiz, J.M. (2003) 'Nucleation of protein crystals', *Journal of Structural Biology*, 142(1), pp 22-31. Available at: https://doi.org/10.1016/S1047-8477(03)00035-2.

Geremia, S. et al. (2006) 'Simulation of diffusion time of small molecules in protein crystals', *Structure*, 14(3), pp 393-400. Available at: https://doi.org/10.1016/j.str.2005.12.007.

Goddard, T.D. et al. (2018) 'UCSF ChimeraX: Meeting modern challenges in visualization and analysis', *Protein Science*, 27(1), pp 14-25. Available at: https://doi.org/10.1002/pro.3235.

Goel, P., Manning, J.A. and Kumar, S. (2015) 'NEDD4-2 (NEDD4L): The ubiquitin ligase for multiple membrane proteins', *Gene*, 557(1), pp 1-10. Available at: https://doi.org/10.1016/j.gene.2014.11.051.

Goldstein, G. et al. (1975) 'Isolation of a polypeptide that has lymphocyte differentiating properties and is probably represented universally in living cells', *Proceedings of the National Academy of Sciences of the USA*, 72(1), pp 11-15. Available at: https://doi.org/10.1073/pnas.72.1.11.

Gong, W. et al. (2015) 'Structure of the HECT domain of human WWP2', *Acta Crystallographica Section: F Structural Biology Communications*, 71(10), pp 1251-1257. Available at: https://doi.org/10.1107/S2053230X1501554X.

Greenfield, N.J. (2007) 'Using circular dichroism spectra to estimate protein secondary structure', *Nature Protocols*, 1(6), pp 2876-2890. Available at: https://doi.org/10.1038/nprot.2006.202.

Groen, E.J.N. and Gillingwater, T.H. (2015) 'UBA1: At the Crossroads of Ubiquitin Homeostasis and Neurodegeneration', *Trends in Molecular Medicine*, 21(10), pp 622-632. Available at: https://doi.org/10.1016/j.molmed.2015.08.003.

Grose, K.R. and Bjeldanes, L.F. (1992) 'Oligomerization of Indole-3-carbinol in Aqueous Acid', *Chemical Research in Toxicology*, 5(2), pp 188-193. Available at: https://doi.org/10.1021/tx00026a007.

Guedes, I.A., de Magalhães, C.S. and Dardenne, L.E. (2014) 'Receptor-ligand molecular docking', *Biophysical Reviews*, 6(1), pp 75-87. Available at: https://doi.org/10.1007/s12551-013-0130-2.

Halgren, T.A. et al. (2004) 'Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening', *Journal of Medicinal Chemistry*, 47(7), pp 1750-1759. Available at: https://doi.org/10.1021/jm030644s.

Harder, E. et al. (2016) 'OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins', *Journal of Chemical Theory and Computation*, 12(1), pp 281-296. Available at: https://doi.org/10.1021/acs.jctc.5b00864.

Harrigan, J.A. et al. (2018) 'Deubiquitylating enzymes and drug discovery: Emerging opportunities', *Nature Reviews Drug Discovery*, 17(1), pp 57-78. Available at: https://doi.org/10.1038/nrd.2017.152.

Harris, C.J. et al. (2011) 'The Design and Application of Target-Focused Compound Libraries', *Combinatorial Chemistry and High Throughput Screening*, 14(6), pp 521-531. Available at: https://doi.org/10.2174/138620711795767802.

Hartshorn, M.J. et al. (2005) 'Fragment-based lead discovery using X-ray crystallography', *Journal of Medicinal Chemistry*, 48(2), pp 403-413. Available at: https://doi.org/10.1021/jm0495778.

Hay, R.T. (2005) 'SUMO: A history of modification', *Molecular Cell*, 18(1), pp 1-12. Available at: https://doi.org/10.1016/j.molcel.2005.03.012.

Hayes, J.D., Kelleher, M.O. and Eggleston, I.M. (2008) 'The cancer chemopreventive actions of phytochemicals derived from glucosinolates', *European Journal of Nutrition*, 47(2), pp 73-88. Available at: https://doi.org/10.1007/s00394-008-2009-8.

Henning, N.J. et al. (2022) 'Deubiquitinase-targeting chimeras for targeted protein stabilization', *Nature Chemical Biology*, 18(4), pp 412-421. Available at: https://doi.org/10.1038/s41589-022-00971-2.

Hershko, A. et al. (1980) 'Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis.', *Proceedings of the National Academy of Sciences of the USA*, 77(4), pp 1783-1786. Available at: https://doi.org/10.1073/pnas.77.4.1783.

Hershko, A., Ciechanover, A. and Rose, I.A. (1979) 'Resolution of the ATP dependent proteolytic system from reticulocytes: A component that interacts with ATP', *Proceedings of the National Academy of Sciences of the USA*, 76(7), pp 3107-3110. Available at: https://doi.org/10.1073/pnas.76.7.3107.

Higgins, J.J. et al. (2008) 'Dysregulation of large-conductance Ca2+-activated K+ channel expression in nonsyndromal mental retardation due to a cereblon p.R419X mutation', *Neurogenetics*, 9(3), pp 219-223. Available at: https://doi.org/10.1007/s10048-008-0128-2.

Hjerpe, R. et al. (2009) 'Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities', *EMBO Reports*, 10(11), pp 1250-1258. Available at: https://doi.org/10.1038/embor.2009.192.

Holdgate, G. et al. (2019) 'Biophysical methods in early drug discovery', *ADMET and DMPK*, 7(4), pp 222-241. Available at: https://doi.org/10.5599/admet.733.

Hou, H. et al. (2019) 'A systematic comparison of sitting and hanging-drop crystallization using traditional and cross-diffusion microbatch crystallization plates', *Journal of Crystal Growth*, 521(1), pp 1-8. Available at: https://doi.org/10.1016/j.jcrysgro.2019.05.011.

Hu, H. et al. (2018) 'Genome-wide identification and analysis of U-box E3 ubiquitin–protein ligase gene family in banana', *International Journal of Molecular Sciences*, 19(12), p 3874. Available at: https://doi.org/10.3390/ijms19123874.

Hu, X. et al. (2021) 'The emerging role of WWP1 in cancer development and progression', *Cell Death Discovery*, 7(1), p 163. Available at: https://doi.org/10.1038/s41420-021-00532-x.

Huang, L. et al. (1999) 'Structure of an E6AP-UbcH7 complex: Insights into ubiquitination by the E2-E3 enzyme cascade', *Science*, 286(5443), pp 1321-1326. Available at: https://doi.org/10.1126/science.286.5443.1321.

Hughes, G. (2019) Development of WWP2 ubiquitin ligase inhibitors using biophysical, structural and synthetic approaches. Ph. D. Thesis. University of East Anglia. Available at: https://ueaeprints.uea.ac.uk/id/eprint/74200/ (Accessed: 22 June 2024).

Hughes, G.R. et al. (2021) 'Frontiers in PROTACs', *Drug Discovery Today*, 26(10), pp 2377-2383. Available at: https://doi.org/10.1016/j.drudis.2021.04.010.

Hughes, J.P. et al. (2011) 'Principles of early drug discovery', *British Journal of Pharmacology*, 162(6), pp 1239-1249. Available at: https://doi.org/10.1111/j.1476-5381.2010.01127.x.

Huibregtse, J.M. *et al.* (1995) 'A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase', *Proceedings of the National Academy of Sciences of the United States of America*, 92(7), pp 2563-2567. Available at: https://doi.org/10.1073/pnas.92.7.2563.

Huynh, K. and Partch, C.L. (2015) 'Analysis of protein stability and ligand interactions by thermal shift assay', *Current Protocols in Protein Science*, 79(1), pp 28901-28914. Available at: https://doi.org/10.1002/0471140864.ps2809s79.

Ingham, R.J. et al. (2005) 'WW Domains Provide a Platform for the Assembly of Multiprotein Networks', *Molecular and Cellular Biology*, 25(16), pp 7092-7106. Available at: https://doi.org/10.1128/mcb.25.16.7092-7106.2005.

Irwin, J.J. et al. (2015) 'An Aggregation Advisor for Ligand Discovery', *Journal of Medicinal Chemistry*, 58(17), pp 7076-7087. Available at: https://doi.org/10.1021/acs.jmedchem.5b01105.

Ishii, K., Noda, M. and Uchiyama, S. (2016) 'Mass spectrometric analysis of protein–ligand interactions', *Biophysics and Physicobiology*, 13(1), pp 87-95. Available at: https://doi.org/10.2142/biophysico.13.0_87.

Ito, T. et al. (2010) 'Identification of a primary target of thalidomide teratogenicity', *Science*, 327(5971), pp 1345-1350. Available at: https://doi.org/10.1126/science.1177319.

Jadhav, A. et al. (2010) 'Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease', *Journal of Medicinal Chemistry*, 53(1), pp 37-51. Available at: https://doi.org/10.1021/jm901070c.

Johnston, R.C. et al. (2023) 'Epik: pKa and Protonation State Prediction through Machine Learning', *Journal of Chemical Theory and Computation*, 19(8), pp 2380-2388. Available at: https://doi.org/10.1021/acs.jctc.3c00044.

Jorgensen, W. and Tirado-Rives, J. (1988) 'The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin', *Journal of American Chemical Society*, 110(6), pp 1657-1666. Available at: https://doi.org/10.1021/ja00214a001.

Jorgensen, W.L., Maxwell, D.S. and Tirado-Rives, J. (1996) 'Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids', *Journal of the American Chemical Society*, 118(45), pp 11225-11236. Available at: https://doi.org/10.1021/ja9621760.

Josephy, P.D., Eling, T. and Mason, R.P. (1982) 'The horseradish peroxidase-catalyzed oxidation of 3,5,3',5'-tetramethylbenzidine. Free radical and charge-transfer complex intermediates', *Journal of Biological Chemistry*, 257(7), pp 3669-3675. Available at: https://doi.org/10.1016/s0021-9258(18)34832-4.

Judson, R.S., Jaeger, E.P. and Treasurywala, A.M. (1994) 'A genetic algorithm based method for docking flexible molecules', *Journal of Molecular Structure: THEOCHEM*, 308(1), pp 191-206. Available at: https://doi.org/10.1016/0166-1280(94)80102-9.

Kadimisetty, K. et al. (2021) 'Tandem Ubiquitin Binding Entities (TUBEs) as Tools to Explore Ubiquitin-Proteasome System and PROTAC Drug Discovery', *Methods in Molecular Biology*, 2365(1), pp 185-202. Available at: https://doi.org/10.1007/978-1-0716-1665-9_10.

Kamadurai, H.B. et al. (2009) 'Insights into Ubiquitin Transfer Cascades from a Structure of a UbcH5B~Ubiquitin-HECTNEDD4L Complex', *Molecular Cell*, 36(6), pp 1095-1102. Available at: https://doi.org/10.1016/j.molcel.2009.11.010.

Kathman, S.G. et al. (2015) 'A Small Molecule That Switches a Ubiquitin Ligase from a Processive to a Distributive Enzymatic Mechanism', *Journal of the American Chemical Society*, 137(39), pp 12442-12445. Available at: https://doi.org/10.1021/jacs.5b06839.

Kathman, S.G., Xu, Z. and Statsyuk, A. V. (2014) 'A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases', *Journal of Medicinal Chemistry*, 57(11), pp 4969-4974. Available at: https://doi.org/10.1021/jm500345q.

Kazi, A. et al. (2009) 'Discovery of a novel proteasome inhibitor selective for cancer cells over non-transformed cells', *Cell Cycle*, 8(12), pp 1940-1951. Available at: https://doi.org/10.4161/cc.8.12.8798.

Kelsall, I.R. et al. (2022) 'HOIL-1 ubiquitin ligase activity targets unbranched glucosaccharides and is required to prevent polyglucosan accumulation', *The EMBO Journal*, 41(8), p 109700. Available at: https://doi.org/10.15252/embj.2021109700.

Kerscher, O., Felberbaum, R. and Hochstrasser, M. (2006) 'Modification of proteins by ubiquitin and ubiquitin-like proteins', *Annual Review of Cell and Developmental Biology*, 22(1), pp 159-180. Available at: https://doi.org/10.1146/annurev.cellbio.22.010605.093503.

Kim, H.C. et al. (2011) 'Structure and function of a HECT domain ubiquitin-binding site', *EMBO Reports*, 12(4), pp 334-341. Available at: https://doi.org/10.1038/embor.2011.23.

Komander, D. (2009) 'The emerging complexity of protein ubiquitination', *Biochemical Society Transactions*, 37(5), pp 937-953. Available at: https://doi.org/10.1042/BST0370937.

Komander, D. and Rape, M. (2012) 'The ubiquitin code', *Annual Review of Biochemistry*, 81(1), pp 203-229. Available at: https://doi.org/10.1146/annurev-biochem-060310-170328.

Komander, D., Clague, M.J. and Urbé, S. (2009) 'Breaking the chains: Structure and function of the deubiquitinases', *Nature Reviews Molecular Cell Biology*, 10(8), pp 550-563. Available at: https://doi.org/10.1038/nrm2731.

Krishnamoorthy, G.K. et al. (2020) 'Isothermal titration calorimetry and surface plasmon resonance analysis using the dynamic approach', *Biochemistry and Biophysics Reports*, 21(1), p 100712. Available at: https://doi.org/10.1016/j.bbrep.2019.100712.

Krönke, J. et al. (2014) 'Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells', *Science*, 343(6168), pp 301-305. Available at: https://doi.org/10.1126/science.1244851.

Kwon, Y.T. and Ciechanover, A. (2017) 'The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy', *Trends in Biochemical Sciences*, 42(11), pp 873-886. Available at: https://doi.org/10.1016/j.tibs.2017.09.002.

Lahlou, M. (2013) 'The Success of Natural Products in Drug Discovery', *Pharmacology and Pharmacy*, 4(3), pp 17-23. Available at: https://doi.org/10.4236/pp.2013.43a003.

Lambert-Smith, I.A., Saunders, D.N. and Yerbury, J.J. (2020) 'The pivotal role of ubiquitin-activating enzyme E1 (UBA1) in neuronal health and neurodegeneration', *International Journal of Biochemistry and Cell Biology* [Internet], 123, e105746. Available at: https://doi.org/10.1016/j.biocel.2020.105746.

Lange, S.M., Armstrong, L.A. and Kulathu, Y. (2022) 'Deubiquitinases: From mechanisms to their inhibition by small molecules', *Molecular Cell*, 82(1), pp 15-29. Available at: https://doi.org/10.1016/j.molcel.2021.10.027.

Lee, J. et al. (2022) 'Discovery of E3 Ligase Ligands for Target Protein Degradation', *Molecules*, 27(19), p 6515. Available at: https://doi.org/10.3390/molecules27196515.

Lee, Y.R. et al. (2019) 'Reactivation of PTEN tumor suppressor for cancer treatment through inhibition of a MYC-WWP1 inhibitory pathway', *Science*, 364(6441), eaau0159. Available at: https://doi.org/10.1126/science.aau0159.

Lee, Y.-R. et al. (2020) 'WWP1 Gain-of-Function Inactivation of PTEN in Cancer Predisposition', *New England Journal of Medicine*, 382(22), pp 2103-2116. Available at: https://doi.org/10.1056/nejmoa1914919.

Lees, J.A., Dias, J.M. and Han, S. (2021) 'Applications of Cryo-EM in small molecule and biologics drug design', *Biochemical Society Transactions*, 49(6), pp 2627-2638. Available at: https://doi.org/10.1042/BST20210444.

Leung, C.S. et al. (2012) 'Methyl effects on protein-ligand binding', *Journal of Medicinal Chemistry*, 55(9), pp 4489-4500. Available at: https://doi.org/10.1021/jm3003697.

Li, Y. et al. (2009) ‘WW domain containing E3 ubiquitin protein ligase 1 targets the full-length ErbB4 for ubiquitin-mediated degradation in breast cancer’, *Oncogene*, 28(33), pp 2948-2958. Available at: https://doi.org/10.1038/onc.2009.162.

Li, Y., Zhou, Z. and Chen, C. (2008) ‘WW domain-containing E3 ubiquitin protein ligase 1 targets p63 transcription factor for ubiquitin-mediated proteasomal degradation and regulates apoptosis’, *Cell Death and Differentiation*, 15(12), pp 1941-1951. Available at: https://doi.org/10.1038/cdd.2008.134.

Liebschner, D. et al. (2017) ‘Polder maps: Improving OMIT maps by excluding bulk solvent’, *Acta Crystallographica Section D: Structural Biology*, 73(2), pp 148-157. Available at: https://doi.org/10.1107/S2059798316018210.

Liebschner, D. et al. (2019) ‘Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix’, *Acta Crystallographica Section D: Structural Biology*, 75(10), pp 861-877. Available at: https://doi.org/10.1107/S2059798319011471.

Lin, J.H. et al. (2013) ‘WWP1 gene is a potential molecular target of human oral cancer’, *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 116(2), pp 221-231. Available at: https://doi.org/10.1016/j.oooo.2013.05.006.

Lin, X.T., Zhang, J. and Xie, C.M. (2023) ‘An optimized protocol to detect protein ubiquitination and activation by ubiquitination assay in vivo and CCK-8 assay’, *STAR Protocols* [Internet], 4(2), e102199. Available at: https://doi.org/10.1016/j.xpro.2023.102199.

Lipinski, C.A. et al. (2001) ‘Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings’, *Advanced Drug Delivery Reviews*, 46(1–3), pp 3-26. Available at: https://doi.org/10.1016/S0169-409X(00)00129-0.

Lv, Z. et al. (2017) ‘Domain alternation and active site remodelling are conserved structural features of ubiquitin E1’, *Journal of Biological Chemistry*, 292(29), pp 12089-12099. Available at: https://doi.org/10.1074/jbc.M117.787622.

Lv, Z. et al. (2018) ‘Crystal structure of a human ubiquitin E1– ubiquitin complex reveals conserved functional elements essential for activity’, *Journal of Biological Chemistry*, 293(47), pp 18337-18352. Available at: https://doi.org/10.1074/jbc.RA118.003975.

Maddika, S. et al. (2011) ‘WWP2 is an E3 ubiquitin ligase for PTEN’, *Nature Cell Biology*, 13(6), pp 728-733. Available at: https://doi.org/10.1038/ncb2240.

Madhavi Sastry, G. et al. (2013) ‘Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments’, *Journal of Computer-Aided Molecular Design*, 27(3), pp 221-234. Available at: https://doi.org/10.1007/s10822-013-9644-8.

Madiraju, C. et al. (2012) ‘TR-FRET-based high-throughput screening assay for identification of UBC13 inhibitors’, *Journal of Biomolecular Screening*, 17(2), pp 163-176. Available at: https://doi.org/10.1177/1087057111423417.

Maki, S. and Hagiwara, M. (2022) ‘Contactless crystallization method of protein by a magnetic force booster’, *Scientific Reports*, 12(1), e17287. Available at: https://doi.org/10.1038/s41598-022-21727-x.

Marblestone, J.G. et al. (2010) ‘Novel approach for characterizing ubiquitin E3 ligase function’, *Journal of Biomolecular Screening*, 15(10), pp 1220-1228. Available at: https://doi.org/10.1177/1087057110380456.

Marblestone, J.G. et al. (2012) ‘Analysis of ubiquitin E3 ligase activity using selective polyubiquitin binding proteins’, *Biochimica et Biophysica Acta*, 1823(11), pp 2094-2097. Available at: https://doi.org/10.1016/j.bbamcr.2012.06.013.

Marciano, D.P., Dharmarajan, V. and Griffin, P.R. (2014) ‘HDX-MS guided drug discovery: Small molecules and biopharmaceuticals’, *Current Opinion in Structural Biology*, 28(1), pp 105-111. Available at: https://doi.org/10.1016/j.sbi.2014.08.007.

Mari, S. et al. (2014) ‘Structural and functional framework for the autoinhibition of nedd4-family ubiquitin ligases’, *Structure*, 22(11), pp 1639-1649. Available at: https://doi.org/10.1016/j.str.2014.09.006.

Marin, I. (2010) ‘Animal HECT ubiquitin ligases: Evolution and functional implications’, *BMC Evolutionary Biology*, 10(1), p 56. Available at: https://doi.org/10.1186/1471-2148-10-56.

Marousis, K.D. et al. (2020) ‘1H, 13C, 15N backbone and side-chain resonance assignment of the native form of UbcH7 (UBE2L3) through solution NMR spectroscopy’, *Biomolecular NMR Assignments*, 14(1), pp 73-78. Available at: https://doi.org/10.1007/s12104-019-09923-9.

Martín-Gago, P. et al. (2017) ‘Covalent Protein Labeling at Glutamic Acids’, *Cell Chemical Biology*, 24(5), pp 589-597. Available at: https://doi.org/10.1016/j.chembiol.2017.03.015.

Maspero, E. et al. (2011) ‘Structure of the HECT:ubiquitin complex and its role in ubiquitin chain elongation’, *EMBO Reports*, 12(4), pp 342-349. Available at: https://doi.org/10.1038/embor.2011.21.

Maspero, E. et al. (2013) ‘Structure of a ubiquitin-loaded HECT ligase reveals the molecular basis for catalytic priming’, *Nature Structural and Molecular Biology*, 20(6), pp 696-701. Available at: https://doi.org/10.1038/nsmb.2566.

Matinyan, S., Filipcik, P. and Abrahams, J.P. (2024) ‘Deep learning applications in protein crystallography’, *Acta Crystallographica Section A: Foundations and Advances*, 80(1), pp 1-17. Available at: https://doi.org/10.1107/S2053273323009300.

Matsumoto, M.L. et al. (2018) ‘Interpreting the language of polyubiquitin with linkage-specific antibodies and mass spectrometry’, *Methods in Molecular Biology*, 1844(1), pp 385-400. Available at: https://doi.org/10.1007/978-1-4939-8706-1_24.

Mattern, M. et al. (2019) 'Using Ubiquitin Binders to Decipher the Ubiquitin Code', *Trends in Biochemical Sciences*, 44(7), pp 599-615. Available at: https://doi.org/10.1016/j.tibs.2019.01.011.

McCoy, A.J. et al. (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(4), pp 658-674. Available at: https://doi.org/10.1107/S0021889807021206.

McCoy, A.J., Sammito, M.D. and Read, R.J. (2022) 'Implications of AlphaFold2 for crystallographic phasing by molecular replacement', *Acta Crystallographica. Section D, Structural Biology*, 78(1), pp 1-13. Available at: https://doi.org/10.1107/S2059798321012122.

McDowell, G.S. and Philpott, A. (2013) 'Non-canonical ubiquitylation: Mechanisms and consequences', *International Journal of Biochemistry and Cell Biology*, 45(8), pp 1833-1842. Available at: https://doi.org/10.1016/j.biocel.2013.05.026.

McGovern, S.L. et al. (2002) 'A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening', *Journal of Medicinal Chemistry*, 45(8), pp 1712-1722. Available at: https://doi.org/10.1021/jm010533y.

McPherson, A. and Gavira, J.A. (2014) 'Introduction to protein crystallization', *Acta Crystallographica Section F: Structural Biology Communications*, 70(1), pp 2-20. Available at: https://doi.org/10.1107/S2053230X13033141.

Meng, E.C. et al. (2023) 'UCSF ChimeraX: Tools for structure building and analysis', *Protein Science*, 32(11), p 4792. Available at: https://doi.org/10.1002/pro.4792.

Meng, X.-Y. et al. (2012) 'Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery', *Current Computer Aided-Drug Design*, 7(2), pp 146-157. Available at: https://doi.org/10.2174/157340911795677602.

Metzger, M.B. et al. (2014) 'RING-type E3 ligases: Master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination', *Biochimica et Biophysica Acta*, 1843(1), pp 47-60. Available at: https://doi.org/10.1016/j.bbamcr.2013.05.026.

Metzger, M.B., Hristova, V.A. and Weissman, A.M. (2012) 'HECT and RING finger families of E3 ubiquitin ligases at a glance', *Journal of Cell Science*, 125(3), pp 531-537. Available at: https://doi.org/10.1242/jcs.091777.

Mohamadi, F. et al. (1990) 'Macromodel—an integrated software system for modelling organic and bioorganic molecules using molecular mechanics', *Journal of Computational Chemistry*, 11(4), pp 440-467. Available at: https://doi.org/10.1002/jcc.540110405.

Mongkol, W. et al. (2015) 'Active Compounds Against Anopheles minimus Carboxypeptidase B for Malaria Transmission-Blocking Strategy', *Journal of Medical Entomology*, 52(6), pp 1322-1332. Available at: https://doi.org/10.1093/jme/tjv133.

Mund, T. et al. (2014) ‘Peptide and small molecule inhibitors of HECT-type ubiquitin ligases’, *Proceedings of the National Academy of Sciences of the USA*, 111(47), pp 16736-16741. Available at: https://doi.org/10.1073/pnas.1412152111.

Murray, A.M. et al. (1998) ‘Production of glutathione-coated microtitre plates for capturing recombinant glutathione S-transferase fusion proteins as antigens in immunoassays’, *Journal of Immunological Methods*, 218(1–2), pp 133-139. Available at: https://doi.org/10.1016/S0022-1759(98)00114-8.

National Center for Biotechnology Information (2024). *PubChem Compound Summary for CID 72571, Redoxal*. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Redoxal (Accessed: 24 May 2024).

National Center for Biotechnology Information (2024). *PubChem Compound Summary for CID 5358027, ethyl 2-[(5,6-dichloro-1H-imidazo[4,5-b]pyrazin-2-yl)sulfanyl]acetate*. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/5358027 (Accessed: 23 May 2024).

Nguyen, H.C., Wang, W. and Xiong, Y. (2017) ‘Cullin-RING E3 ubiquitin ligases: Bridges to destruction’, *Sub-Cellular Biochemistry*, 83(1), pp 323-347. Available at: https://doi.org/10.1007/978-3-319-46503-6_12.

Nguyen, H.H. et al. (2008) ‘The dietary phytochemical indole-3-carbinol is a natural elastase enzymatic inhibitor that disrupts cyclin e protein processing’, *Proceedings of the National Academy of Sciences of the USA*, 105(50), pp 19750-19755. Available at: https://doi.org/10.1073/pnas.0806581105.

Niphakis, M.J. and Cravatt, B.F. (2024) ‘Ligand discovery by activity-based protein profiling.’, *Cell chemical biology*, 31(9), pp. 1636–1651. Available at: https://doi.org/10.1016/j.chembiol.2024.08.006.

Novelli, G. et al. (2021) ‘Inhibition of HECT E3 ligases as potential therapy for COVID-19’, *Cell Death and Disease*, 12(4), e310. Available at: https://doi.org/10.1038/s41419-021-03513-1.

Ogasawara, D. et al. (2024) ‘Chemical tools to expand the ligandable proteome: diversity-oriented synthesis-based photoreactive stereoprobes’, *bioRxiv*, 31(12), pp 2138-2155. Available at: https://doi.org/10.1016/j.chembiol.2024.10.005.

Ogunjimi, A.A. et al. (2010) ‘The ubiquitin binding region of the smurf HECT domain facilitates polyubiquitylation and binding of ubiquitylated substrates’, *Journal of Biological Chemistry*, 285(9), 6308-6315. Available at: https://doi.org/10.1074/jbc.M109.044537.

Ohori, M. et al. (2007) ‘Role of a cysteine residue in the active site of ERK and the MAPKK family’, *Biochemical and Biophysical Research Communications*, 353(3), pp 633-637. Available at: https://doi.org/10.1016/j.bbrc.2006.12.083.

Ohtake, F. et al. (2015) ‘Ubiquitin acetylation inhibits polyubiquitin chain elongation’, *EMBO reports*, 16(2), pp 192-201. Available at: https://doi.org/10.15252/embr.201439152.

Olsen, S.K. and Lima, C.D. (2013) 'Structure of a Ubiquitin E1-E2 Complex: Insights to E1-E2 Thioester Transfer', *Molecular Cell*, 49(5), pp 884-896. Available at: https://doi.org/10.1016/j.molcel.2013.01.013.

Omar, H.A. et al. (2009) 'Targeting of the Akt-nuclear factor-κB signaling network by [1-(4-chloro-3-nitrobenzenesulfonyl)-1H-indol-3-yl]-methanol (OSU-A9), a novel indole-3-carbinol derivative, in a mouse model of hepatocellular carcinoma', *Molecular Pharmacology*, 76(5), pp 957-968. Available at: https://doi.org/10.1124/mol.109.058180.

Otten, E.G. et al. (2021) 'Ubiquitylation of lipopolysaccharide by RNF213 during bacterial infection', *Nature*, 594(7861), pp 111-116. Available at: https://doi.org/10.1038/s41586-021-03566-4.

Otvos, L. and Wade, J.D. (2014) 'Current challenges in peptide-based drug discovery', *Frontiers in Chemistry* [Internet], 2, e62. Available at: https://doi.org/10.3389/fchem.2014.00062.

Pantoliano, M.W. et al. (2001) 'High-density miniaturized thermal shift assays as a general strategy for drug discovery', *Journal of Biomolecular Screening*, 6(6), pp 429-440. Available at: https://doi.org/10.1089/108705701753364922.

Passmore, L.A. and Barford, D. (2004) 'Getting into position: The catalytic mechanisms of protein of protein ubiquitylation', *Biochemical Journal*, 379(3), pp 513-525. Available at: https://doi.org/10.1042/BJ20040198.

Pastushok, L. et al. (2005) 'A single Mms2 "key" residue insertion into a Ubc13 pocket determines the interface specificity of a human Lys63 ubiquitin conjugation complex', *Journal of Biological Chemistry*, 280(18), pp 17891-17900. Available at: https://doi.org/10.1074/jbc.M410469200.

Patani, G.A. and LaVoie, E.J. (1996) 'Bioisosterism: A rational approach in drug design', *Chemical Reviews*, 96(8), pp 3147-3176. Available at: https://doi.org/10.1021/cr950066q.

Pei, H. et al. (2019) 'Small molecule PROTACs: An emerging technology for targeted therapy in drug discovery', *RSC Advances*, 9(30), pp 16967-16976. Available at: https://doi.org/10.1039/c9ra03423d.

Pettersen, E.F. et al. (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science*, 30(1), pp 70-82. Available at: https://doi.org/10.1002/pro.3943.

Pflugrath, J.W. (2015) 'Practical macromolecular cryocrystallography', *Acta Crystallographica Section F: Structural Biology Communications*, 71(6), pp 622-642. Available at: https://doi.org/10.1107/S2053230X15008304.

Pirozzi, G. et al. (1997) 'Identification of novel human WW domain-containing proteins by cloning of ligand targets', *Journal of Biological Chemistry*, 272(23), pp 14611-14616. Available at: https://doi.org/10.1074/jbc.272.23.14611.

Plant, P.J. et al. (2000) ‘Apical membrane targeting of Nedd4 is mediated by an association of its C2 domain with annexin XIIIb’, *Journal of Cell Biology*, 149(7), pp 1473-1484. Available at: https://doi.org/10.1083/jcb.149.7.1473.

Powers, J.C. et al. (2002) ‘Irreversible inhibitors of serine, cysteine, and threonine proteases’, *Chemical Reviews*, 102(12), pp 4639-4750. Available at: https://doi.org/10.1021/cr010182v.

Prado, N.J. et al. (2022) ‘Anti-inflammatory, antioxidant, antihypertensive, and antiarrhythmic effect of indole-3-carbinol, a phytochemical derived from cruciferous vegetables’, *Heliyon* [Internet], 8(2), e08989. Available at: https://doi.org/10.1016/j.heliyon.2022.e08989.

Pruneda, J.N. et al. (2011) ‘Ubiquitin in motion: Structural studies of the ubiquitin-conjugating enzyme∼ubiquitin conjugate’, *Biochemistry*, 50(10), pp 1624-1633. Available at: https://doi.org/10.1021/bi101913m.

Qian, Y.W. et al. (2012) ‘P28GANK prevents degradation of Oct4 and promotes expansion of tumor-initiating cells in hepatocarcinogenesis’, *Gastroenterology*, 142(7), pp 1547-1558. Available at: https://doi.org/10.1053/j.gastro.2012.02.042.

Quach, D. et al. (2021) ‘Strategic Design of Catalytic Lysine-Targeting Reversible Covalent BCR-ABL Inhibitors**’, *Angewandte Chemie - International Edition*, 60(31), pp 17131-17137. Available at: https://doi.org/10.1002/anie.202105383.

Quirit, J.G. et al. (2017) ‘Indole-3-carbinol (I3C) analogues are potent small molecule inhibitors of NEDD4-1 ubiquitin ligase activity that disrupt proliferation of human melanoma cells’, *Biochemical Pharmacology*, 127(1), pp 13-27. Available at: https://doi.org/10.1016/j.bcp.2016.12.007.

Reyes-Hernández, O.D. et al. (2023) ‘3,3′-Diindolylmethane and indole-3-carbinol: potential therapeutic molecules for cancer chemoprevention and treatment via regulating cellular signalling pathways’, *Cancer Cell International* [Internet], 23, e180. Available at: https://doi.org/10.1186/s12935-023-03031-4.

Rhodes, G. (2006) ‘*Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*’ 3rd Ed, Oxford: Academic Press. Available at: https://doi.org/10.1016/B978-0-12-587073-3.X5000-4.

Rigby, J. (2024) ‘The Synthesis and Evaluation of WWP2 E3 Ubiquitin Ligase Inhibitors as Anti-Cancer Leads’. Ph. D. Thesis (unpublished). University of East Anglia.

Rizo, J. and Sudhof, T.C. (1998) ‘C2-domains, structure and function of a universal Ca2+-binding domain’, *Journal of Biological Chemistry*, 273(26), pp 15879-15882. Available at: https://doi.org/10.1074/jbc.273.26.15879.

Roos, K. et al. (2019) ‘OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules’, *Journal of Chemical Theory and Computation*, 15(3), pp 1863-1874. Available at: https://doi.org/10.1021/acs.jctc.8b01026.

Rossi, M. et al. (2014) 'High throughput screening for inhibitors of the HECT ubiquitin E3 ligase ITCH identifies antidepressant drugs as regulators of autophagy', *Cell Death and Disease*, 5(5), e1203. Available at: https://doi.org/10.1038/cddis.2014.113.

Rotin, D. and Kumar, S. (2009) 'Physiological functions of the HECT family of ubiquitin ligases', *Nature Reviews Molecular Cell Biology*, 10(6), pp 398-409. Available at: https://doi.org/10.1038/nrm2690.

Sakamaki, J. ichi et al. (2022) 'Ubiquitination of phosphatidylethanolamine in organellar membranes', *Molecular Cell*, 82(19), pp 3677-3692. Available at: https://doi.org/10.1016/j.molcel.2022.08.008.

Sakamoto, K.M. et al. (2001) 'Protacs: Chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation', *Proceedings of the National Academy of Sciences of the USA*, 98(15), pp 8554-8559. Available at: https://doi.org/10.1073/pnas.141230798.

Sakata, E. et al. (2010) 'Crystal Structure of UbcH5b∼Ubiquitin Intermediate: Insight into the Formation of the Self-Assembled E2∼Ub Conjugates', *Structure*, 18(1), pp 138-147. Available at: https://doi.org/10.1016/j.str.2009.11.007.

Sanarico, A.G. et al. (2018) 'The E3 ubiquitin ligase WWP1 sustains the growth of acute myeloid leukaemia', *Leukemia*, 32(4), pp 911-919. Available at: https://doi.org/10.1038/leu.2017.342.

Sato, Y. et al. (2008) 'Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains', *Nature*, 455(7211), pp 358-362. Available at: https://doi.org/10.1038/nature07254.

Schäfer, A., Kuhn, M. and Schindelin, H. (2014) 'Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules', *Acta Crystallographica Section D: Biological Crystallography*, 70(5), pp 1311-1320. Available at: https://doi.org/10.1107/S1399004714002910.

Scheffner, M. and Kumar, S. (2014) 'Mammalian HECT ubiquitin-protein ligases: Biological and pathophysiological aspects', *Biochimica et Biophysica Acta*, 1843(1), pp 64-74. Available at: https://doi.org/10.1016/j.bbamcr.2013.03.024.

Scheffner, M. and Staub, O. (2007) 'HECT E3s and human disease', *BMC Biochemistry*, 8(1), S6. Available at: https://doi.org/10.1186/1471-2091-8-S1-S6.

Schlander, M. et al. (2021) 'How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment', *PharmacoEconomics*, 39(11), pp 1243-1269. Available at: https://doi.org/10.1007/s40273-021-01065-y.

Schmidt, T.G.M. and Skerra, A. (2007) 'The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins', *Nature Protocols*, 2(6), pp 1528-1535. Available at: https://doi.org/10.1038/nprot.2007.209.

Schulman, B.A. and Wade Harper, J. (2009) 'Ubiquitin-like protein activation by E1 enzymes: The apex for downstream signalling pathways', *Nature Reviews Molecular Cell Biology*, 10(5), pp 319-331. Available at: https://doi.org/10.1038/nrm2673.

Scott, D.C. et al. (2014) 'Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8', *Cell*, 157(7), pp 1671-1684. Available at: https://doi.org/10.1016/j.cell.2014.04.037.

Seo, S.R. et al. (2004) 'The novel E3 ubiquitin ligase Tiul1 associates with TGIF to target Smad2 for degradation', *EMBO Journal*, 23(19), pp 3780-3792. Available at: https://doi.org/10.1038/sj.emboj.7600398.

Shah, S.S. and Kumar, S. (2021) 'Adaptors as the regulators of HECT ubiquitin ligases', *Cell Death and Differentiation*, 28(2), pp 455-472. Available at: https://doi.org/10.1038/s41418-020-00707-6.

Shao, G. et al. (2018) 'The E3 ubiquitin ligase NEDD4 mediates cell migration signaling of EGFR in lung cancer cells', *Molecular Cancer*, 17(1), p 24. Available at: https://doi.org/10.1186/s12943-018-0784-2.

Shi, L. and Zhang, N. (2021) 'Applications of solution nmr in drug discovery', *Molecules*, 26(3), p 576. Available at: https://doi.org/10.3390/molecules26030576.

Shi, Y. and von Itzstein, M. (2019) 'How size matters: Diversity for fragment library design', *Molecules*, 24(15), p 2838. Available at: https://doi.org/10.3390/molecules24152838.

Shimizu, Y., Okuda-Shimizu, Y. and Hendershot, L.M. (2010) 'Ubiquitylation of an ERAD Substrate Occurs on Multiple Types of Amino Acids', *Molecular Cell*, 40(6), pp 917-926. Available at: https://doi.org/10.1016/j.molcel.2010.11.033.

Shivakumar, D. et al. (2010) 'Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field', *Journal of Chemical Theory and Computation*, 6(5), pp 1509-1519. Available at: https://doi.org/10.1021/ct900587b.

Shrestha, R.K. et al. (2014) 'Insights into the mechanism of deubiquitination by jamm deubiquitinases from cocrystal structures of the enzyme with the substrate and product', *Biochemistry*, 53(19), pp 3199-3217. Available at: https://doi.org/10.1021/bi5003162.

Sigismund, S. and Polo, S. (2016) 'Strategies to detect endogenous ubiquitination of a target mammalian protein', *Methods in Molecular Biology*, 1449(1), pp 143-151. Available at: https://doi.org/10.1007/978-1-4939-3756-1_6.

Sluimer, J. and Distel, B. (2018) 'Regulating the human HECT E3 ligases', *Cellular and Molecular Life Sciences*, pp 3121-3141. Available at: https://doi.org/10.1007/s00018-018-2848-2.

Song, L. and Luo, Z.Q. (2019) 'Post-translational regulation of ubiquitin signaling', *Journal of Cell Biology*, 218(6), pp 1776-1786. Available at: https://doi.org/10.1083/JCB.201902074.

Soond, S.M. and Chantry, A. (2011) 'Selective targeting of activating and inhibitory Smads by distinct WWP2 ubiquitin ligase isoforms differentially modulates TGFB signalling and EMT', *Oncogene*, 30(21), pp 2451-2462. Available at: https://doi.org/10.1038/onc.2010.617.

Soond, S.M. et al. (2013) 'Novel WWP2 ubiquitin ligase isoforms as potential prognostic markers and molecular targets in cancer', *Biochimica et Biophysica Acta*, 1832(12), pp 2127-2135. Available at: https://doi.org/10.1016/j.bbadis.2013.08.001.

Stephen, A.G. et al. (1997) 'Identification of a region within the ubiquitin-activating enzyme required for nuclear targeting and phosphorylation', *Journal of Biological Chemistry*, 272(16), pp 10895-10903. Available at: https://doi.org/10.1074/jbc.272.16.10895.

Sterling, T. and Irwin, J.J. (2015) 'ZINC 15 - Ligand Discovery for Everyone', *Journal of Chemical Information and Modelling*, 55(11), pp 2324-2337. Available at: https://doi.org/10.1021/acs.jcim.5b00559.

Stewart, M.D. et al. (2016) 'E2 enzymes: More than just middle men', *Cell Research*, 26(4), pp 423-440. Available at: https://doi.org/10.1038/cr.2016.35.

Stroganov, O. V et al. (2008) 'Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening.', *Journal of Chemical Information and Modelling*, 48(12), pp. 2371–2385. Available at: https://doi.org/10.1021/ci800166p.

Sun, A. et al. (2014) 'Nedd4-1 is an exceptional prognostic biomarker for gastric cardia adenocarcinoma and functionally associated with metastasis', *Molecular Cancer*, 13(1), p 248. Available at: https://doi.org/10.1186/1476-4598-13-248.

Sun, D. et al. (2022) 'Why 90% of clinical drug development fails and how to improve it?', *Acta Pharmaceutica Sinica. B*, 12(7), pp 3049-3062. Available at: https://doi.org/10.1016/j.apsb.2022.02.002.

Sun, H., Tawa, G. and Wallqvist, A. (2012) 'Classification of scaffold-hopping approaches', *Drug Discovery Today*, 17(7-8), pp 310-324. Available at: https://doi.org/10.1016/j.drudis.2011.10.024.

Sutanto, F., Konstantinidou, M. and Dömling, A. (2020) 'Covalent inhibitors: A rational approach to drug discovery', *RSC Medicinal Chemistry*, 11(8), pp 876-884. Available at: https://doi.org/10.1039/d0md00154f.

Tait, S.W.G. et al. (2007) 'Apoptosis induction by Bid requires unconventional ubiquitination and degradation of its N-terminal fragment', *Journal of Cell Biology*, 179(7), pp 1453-1466. Available at: https://doi.org/10.1083/jcb.200707063.

Tanramluk, D. et al. (2009) 'On the origins of enzyme inhibitor selectivity and promiscuity: A case study of protein kinase binding to staurosporine', *Chemical Biology and Drug Design*, 74(1), pp 16-24. Available at: https://doi.org/10.1111/j.1747-0285.2009.00832.x.

Tatham, M.H. et al. (2011) ‘Comparative proteomic analysis identifies a role for SUMO in protein quality control’, *Science Signaling*, 4(178). Available at: https://doi.org/10.1126/scisignal.2001484.

Taylor, G. (2003) ‘The phase problem.’, *Acta Crystallographica. Section D: Biological Crystallography*, 59(11), pp. 1881–1890. Available at: https://doi.org/10.1107/s0907444903017815.

Theerawatanasirikul, S. et al. (2023) ‘Small Molecules Targeting 3C Protease Inhibit FMDV Replication and Exhibit Virucidal Effect in Cell-Based Assays’, *Viruses*, 15(9), p 1887. Available at: https://doi.org/10.3390/v15091887.

Theodorou, V. et al. (2007) ‘A simple method for the alkaline hydrolysis of esters’, *Tetrahedron Letters*, 48(46), pp 8230-8233. Available at: https://doi.org/10.1016/j.tetlet.2007.09.074.

Thomas, S.E. et al. (2017) ‘Structural Biology and the Design of New Therapeutics: From HIV and Cancer to Mycobacterial Infections: A Paper Dedicated to John Kendrew’, *Journal of Molecular Biology*, 429(17), pp 2677-2693. Available at: https://doi.org/10.1016/j.jmb.2017.06.014.

Tian, M. et al. (2019) ‘A cell-based high-throughput screening method based on a ubiquitin-reference technique for identifying modulators of E3 ligases’, *Journal of Biological Chemistry*, 294(8), pp 2880-2891. Available at: https://doi.org/10.1074/jbc.RA118.003822.

Tian, X. et al. (2023) ‘NEDD4 E3 ubiquitin ligases: Promising biomarkers and therapeutic targets for cancer’, *Biochemical Pharmacology* [Internet], 214, 115641. Available at: https://doi.org/10.1016/j.bcp.2023.115641.

Tokareva, O.S. et al. (2023) ‘Recognition and reprogramming of E3 ubiquitin ligase surfaces by α-helical peptides’, *Nature Communications*, 14(1), 6992. Available at: https://doi.org/10.1038/s41467-023-42395-z.

Tokgöz, Z. et al. (2012) ‘E1-E2 interactions in ubiquitin and Nedd8 ligation pathways’, *Journal of Biological Chemistry*, 287(1), pp 311-321. Available at: https://doi.org/10.1074/jbc.M111.294975.

Tong, Y.C. (1981) ‘lH-Irnidazo[4,5-b]pyrazines.1 11. 2-Thiols and Derivatives (1,2)’, *Journal of Heterocyclic Chemistry*, 18(751), pp 22–24. Available at: https://doi.org/10.1002/jhet.5570180423.

Torres, P.H.M. et al. (2019) ‘Key topics in molecular docking for drug design’, *International Journal of Molecular Sciences*, 20(18), 4574. Available at: https://doi.org/10.3390/ijms20184574.

Tropsha, A. et al. (2024) ‘Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR’, *Nature Reviews Drug Discovery*, 23(2), pp 141-155. Available at: https://doi.org/10.1038/s41573-023-00832-0.

Truongvan, N. et al. (2022) ‘Structures of UBA6 explain its dual specificity for ubiquitin and FAT10’, *Nature Communications*, 13(1), 4789. Available at: https://doi.org/10.1038/s41467-022-32040-6.

Tucker, D.L. and Rule, S.A. (2015) ‘A critical appraisal of ibrutinib in the treatment of mantle cell lymphoma and chronic lymphocytic leukemia’, *Therapeutics and Clinical Risk Management*, 11(1), pp 979-990. Available at: https://doi.org/10.2147/TCRM.S73559.

Vagenende, V., Yap, M.G.S. and Trout, B.L. (2009) ‘Mechanisms of protein stabilization and prevention of protein aggregation by glycerol’, *Biochemistry*, 48(46), pp 11084-11096. Available at: https://doi.org/10.1021/bi900649t.

Vagin, A.A. et al. (2004) ‘REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use’, *Acta Crystallographica Section D: Biological Crystallography*, 60(12), pp 2184-2195. Available at: https://doi.org/10.1107/S0907444904023510.

Van Montfort, R.L.M. and Workman, P. (2017) ‘Structure-based drug design: Aiming for a perfect fit’, *Essays in Biochemistry*, 61(5), pp 431-437. Available at: https://doi.org/10.1042/EBC20170052.

Vanderlaag, K. et al. (2006) ‘Inhibition of breast cancer cell growth and induction of cell death by 1,1-bis(3′-indolyl)methane (DIM) and 5,5′-dibromoDIM’, *Cancer Letters*, 236(2), pp 198-212. Available at: https://doi.org/10.1016/j.canlet.2005.05.036.

Varshavsky, A. (2005) ‘Ubiquitin fusion technique and related methods’, *Methods in Enzymology*, 339(1), pp 777-799. Available at: https://doi.org/10.1016/s0076-6879(05)99051-4.

Verdecia, M.A. et al. (2003) ‘Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase’, *Molecular Cell*, 11(1), pp 249-259. Available at: https://doi.org/10.1016/S1097-2765(02)00774-8.

Vijay-kumar, S., Bugg, C.E. and Cook, W.J. (1987) ‘Structure of ubiquitin refined at 1.8 Å resolution’, *Journal of Molecular Biology*, 194(3), pp 531-544. Available at: https://doi.org/10.1016/0022-2836(87)90679-6.

Vonrhein, C. et al. (2011) ‘Data processing and analysis with the autoPROC toolbox’, *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp 293-302. Available at: https://doi.org/10.1107/S0907444911007773.

Wade Harper, J., Ordureau, A. and Heo, J.M. (2018) ‘Building and decoding ubiquitin hains for mitophagy’, *Nature Reviews Molecular Cell Biology*, 19(1), pp 93-108. Available at: https://doi.org/10.1038/nrm.2017.129.

Wahl, L.C. et al. (2019) ‘Smad7 Binds Differently to Individual and Tandem WW3 and WW4 Domains of WWP2 Ubiquitin Ligase Isoforms’, I*nternational Journal of Molecular Sciences*, 20(19), 4682. Available at: https://doi.org/10.3390/ijms20194682.

Wan, L. et al. (2019) 'NEDD4 expression is associated with breast cancer progression and is predictive of a poor prognosis', *Breast Cancer Research*, 21(1), p 148. Available at: https://doi.org/10.1186/s13058-019-1236-7.

Wan, W. et al. (2021) 'High-Throughput Screening of an FDA-Approved Drug Library Identifies Inhibitors against Arenaviruses and SARS-CoV-2', *ACS Infectious Diseases*, 7(6), pp 1409-1422. Available at: https://doi.org/10.1021/acsinfecdis.0c00486.

Wang, K. et al. (2020) 'WWP2 regulates proliferation of gastric cancer cells in a PTEN-dependent manner', *Biochemical and Biophysical Research Communications*, 521(3), pp 652-659. Available at: https://doi.org/10.1016/j.bbrc.2019.10.179.

Wang, X. et al. (2022) 'High Expression of WWP1 Associates with Tumor Progression in Papillary Thyroid Cancer', *Cancer Biotherapy and Radiopharmaceuticals*, 37(4), pp 313-323. Available at: https://doi.org/10.1089/cbr.2020.4148.

Wang, X.S. et al. (2023) 'The unifying catalytic mechanism of the RING-between-RING E3 ubiquitin ligase family', *Nature Communications*, 14(1), 168. Available at: https://doi.org/10.1038/s41467-023-35871-z.

Wang, Y. et al. (2024) 'The role of WWP1 and WWP2 in bone/cartilage development and diseases', *Molecular and Cellular Biochemistry* [Preprint]. Available at: https://doi.org/10.1007/s11010-023-04917-7.

Wang, Z. et al. (2019) 'A multi-lock inhibitory mechanism for fine-tuning enzyme activities of the HECT family E3 ligases', *Nature Communications*, 10(1), 3162. Available at: https://doi.org/10.1038/s41467-019-11224-7.

Watt, J.E. et al. (2018) 'Discovery of Small Molecule WWP2 Ubiquitin Ligase Inhibitors', *Chemistry - A European Journal*, 24(67), pp 17677-17680. Available at: https://doi.org/10.1002/chem.201804169.

Watts, K.S. et al. (2014) 'Macrocycle conformational sampling with macromodel', *Journal of Chemical Information and Modelling*, 54(10), pp 2680-2696. Available at: https://doi.org/10.1021/ci5001696.

Wauer, T. et al. (2015) 'Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis', *The EMBO Journal*, 34(3), pp 307-325. Available at: https://doi.org/10.15252/embj.201489847.

Weber, J., Polo, S. and Maspero, E. (2019) 'HECT E3 ligases: A tale with multiple facets', *Frontiers in Physiology* [Internet], 10, e370. Available at: https://doi.org/10.3389/fphys.2019.00370.

Wei, D. and Sun, Y. (2010) 'Small RING finger proteins RBX1 and RBX2 of SCF E3 ubiquitin ligases: The role in cancer and as cancer targets', *Genes and Cancer*, 1(7), pp 700-707. Available at: https://doi.org/10.1177/1947601910382776.

Wen, W. et al. (2017) ‘Inhibition of NEDD4 inhibits cell growth and invasion and induces cell apoptosis in bladder cancer cells’, *Cell Cycle*, 16(16), pp 1509-1514. Available at: https://doi.org/10.1080/15384101.2017.1338220.

Weng, J.R. et al. (2007) ‘A potent indole-3-carbinol-derived antitumor agent with pleiotropic effects on multiple signalling pathways in prostate cancer cells’, *Cancer Research*, 67(16), pp 7815-7824. Available at: https://doi.org/10.1158/0008-5472.CAN-07-0794.

Weng, J.R. et al. (2009) ‘OSU-A9, a potent indole-3-carbinol derivative, suppresses breast tumor growth by targeting the Akt-NF-κB pathway and stress response signaling’, *Carcinogenesis*, 30(10), pp 1702-1709. Available at: https://doi.org/10.1093/carcin/bgp202.

Weng, M. et al. (2017) ‘The E3 ubiquitin ligase NEDD4 is translationally upregulated and facilitates pancreatic cancer’, *Oncotarget*, 8(12), pp 20288-20296. Available at: https://doi.org/10.18632/oncotarget.15446.

Wenzel, D.M. et al. (2011) ‘UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids’, *Nature*, 474(7349), pp 105-108. Available at: https://doi.org/10.1038/nature09966.

Wheeler, S.E. et al. (2010) ‘Probing substituent effects in aryl-aryl interactions using stereoselective diels-alder cycloadditions’, *Journal of the American Chemical Society*, 132(10), pp 3304-3311. Available at: https://doi.org/10.1021/ja903653j.

Wickliffe, K.E. et al. (2011) ‘The mechanism of linkage-specific ubiquitin chain elongation by a single-subunit E2’, *Cell*, 144(5), pp 769-781. Available at: https://doi.org/10.1016/j.cell.2011.01.035.

Wijk, S.J.L. and Timmers, H.T.M. (2010) ‘The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins’, *The FASEB Journal*, 24(4), pp 981-993. Available at: https://doi.org/10.1096/fj.09-136259.

Winter, G. (2010) ‘Xia2: An expert system for macromolecular crystallography data reduction’, *Journal of Applied Crystallography*, 43(1), pp 186-190. Available at: https://doi.org/10.1107/S0021889809045701.

Winter, G. et al. (2022) ‘DIALS as a toolkit’, *Protein Science*, 31(1), pp 232-250. Available at: https://doi.org/10.1002/pro.4224.

Wood, J.D. et al. (1998) ‘Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins’, *Molecular and Cellular Neurosciences*, 11(3), pp 149-160. Available at: https://doi.org/10.1006/mcne.1998.0677.

Wu, G. et al. (2003) ‘DeltaNp63alpha and TAp63alpha regulate transcription of genes with distinct biological functions in cancer and development’, *Cancer Research*, 63(10), pp 2351-2357. Available at: https://pubmed.ncbi.nlm.nih.gov/12750249/ (Accessed: 23 June 2024).

Wu, Z. et al. (2015) ‘Knockdown of WWP1 inhibits growth and invasion, but induces apoptosis of osteosarcoma cells’, *International Journal of Clinical and Experimental Pathology*, 8(7), pp 7869-7877.

Xu, J. et al. (2024) ‘NEDD4 and NEDD4L: Ubiquitin Ligases Closely Related to Digestive Diseases’, *Biomolecules*, 14(5), p. 577. Available at: https://doi.org/10.3390/biom14050577.

Yamaguchi, S., Kaneko, M. and Narukawa, M. (2021) ‘Approval success rates of drug candidates based on target, action, modality, application, and their combinations’, *Clinical and Translational Science*, 14(3), p 1113-1122. Available at: https://doi.org/10.1111/cts.12980.

Yang, R. et al. (2016) ‘Elevated expression of WWP2 in human lung adenocarcinoma and its effect on migration and invasion’, *Biochemical and Biophysical Research Communications*, 479(2), pp 146-151. Available at: https://doi.org/10.1016/j.bbrc.2016.07.084.

Ye, Y. and Rape, M. (2009) ‘Building ubiquitin chains: E2 enzymes at work’, *Nature Reviews Molecular Cell Biology*, 10(11), pp 755-764. Available at: https://doi.org/10.1038/nrm2780.

Ye, Y. et al. (2011) ‘Polyubiquitin binding and cross-reactivity in the USP domain deubiquitinase USP21’, *EMBO Reports*, 12(4), pp 350-357. Available at: https://doi.org/10.1038/embor.2011.17.

Yin, Q. et al. (2009) ‘E2 interaction and dimerization in the crystal structure of TRAF6’, *Nature Structural and Molecular Biology*, 16(6), pp 658-666. Available at: https://doi.org/10.1038/nsmb.1605.

You, S. et al. (2024) ‘E3 ubiquitin ligase WWP2 as a promising therapeutic target for diverse human diseases’, *Molecular Aspects of Medicine*, 96(1), e101257. Available at: https://doi.org/10.1016/j.mam.2024.101257.

Zhang, K. et al. (2019) ‘Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution’, *Nature Communications*, 10(1), p 5511. Available at: https://doi.org/10.1038/s41467-019-13494-7.

Zhang, L. et al. (2015) ‘WWP1 as a potential tumor oncogene regulates PTEN-Akt signaling pathway in human gastric carcinoma’, *Tumor Biology*, 36(2), pp 787-798. Available at: https://doi.org/10.1007/s13277-014-2696-0.

Zhang, W. et al. (2016) ‘System-Wide Modulation of HECT E3 Ligases with Selective Ubiquitin Variant Probes’, *Molecular Cell*, 62(1), pp 121-136. Available at: https://doi.org/10.1016/j.molcel.2016.02.005.

Zhang, W. et al. (2016) ‘System-Wide Modulation of HECT E3 Ligases with Selective Ubiquitin Variant Probes’, *Molecular Cell*, 62(1), pp 121-136. Available at: https://doi.org/10.1016/j.molcel.2016.02.005.

Zhang, X.F. et al. (2015) ‘Overexpression of WWP1 promotes tumorigenesis and predicts unfavorable prognosis in patients with hepatocellular carcinoma’, *Oncotarget*, 6(38), pp 40920-40933. Available at: https://doi.org/10.18632/oncotarget.5712.

Zhang, Y. et al. (2020) ‘E3 ubiquitin ligase nedd4 family-regulatory network in cardiovascular disease’, *International Journal of Biological Sciences*, 16(14), pp 2727-2740. Available at: https://doi.org/10.7150/ijbs.48437.

Zhao, D. et al. (2012) ‘TAZ antagonizes the WWP1-mediated KLF5 degradation and promotes breast cell proliferation and tumorigenesis’, *Carcinogenesis*, 33(1), pp 59-67. Available at: https://doi.org/10.1093/carcin/bgr242.

Zheng, H. et al. (2018) ‘NEDD4 promotes cell growth and motility in hepatocellular carcinoma’, *Cell Cycle*, 17(6), pp 728-738. Available at: https://doi.org/10.1080/15384101.2018.1440879.

Zheng, N. et al. (2002) ‘Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex’, *Nature*, 416(6882), pp 703-709. Available at: https://doi.org/10.1038/416703a.

Zheng, W. et al. (2023) ‘Identification of the HECT domain binding of indole-3-carbinol (I3C) derivatives for breast cancer therapy’, *Phytochemistry Letters*, 54(1), pp 7-13. Available at: https://doi.org/10.1016/j.phytol.2023.01.002.

Zhi, X. and Chen, C. (2012) ‘WWP1: A versatile ubiquitin E3 ligase in signaling and diseases’, *Cellular and Molecular Life Sciences*, 69(9), pp 1425-1434. Available at: https://doi.org/10.1007/s00018-011-0871-7.

Zhu, K. *et al.* (2017) ‘Allosteric auto-inhibition and activation of the Nedd4 family E3 ligase Itch’, *EMBO reports*, 18(9), pp 1618-1630. Available at: https://doi.org/10.15252/embr.201744454.

Zhu, K. et al. (2024) ‘Ubiquitylation of nucleic acids by DELTEX ubiquitin E3 ligase DTX3L.’, *EMBO reports*, 25(10), pp. 4172–4189. Available at: https://doi.org/10.1038/s44319-024-00235-1.

Zweifach, A. (2020) ‘The National Cancer Institute’s Plated Compound Sets Can Be a Valuable Resource for Academic Researchers’, *SLAS Discovery*, 25(1), pp 2-6. Available at: https://doi.org/10.1177/2472555219873557.