Discovery of Keap1 Inhibitors for Reduction of Oxidative Stress in Amyotrophic Lateral Sclerosis

The University of Sheffield

Department of Chemistry and the Sheffield Institute for Translational Neuroscience (SITraN)

Submitted for the degree of Doctor of Philosophy (PhD)

The University Of Sheffield.

Matthew Sellwood

Supervisors: Prof. Dame. Pamela Shaw, Prof. Beining Chen and Dr Richard Mead

April 2016

Abstract

ALS is a devastating neurodegenerative disease for which there is no cure, and new therapies are urgently needed. Oxidative stress is an important mechanism in the pathogenesis of ALS, and can be reduced by activation of the Nrf2-ARE pathway. This pathway is an attractive target for drug development. Keap1 is the cytoplasmic inhibitor of Nrf2 and disruption to the normal function of Keap1 leads to an increased antioxidant response. Current molecules that inhibit the Keap1-Nrf2 interaction react with sulfhydryl groups on Keap1. However, these reactive molecules are prone to side effects and are difficult to optimise using medical chemistry. Therefore, a direct inhibitor of the Nrf2-Keap1 interaction is warranted. This project explores small molecule inhibitors of the Nrf2-Keap1 interaction using biochemical screening and *in-silico* design.

An HTS fluorescence polarisation assay was developed which probes the interaction between full length Keap1 and a FITC tagged ETGE-motif Nrf2 peptide. Using the assay, binding constants for labelled and unlabelled peptides were established and two large compound libraries were screened. The screens produced a number of hits with molecules able to show activity in a number of different biophysical assays. Whilst the compounds showed no cellular potency, the molecules had potential for optimisation of physiochemical properties through medicinal chemistry. Computational docking analysis of the compounds was performed to define the binding orientation of the molecules to Keap1. Analysis showed two potential binding orientations, a chemistry plan was then developed to lead future optimisation of the molecules.

Full length recombinant expressed mammalian Keap1 protein was characterised. This showed that Keap1 is potentially able to form a number of oligomeric forms with dimer, trimer and tetramer being the most highly observed forms. This could elucidate a potential mechanism of action for some reactive inhibitors of Keap1, and a new target mechanism for small molecule inhibitors.

Acknowledgments

I would like to thank my supervisors for helping guide me through this project. Firstly, Beining who without her support, optimism and guidance I would not have been able to start this PhD yet alone finish it. Secondly thanks go to Richard who had all the time and patience to answer my stupid questions and talk through ideas, and without whose expertise the project would not have been possible. Finally, I would like to thank Pam who has supported me without question, introduced me to the key researcher-patient link, reminding us all why we peruse our research and created the Sheffield Institute for Translational Neuroscience whose community spirit and support I could not have been without.

This leads me on to thank everyone within the Chen group, in particular Matt J, Ashlie, Jenny L, Jenn and Sonny who showed me it was okay to laugh when things go wrong and that it is possible to get there in the end! I would also like to thank all the people from the Shaw and Mead groups and everybody within SITraN with special thanks going to Nazia, Adrian and Guillaume for helping with too many things to mention. I also thank my friends from SITraN who have helped me over the years, in particular (in no particular order) Matt, Jenn, Sarah, Laura, Alex, Chris, Cassy, Dave, and Jodie for their close friendship, advice and support.

I would also like to thank Svet from the Biological Sciences EM facility, Simon from the Department of Chemistry Mass Spectrometry suite, and the entire team at the European Lead Factory for their huge contribution to the project.

I also have to mention the amazing amount of baked goods and fantastic nights out I have enjoyed over the years so to all those who made my time more enjoyable I thank you!

Finally, I would like to thank my wife Eve for supporting and putting up with me on the many times I had to work late, or have been stressed over my work. You have got me through and have always be believed in me so thank you!

Table of Contents

Abstract I

Acknowledgments III

Abbreviations VIII

List of Figures XIV

List of Tables XX

1. Introduction 1

1.1 What is Amyotrophic Lateral Sclerosis? 2

1.1.1 Introduction 2

1.1.2 Clinical features of ALS 2

1.1.3 Symptoms and management 3

1.2 Causes of Motor Neuron Disease 5

1.2.1 Oxidative Stress 5

1.2.2 Excitotoxicity 9

1.2.3 Mitochondrial dysfunction 11

1.2.4 Defects of RNA processing 12

1.2.5 Genetic Factors: C9ORF72 13

1.2.6 Protein Aggregation 16

1.2.7 Neuroinflammation 16

1.2.8 Non neuronal cells 17

1.2.9 Summary 17

1.3 Current therapeutic strategies 18

1.3.1 Initial development of therapeutics within ALS 18

1.3.2 Clinical Trials in Amyotrophic Lateral Sclerosis 20

1.4 The Keap1-Nrf2-ARE pathway 22

1.4.1 Dysregulation of the Nrf2-Keap1 pathway in ALS 22

1.4.2 Basic mechanism of the pathway 23

1.4.3 Components of the pathway 23

1.4.4 Keap1-Nrf2-ARE pathway mechanism 25

1.4.5 Disrupting the Keap1-Nrf2-ARE pathway 28

1.5 Project Aims and Objectives 35

1.5.1 Design of a high throughput screening assay to probe the interaction of Nrf2 and Keap1 35

1.5.2 Utilising computational methods to screen and design molecular inhibitors of the interaction between Keap1 and Nrf2 35
1.5.3 Screening *in-silico* selected compounds and selected molecular libraries in high throughput assays 35
2. Materials and Methods 37
2.1 Materials 38
2.1.1 List of commonly used chemicals 38
2.1.2 Solutions 39
2.1.3 Antibodies 42
2.2 Protein Synthesis and Analysis 43
2.2.1 Production of Keap1 cDNA plasmid 43
2.2.1 Cell Culture 43
2.2.2 Optimisation of Cell Transfection 43
2.2.3 Large scale Cell Transfection of Keap1 44
2.2.4 Cell Lysis 44
2.2.5 Isolation and Purification of Keap1 44
2.2.6 SDS-polyacrylamide-gel electrophoresis 45
2.2.7 Western Blotting Analysis 45
2.2.8 BCA Assay for Protein Concentration Determination 45
2.3 Fluorescence Polarisation Assay 47
2.3.1 Fluorescence Polarisation Assay optimisation method 47
2.3.2 Fluorescence Polarisation Assay - General Screening method 47
2.4 ARE Assay 48
2.4.1 Cell Culture 48
2.4.2 ARE reporter assay 48
2.5 Protein Characterisation 49
2.5.1 Size Exclusion Chromatography 49
2.5.2 Cross-Linking of Keap1 49
2.5.3 Isothermal Calorimetry 50
2.5.4 Protein Liquid Chromatography Mass Spectrometry 50
2.5.5 Electron Microscopy 50
2.6 Computational Studies 52
2.6.1 Preparation of a virtual high-throughput (vHTS) library 52
2.6.2 Preparing Keap1 crystal structures for in-silico analysis. 52
2.6.3 Preparing pharmacophore models 52
2.6.4 Screening database against pharmacophore 53

2.6.5 Preparing ligands for docking 53
2.6.6 Docking 53
2.6.7 Predictive ADME 56
2.6.8 Predictive Toxicity 57
2.6.9 Data mining 57
2.6.10 Pipeline pilot analysis of HTS screening data 57
2.6.11 Homology modelling 58

3. Development of an HTS *in-vitro* screening assay 59

3.1 Introduction 60

3.1.1 Introduction to HTS 60
3.1.2 Fluorescent HTS techniques 60
3.1.3 Current fluorescence based inhibition assays for the Keap1-Nrf2 interaction 62
3.1.4 A pharmacologically relevant fluorescent polarisation assay 66
3.1.5 Previous work on the project 66

3.2 Results and Discussion 70

3.2.1 Recombinant Expression of Keap1 70
3.2.2 Miniaturisation of HTS assay 77
3.2.3 Further assay exploration and development 85

3.3 Conclusions 91

4. High throughput screening to find inhibitors of the Keap1-Nrf2 interaction. ... 92

4.1 Introduction 93

4.2 Results and Discussion 93

4.2.1 Screening of the MRCT Protein-Protein inhibitor (PPI) Library 93
4.2.2 Screening of European Lead Factory Library 100

4.3 Conclusion 131

5. In-silico techniques for finding potential ligands for protein binding 133

5.1 Introduction 134

5.1.1 Drug-Receptor interactions 134
5.1.2 CADD 136
5.1.3 CADD Techniques 140
5.1.4 Software packages 144

5.2 Results and Discussion 149

5.2.1 Initial Screening 149
5.2.2 Second *in-silico* screen 161

5.3 Conclusions and Future Work 172

6. Characterisation of mammalian Keap1 protein. 173
6.1 Introduction 174
6.2 Results and Discussion 176
6.2.1 Size Exclusion Chromatography 176
6.2.2 Could Keap1 form tetramers and aggregates? A literature and *in-silico* exploration 180
6.2.3 Cross Linking of Keap1 188
6.2.4 Mass spectrometry 192
6.2.5 Electron microscopy 193
6.3 Discussion and Conclusion 199
7. General Discussion and Conclusions 201
Appendix 211
Bibliography 224

Abbreviations

AD	Autosomal Dominant
ADMET	Adsorption, Distribution, Metabolism, Excretion and Toxicity
ALS	Amyotrophic Lateral Sclerosis
ALS2	Alsin
AMPA	2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid
ANG	Angiogenin
AR	Autosomal Recessive
ARE	Antioxidant Response Element
ATRA	All trans retinoic acid
BAD	Bcl2-Associated Death promoter
BBB	Blood-Brain Barrier
BDNF	Brain-derived neurotrophic factor
BSA	Bovine Serum Albumin
BTB	Broad-Tramtrack-Brick-a-brack
bZIP	Basic Leucine Zipper
C9ORF72	Chromosome 9 Open Reading Frame 72
CADD	Computer Aided Drug Design
cDNA	Complimentary DNA
CHO	Chinese Hamster Ovary
CHMP2B	Chromatin Modifying Protein 2B
cLogP	Calculated log of the partition coefficient between n-octanol and water
CNS	Central Nervous System
CSF	Cerebrospinal Fluid

CTR	C-Terminal Region
Cul3	Cullin-3
DAO	D-amino Acid Oxidase
DENN	Differentially Expressed in Normal and Neoplasia
DGR	Double Glycine Repeat
DMEM	Dulbecco's modified Eagle's minimal essential medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DS	Discovery Studio
DTT	Dithiothreitol
EAAT2	Excitatory amino acid transporter 2
EC_{50}	Effective concentration at 50 % of maximal effect
eGFP	Enhanced Green Fluorescent Protein
ELF	European Lead Factory
ER	Endoplasmic Reticulum
FALS	Familial Amyotrophic Lateral Sclerosis
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
FIG4	Polyphosphoinositide phosphatase
FP	Fluorescence Polarisation
FRET	Förster resonance energy transfer
FTD	Fronto-temporal dementia
FUS	Fused in Sarcoma
GFA	Genetic Function Approximation

GLC	Germ Cell-less
GluR2	AMPA receptor subunit Glutamate Receptor 2
GST	Glutathione S-transferase
GTP	Guanosine-5'-triphosphate
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HEK293	Human Embryonic Kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
HSP	Heat Shock Protein
HTS	High Throughput Screening
IC_{50}	Inhibitor concentration at 50 % of maximal effect
ITC	Isothermal Titration Calorimetry
IVR	Intervening Region
Keap1	Kelch-Like ECH-Associated Protein 1
LB	Lysogeny Broth
LCMS	Liquid Chromatography Mass Spectrometry
LMN	Lower Motor Neruon
Maf	V-maf musculoaponeurotic fibrosarcoma oncogene homolog
MAPT	Microtubule-associated protein Tau
MLPCN	Molecular Libraries Probe Production Centres Network
MND	Motor Neurone disease
mPSA	Molecular Polar Surface Area
MRCT	Medical Research Council Technology

mRNA	Messenger ribonucleic acid
mSOD1	Mutant Superoxide Dismutase 1
MST	Microscale Thermophoresis
MW	Molecular Weight
ND	Not Determined
NES	Nuclear Export Signal
NMDA	N-Methyl-D-aspartic acid
NOx	Nitrogen Oxides
Nrf2	Nuclear factor erythroid 2-related factor 2
NSC34	Neuroblastoma spinal cord
NTP	National Toxicity Profile
NTR	N-Terminal Region
OPTN	Optinneurin
PBS	Phosphate-buffered saline
PD	Parkinson's Disease
PDB	Protein Data Bank
pEC_{50}	The negative logarithm of the EC_{50} (see EC_{50})
PEI	Polyethylenimine
PFN1	Profilin 1
PI	Phosphatase inhibitor
pIC_{50}	The negative logarithm of the IC_{50} (see IC_{50})
PLS	Primary Lateral Sclerosis
PMA	Progressive Muscular Atrophy
PMSF	Phenylmethylsulfonyl fluoride

PPB	Plasma Protein Binding
PVDF	Polyvinylidene difluoride
QSTR	Quantitative Structural Toxicity Relationship
QSAR	Quantitative SAR (see SAR)
Rac1	Ras-related C3 botulinum toxin substrate 1
RARα	Retinoic acid receptor alpha
rKeap1	Recombinant Keap1 (see Keap1)
RIPA	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
Rpm	Rotations per minute
RT	Room Temperature (293 K)
R&D	Research and Development
SALS	Sporadic Amyotrophic Lateral Sclerosis
SADD	Structure-Aided Drug Design
SAR	Structure-Activity Relationship
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SETX	Senataxin
SIGMAR1	σ Non-opioid Receptor 1
SMA	Spinal Muscular Atrophy
SMN1	Survival Motor Neuron 1

SOC	Super Optimal Broth
SOD1	Superoxide Dismutase 1
SPG11	Spatacsin
SPOP	Speckle-type POZ protein
SQSTM1	Sequestosome 1
SW	Signal Window
TBS	Tris-Buffered Saline
TBST	Tris-Buffered Saline with Tween 20
TCEP	Tris(2-carboxyethyl)phosphine
TDP-43	TAR DNA-binding Protein 43
TEM	Transmission Electron Microscopy
UBQLN2	Ubiquilin 2
UMN	Upper Motor Neuron
UPS	Ubiquitin protease system
UV	Ultraviolet
VAPB	Vesicle-Associated membrane Protein-associated protein B
VCP	Valosin containing protein
vHTS	Virtual High Throughput Screening
WT SOD1	Wild type SOD1 (see SOD1)
Z’	Z-Prime Factor
ZINC	ZINC is not commercial

List of Figures

Figure 1: Atomic resolution crystal structure of dimeric SOD1 6

Figure 2: Overall equation for conversion of superoxide radicals. 7

Figure 3: Diagram illustrating the effects of oxidative stress on cells. 9

Figure 4: Diagram illustrating glutamatergic neurotransmission. 10

Figure 5: The chemical structures of vitamin E, gabapentin and memantine 19

Figure 6: The chemical structure of Riluzole (Rilutek) 19

Figure 7: The chemical structure of Dexpramipexole 19

Figure 8: The 5 domains of Keap1 24

Figure 9: The 6 domains of Nrf2 with key amino acids shown in blue. 25

Figure 10: The Keap1-Nrf2-ARE pathway 27

Figure 11: Sulfhydryl nucleophilic attack on an electrophilic Michael acceptor 29

Figure 12: The chemical structure of curcumin 29

Figure 13: Further known electrophilic inducers of the Keap1-Nrf2 pathway. 30

Figure 14: Two representations of the crystal structure of the Kelch domain of Keap1 31

Figure 15: Reported optimisation of a tetrahydroisoquinoline 32

Figure 16: Hit and optimised compounds yielded from the HTS screen of Marcotte et al. 33

Figure 17: Crystal structures of Keap1 Kelch domain with compounds 16 and 15 33

Figure 18: Structure of a carbazone found to disrupt the Nrf2-Keap1 interaction 34

Figure 19: Jablonski diagram to illustrate fluorescence. 60

Figure 20: Showing fluorescence resonance energy transfer between a donor molecule (D) and an acceptor molecule (A). 62

Figure 21: Keap1 FRET assay 63

Figure 22: Basics of fluorescence polarisation 64

Figure 23: Optimisation of peptide concentration and buffer compositions. 68

Figure 24: Titration of two commercially available Keap1 protein samples 69

Figure 25: Plasmid map of the Keap1 cDNA 71

Figure 26: Transfection optimisation. 73

Figure 27: Purity quantification of purified Keap1. 74

Figure 28: Western blot of an SDS-PAGE gel showing purified Keap1 75

Figure 29: Representative coomassie stained SDS-PAGE gel showing eluted fractions from the Keap1 affinity purification step 75

Figure 30: Activity comparison of purchased and synthesised protein. 76

Figure 31: Peptide titration curves at various assay volumes 77

Figure 32: Protein titration curves 78

Figure 33: X-ray crystal structure of Keap1 and ETGE peptide 79

Figure 34: FP competition binding assay using C-terminal fluorescein labelled ETGE peptide 80

Figure 35: Protein titration curves to calculate non-specific binding. 81

Figure 36: Chemical structure of the test compound used as a standard agonist 83

Figure 37: Plate layout for Z' and SW determination 84

Figure 38: Results from Z' and SW assay. 84

Figure 39: Graphs illustrating well data (A), row averages (B) and column averages (C) 85

Figure 40: Optimisation of DLG peptide 86

Figure 41: Partial amino acid sequence of Nrf2 86

Figure 42: Peptide titration curve of N-terminal labelled DLGex Nrf2 peptide. 87

Figure 43: Keap1 protein titration curve using an extended sequence length 4 nM N-terminal labelled DLG Nrf2 peptide (DLG-Ex). 88

Figure 44: Chemical structure of "Hit1" 88

Figure 45: Comparison of IC_{50} values in the presence and absence of the reducing agent DTT using the control compound "Hit1". 89

Figure 46: Activators of Nrf2 that work through a classical mechanism of covalent modification of Keap1. 90

Figure 47: Graph showing how varying concentration of classical activators fails to have any effect on reduction of fluorescence polarisation signal 90

Figure 48: Screening plate layout based on a 384 well plate format. 94

Figure 49: Graph illustrating the percentage inhibition of each compound in the MRCT library 95

Figure 50: Pipeline protocol written to identify active compounds from the screening data 96

Figure 51: Flowchart to show a simplified view of how screening data is processed 97

Figure 52: Dose response fluorescence polarisation assay for MRCT00160026. 98

Figure 53: Chemical structure of MRT00160026 98

Figure 54: Fluorescence polarisation assay to test MRCT analogues for potential inhibition of fluorescently labelled Nrf2 peptide to Keap1 98

Figure 55: Cellular ARE assay data for the MRCT compound structural analogues 99

Figure 56: Comparison between the new batch of peptide and the old batch of peptide used in the MRCT screen 102

Figure 57: Coomassie stained SDS-PAGE gel showing protein expression of lysates of different expression vectors for Keap1 protein 103

Figure 58: Comparison of protein titration curves 104

Figure 59: Flow chart detailing the triage of compounds through the screening cascade. 106

Figure 60: Structures of hits from the ELF screen 107

Figure 61: Graphs illustrating dose response curves for the 6 resynthesised ELF compounds 109

Figure 62: Graphs illustrating the results of the ELF compounds in the ARE reporter assay 111

Figure 63: Hit compounds activate Nrf2-ARE directed transcription. 112

Figure 64: How a MST machine functions 115

Figure 65: Illustrates MST titration curves for various compounds against a 200 nM concentration of Keap1 protein 116

Figure 66: Use of MST to show ELF compounds participate in competitive inhibition of Keap1 118

Figure 67: MST assay curves showing binding of the ETGE peptide to Keap1. 119

Figure 68: Diagram illustrating the basic layout of an ITC machine 120

Figure 69: Isothermal titration calorimetry traces for ETGE peptide 121

Figure 70: Isothermal titration calorimetry traces for ELF2 123

Figure 71: Top two docking poses achieved for ELF1 and ELF2 125

Figure 72: Docking poses superimposed onto the co-crystal structure of compound 16 from PDB ID: 4IQY 127

Figure 73: Illustration of potential modifications that could be made to the hit molecular series 130

Figure 74: Different types of hydrogen bond commonly seen in receptor-ligand complexes 135

Figure 75: Venn diagram illustrating the difficulty in finding a potential drug candidate. 137

Figure 76: A flow chart representing a classical approach to structure-aided drug design 139

Figure 77*:* A Typical Pharmacophore model 141

Figure 78: Filter for High Drug Likeness 144

Figure 79: Screenshot of the Discovery Studio interface 146

Figure 80: Screenshot of the pipeline pilot user interface 148

Figure 81: Pipeline Pilot Protocol for creating a drug-like subset of the ZINC database. 150

Figure 82: The crystal structure of KEAP1 DC domain 153

Figure 83: Pharmacophore model of interactions between Keap1 DC domain and Nrf2 peptide 154

Figure 84: Comparison of scoring functions to look for correlation in the data set of 686 molecules selected for docking to the Keap1 crystal structure 155

Figure 85: Dose response fluorescence polarisation assay testing available analogues from in-silico screening program. 156

Figure 86: Comparison of hydrophobic protein surfaces to analyse available binding pockets 157

Figure 87: *in-silico* analysis of the binding modes of Keap1 to peptides and small molecules... 159

Figure 88: Further analysis of published co-crystal structures 160

Figure 89: Pharmacophore generated from interactions between the ETGE peptide and Keap1 using PDB structure 2FLU 162

Figure 90: Pharmacophore generated from interactions between the compound "compound 16" and Keap1 using PDB structure 4IQK 162

Figure 91: Pharmacophore generated from interactions between the compound "Hit 1" and Keap1 using PDB structure 4L7D 163

Figure 92: Comparison of the docking pose of compound 16 (grey) with the experimental data from the crystal structure (green) 164

Figure 93: Comparison of the top three scoring docking poses of compound 16 (grey) with the experimental data from the crystal structure (green) 164

Figure 94: Dose response fluorescent polarisation assay of compounds selected from *in-silico* screen 166

Figure 95: Dose response graphs of MS50 and MS57 showing how autofluorescence can cause either an increase or decrease in fluorescence polarisation signal. 167

Figure 96: ARE reporter cell line assay. 169

Figure 97: Selected *in-silico* compounds tested in an 8-point dose response in both cell lines. 171

Figure 98: Representative gel filtration trace of 100 µg Keap1 injection with 1 mM DTT..... 176

Figure 99: Western blot images of SDS-PAGE gels loaded with 15 µL of each sample from the reducing gel filtration fractions and probed with Anti-Keap1 antibody 177

Figure 100: Coomassie stained SDS-PAGE gel of the Keap1 batch used for Gel Filtration column chromatography 178

Figure 101: Representative gel filtration trace of 100 µg Keap1 injection without DTT 179

Figure 102: Western blot of non-reducing gel filtration fractions of equal volume loading imaged with Anti-Keap1 antibody 180

Figure 103: Raw electron micrograph images with different Euler angles 180

Figure 104: SPOP dimerisation 181

Figure 105: Sequence alignment of SPOP and Keap1 182

Figure 106: Structural superposition of the BTB domains of SPOP (PDB ID: 4J8Z) and Keap1 (PDB ID: 4CXI) 183

Figure 107: Secondary structure prediction for residues contained within the suggested interaction domain region 183

Figure 108: Homology model of Keap1 dimer 184

Figure 109: Dimeric model of Keap1 (residues 50-315) indicating how Cul3 (orange) may bind the Keap1 (pink) shown from two viewpoints 185

Figure 110: Various orientations of a monomeric unit of the Keap1 homology model are shown with different binding partners. 187

Figure 111: Western blot analysis of a SDS-PAGE gel of Keap1 after cross-linking with glutaraldehyde imaged with anti-Keap1 antibody 188

Figure 112: Crosslinking of Keap1 in the presence of reactive Keap1 inhibitors 190

Figure 113: Mass spectrometry analysis of full length recombinant mammalian Keap1 protein 193

Figure 114: Representative electron micrographs images of recombinantly expressed human Keap1 protein at 21000X 195

Figure 115: Representative electron micrographs images of recombinantly expressed human Keap1 protein at 52000X 197

Figure 116: Selected electron microscopy images of negatively stained Keap1 particles at 52000X. 198

Figure 117: Hypothesised structure of tetrameric Keap1 consisting of two Keap1 dimers 198

List of Tables

Table 1: Symptomatic management in patients with motor neuron disease....4

Table 2: Genes associated with ALS....15

Table 3: Molecules currently in clinical trials for the treatment of ALS....20

Table 4: Antibodies used in Western Blotting and their concentrations....42

Table 5: Cell transfection optimisation conditions....44

Table 6: GOLD settings used during docking of Keap1 with each ligand in the defined binding pocket....54

Table 7: LibDock settings used during docking of Keap1 with each ligand in the defined binding pocket....55

Table 8: CDocker settings used during docking of Keap1 with each ligand in the defined binding pocket....56

Table 9: Advantages and disadvantages of different fluorescent techniques for use in HTS....61

Table 10: Comparison of published FP assays.....65

Table 11: Buffer Composition....67

Table 12: Illustrates the molecular properties of the hits....106

Table 13: Comparison of properties for drug-and lead-like properties.....150

Table 14: Review of available crystal structures.....151

Table 15: Table showing the calculated rates of formation and decay for species of Keap1 formed through crosslinking of the protein....192

Table 16: Summary table of the three main hit compounds identified from the screen of the ELF library....205

1. Introduction

1.1 What is Amyotrophic Lateral Sclerosis?

1.1.1 Introduction

Amyotrophic Lateral Sclerosis (ALS) or motor neuron disease (MND) is one of several diseases that are responsible for degeneration of the motor system within the central nervous system (CNS). Other diseases within the same class include hereditary spastic paraplegia, and spinal muscular atrophy (SMA). ALS is characterised by the degeneration of upper and lower motor neurons within the spinal cord, cranial nerve motor nuclei and the cortex of the brain.[1] The condition of motor neuron disease was first described in 1824 by Charles Bell.[2] However, it was not until 1869 that Jean-Martin Charcot fully classified the disease by linking the symptoms that were described in clinic to results seen from autopsy.[2]

Motor neuron disease is known by several other names. It is known commonly as Lou Gehrig's disease in the United States, after the famous major league baseball player, who played for the New York Yankees between 1923 and 1939 before contracting the disease.[3] Another common and more formal name for the disease is Amyotrophic Lateral Sclerosis (ALS), the term that was coined by Charcot when he first classified the disease in 1874.[4]

Most cases of ALS are sporadic (SALS) with an incidence rate of around 2 per 100,000 people per year. This form of ALS has no known cause. A much smaller proportion of cases (around 5%) of ALS are familial (FALS), with causes linked to genetic mutations passed down through families.[5,6]

This devastating disease leads to the neurodegeneration of the upper and lower motor neurons. This causes muscle weakness and atrophy and leads to a loss of mobility in limbs and problems with breathing, speaking, eating and swallowing. Patients with the disease can expect to survive for a median of 3-5 years after diagnosis.[7]

1.1.2 Clinical features of ALS

The symptoms of ALS are presented through the effect of loss of motor neurons in several places. The loss of lower motor neurons (LMNs) leads to progressive muscle weakness and atrophy. The loss of upper motor neurons (UMNs) can produce stiffness, exaggerated reflexes and involuntary movements. In addition, the loss of prefrontal cortical neurons can be affected; these are involved in organising the work of the upper and lower motor neurons. Loss of these neurons can lead to cognitive impairment including executive dysfunction, and a change in the patients awareness within social situations and subsequent social behaviour.[8, 9]

Whilst ALS is characterised through the involvement of both upper and lower motor neurons, disease involving only the UMNs is known as primary lateral sclerosis, and disease involving only the LMNs is known as progressive muscular atrophy. Another similar form of the disease is when the disease is restricted to the bulbar muscles; this is known as progressive bulbar palsy. It is quite common for the disease to start within the bulbar muscles and to then progress into classic ALS to affect a larger group of muscles.[10]

Patients with bulbar-onset ALS account for about 25% of patients, whilst those with limb-onset and with initial trunk or respiratory forms represent 70% and 5% respectively.[9] ALS progression is relentless with 50% of patients dying within 30 months of onset. A further 20% of patients survive between 5 and 10 years, with the majority of patients dying from respiratory failure.[11]

1.1.3 Symptoms and management

There is no current cure for ALS. Therefore, most treatment is focused on management of symptoms. The table below (Table 1) describes the symptoms associated with the disease; these symptoms have a devastating effect on both the lives of the patients, families and friends of those suffering from the disease.

Table 1: Symptomatic management in patients with motor neuron disease (reproduced with permission from Talbot, 2009).[11]

Symptom	Frequency within ALS	How symptom is managed
Hypersalivation and drooling	Occurs mostly within patients with bulbar-onset ALS (around 25%)	Thin secretions: Hyoscine patches, amitriptyline, glycopyrronium, botulinum toxin. Thick secretions: pineapple or papaya juice, nebulised saline, β-blockers, carbocisteine
Fragmented nocturnal sleep and daytime drowsiness (somnolence)	Occurs in 40-50% of patients	Through use of an intermittent positive pressure ventilation mask to aid nocturnal breathing
Spasticity, cramps and involuntary movements (fasciculation).	Normally only lasts a small amount of time and recedes as the disease progresses. Not necessary to treat in most.	Exercise and physiotherapy, baclofen or tizanidine, gabapentin, botulinum toxin
Dyspnoea	Normally a minority towards the end of life	Oramorph, with consent of the patient.
Neck weakness, head drop	10-20%	Neck supporting collars
Emotional lability	10% but only lasts a small time	Amitriptyline, selective serotonin reuptake inhibitors e.g. citalopram
Nutritional insufficiency	Affects the majority of patients but at different stages	Around 60-70% of patients will require enteral feeding.

1.2 Causes of Motor Neuron Disease

The neurodegeneration process associated with the disease is thought to be multifactorial, initiated through a number of mechanisms including oxidative stress, excitotoxicity, and mitochondrial dysfunction to name a few.[12]

1.2.1 Oxidative Stress

Oxidative stress is where an imbalance occurs between the systemic levels of reactive oxygen species (ROS) and the body's natural biological ability to detoxify the reactive species. This leads to a build-up of ROS such as peroxides and free radicals which causes damage to all components of cells.[13,14] This damage can be a trigger of apoptosis or even necrosis.[15] Oxidative stress has been widely implicated in many diseases including cancer, Parkinson's disease and Alzheimer's disease, in addition to ALS.

In ALS, patient samples from both SALS and FALS related cases showed increased amounts of oxidative damage to proteins, DNA and lipids.[16–18] The amount of reactive oxygen species within the body is known to increase with age, and is the main cause of ageing. This increase is mainly due to leakage of ROS from the mitochondrial respiratory chain. Hence, few cases of ALS are diagnosed before the age of 50.[19–21]

In ALS there has been particular interest in oxidative stress since mutations in the gene encoding for superoxide dismutase 1 (SOD1) was shown to account for 20% of FALS cases.[22] Other genes with mutations in processes such as excitotoxicity and RNA processing have been implicated in ALS and will be discussed in more detail later. These include genes encoding for TAR DNA-binding protein (TDP-43), charged multivescular body protein 2B (CHMP2B), vesicle-associated membrane protein B (VAPB), angiogenin, fused in sarcoma/translated in liposarcoma (FUS/TLS), and chromosome 9 open reading frame 72 (C9orf72).[23]

1.2.1.1 SOD1

SOD1 is one of three superoxide dismutases that are responsible for detoxifying free superoxide radicals within the body.

SOD1 binds copper and zinc ions (Figure 1). Other metal binding superoxide dismutases include dismutases that bind manganese, iron and nickel ions. Normally SOD1 catalyses the removal of superoxide radicals and converts them into molecular oxygen and hydrogen peroxide in a two-step redox reaction with the copper undergoing the redox process in the active site of the SOD1 dimer (Figure 2).

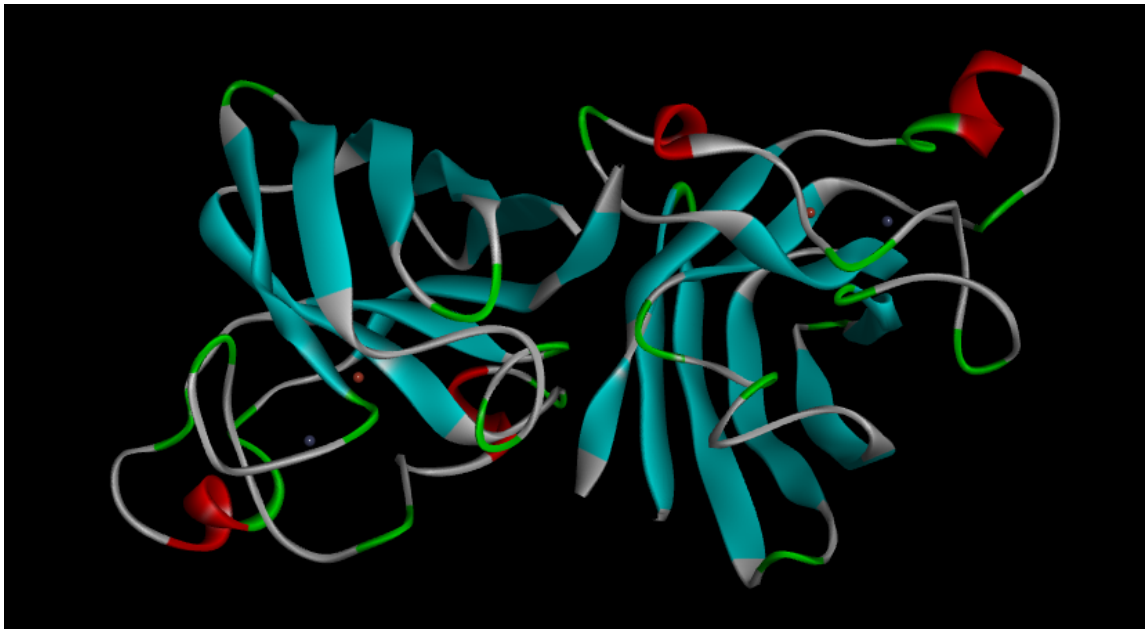


Figure 1: Atomic resolution crystal structure of dimeric SOD1 with Zinc (gray spheres) and Copper ions (brown spheres) (PDB: 2V0A).[24]

With 20% of all FALS cases linked to SOD1 mutations, further investigation into cellular and mouse models of SOD1-related ALS revealed the occurrence of oxidative damage.[23] In addition, the SOD1 protein itself was found to be particularly susceptible to oxidative post-translational modification.[25]

Over 150 mutations in the SOD1 gene have now been linked to familial ALS. Interestingly, there appears to be a geographic distribution in SOD1 mutations, with different geographic locations having differing prevalence of mutations. The D90A mutation is the most common both globally and within Europe. However, in the United States the A4V mutation is the most prevalent amongst SOD1 mutation.[26]

The clear correlation between SOD1 mutations and ALS revealed a link between oxidative stress and disease onset. Initial mechanisms were proposed based on altered dismutase activity.[22,27] However, further investigation has shown that the mechanism is not as simple as it appears. Some mutations do show a decreased dismutase activity of SOD1 (A4V and G85R[28]) with other mutations showing no significant decrease in activity (G37R and G93A[29,30]). Furthermore, a SOD1 knockout mouse did not develop ALS, and varying levels of wild-type SOD1 also had no effect on mSOD1 related ALS.[31,32] This supports the view that loss of SOD1 activity is not the sole cause of ALS and that it is an increase or gain in an unknown function that causes toxicity of mSOD1.

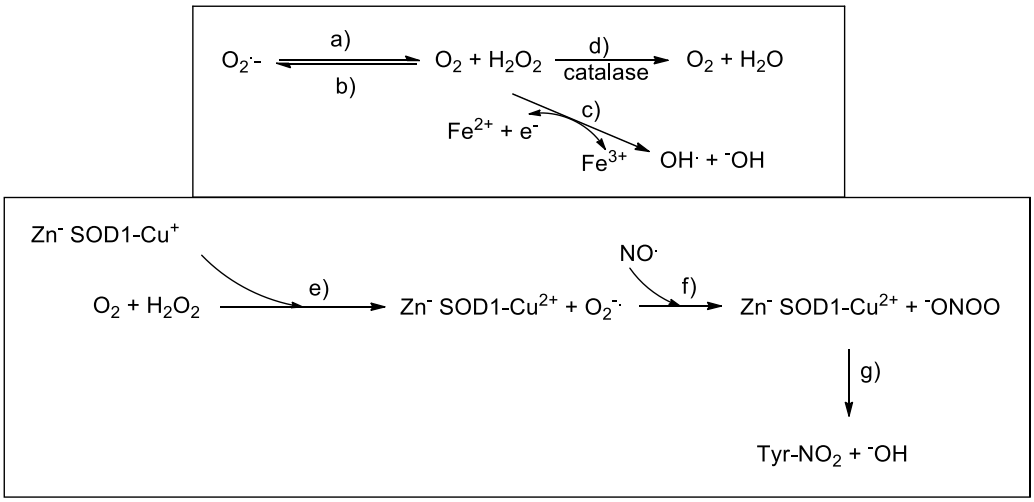


Figure 2: Overall equation for conversion of superoxide radicals.[33] SOD1 has been shown to act as a peroxidase, through either catalysing the reverse of the normal reaction (b) or through taking the hydrogen peroxide produced by (a) to use in the Fenton reaction (c) to produce hydroxyl radicals.[34,35] Zinc-deficient (Zn^-) SOD1, when reduced (to the Cu^+ species) produces superoxide (e) which goes on to form peroxynitrite through a reaction with nitric acid within the active site (f), which causes tyrosine nitration (g).[36,37]

As seen in Figure 2 SOD1 has been shown to act as a peroxidase, through either catalysing the reverse of the normal reaction (b) or through taking the hydrogen peroxide produced by (a) to use in the Fenton reaction (c) to produce hydroxyl radicals.[34,35] Evidence of increased peroxidase activity was found to be present in A4V and G93A SOD1 mutants.[38,39] However, another study found no difference between the peroxidase activity of wild-type SOD1 to that of mutant SOD1.[40] Additionally, it has been shown that some SOD1 mutants have reduced zinc binding capacity, this lead to speculation that molecules other than superoxide are present that are able to react with the oxidised copper in the active site.[41,42] This zinc-deficient (Zn^-) SOD1, when reduced (to the Cu^+ species) produces superoxide (e) which goes on to form peroxynitrite through a reaction with nitric acid within the active site (f), which causes tyrosine nitration (g).[36,37] Evidence supporting this demonstrated that zinc deficient mice with the G93A SOD1 mutation showed increased disease progression, whereas having a zinc supplement led to delayed progression and death.[43]

Despite evidence for occurrence of the reactions (b-g) none of these SOD1 catalysed reactions are likely to be the main contributing factor in SOD1 mutant related oxidative stress. All of the above reactions require a copper ion to be available in the active site. However, evidence has shown that deactivated SOD1 is still capable of causing motor neuron degeneration in G93A, G37R and G85R mSOD1 mice, both through copper loading depletion and mutation of the residues in the active site, showing that copper ions are not required for SOD1 related disease progression.[44,45]

Whilst there is evidence clearly showing that mutations in SOD1 are responsible for some cases of FALS, a body of evidence shows that knockdown of SOD1 in mice does not lead to development of ALS, and induction of oxidative stress is not related to SOD1 catalytic activity. This has led to the conclusion that SOD1 induces oxidative stress through an as yet unknown toxic gain of function mechanism.

1.2.1.2 An alternate mechanism:

Microarray studies of motor neuronal cells expressing mutant SOD1 have shown that genes involved in antioxidant response were down-regulated. These included nuclear erythroid 2-related factor 2 (Nrf2), peroxiredoxins, and several glutathione S-transferases.[46] Nrf2 is known as the master regulator of the antioxidant response and will be discussed in more detail later.

Another possible mSOD1 oxidative stress mechanism has been shown, involving microglia. Evidence suggests that mSOD1 increases microglial superoxide production by NADPH oxidases (Nox). Nox enzymes are transmembrane proteins which reduce oxygen to superoxide when activated.[47] Mutant SOD1 was found to lock Rac1 (Ras-related C3 botulinum toxin substrate 1) in its active GTP-bound form in the Nox2 complex. This has the effect of prolonging production of ROS, leading to damage of nearby neurons.[48,49] Nox2 expression is also increased in the CNS of mSOD1 mice and patients with ALS, and the targeted deletion of the Nox2 gene in SOD1 mutant mice leads to an extension in their survival.[50,51]

Many studies have shown that oxidative stress is highly involved in many of the processes that contribute to injury of motor neurons. These include excitotoxicity, mitochondrial impairment, protein aggregation, ER stress and changes in astrocyte and microglia signalling. This is represented in Figure 3. Therefore, if oxidative stress was reduced it could have beneficial effects across many of the multiple mechanisms that contribute to motor neuron degeneration.

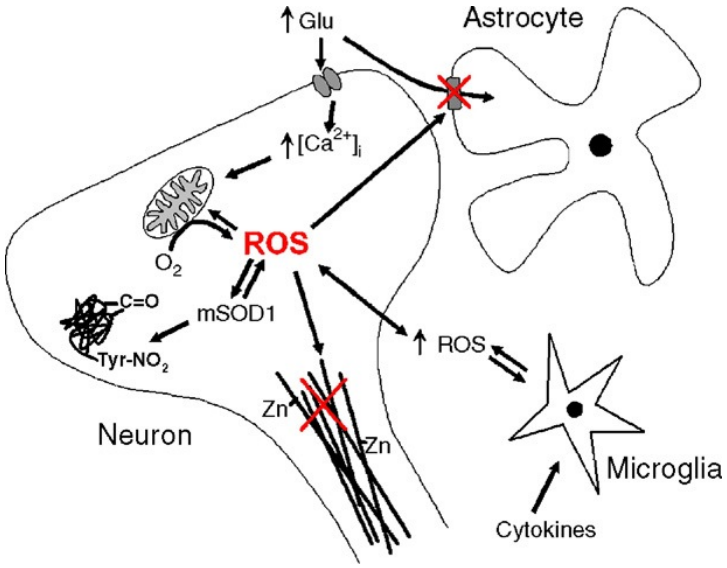


Figure 3: Diagram illustrating the effects of oxidative stress on cells. Reproduced with permission from Barber et al. (2010)[23] Oxidative stress can influence multiple factors in neurodegeneration in ALS. Excitotoxicity increases amounts of intracellular calcium levels that are buffered by mitochondria, this leads to increased ROS production. Inhibition of glutamate uptake through the EAAT2 transporter in glial cells is also caused by ROS. Furthermore, ROS can cross the cell membrane and activate microglia. The activated microglia respond by releasing cytokines and more ROS. mSOD1 catalysed oxidative reactions increases the production of highly reactive peroxynitrite and hydroxyl radicals. This causes nitration and aggregation of proteins including mSOD1 itself, and may also inhibit neurofilament assembly and cytoskeletal transport. Zinc binding to neurofilaments could deplete zinc binding to mSOD1 and exacerbate aberrant SOD1 chemistry.

1.2.2 Excitotoxicity

Excitotoxicity is where the excitatory action of glutamate becomes toxic either due to an increased level of synaptic glutamate, or by increased sensitivity of the postsynaptic neurons to glutamate.[52] This leads to an excessive stimulation of neurotransmitters which results in neuronal injury and death.[53]

Human motor neurons have a high concentration of glutamate receptors, and glutamate plays a key role in the major excitatory neurotransmission process within the human CNS (central nervous system).[54,55] In normal glutamatergic neurotransmission (Figure 4) glutamate is released from the presynaptic neuronal terminal and activates receptors on the postsynaptic neuron. Reuptake transporter proteins, located mainly on perisynaptic astrocytes remove the glutamate molecules from the synaptic cleft and terminate the excitatory signal. The most abundant of the reuptake proteins is known as excitatory amino acid transporter 2 (EAAT2).

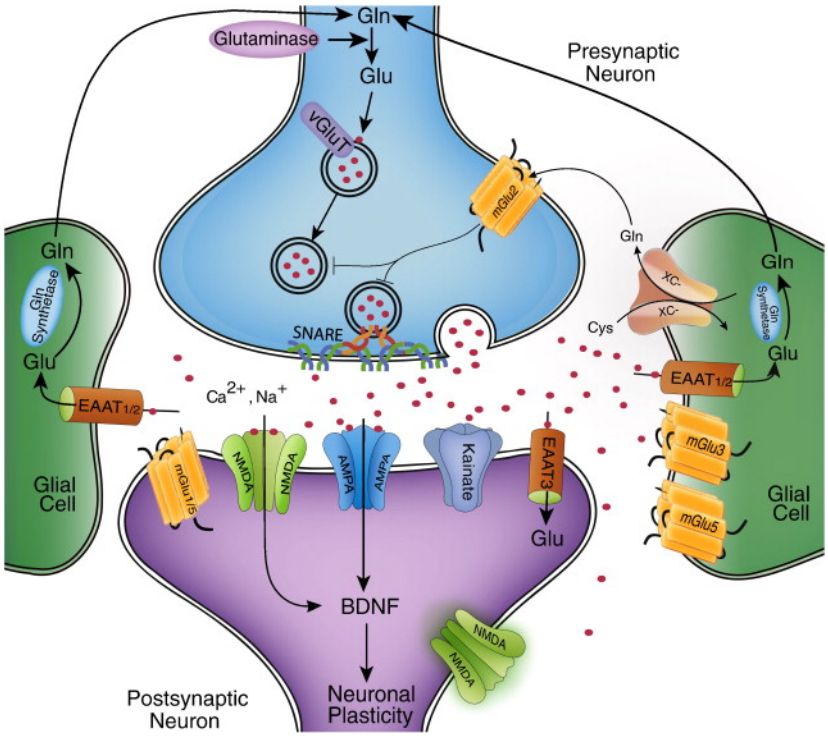


Figure 4: Diagram illustrating glutamatergic neurotransmission. During Normal glutamatergic neurotransmission when there is excessive simulation of the glutamatergic receptors the system needs to react to control the levels of glutamate to reduce excitotoxic damage to the CNS. Glutamine (Gln) is converted by glutaminase to glutamate (Glu). Glu is put into presynaptic vesicles by the vesicular Glu transporter (vGluT) proteins that is then released into the synapse through vascular interactions with SNARE proteins in a voltage-dependent manner. The released Glu is taken up by excitatory amino acid transporters (EAATs), which are expressed mainly on astroglia. In astrocytes Glu is then recycled into Gln by Gln synthetase after which it is exported extracelluarly to be taken up again by neurons. Glu receptors such as NMDA, AMPA, AMKA, and mGluRs are also present on the presynaptic and postsynaptic neurons in addition to expression on glial cells. These receptors are split into ionotropic and metabotropic receptors and respond differently to Glu simulation, based on their receptor subtype, local and any interactions they may have with any scaffolding and signaling proteins. Simulation of these receptors can lead to rapid ionotropic effects and synaptic plasticity, leading to long-term potentiation and long-term depression. (reproduced with permission from Niciu et al. 2012).[56]

Mammalian motor neurons are susceptible to toxic effects following activation of N-methyl-D-aspartate (NMDA) or non-NMDA receptors.[57] This is because of a relatively low expression of the GluR2 subunit component of the AMPA receptor causing the cell to become more permeable to Ca^{2+}.[58] In excitotoxicity, when there is a large concentration of glutamate or other excitotoxins such as NMDA and kainic acid, the glutamate receptors such as the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor and the NMDA receptor are over-activated. This causes a depolarisation-mediated influx of water, sodium and chloride ions (Na^+, Cl^-), causing a reversible neuronal swelling.[59] This process is then followed by high levels of calcium (Ca^{2+}) entering the cell, which activates a number of enzymes.[60] These enzymes including phospholipases, proteases, calpain, and endonucleases then progress to damage the structural components of the cell.[61]

Interestingly, activation of glutamate receptors is an important mechanism in generating intracellular free radicals.[62] Activation of both N-methyl-D-aspartate (NMDA) and nonNMDA glutamate receptors leads to increased production of superoxide radicals.[63] As indicated above, oxidative stress can cause damage to intracellular DNA, proteins, and lipids. Moreover, glutamate can directly interfere with the neurons ability to withstand oxidative stress by depleting intracellular glutathione.[64]

Whilst there is clear evidence linking excitotoxicity as a causal mechanism for motor neuron injury in ALS, evidence suggesting that it is a **primary** disease mechanism is lacking. Patients with ALS have been found to have increased levels of glutamate in the cerebrospinal fluid (CSF).[65] In addition some patients also displayed reduced levels of EAAT2 expression in affected areas of the CNS.[66] However, it is unknown as to whether it is a cause or consequence of neuronal loss.[67]

1.2.3 Mitochondrial dysfunction

Mitochondria are extremely important membrane-enclosed organelles that are found in most eukaryotic cells.[68] They are responsible for intracellular energy production (through generating most of the ATP needed for the cell's survival), in addition to calcium homeostasis and apoptosis control amongst others.[69] Mitochondrial dysfunction has been linked to ALS pathogenesis through several bodies of evidence.

Firstly, the deregulation of energy metabolism has been implicated in motor neuron dysfunction in ALS. Tissue from ALS patients and from mSOD1 models showed defective respiratory chain function that is associated with oxidative damage to mitochondrial proteins and lipids.[70,71]

Secondly, in mutant SOD1 (mSOD1) transgenic mice, mutated SOD1 protein was found to aggregate in the mitochondrial intermembrane space.[72] In addition, as the mice aged it was found that the mutated protein increased its adherence to the mitochondrial outer membrane.[73] The increase in deposition of SOD1 is thought to lead to organelle dysfunction within the spinal cord through affecting protein import, ionic homeostasis, mitochondrial motility, mitochondrial fission/fusion, or regulation of apoptosis.[74,75]

Thirdly, calcium homeostasis is an important part of mitochondrial function. Calcium buffering was found to be reduced in mitochondria purified from the central nervous system (CNS) of mutant SOD1 mice.[76] This could lead to motor neurons becoming more susceptible to the change in calcium homeostasis that occurs in excitotoxicity. Moreover, ALS models have shown that endoplasmic reticulum (ER) stress is also present.[77] ER stress is thought to disrupt the ER-mitochondria calcium exchange mechanism.

The damage to mitochondria is thought to contribute to the mechanism of motor neuron cell death in ALS which is predicted to involve the activation of caspases and apoptosis.[78] Further support for mitochondrial dysfunction comes from evidence showing altered mitochondrial morphology. Altered morphology has been observed in patients with ALS where both skeletal muscle and spinal motor neurons showed morphologically altered mitochondria.[72] It was also seen in the motor neuronal NSC-34 cell line when mutant SOD1 was expressed.[79]

1.2.4 Defects of RNA processing

RNA processing was first implicated in motor neuron disease in the 1990's when mutations in survival motor neuron protein 1 (SMN1) were identified as being a cause for spinal muscular atrophy.[80] Since then a protein known as TAR DNA-binding protein of 43kDa (TDP-43) was discovered to be a major component of ubiquitinated cytoplasmic inclusions which develop as motor neurons become injured in ALS.[81] This brought attention to changes in RNA processing as a potential pathophysiological mechanism in ALS.

1.2.4.1 TDP-43

In 2008 mutations in the TARDBP gene encoding TDP-43 were identified in some patients with ALS.[82] It has been found that mutations in the TARDBP gene are responsible for around 4% of FALS cases and 1.5% of SALS cases.[83]

TDP-43 protein is normally found localised within the nucleus and functions during RNA-processing in transcriptional regulation, alternative splicing and microRNA processing.[84] However, in patients with ALS, TDP-43 is predominantly found in the cytoplasm of neuronal cells (BE-M17 neuroblastoma cells).[85] Moreover, cases of ALS were found to have cytoplasmic inclusions with positive TDP-43 immunostaining in both glial and neuronal cells. However mSOD1, and fused in sarcoma (FUS) related ALS cases did not show the same TDP-43 positive inclusions. In addition, biochemical studies showed that the TDP-43 within the cytoplasm was of abnormal variants of the protein.[83] The majority of mutations are contained within exon 6 of the TARDBP gene. Exon 6 encodes the glycine-rich domain of TDP-43 which is responsible for protein-protein interactions and nuclear transport.[83] Recent studies have also shown that in fibroblast cell lines from ALS patients with mutations of TARDBP there was a loss of nuclear expression of TDP-43 and extensive changes in splicing, including in genes also implicated in ALS. This lead to the conclusion that mutant TDP-43 dysregulates the alternative splicing of mRNA.[84]

1.2.4.2 FUS

FUS is another DNA-binding protein involved in transcriptional regulation, micro RNA and RNA processing and RNA transport. It has been found that mutations in the FUS gene account for around 4% of FALS and less than 1% of SALS cases.[83] In FUS most of the mutations are located within exons 13 to 15. These encode for the arginine and glycine-rich region and the nuclear localisation signal of FUS. These mutations in FUS disrupt the transportin-mediated nuclear localisation of FUS as well as triggering the formation of FUS-containing cytoplasmic stress granules.[86,87]

1.2.4.3 Other genes involved in RNA processing

Additional evidence that dysfunctional RNA metabolism is important in ALS comes from the identification of mutations in angiogenin (ANG) and senataxin (SETX), a DNA-RNA helicase associated with juvenile-onset FALS.[88,89] ANG is known to regulate ribosomal RNA transcription and acts as a transfer RNA-specific ribonuclease. ANG is thought to normally prevent cell death through inhibiting the translocation of the apoptosis-inducing factor into the nucleus.[90] The SETX protein is a DNA and RNA helicase which forms part of large ribonucleoprotein complex and is involved in DNA repair as a consequence of oxidative stress, and RNA processing.[88] The exact mechanism by which the SETX mutant protein causes ALS is not known. Mutations in ANG are likely to affect the normal function of the protein, as overexpressing ANG in mSOD1 mice extends their lifespan.[90]

All of the above, together with the evidence showing that biomarkers of RNA oxidation are present in ALS and mSOD1 mice suggests that dysregulated RNA processing may play an important part in injury to motor neurons and in ALS.[91]

1.2.5 Genetic Factors: C9ORF72

With the realisation that genetic factors are an important in many of the pathological mechanisms in ALS in both SALS and FALS the discovery of gene mutations (such as those in SOD1 and FUS) related to cases of ALS has led to new important avenues to research in ALS. Recently a hexanucleotide repeat expansion has been identified in an intron of chromosome 9 open reading frame 72 (C9ORF72) as the cause for cases of both SALS and FALS linked to the 9q21-q22 chromosomal locus.[92,93] C9ORF72 is of particular significance due to its pathogenic mechanism that links ALS with frontotemporal dementia (FTD).[94] This has led to suggestions that there is the potential to develop a joint therapy that can treat both diseases.[95]

Many other genetic factors are implicated in ALS and are important in many areas of ALS pathogenesis, as can be seen elsewhere in this review. A full table of known genetic subtypes in ALS is listed below (Table 2).

Table 2: Genes associated with ALS *(reproduced from Ferraiuolo et. al, 2011).*[84]

	Genetic subtype	Chromosomal locus	Gene	Onset /inheritance	Reference
Oxidative stress					
	ALS1	21q22	Superoxide dismutase 1 (*SOD1*)	Adult/AD	Rosen (1993)[22]
RNA processing					
	ALS4	9q34	Senataxin (*SETX*)	Juvenile/AD	Chen *et al.* (2004)[88]
	ALS6	16p11.2	Fused in sarcoma (*FUS*)	Adult/AD	Kwiatkowski *et al.* (2009)[96] Vance *et al.* (2009)[97]
	ALS9	14q11.2	Angiogenin (*ANG*)	Adult/AD	Greenway *et al.* (2006)[89]
	ALS10	1p36.2	TAR DNA-binding protein (*TARDBP*)	Adult/AD	Sreedharan *et al.* (2008)[82]
Endosomal trafficking and cell signalling					
	ALS2	2q33	Alsin (*ALS2*)	Juvenile/AR	Yang *et al.* (2001)[98]
	ALS11	6q21	Polyphosphoinositide phosphatase (*FIG4*)	Adult/AD	Chow *et al.* (2009)[99]
	ALS8	20q13.3	Vesicle-associated membrane protein-associated protein B (*VAPB*)	Adult/AD	Nishimura *et al.* (2004)[100]
	ALS12	10p13	Optineurin (*OPTN*)	Adult/AD and AR	Maruyama *et al.* (2010)[101]
Glutamate excitotoxicity					
	ND	12q24	d-amino acid oxidase (*DAO*)	Adult/AD	Mitchell *et al.* (2010)[102]
Ubiquitin/protein degradation					
	ND	9p13–p12	Valosin-containing protein (*VCP*)	Adult/AD	Johnson *et al.* (2010)[103]
	ALSX	Xp11	Ubiquilin 2 (*UBQLN2*)	Adult/X-linked	Deng *et al.* (2011)[104]
Cytoskeleton					
	ALS18	17p13.3	Profilin 1 (PNF1)	Adult/AD	Wu et al. (2013)
	ALS–dementia–PD	17q21	Microtubule-associated protein tau *(MAPT)*	Adult/AD	Hutton *et al.* (1998)[105]
Other genes					
	ALS5	15q15–q21	Spatacsin (*SPG11*)	Juvenile/AR	Orlacchio *et al.* (2010)[106]
	ALS–FTD	9p13.3	σ Non-opioid receptor 1 (*SIGMAR1*)	Adult/AD Juvenile/AR	Luty *et al.* (2010)[107] Al-Saif *et al.* (2011)[108]
	ALS–FTD	9q21–q22	Chromosome 9 open reading frame 72 (*C9ORF72*)	Adult/AD	Hosler *et al.* (2000)[109] Renton *et al.* (2011)[93] De Jesus-Hernandez *et al.* (2011)[92]
Unknown genes					
	ALS3	18q21	Unknown	Adult/AD	Hand *et al.* (2002)[110]
	ALS7	20ptel–p13	Unknown	Adult/AD	Sapp *et al.* (2003)[111]

Abbreviations: AD, autosomal dominant; ALS, amyotrophic lateral sclerosis; AR, autosomal recessive; FTD, frontotemporal dementia; PD, Parkinson disease; ND, not determined.

1.2.6 Protein Aggregation

Protein aggregation is highly implicated in neurodegenerative diseases including ALS.[112] Previous research has shown that skein-like and spherical ubiquitnated protein inclusions were identified to be an important feature in ALS.[113] Several different proteins have been associated with forming inclusions in ALS, including SOD1, and TDP-43.

Aggregates of SOD1 have been found in patients with FALS, additionally mSOD1 cellular and mouse models also show SOD1 inclusions.[114] Further evidence for SOD1 inclusions was found when monoclonal antibodies which were specific for epitopes of misfolded SOD1 showed the presence of the misfolded protein primarily within degenerating motor neurons.[115]

TDP-43 was originally identified as a major protein constituent of the inclusions in ALS, as mentioned previously, TDP-43 is normally found within the nucleus but in cases of ALS TDP-43 is found as inclusions in the cytoplasm.[81,85] The distribution of TDP-43 to the cytoplasm is an early event in the pathogenesis of ALS, and thus counteracting this may present a possible therapeutic approach to treatment of ALS.[116]

1.2.7 Neuroinflammation

It has long been known that the activation of microglia occurs in murine and human variants of ALS.[117–119] This suggests that neuroinflammation may play a part in the pathogenesis of ALS. Further evidence supporting this came from research showing that expression of particular cytokines and chemokines is changed in human ALS patients and in ALS animal models.[120–125] Although the majority of studies of cytokine expression have been performed in mSOD1 mice models due to the ease-of-access of CNS tissue in comparison to human tissue,[120,126] some studies with human samples have also been tested. Levels of inflammatory cytokines are increased in human ALS cerebrospinal fluid and blood serum.[121,127]

Several studies implicate tumour necrosis factor-α (TNFα) in ALS with studies showing elevated levels in blood serum from ALS patients.[121,122] Furthermore, TNFα is one of the earliest gene products to be upregulated in spinal cord tissue from mSOD1 with the G93A mutation.[126,128] Studies showing that primary mSOD1 microglia have over-expression of TNFα after simulation by lipopolysaccharide *in-vitro* gives further evidence that TNFα has a role to play in the neuroinflammation involved in ALS.[129] However, recent evidence has questioned the role of TNF after studies showing that crossing $SOD1^{G93A}$ mice with TNF knockout mice didn't affect progression of the

disease.[130] Nevertheless the role of increased TNFα in humans with ALS is still a subject of debate and research.

Astrocyte activation has also been shown to be important in neuroinflammation. Astrocytes cultured from 7 day old mice transgenic for human G93A SOD1 showed an increase in levels of prostaglandin E_2, leukotriene B_4, and reactive oxygen and nitrogen species under both basal and activating conditions.[131] In addition glial cells have also been found to have enhanced cytotoxic potential in mSOD1 mice after microglial activation.[129]

1.2.8 Non neuronal cells

It is important to note that motor neuron death is not cell-autonomous and needs contributions from surrounding neuronal cells with the most likely candidate being neighbouring glial cells. The evidence supporting this shows that in mSOD1 mice motor neurons with mSOD1 survive longer when surrounded by a wild type glia environment, compared with when mSOD1 was expressed in both neurons and the surrounding glial cells. Additionally the disease progression can be linked to amount of glial mSOD1 expressed.[132]

1.2.9 Summary

In summary, ALS is a multifactorial disease of which a major component is oxidative stress, which has been shown to have the potential to influence several factors within neurodegereration and ALS, including excitotoxicty, mitochondrial dysfunction, and protein aggregation. This gives hope that a therapeutic agent targeting reduction in oxidative stress could bring benefits to patients with ALS.

1.3 Current therapeutic strategies

The need for a new therapy for ALS is of great importance as the only current Food and Drug Administration (FDA) approved drug for ALS is Riluzole (Figure 6) which has the effect of extending life expectancy of ALS patents by around 2-4 months on the typical 2-3 year survival rate.[133]

Despite recent advances in the field of ALS, current therapeutic strategies have been hampered by the lack of understanding in ALS etiology. However, much research has been performed to discover new targets and therapies for ALS due to the invariably fatal outcome of the disease, the hope is that through understanding the neurobiology of ALS new therapies can be developed for ALS and also for other neurodegenerative diseases.

1.3.1 Initial development of therapeutics within ALS

Initial development of treatments for ALS focused on antioxidants and glutamate receptor agonists. This was because at the time it was known that human nervous system tissue in ALS patients showed oxidative stress, protein aggregation and loss of homeostasis due to in part, defective glutamate transport.[134] Therefore, efforts were focused on existing compounds known to have antioxidant scavenging properties, such as vitamin E, or glutamate receptor agonists such as N-Methyl-D-aspartate.[135,136] This approach was later supported when SOD1 mutations were identified in FALS, and when the first SOD1 transgenic mouse models showed a loss in the glial glutamate transporter EAAT2.[137]

The results of these studies gave promising results with Riluzole and gabapentin, the two anti-glutamate agents prolonging survival and slowing the progression of the disease in G93A mSOD1 mice.[136] However only Riluzole progressed further with clinical trials of vitamin E, gabapentin, and memantine (a NMDA receptor antagonist) failing to show any significant benefits in ALS patients, although in the case of vitamin E this is thought to be partly due to the trial being underpowered (Figure 5).[138–140] Clinical trials of Riluzole (Figure 6) showed the drug slowed the progression of SALS, later studies showed that Riluzole improved motor function in a transgenic mice model of ALS.[135,141] The reason for other anti-glutamate agents failing in clinical trials is unknown, however it is thought that Riluzole may have multiple neuroprotective effects extending beyond inhibition of pre-synaptic glutamate release.[142]

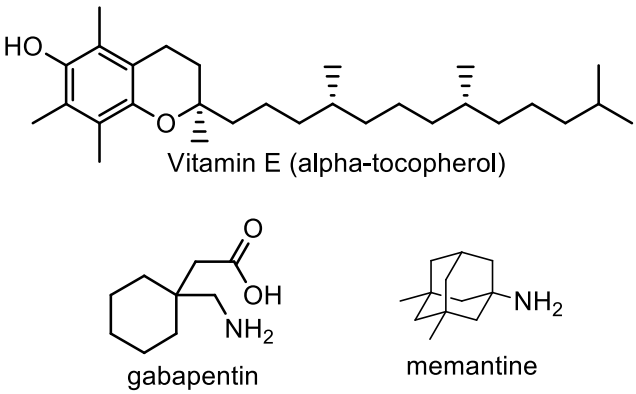


Figure 5: The chemical structures of vitamin E, gabapentin and memantine

Further clinical trials were conducted with a 100 mg/day dose showing positive effects in tests measuring quality of life and tracheostomy-free survival, giving an average of an additional 3 months life expectancy to patients with ALS.[133] This led to Riluzole being approved by the United States Food and Drug Administration (FDA) under the trade name Rilutek and is still the only approved disease modifying agent for the treatment of ALS.

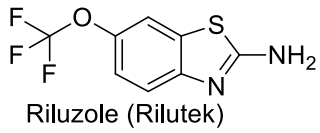


Figure 6: The chemical structure of Riluzole (Rilutek)

There have been more recent efforts in developing treatments based on the anti-excitotoxin approach. Dexpramipexole (Figure 7) a benzothiazole, is structurally similar to riluzole and the enantiomer of pramipexole which is a dopamine agonist for the treatment of Parkinson's disease.[143]

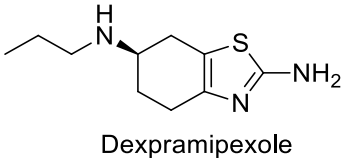


Figure 7: The chemical structure of Dexpramipexole

Dexpramipexole was thought to protect motor neurons from glutamate excitotoxicity possibly through an effect in mitochondria.[144] Clinical trials of dexpramipexole in patients showed that the drug was well tolerated, with linear pharmacokinetics, and phase II clinical trials in ALS patients showed a slowing of ALS disease progression.[145,146] However, a recent conclusion of a phase III clinical trial failed to show efficacy in

function or survival so development has been discontinued.[147] This has resulted in very few drug candidates left in late stage clinical trials for ALS.

1.3.2 Clinical Trials in Amyotrophic Lateral Sclerosis

Due to the long history of oxidative stress in ALS physicians have often recommended patients take dietary antioxidants.[33] However, a Cochrane review of trials of antioxidant treatments found no beneficial effect in ALS. Results of this should however be treated with caution as it was stated in the review that the trials were generally of poor methodological quality with weak statistical power.[148] However, many of these may have also failed due to the difficulty in penetrating the blood-brain barrier (BBB) at a pharmacologically relevant dose.[149]

Current on-going clinical trials include a variety of different approaches. These range from therapies for managing symptoms of the disease to treatments that can slow or cure the progression of the disease. A list of current clinical trials and progress is listed below in Table 3.

Table 3: Molecules currently in clinical trials for the treatment of ALS

Drug	Progress	Details
Tirasemtiv	Finished Phase IIb	Tirasemtiv, a muscle troponin-A activator, delays muscle fatigue in patients with ALS and increases the force their muscles can generate.[150] The end of the study in 2015 found no difference in disease progression. [151]
GSK1223249	Phase II	GSK1223249 is known to target neurite outgrowth inhibitor A (Nogo-A). Nogo-A is known to impair neurone regeneration and has been found in increased levels in patients with ALS. [150]
NP001	Soon to enter Phase III	Successful phase IIb trials showed that the drug was well tolerated and had potential MND benefits. NP001 is thought to regulate macrophage activation through transforming them from a

		neurotoxic state to a neuroprotective state.[152]
Neuralstem and Brainstorm	Phase II	Neuralstem uses neural stem cells whereas brainstorm uses bone-marrow stem cells that are then reprogrammed to neuron-supporting cells. The treatments are aimed at promoting neuron growth and regrowth.[153]
Arimoclomol (CytRx)	Phase II/III	Arimoclamol (CytRx) enhances the heat shock response, a natural cell defence mechanism. Heat shock proteins (HSP's) help to prevent aggregation of proteins, and help to prevent cell death in motor neurons.[150]
sNN0029	Entering Phase II	sNN0029 is a drug that contains vascular endothelial growth factor (VEGF). Animal studies showed that decreasing the levels of VEGF in mSOD1 mice decreased both the age at onset of the disease and their life span.[12]
GM604	Phase II	GM604 is a small hexapeptide that is derived from a motoneuronotrophic factor. Neurotrophic factors are important in the development of the nervous system so GM604 may help to protect neurons.[154]
ISIS SOD1Rx	Phase I completed and found that ISIS SODRx was well tolerated.	SOD1Rx is an antisense drug treatment. Treatment of mSOD1 rats with SOD1Rx showed a reduction in disease progression of ALS.[155,156]

1.4 The Keap1-Nrf2-ARE pathway

As mentioned previously it is still unclear as to whether antioxidant therapies can be effective therapies in ALS. There is widespread evidence that oxidative damage occurs in ALS and although the cause of increased ROS is not yet known, antioxidant treatments remain one of the key potential areas for drug development.

A recent paper described the Kelch-like ECH-associated protein (Keap1)-nuclear factor erythroid 2-related factor 2 (Nrf2)-antioxidant response element (ARE) pathway as "one of the most important cellular defensive mechanisms against oxidative stress and xenobiotic damage".[157] The pathway is ubiquitously expressed within the body with the highest concentrations (from highest to lowest) in the kidney, muscle, lung, heart, liver, and brain.[158] It has two main roles in the antioxidant response. Firstly, it is responsible for transcriptionally regulating a large number of cytoprotective proteins,[159–162] secondly, activation of this pathway can regulate the cellular redox state and metabolism.[163–165]

1.4.1 Dysregulation of the Nrf2-Keap1 pathway in ALS

Increased age is a risk factor in development of ALS, due to increased mitochondrial leakage of ROS. In 24 month aged rats, Nrf2 levels have been found to be substantially decreased compared to young 3 month old rats.[166] More specifically, the Nrf2-Keap1 pathway has been shown as being directly implicated in both animal models of ALS and post-mortem tissue of patients with ALS.

In-vivo, levels of Nrf2 protein in motor neurons and lumbar spinal cord in mSOD1 transgenic mice were found to be dramatically increased, whilst protein levels of Keap1 were found to be decreased. However, downstream gene expression of ARE regulated genes was only modestly increased, suggesting the pathway is dysregulated.[167] Overexpression of Nrf2 in astrocytes had a protective effect against neurodegeneration in mouse models of ALS, with decreased glial activation also shown.[168] Finally, knockout Nrf2 mice with the G93A SOD1 mutation had accelerated astrocytic proliferation and motor neuron loss.[169]

Dysregulation has also been shown in ALS post-mortem brain and spinal cord tissue. Levels of Nrf2 were found to be reduced at both the RNA and protein level, whilst levels of Keap1 mRNA were found to be increased without a similar increase in protein levels.[170] A study of various neurodegenerative diseases, which included ALS found that Keap1 was found in p62-containing neuronal cytoplasmic inclusions.[171] Accumulation of nuclear Nrf2 and elevated levels of oxidative stress has also been found to be in motor

neuron-like cells expressing mutant TDP-43.[172] Transgenic mice expressing the TDP-43 A315T mutant showed increases in the level of the Nrf2 target gene HO-1.[173]

More recently, another RNA binding protein, RBM45 a predominantly nuclear protein in neural development has been linked to ALS. A proteomic study of CSF samples from ALS and control patents showed increased levels of RBM45 in patient samples, whilst motor neurons isolated from ALS spinal cord showed co-localisation of RBM45 and TDP-43 in cytoplasmic inclusions.[174] RBM45 was shown to bind Keap1 and interfere with proteasomal degradation of the protein leading to stabilisation. This results in the protective antioxidant pathway being impeded, leading to an increase oxidative stress and induced cellular toxicity. This mechanism may explain how the pathway could be dysregulated in ALS.[175]

1.4.2 Basic mechanism of the pathway

Under the normal reducing (basal) conditions of cell growth Keap1 targets Nrf2 for ubiquitination along with association to Cul3 (cullin E3 ubiquitin ligase). Ubiquitination of Nrf2 leads to subsequent proteolysis of Nrf2 via the 26S proteasome, this leads to Nrf2 having a half-life of around 20 minutes under basal conditions.[176,177] However, under exposure of a cell to oxidative stress, Nrf2 stops undergoing Keap1-mediated degradation. Newly synthesised Nrf2 is able to transfer from the cytoplasm into the nucleus where it activates ARE-dependent gene expression of over 100 protective genes.[178,179] Amongst the genes unregulated by Nrf2 are heme oxygenase 1 (HO1) and NAD(P)H quinone oxidoreductase 1 (NQO1), these are often used as biomarkers in monitoring Nrf2-ARE activation. Interestingly Nrf2 itself contains an ARE sequence so regulates its own expression levels.[180] Activation of Nrf2 leads to increased levels of Nrf2 in the cytosol ensuring larger amounts are available for future activation.[33]

1.4.3 Components of the pathway

1.4.3.1 Keap1

Keap1 is a cysteine rich 69-kDa protein containing 627 amino acids, of which 27 are cysteine. The protein is localised in the cytoplasm and is responsible for regulating protein turnover of Nrf2. Whilst a full length crystal structure has not been solved, single particle electron microscopy has shown Keap1 to be a cherry-bob-like structure.[181] The structure of Keap1 can be broken down into 5 regions or domains (see Figure 8): the N-terminal region (NTR); the Broad complex, Tramtrack, and Bric-a-Brac sections make up the BTB domain; the intervening region (IVR); the Kelch domain, a double glycine repeat (DGR); and the C-terminal region (CTR).[181]

The BTB, IVR, and DGR domains are the most important for the function of Keap1. It is known that the BTB domain is involved in the homodimerisation of the protein and also involved with the interaction to the cytosolic Cullin (Cul3) based ubiquitin E3 ligase complex, the ligase responsible for Nrf2 ubiquitination.[181–183] The IVR domain is known to contain the majority of the key reactive cysteine residues that act as sensors towards oxidation and the nuclear export signal (NES) sequence.[181,184] The DGR or Kelch domain is responsible for Nrf2 binding, and consists of 6 Kelch repeats, it also contains some of the reactive cysteine residues believed to be involved in redox sensing.[185] Of these cysteine residues seven of them are known to be highly reactive towards ROS and electrophiles. These cysteine residues are thought to act as redox sensors and are very important in the regulation of Nrf2.[186,187]

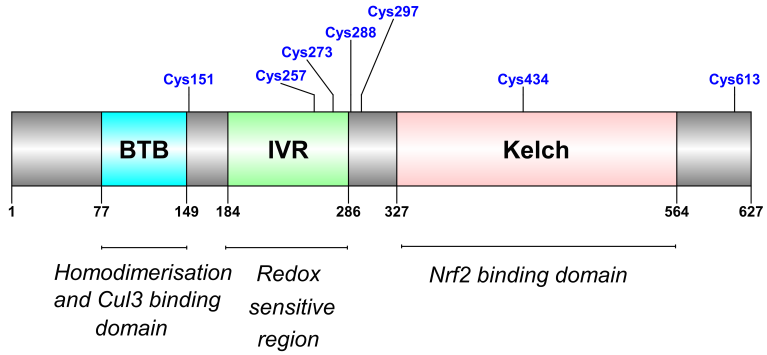


Figure 8: The 5 domains of Keap1 (reproduced from Magesh et al. 2012)[157] From left to right, N-terminal region (NTR), broad tramtrack bric-a-brac (BTB) domain, Intervening Region (IVR), double glycine repeat (DGR, or Kelch domain), C-terminal region (CRT). Key cysteine residues involved in the redox sensing of Keap1 are shown in blue along the length of the protein.

1.4.3.2 Nrf2

Nrf2 is a 605 amino acid basic leucine Zipper (bZip), cap "n" collar (CNC) transcription factor. Nrf2 is made up of 6 distinct domains (Neh1 to Neh6). The Nrf2-ECH homolog-1 (Neh1) of Nrf2 has a basic leucine zipper (bZip) motif known to be responsible for the binding to DNA as a heterodimer through which Nrf2 can bind with the small musculoaponeurotic fibrosarcoma (Maf) protein.[162] Small Maf proteins are basic leucine zipper-type transcription factors that are able to bind to DNA as well as Nrf2 and regulate gene expression. There are three types of small Maf proteins, sMafF, sMafG, and sMafK. The sMaf is crucial to Nrf2 function as a triple sMaf knockout failed to show induction of Nrf2 regulated genes when stimulated.[188] This Maf-Nrf2 heterodimer then is able to bind to ARE activating gene expression and antioxidant response.[159] The domain that binds Keap1 is the N-terminal Neh2 domain.[160] This domain contains two motifs that bind the Kelch domain on Keap1, the low affinity DLG motif and the high

affinity ETGE motif.[185] Neh3 binds to CHD6 (a chromo-ATPase/helicase DNA binding protein), and may affect the transcriptional activity towards transcription of ARE-dependant genes.[189] Neh4 and Neh5 control the transactivational activity synergistically through binding to a transcriptional co-activator.[190] Neh6 however controls the Keap1-independent negative regulation of Nrf2.[191] Keap1-indepentant negative regulation of Nrf2 is mediated through GSK-3 mediated phosphorylation of S335 and S338 of the Neh6 domain. This enables recruitment of SCF/β-TrCP which enables ubiquitination of Nrf2 on K317 and K322 via a Cul1 mediated mechanism.[192,193]

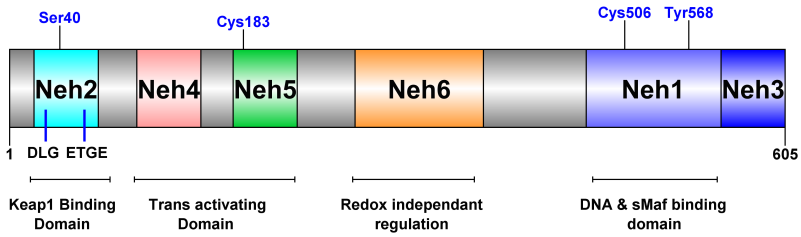


Figure 9: The 6 domains of Nrf2 with key amino acids shown in blue.[157] Key domains include Neh2 which binds Keap1, Neh4 and Neh5 which are responsible for trans activation of Nrf2, Neh6 which contains the redox independent regulation of Nrf2 through GSK-3 mediated phosphorylation, and Neh1 which binds DNA and the small Maf protein.

1.4.3.3 ARE

ARE, is a *cis*-regulatory element which contains specific DNA sequences that are also contained within genes that encode for the cytoprotective proteins and detoxifying enzymes. ARE cannot be represented as a single sequence because different ARE sequences are found in different genes.[194] However, sequences of AREs do contain some similarities within the typical functional length of 16 nucleotides. The consensus sequence can be represented as 5'-T $^{A}/_{C}$A *n n* $^{A}/_{G}$T G A $^{C}/_{G}$*n n n* G C $^{A}/_{G}$-3', where *n* is the variable that can indicate any nucleotide.[157] As stated previously, under conditions of oxidative stress ARE-dependent gene expression is activated by Nrf2-Maf heterodimer. However, a deactivator of the ARE pathway is also known. Bach1 (BTB and CNC homology-1) negatively regulates certain ARE-dependent genes. Bach1 is able to bind to ARE and form a dimer with Maf, this then prevents Nrf2 from binding to DNA.[195]

1.4.4 Keap1-Nrf2-ARE pathway mechanism

Under normal (basal) conditions Nrf2 is degraded via Keap1 through a specific ubiquitin-26S proteasomal pathway, and is rapidly degraded with a half-life of <20 minutes.[182,196]

The proposed mechanism through which Nrf2 is degraded is as follows (see Figure 10). The Keap1 homodimer (formed through the BTB domain)[183] associates with Nrf2

through the Kelch domains of Keap1 via a hinge and latch model, where the higher affinity ETGE motif is the hinge and the lower affinity DLG motif is the latch.[197] The Keap1-Nrf2 complex now also associated with Cul3 brings Nrf2 in close proximity to the Cul3-based E3 ligase complex leading to polyubiquitination of the lysine rich region of Nrf2 between the DLG and ETGE motifs.[176,198] The polyubiquitinated Nrf2 is then degraded by the 26S proteasome.[199] The Nrf2-Keap1-Cul3 stoichiometry has been a subject of debate within the field; previous research had shown a 1:2:1 stoichiometry, however recent research using crystallographic analysis of a variety of BTB-Kelch E3 ligases suggested a 1:2:2 complex.[200] Additional studies using co-transfection of two differently tagged Cul3 proteins with Keap1 in HEK293T cells gave more evidence towards a 1:2:2 stoichiometry.[201]

Under exposure to stressed conditions, such as in the presence of oxidative stress or electrophiles the redox sensitive cysteine residues on Keap1 are thought to induce a conformational change in Keap1, thus preventing the ubiquitination of Nrf2. This allows the newly synthesised, stable Nrf2 (half-life of 200 minutes) to pass into the nucleus and activate transcription via ARE binding.[176,198] Several different mechanisms have been proposed to explain how the Keap1 mediated degradation of Nrf2 is prevented. The most prominent mechanism in the literature is a "hinge and latch" mechanism, with dissociation of Nrf2. However, other mechanisms have also been proposed including the dissociation of Keap1 and Cul3, and a non-dissociation mechanism of Nrf2, where the complex is blocked and unable to release Nrf2, leaving newly synthesised Nrf2 to translocate to the nucleus, recruit the small maf protein and activate ARE transcription.[162,197]

In the "hinge and latch model" the hinge (ETGE) keeps Nrf2 bound to Keap1, and the latch (DLG) keeps the Neh2 domain in place for ubiquitination.[185] Under oxidative stress, oxidation of the cysteine residues in the IVR domain causes a conformational change in Keap1 which disrupts the DLG motif interaction.[176,202] This causes the latch to release from the complex with the ETGE motif hinge remaining attached due to the higher binding affinity, the Neh2 domain then cannot be ubiquitinated and Nrf2 escapes degradation.[185] This Nrf2 along with any newly synthesised Nrf2 can then translocate unhindered to the nucleus where it can bind to AREs and activate gene transcription via heterdimerisation with small Mafs.

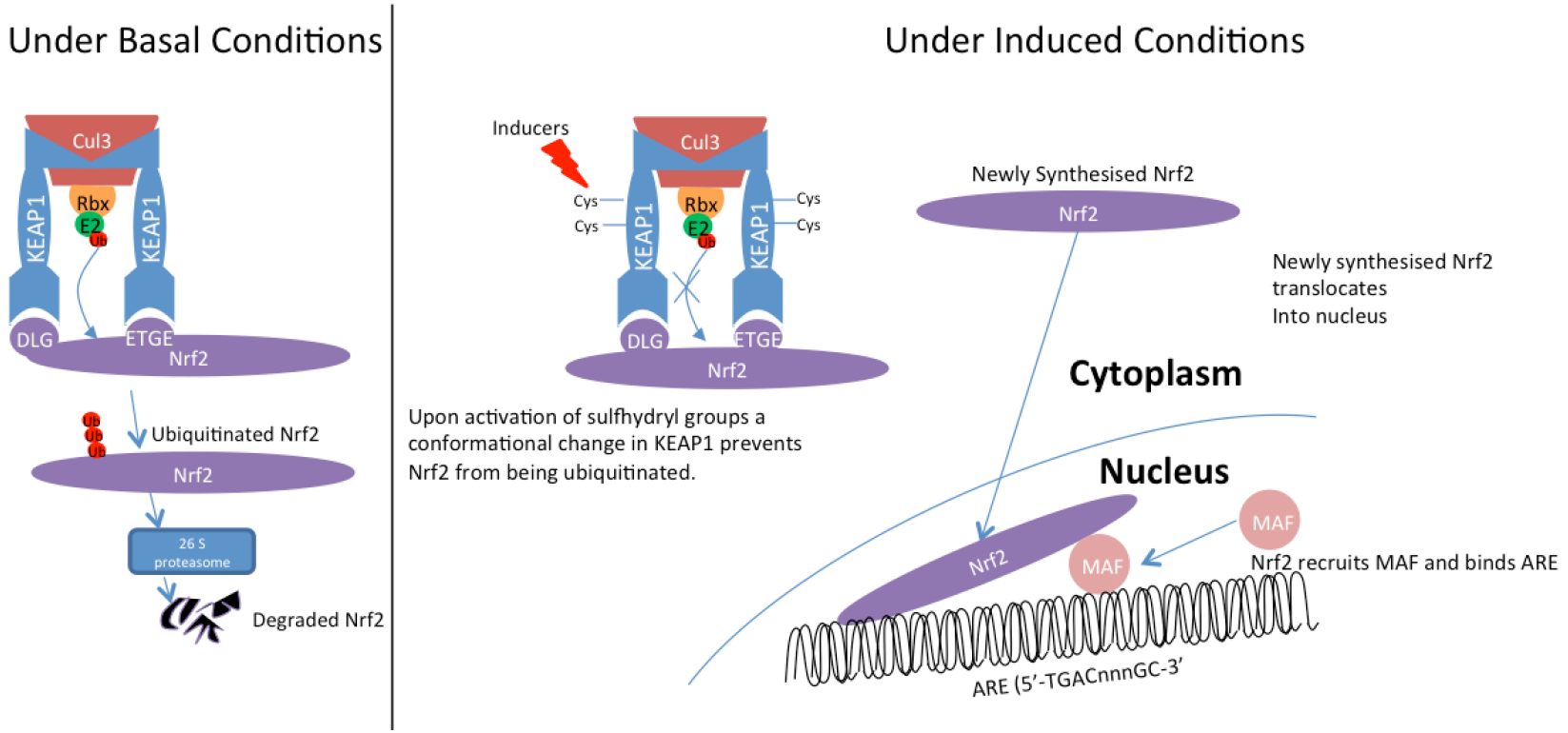


Figure 10: The Keap1-Nrf2-ARE pathway[157] Under basal conditions the Keap1 homodimer associates with Nrf2 through the Kelch domains of Keap1. The Keap1-Nrf2 complex associated with Cul3 brings Nrf2 in close proximity to the Cul3-based E3 ligase complex leading to polyubiquitination of the lysine rich region of Nrf2 between the DLG and ETGE motifs.[176,198] The polyubiquitinated Nrf2 is then degraded by the 26S proteasome.[199] However, under induced conditions such as oxidative stress the redox sensitive cysteine residues on Keap1 are activated and are thought to induce a conformational change in Keap1, thus preventing the ubiquitination of Nrf2. This allows the newly synthesised Nrf2 to pass into the nucleus where it recruits the small Maf protein and activate transcription via ARE binding.

In the dissociation mechanism inducing conditions covalently modify the cysteine residues in the Cul3-binding BTB region of Keap1; this gives rise to an unfavourable steric clash between Keap1 and Cul3 caused by a conformational change in Keap1. Research suggests that it is the conformational change in Keap1 rather than the dissociation from Cul3 that leads to the stabilisation of Nrf2 and prevention of ubiquitination.[203,204]

After normal redox conditions have been restored Nrf2 is transported back into the cytoplasm where it undergoes normal Keap1-mediated degradation.[205]

1.4.5 Disrupting the Keap1-Nrf2-ARE pathway

Disruption of the Keap1-Nrf2 interaction has gained interest recently in terms of treatment of diseases where an oxidative stress related mechanism is implicated.[157] There have been a number of modulators of the Nrf2-Keap1-ARE pathway reported previously in the literature, however most react via an indirect inhibition of the Keap1-Nrf2 interaction and are known as indirect Nrf2 inducers.[157] These compounds are believed to react via formation of covalent adducts (through oxidation or alkylation) with the sulfhydryl groups of the reactive cysteine in Keap1, and many contain electrophilic groups.[206] More recently, within the duration of this project a number of direct Keap1-Nrf2 compounds have been discovered, these molecules work through inhibition of the protein-protein interaction between Keap1 and Nrf2.[207–212]

1.4.5.1 Indirect Nrf2 inducers

Indirect Nrf2 inducers can be split into ten categories based on their chemical structure:

1. Michael Acceptors
2. Oxidizable Phenols and Quinones
3. Isothiocyanates and Sulfoxythiocarbamates
4. Dithiolethiones and Diallyl Sulfides
5. Trivalent Arsenicals
6. Vicinal Dimercaptans
7. Selenium-Based compounds
8. Polyenes
9. Hydroperoxides
10. Miscellaneous Inducers

However, by far the largest group is the electrophilic Michael acceptors. A Michael acceptor, otherwise known as an α,β-unsaturated carbonyl compound is a Lewis acid. Michael acceptors are able to react with the cysteine thiolates on Keap1. Thiolates are soft bases and react through a Michael addition reaction (see Figure 11).[213]

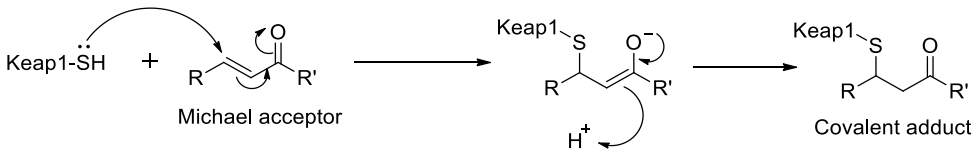


Figure 11: Sulfhydryl nucleophilic attack on an electrophilic Michael acceptor

Michael acceptors are commonly found in phytochemicals such as curcuminoids, coumarins, flavonoids, and chalcones. Curcumin (Figure 12) the chemical responsible for the yellow colour in curry has two Michael acceptors that can react readily with the sulfhydryl groups.[214]

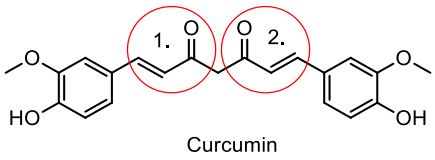


Figure 12: The chemical structure of curcumin with the two Michael acceptors highlighted (1) and (2).

Curcumin, in particular has been found to effectively activate the Keap1-Nrf2-ARE signalling pathway.[215,216]

There have been a number of other potent electrophilic inducers published in the literature that have been known to activate the Nrf2-Keap1 pathway. Notably, dimethyl fumarate (DMF) has recently been approved as a therapeutic in multiple sclerosis under the Biogen Idec name Tecfidera® (Figure 13) **(1).** Bardoxolone methyl **(3)** is another very potent Nrf2 inducer, with a therapeutic index much higher than other electrophiles such as DMF and Sulforaphane **(3)**.[217] Bardoxolone methyl is currently in clinical trials for pulmonary arterial hypertension in the US, as well as for diabetic neuropathy in Japan. In neurodegeneration the key to finding a potent drug is first penetrating the CNS. S(+)-Apomorphine **(4)** was found to be a CNS penetrating activator of the Nrf2 pathway, and can be classed under the oxidizable phenol class of activators which work through oxidation of the 1,2-diphenol group to a Michael acceptor containing 1,2 benzoquinone.[157,218]

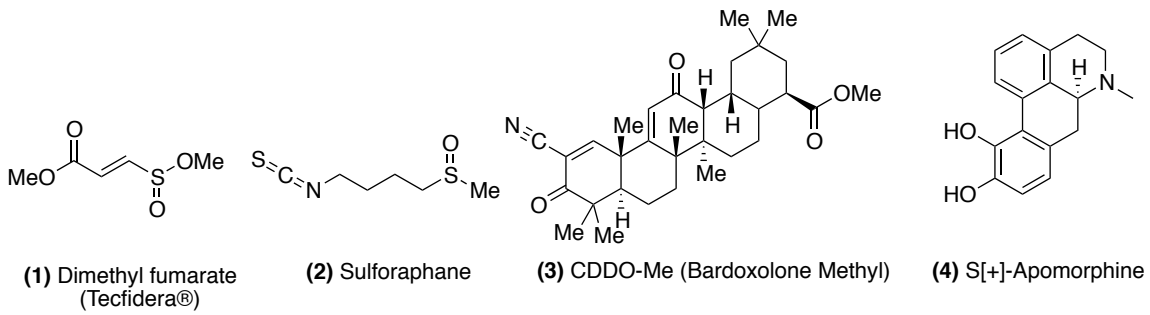


Figure 13: Further known electrophilic inducers of the Keap1-Nrf2 pathway.

However, whilst some electrophilic inducers are being progressed successfully to the clinic there are many documented downsides. In particular, for these molecules a therapeutic window can be difficult to achieve. This is due to many reactive Nrf2-ARE electrophilic inducers having a bell shaped dose response curve where they show a beneficial cellular response at low concentrations before their "off-target" effects are seen and unwanted cytotoxicity is seen at higher concentrations.[217,219] For example, Tecfidera® has been shown to have side effects including severe flushing, abdominal pain, diarrhea, nausea, vomiting, rash, itching, redness, and indigestion.[220] Bardoxolone methyl for example has been shown to react with over 500 molecular targets and many of the electrophilic reactors of Keap1 are also known to share structures with members of the pan-assay interference compounds (PAINS).[221–223]

1.4.5.2 Direct Keap1-Nrf2 inhibitors

As discussed previously (1.4.5.1) most inducers of the Keap1-Nrf2-ARE pathway work through electrophilic attack on the Keap1 sulfhydryl groups. However the off-target effect of these chemically reactive molecules has led to research into direct inhibitors of the Keap1-Nrf2 interaction. This has narrowed into two approaches, the inhibition of the Keap1-Cul3 interaction, and inhibition of the Keap1-Nrf2 direct interaction.[224] However, there remains some concern over disrupting the Keap1-Cul3 interaction because there are several BTB-Kelch substrate adaptor proteins that associate through similarly conserved BTB domains.[225]

As for the interaction between Keap1 and Nrf2, this can be modelled through recently published high resolution crystal structures of the Kelch domain of Keap1 with and without Nrf2 derived peptides.[226–229] It has been shown that the Kelch domain forms a six bladed β-propeller structure that is highly symmetric (Figure 14). Each blade consists of four β-sheets and highly conserved kelch repeats. All six blades are known to contribute to the binding of Nrf2.[226,229]

1.4.5.2.1 Peptide Inhibitors of Keap1

Several Nrf2 based peptides have been reported in the literature to perturb the Nrf2:Keap1 interaction. The 16mer peptide AFFAQLQLDEETGEFL (amino acids 69-84 of Nrf2) and the 14mer peptide LQLDEETGEFLPIQ (amino acids 74-87 of Nrf2), both of which contain the Nrf2 ETGE motif were shown to displace Nrf2 from the Keap1-Nrf2 complex with a comparable K_D of 20nM by isothermal calorimetry (ITC).[229] This compared well with the published values of the Neh2 domain binding to full-length Keap1 protein of 5-9 nM.[229] However, a shorter 10mer peptide LDEETGEEFLP (amino acids 76-85) whist still effective at displacement of Nrf2 was not quite as effective due to the lack of two key backbone interactions that can stabilise the longer peptides.[229] The minimum peptide that is required for Keap1 binding before significant loss in affinity is seen is 7 amino acids in both the ETGE and DLG.

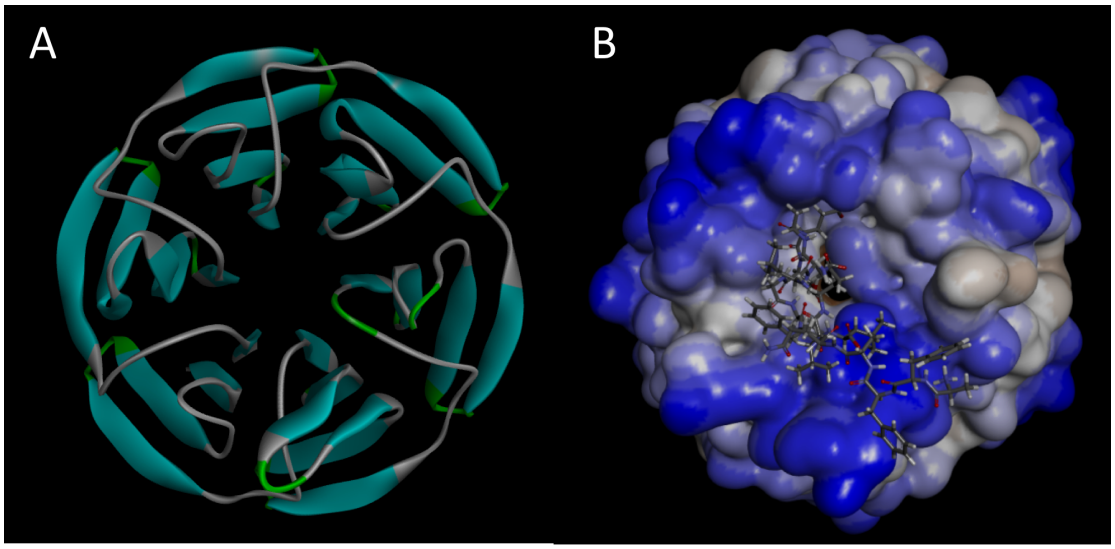


Figure 14: Two representations of the crystal structure of the Kelch domain of Keap1 (PDB: 2FLU)[229]. A) Tertary structure of Keap1 showing β-propeller structure. B) 16mer Nrf2 ETGE peptide cocrystallised in binding pocket with a hydrophobicity plot shown on the protein

1.4.5.2.2 Small molecule Keap1 inhibitors

Recently, several high throughput screening (HTS) campaigns have yielded a number of small molecule direct inhibitors of the Keap1-Nrf2 interaction. Hu et al. discovered a tetrahydroisoquinoline (THIQ) in a high throughput screen of the MLPCN library.[211] They then went on to explore the structure activity relationship (SAR) around the initial compound, first optimising the stereochemistry to elucidate the most active stereoisomer (SRS-5) from their racemic screening hit LH601 (Figure 15).[211] A co-crystal structure of the initial screening hit with the kelch domain of Keap1 then lead to

further exploration of the SAR around the compound leading to compound 59.[212] Jnoff et al. explored a variety of different substitutions around the molecule finding that the phthalimide group could be replaced with isoxindole whilst maintaining similar potencies. The group also found that the cyclohexane ring could be replaced with cyclopentane. Additionally, the carboxylic acid was found to be important for potency and any deviation away from the group resulted in a reduction in potency, with the exception of the carboxylic acid bioisostere tetrazole, which only had a modest decrease in potency. Finally, exploration into substitutions off the phenyl ring lead to an increase in potency with a methyl group at the 5-position. However, whilst substitution off the 5-position lead to some hope that these molecules could be further modified to increase potency the rest of the SAR suggests that space to gain further potency may be limited.[212]

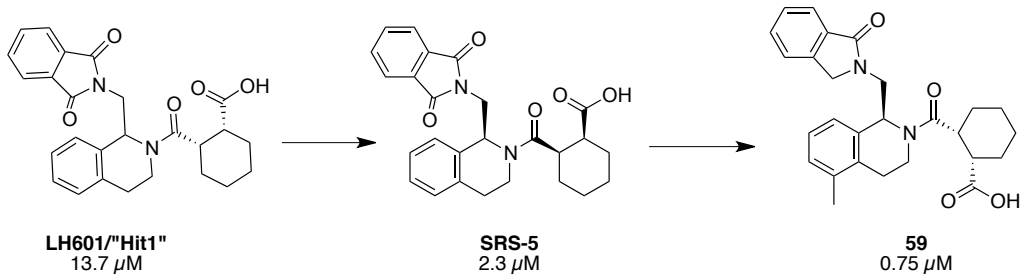


Figure 15: Reported optimisation of a tetrahydroisoquinoline

A separate screen using the Evotec Lead Discovery library, undertaken by Marcotte et al.[210] yielded two hits: a napthaline and a thiopyrimidine (Figure 16). Crystallographic studies of both compounds 15 and 16 suggested that both bind the Keap1 kelch domain. Interestingly in crystalisation studies two molecules of compound 15 appeared to bind to the kelch domain (Figure 17). However, no further development has been seen on this compound, probably due to the low potency of the compound, possible driven by its inability to bind to the hinge binding region (1) in the centre of the active site. In addition, the lack of efficacy in a luciferase ARE-reporter assay, may have been a contributing factor to the lack of development. Compound 16 however was shown to bind the kelch domain with the napthaline core protruding into the centre-binding pocket 1, stabilising the binding conformation (Figure 17). Compound 16 was also able to show some efficacy in the reporter assay, but with a much lower efficacy when compared to the reactive inducer DMF.[210]

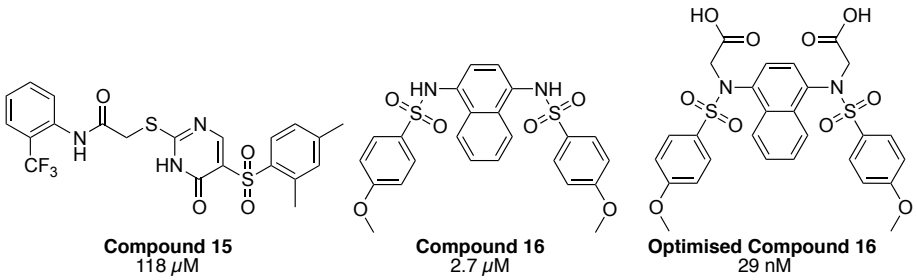


Figure 16: Hit and optimised compounds yielded from the HTS screen of Marcotte et al.[210]

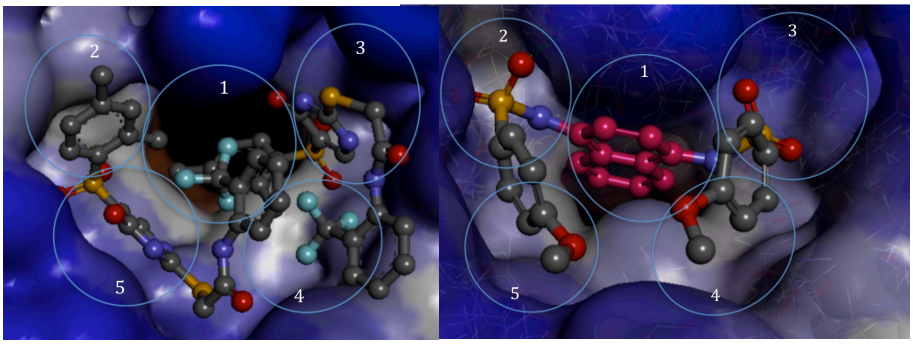
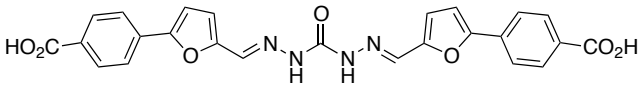


Figure 17: Crystal structures of Keap1 Kelch domain with compounds 16 and 15.[210]

Following the crystal structure of compound 16 elucidating the binding mode of compound 16 several studies were performed on the molecule. In one study molecular modelling was used to optimise compound 16 to utilise the relatively unused hydrophilic space provided by binding pockets 2 and 3.[209] The group added two acetate functionalities to give optimised compound 16 (Figure 16) with an improved EC_{50} of 29 nM. However, there is the possibility that the increased polarity from the acetate groups may inhibit membrane permeability, although the group were able to show western blot induction of Nrf2 target genes after 24 h.[209] However, the development of any analogues of compound 16 is further limited due to the carcinogenic and mutagenic nature of the diaminonapthaline cores, which has been well documented.[230] Diaminonapthaline cores have also been shown to have undesirable induction of CYP1A, and can bind to the cytosolic aryl-hydrocarbon (Ah) receptor.[230]

Structure-based virtual screening of compound 16 has also been utilised to find several compounds similar to compound 16.[207] The compounds identified were tested and found to have a range of activities from 3.6 μM to 75.2 μM. However, an issue with this study was recently identified as the majority of the structural classes identified by the study were classed as PAINS, and therefore they may have activity for additional targets other than Nrf2.[222,223]



Specs no. AN-465/14458038
9.8 μM

Figure 18: Structure of a carbazone found to disrupt the Nrf2-Keap1 interaction.[208]

The same group also identified a carbazone (Figure 18) using structure-based screening of the Specs database. The group selected compounds with similar properties to the crystal structures of the ETGE and DLG Nrf2 peptides when bound to Keap1, such as molecules needing to have a net charge of ≤-1.[208] They then screened the compounds against a pharmacophore of the ETGE peptide made up of two hydrogen bond acceptors (HBA) one hydrogen bond donor (HBD), and three negative ionisable centres leading to the carbazome structure with an IC_{50} of 9.8 μM. However, the presence of the two carboxylic acids and carbazone groups limit the molecules cellular permeability, and hence has led to its documented low cellular activity.[208]

Despite several published screens leading to tool compounds, many of these compounds have problems with either low cellular activity due to low permeability, or toxicologically active cores meaning that they are unsuitable for development into brain penetrant treatments. Hence a gap still exists for a brain penetrant inhibitor of the interaction to be discovered.

1.4.5.1 Nrf2 Inhibitors

Few Nrf2 molecules are known to inhibit Nrf2 compared to the large number of inducers. However, several have been identified including ascorbic acid (vitamin C), all-trans retinoic acid (ATRA), and Luteolin.

Vitamin C is a known antioxidant and has been found to reduce the levels of the Nrf2/DNA complex at the ARE of the germ cell-less (GCL) gene promoter.[231] ATRA was found to reduce the ability of Nrf2 to mediate the induction of ARE-driven genes both *in vivo* and *in vitro*. Nrf2 was found to form a complex with the retinoic acid receptor alpha (RARα), this leads to reduced binding or Nrf2 to ARE due to the Nrf2-RARα complexes being unable to bind ARE.[232] Luteolin is a flavonoid that has been shown to inhibit ARE-driven gene expression, and inhibit Nrf2. In A549 cells (a non-small-cell lung cancer cell that contain constitutively active Nrf2) luteolin was found to decrease Nrf2 levels at mRNA and protein level. It also showed a reduction in binding of Nrf2 to ARE, down-regulation of ARE-driven genes and depletion of reduced glutathione.[233]

1.5 Project Aims and Objectives

As previously discussed, there is an unmet need for therapeutics in ALS, with current therapies mainly focussing on palliative care, with only one approved drug Riluzole extending the lifespan of some patients by only a few months. This unmet need leaves the door open for new therapies in ALS.

The Nrf2-Keap1 pathway has been shown to be an attractive therapeutic target in reducing oxidative stress in neurodegenerative diseases. In particular, inhibiting the function of Keap1 and its interaction with Nrf2 has shown particular promise with reactive Keap1 modulator Tecfidera® becoming approved for the treatment of multiple sclerosis. Therefore, this objective of this project will be to find a non-reactive activator of Nrf2 through direct inhibition of the Nrf2-Keap1 protein-protein interaction. In particular, the project will have several aims to achieve this goal:

1.5.1 Design of a high throughput screening assay to probe the interaction of Nrf2 and Keap1

This project will aim to design an assay that is robust, relatively cost efficient, reproducible, and capable of screening large numbers of molecules. This will enable large screening libraries to be tested for inhibition of the interaction.

1.5.2 Utilising computational methods to screen and design molecular inhibitors of the interaction between Keap1 and Nrf2

In order to reduce costs and improve efficiency associated with medicinal chemical development of molecules *in-silico* analysis of data is becoming more mainstream within the drug discovery industry. This project will aim to utilise *in-silico* methodologies to enhance and aid the discovery of small molecule inhibitors.

In addition, the study will aim to investigate and implement *in-silico* algorithms to predict the physiological properties of molecules developed and tested. Particular focus will be given to increasing the likelihood of molecules to crossing the blood-brain barrier. As if a molecule is to be successful drug in treatment for ALS it will need to be able to be delivered to the site of action for the disease.

1.5.3 Screening *in-silico* selected compounds and selected molecular libraries in high throughput assays

Using the established high throughput screening assay the project will aim to screen, both compounds selected from computational screening and large high throughput screening libraries. Any compounds shown to be active in the primary high throughput screening assay will then be able to be investigated further in secondary screening

assays, which should aim to include a cellular model to analyse ARE activation. Successful compounds can then be investigated using *in-silico* design to select compounds for synthesis.

2. Materials and Methods

2.1 Materials

2.1.1 List of commonly used chemicals

Any other reagents are specified in the text

- HEPES (Melford, UK)
- Sodium Chloride (NaCl) (Sigma, UK)
- NP-40 (Calbiochem, UK)
- Sodium deoxycholate (Sigma, UK)
- Sodium dodecyl sulphate (SDS) (Sigma, UK)
- Tris-HCl (Sigma, UK)
- Tris-Base (Sigma, UK)
- Glycine (Sigma, UK)
- cOmlete mini protease inhibitor cocktail tablet (Roche Diagnostics)
- Tween 20 (Sigma, UK)
- Sodium hydrogen phosphate (Na_2HPO_4) (Sigma, UK)
- Potassium hydrogen phosphate (KH_2PO_4) (Sigma, UK)
- Potassium Chloride (KCl) (Sigma, UK)
- Coomassie Brilliant Blue R-250 (Sigma, UK)
- Analytical grade methanol (Fisher, UK)
- Glacial acetic acid (Fisher, UK)
- Glycerol (Sigma, UK)
- Bromophenyl Blue (Sigma, UK)
- 2-mercapto ethanol (Sigma, UK)
- Dithiothreitol (DTT) (Sigma, UK)
- DMSO (Sigma, UK)
- Phenylmethyanesulfonyl fluoride (PMSF) (Sigma, UK)
- TCEP (Sigma, UK)

rKeap1 protein was either purchased from Origene (Rockville, MD, USA) or Novus biologicals (Cambridge, UK), expressed in-house (3.4), or kindly donated from the Structural Genomics Consortium (SGC) (Oxford, UK).

All peptides were purchased from Eurogentec (Southampton, UK).

Compound libraries were kindly provided by the MRCT (London, UK), and the European Lead Factory (Oss, Netherlands). Other screening compounds were purchased either from Chembridge (San Diego, CA, USA), or ChemDiv (San Diego, CA, USA).

2.1.2 Solutions

- **Buffer B**

  150 mM NaCl

  10 mM HEPES

  0.1 % (v/v) NP-40

  pH 7.6

- **Buffer C**

  150 mM NaCl

  10 mM HEPES

  pH 7.6

- **RadioImmunoPrecipitation Assay (RIPA) Buffer**

  150 mM NaCl

  1.0 % (v/v) NP-40

  0.5 % Sodium deoxycholate

  0.1 % SDS

  50 mM Tris-HCl

  cOmlete mini protease inhibitor cocktail tablet (Roche Diagnostics), 1 tablet/10 ml buffer

  2 mM PMSF

  pH 8.0

- **1M RIPA Buffer**

  1 M NaCl

  1.0 % (v/v) NP-40

  0.5 % Sodium deoxycholate

  0.1 % SDS

  50 mM Tris-HCl

  cOmlete mini protease inhibitor cocktail tablet (Roche Diagnostics), 1 tablet/10 ml buffer

  2 mM PMSF

  pH 8.0

- **Tris-buffered saline-Tween (TBST)**

  50 mM Tris

  150 mM NaCl

  0.1 % (v/v) Tween 20

  pH 7.6

- **Tris-buffered saline (TBS)**

  50 mM Tris

  150 mM NaCl

  pH 7.6

- **PBS**

  3.2 mM Na_2HPO_4

  0.5 mM KH_2PO_4

  1.3 mM KCl

  135 mM NaCl

  pH 7.4

- **Coomassie Stain**

  0.1 % Coomassie Brilliant Blue R-250

  50 % methanol

  10 % glacial acetic acid

- **SDS Running Buffer**

  190 mM Glycine

  25 mM Tris Base

  0.1 % SDS

  pH 8.3

- **Novex running buffer**

  50 mL NuPAGE® Tris-Acetate SDS Running Buffer (20X) (Invitrogen, UK)

  950 mL dH_2O

  0.5 mL NuPAGE® Antioxidant (Added to upper buffer chamber only) (Invitrogen, UK)

- **Transfer Buffer**

  190 mM Glycine
  25 mM Tris base
  0.1 % SDS
  20 % (v/v) Methanol
  pH 8.3

- **Blocking Buffer**

  10 % (w/v) powdered milk in TBST

- **Western blotting incubation solution**

  The required calculated concentration of the antibody diluted in 1 % (w/v) powdered milk in TBST

- **Laemelli sample buffer (4x)**

  277.8 mM Tris-HCl, pH 6.8
  44.4 % (v/v) Glycerol
  4.4 % SDS
  0.02 % bromophenyl blue
  200 mM 2-mercaptoethanol

- **SEC Buffer**

  20 mM Tris,
  150 mM NaCl,
  1 mM DTT,
  pH 7.5

- **SEC Buffer (non reducing)**

  20 mM Tris,
  150 mM NaCl,
  pH 7.5

2.1.3 Antibodies

Table 4: Antibodies used in Western Blotting and their concentrations

Antibody	Species	Dilution and conditions	Supplier and catalogue number
Keap1	Rabbit polyclonal	1/1000 1 hour RT	Sigma, UK HPA005558
FLAG	Mouse monoclonal	1/1000 1 hour RT	University of Sheffield, UK
Alpha Tubulin	Mouse monoclonal	1/ 2500 1 hour RT	Fisher, UK 62204
Goat-Anti rabbit HRP secondary	Goat polyclonal	1/5000 1 hour RT	Dako, UK P 0448
Goat-Anti mouse HRP secondary	Rabbit polyclonal	1/5000	Dako, UK

2.2 Protein Synthesis and Analysis

2.2.1 Production of Keap1 cDNA plasmid

Lyophilised human Keap1 cDNA cloned into the pCMV6-Entry vector with a Myc and Flag tag at the C-terminus (Origene, RC202189) was re-suspended in 100 µL nuclease free water. The plasmid (33.7 ng) was then added gently to 10 µL of competent *E-coli* DH5-α cells (Invitrogen) and incubated for 45 min at 0 °C. The solution was then heat shocked for 45 s at 42 °C, added to 500 µL of SOC outgrowth medium (NEB, Hitchin, UK) and shaken at 220 rpm at 37 °C for 30 min. The media was then spread evenly onto a fresh kanamycin treated 10 cm LB agar plate and left to incubate O/N at 37 °C. Single colonies were then selected and suspended in 10 mL of kanamycin treated lysogeny broth (LB) per colony and incubated O/N at 37 °C with shaking at 220 rpm. A glycerol stock of each culture was prepared using 500 µL of culture and 500 µL of 30% glycerol for each culture and frozen at -80 °C. Bacterial cultures were then pelleted using centrifugation at 4500 rpm. DNA was extracted and purified using a Nucleospin Plasmid Quickspin miniprep kit (Machery-Nagel) according to manufacturers instructions to yield approximately 35 µg of plasmid per culture. Plasmid concentrations were determined using a Nanodrop 2000 spectrophotometer (ThermoFisher, UK). DNA sequencing at the University of Sheffield Medical School Core Sequencing Facility using the Sanger method on an Applied Biosystems' 3730 DNA Analyser confirmed the sequence of the plasmid (See Appendix 1). With the following primers used in sequencing reactions: CMV-Forward:5'-CGCAAATGGGCGGTAGGCGTG-3', BGH-Reverse:5'-TAGAAGGCACAGTCGAGG-3'.

Additional plasmid preparations were performed using glycerol stocks grown to culture volumes of up to 500 mL and purified using Plasmid Maxi Kits (QIAGEN, UK) according to the manufacturers instructions.

2.2.1 Cell Culture

Human embryonic kidney cells (HEK293T) were cultured in Dulbecco's modified eagle medium (DMEM) (Sigma, UK) containing 10 % Fetal bovine serum (FBS) (Sigma, UK), and 50 U/ml penicillin/streptomycin (Lonza, UK).

2.2.2 Optimisation of Cell Transfection

HEK293T cells were cultured in a 6 well plate to a confluence of 60 % and transfected using poly(ethyleneimine) (PEI) (Sigma, UK) at different ratios to Keap1 plasmid as indicated in the table below. DNA and Opti-MEM Reduced Serum Media (Life Technologies, Fisher, UK) were first mixed together gently, PEI was then added and the

mixture vortexed for 10 seconds. The mixture was then incubated for 10 minutes at room temperature before adding to the cells. The cells were incubated for either 48 hours or 72 hours post transfection at 37 ^{o}C, at an atmosphere of 5 % CO_2.

Table 5: Cell transfection optimisation conditions

Optimem	Keap1 pCMV6	PEI	Ratio
200 µL	2 µg	4 µg	1:2
200 µL	2 µg	8 µg	1:3
200 µL	2 µg	12 µg	1:4

2.2.3 Large scale Cell Transfection of Keap1

HEK293T cells were seeded at a density of 0.5 x 10^6 cells/plate and cultured for 72 hours in 150 cm plates to a density of 60 %. Cells were transfected using 120 µg PEI (Sigma, UK), 30 µg DNA, and 3 mL of Opti-MEM Reduced Serum Media per plate. DNA and Opti-MEM Reduced Serum Media (Life Technologies, Fisher, UK) were first mixed together gently, PEI was then added and the mixture vortexed for 10 seconds. The mixture was then incubated for 10 minutes at room temperature before adding to the cells. The cells were incubated for 72 hours post transfection at 37 ^{o}C, at an atmosphere of 5 % CO_2.

2.2.4 Cell Lysis

Cells were lysed at 4 ^{o}C in RIPA buffer (Sigma, UK), containing 2 mM PMSF (Sigma, UK) and Protease Inhibitor Cocktail tablets (Roche, UK) used according to manufacturers instructions, the cell lysate was made up to a salt concentration of 1 M NaCl (to reduce non specific binding during purification) and centrifuged at 15,000 x g for 1 hour at 4 ^{o}C, and the supernatant collected.

2.2.5 Isolation and Purification of Keap1

The cell lysate was purified at 4 ^{o}C using ANTI-FLAG® M2 affinity beads (Sigma, UK). Briefly, at 4 ^{o}C 500 µL of beads were added to a 10 mL column, rinsed with TBS (2 x 10 mL) followed by 3 x 1 mL 0.1 M glycine HCl, pH 3.5, washed with TBS (2 x 10 mL). The lysate was then flowed over the beads 3 times at a flow rate of 1 mL/min. Beads were then washed with 10 mL of RIPA buffer containing 1M NaCl, followed by 10 mL of RIPA buffer (with standard NaCl concentration), and 10 mL of Buffer B. Protein was then eluted using 300 µL fractions of 3X FLAG peptide (Sigma, UK) at 100 µg/mL in Buffer B.

The 3X FLAG peptide represents where the FLAG motif is repeated 3 times within the peptide sequence, the full sequence of the synthetic peptide is: H_2N-Met-Asp-Tyr-Lys-Asp-His-Asp-Gly-Asp-Tyr-Lys-Asp-His-Asp-Ile-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys-COOH. Excess peptide was then removed from the protein using Slide-A-Lyzer MINI Dialysis Devices (ThermoFisher) into fresh Buffer B as per manufacturers instructions.

2.2.6 SDS-polyacrylamide-gel electrophoresis

Protein samples (prepared in Laemmli buffer followed by boiling at 95°C for 5 min) were separated on a 1.5 mm, 4-15% mini SDS-PAGE gels (Mini-PROTEAN® TGX™, 12-well, 20 µl, Bio-Rad) using the Bio-Rad Tetracell system at 200 V for 30 min. During electrophoresis proteins are denatured in the presence of SDS and 2-mercaptoethanol, causing them to acquire uniform charge-to-mass ratios which are proportional to their molecular weights. Protein sizes were determined through comparing the migration of the protein bands to a molecular mass standard (ColorPlus prestained protein marker, 7-175 kDa, NEB). Location of protein bands was determined by either staining with colloidal coomassie, or by western blotting.

2.2.7 Western Blotting Analysis

Proteins were transferred from the SDS-PAGE gel onto a PVDF membrane for 120 min at 250 mA using standard wet-transfer apparatus (Bio-Rad). The membrane was rinsed with distilled H_2O and protein bands detected with a Ponceau-S stain. Ponceau stain was removed with 3x washes with TBST. The membrane was then incubated in blocking buffer for 1 h at RT to reduce unspecific binding followed by incubation with primary antibody, diluted in 10 % blocking buffer for 1 h at room temperature. Next, the membranes were washed (3 x 5 min incubation in TBST) and incubated with appropriate secondary antibodies diluted in 10 % blocking buffer for 1 h at RT. The membranes were washed as previously and the protein bands were detected using chemiluminescent substrate (Geneflow), and visualized using the G:BOX chemiluminecence camera system (Syngene), and GeneSnap software (Syngene). Size and densitometry analysis was performed using GeneTools software (Syngene).

2.2.8 BCA Assay for Protein Concentration Determination

Bicinchoninic acid (BCA) protein assays rely on the formation of a Cu^{2+} protein complex under alkaline conditions, followed by the reduction of the Cu^{2+} to Cu^{1+} by the protein. The amount of Cu^{1+} is then detected by BCA which converts from light-blue in the presence of Cu^{2+} to dark purple when it is able to form a chelation complex to Cu^{1+}. The amount of reduction and therefore change in colouration is proportional to the amount of protein present.

200 µL of BCA working reagent (Pierce, ThermoFisher, UK) was mixed with 10 µl of the sample, incubated for 30 min at 37 °C and absorbance was determined at 562 nm. A calibration curve was established each time a protein assay was performed with Albumin standard (BSA, Pierce, ThermoFisher, UK) of known concentrations of 2000, 1500, 1000, 750, 500, 250, 125, 25 and 0 µg/mL. The concentration of each sample was determined using the absorbance of each sample at 562 nm measured using the Pherastar FS platereader (BMG Labtech) in relation to the standard curve.

2.3 Fluorescence Polarisation Assay

2.3.1 Fluorescence Polarisation Assay optimisation method

Fluorescence polarisation assays were carried out on a Pherastar FS platereader (BMG Labtech, Ortenberg, Germany) using an optic module optimised for fluorescein which has a single polarised filter for excitation at 485±12 nm with two oppositely polarised filters for emission at 520±30 nm. The data were recorded using the Pherastar software (BMG Labtech, Ortenberg, Germany), and analysed using Prism 6 (GraphPad Software, San Diego, CA, USA, www.graphpad.com). Different parameters were optimised including, peptide, peptide concentration, protein concentration, buffer composition, and assay volume. All parameters were tested in triplicate in a black, non-binding, 384 well plate (Greiner Bio-One, Stonehouse, UK). Protein was dispensed into plates using a Starlab StarPet E electronic multichannel. Peptide and compounds were dispensed in DMSO using an ECHO550 acoustic dispenser. Plates were incubated at 25 ^{o}C for 10 minutes before reading.

2.3.2 Fluorescence Polarisation Assay - General Screening method

Fluorescence polarization assays were carried out on a Pherastar FS platereader (BMG Labtech, Germany) using Optic module "FP 485, 520, 520". The data were recorded in Pherastar software (BMG Labtech, Germany) and analysed using Pipeline Pilot software (Accelrys, USA). Compounds were either screened at 30 µM concentration, (single point), or at a seven point semilog dose response curve at concentrations from 1 mM to 1 µM in in a black, non-binding, 384 well plate (Greiner Bio-One, UK). Working assay buffer used was 10 mM HEPES, 150 mM NaCl, 0.1 % NP-40 alternative, pH 7.6 (Buffer B). Protein was dispensed in assay buffer at a concentration of 90 nM at a 5 µL volume into screening plates using a Matrix Platemate liquid handler (96 well head) with 30 µL Matrix D.A.R.T tips (ThermoFisher). Controls, peptide and samples were dispensed in DMSO using an ECHO550 acoustic dispenser (Labcyte, USA). Plates were incubated at 25 ^{o}C for 10 minutes before reading. Non-linear regression was used to fit a sigmoidal dose-response curve on a semi-Log plot to calculate the EC50 using GraphPad Prism (GraphPad Software, USA, www.graphpad.com).

2.4 ARE Assay

2.4.1 Cell Culture

The reporter cell lines were generated prior to the start of the project as previously described.[218] Chinese Hamster Ovary (CHO), or C6 (rat) astrocytic lines were routinely maintained in DMEM supplemented with 10% FBS and penicillin/streptomycin. The ARE-TK-eGFP and TK-eGFP reporter constructs were a kind gift from William E Fahl McArdle Laboratory for Cancer Research, University of Wisconsin. The TK-eGFP reporter construct consists of a 123bp thymidine kinase promoter inserted in the multiple cloning site of peGFP (Clontech, USA). The ARE-TK-eGFP also contains four repeats of a 41bp GST ARE motif (TAGCTTGGAAATGACATTGCTAATCGTGACAAAGCAACTTT) 3' to the TK promoter. These plasmids were transfected into CHO and C6 cell lines using Lipofectamine 2000 (Invitrogen) and following 10-14 days of selection in 0.5mg/ml G418 they were expanded and selected for basal eGFP expression using fluorescence activated cell sorting (BD, FACSAria) with two sequential cell sorts for each cell line. These mixed populations of stable transfectants with basal eGFP expression were used in subsequent assays and designated 4xARE-TK-GFP for the ARE containing line and TK-eGFP for the control cell line.

2.4.2 ARE reporter assay

Assay was performed as previously described.[218] Briefly 24 hours prior to the assay the TK-eGFP CHO ARE reporter cell line was seeded into a 384 well plate (Greiner Bio-one, µClear, Black) in rows A-H at a plating density of 15 x 10^4 cells/well in 50 µL culture media (as described above) supplemented with 0.5 mg/ml G418. The TK-eGFP CHO cell line was seeded at the same density in rows I-P. After 24 hours the media was removed manually, the wells gently washed (50 µL PBS/well) and replaced with serum free media. Compounds were then dosed in triplicate at a 7 point semilog concentration curve ranging in concentrations from 300 µM to 0.3 µM onto both the TK-eGFP CHO ARE reporter cell line, and the TK-eGFP CHO control cell line and incubated for 24 hours. eGFP fluorescence (ARE induction) was then measured at Ex485nm/Em530nm. Non-linear regression was used to fit a sigmoidal dose- response curve on a semi-Log plot to calculate the EC50 using GraphPad Prism (GraphPad Software, USA, www.graphpad.com).

The reporter assay was performed similarly in C6 astrocyte cell lines.

2.5 Protein Characterisation

2.5.1 Size Exclusion Chromatography

Size exclusion chromatography (SEC) was performed at 25 °C, using a 300 × 10 mm Superdex® SEC 200 10/300 GL column (GE Healthcare, USA), equilibrated with 20 mM Tris pH 7.5, 150 mM NaCl, and either 0 mM (non-reducing) or 1 mM DTT (reducing) depending on requirements. Several molecular weight standards were used for calibration, including carbonic anhydrase (29 kDa), albumin (66 kDa), alcohol dehydrogenase (150 kDa), apoferritin (443 kDa), and blue dextran (2000 kDa) (Sigma, UK).

Keap1 (100 µg) was diluted to 500 µL sample volume in the appropriate running buffer and left at 25 °C for 10 minutes before loading. The sample was then loaded on the column. Separation was monitored by UV spectrophotometry at 280 nm and the resulting fractions (500 µL volumes) were analysed by SDS–PAGE using western blotting analysis.

The retention volumes for each of the standards and samples were measured and used to calculate the partition coefficients, K_{av}:

$$K_{av}=(V_r-V_o)/(V_c-V_o)$$

Where V_r is the retention volume, V_o is the void volume (determined by blue dextran standard), and V_c is the geometric bead volume for the column. The K_{av} for each standard was plotted against the log of the molecular weight in order to generate a standard curve, which was then used to calculate the approximate molecular weight for each experimental sample.

2.5.2 Cross-Linking of Keap1

Cross-linking was performed as previously described.[234] Cross-linking of Keap1 (0.25 µM) was carried out at RT with 0.05, 0.1 or 0.5 % glutaraldehyde in Buffer B containing 2.5 mM TCEP at pH 8.0. Keap1 was first incubated in Buffer B before addition of glutaradehyde. The reaction was monitored over time, with 1M Tris-HCl (pH 8.0) being used to terminate the reaction at 0, 3, 10, 30, 60 and 120 minute intervals. Samples were then run on SDS-PAGE followed by western analysis using a modified protocol. The samples (made up in NuPage LDS sample buffer (Life Technologies) and heated to 70 ^{o}C for 10 minutes) were run on a 3-8 % Tris-Acetate Gel (Life Technologies) using the Novex system (Life Technologies). Protein sizes were determined through comparing the migration of the protein bands to a molecular mass standard (Hi-Mark pre-stained

protein standard, Life Technologies). Proteins were then transferred and imaged using standard protocols.

2.5.3 Isothermal Calorimetry

Isothermal Titration Calorimetry (ITC)—Kinetic studies of the binding interactions between Keap1 and different ligands were established using the Nano ITC (TA Instruments, USA). All samples were made up using Buffer C. Experiments were carried out at 25 °C. The syringe contained 1 mM test compound that was injected using a total volume of 62.5 µL in 25 x 2.5 µL injections. The cell contained a total volume of 190 µL of Keap1 at 60 µM concentration. NanoAnalyse software version 3.5 (TA Instruments, USA) was used to determine the dissociation constants (K_D). All measurements were repeated at least twice.

2.5.4 Protein Liquid Chromatography Mass Spectrometry

A 1 µL sample volume of 0.5 mg/mL Keap1 protein in buffer C (with addition of 1 mM TCEP) was analysed by LC-MS analysis using an Agilent 1260 Infinty LC instrument attached to a Agilent 6530 Q-ToF detector. The sample was injected onto a Agilent Extended C18 2.1 mm x 50 mm liquid chromatography column and separated using a gradient of 5 % solvent A to 95 % solvent B over 10 minutes at a flow rate of 0.4 ml/min (solvent A, 0.1 % formic acid (aqueous solution), solvent B, Acetonitrile/0.1 % formic acid).

Mass spectra were taken in electrospray ionization (ESI) +ve mode with source parameters as follows: Drying Gas temperature 350 ^{o}C, 11 L/min, Nebuliser 45 psig, Capillary voltage 400 v.

2.5.5 Electron Microscopy

In Transmission Electron Microscopy (TEM) an electron gun emits a beam of electrons which is focussed onto a camera detector by magnetic fields. As the beam passes through a sample placed within the path of the beam the electron dense regions of the sample prevent electrons from passing, creating regions of absent signal on the camera. From this information a 2D image of the electron dense material can be created.

Copper grids (Agar Scientific, UK) were carbon coated and glow discharged. 5 µL of Keap1 sample either in Buffer B (non-reduced) or Reducing Buffer B at 0.02 µg/µL were applied to the grid and left to absorb for 60 seconds. Unattached protein sample was removed by blotting. Grids were then exposed to 0.75 % (w/w) uranyl formate (aq) (steryl filtered using a 0.2 µm membrane) for 30 seconds and blotted dry. Excess liquid was removed by vacuum. Stained electron micrographs (EM) were then imaged using a

Philips CM200 high resolution electron microscope. EM were captured at either 21000x or 52000x magnification.

2.6 Computational Studies

All computational studies were performed on a Dell XPS L502X (Intel Core i7-2720QM, 8 GB DDR-3 RAM @ 1333 MHz, 2 GB NVIDIA GeForce GT 540M Graphics)

2.6.1 Preparation of a virtual high-throughput (vHTS) library

Initial Filter:

The complete ZINC database, a free database of commercially-available compounds for virtual screening was downloaded in August 2012, and contained around 20 million compounds.[235] Pipeline Pilot 8.5 was then used to filter the database using a drug-likeness filter derived from combining the works of Murcko, Ghose, Mozziconacci, Oprea, Lipinksi and Palm (Figure 81).[236–240] This then lead to a drug-like subset of the ZINC library containing 3732645 compounds.

Second Filter:

The "In Stock" subset of the complete ZINC database was downloaded in October 2014 and contained around 12.7 million compounds.[235] Pipeline Pilot 8.5 was then used to filter the database using the criteria of a mPSA of less than 80, a molecular weight of between 350 and 800 and a CLogP of between 1.5 and 2.5. Discovery Studio 3.5 was then used to predict brain penetrance using the "ADMET profiling" tool. The results file was then filtered to remove molecules with a value of more than 1 to give only those with a predicted brain penetrance of "high" or greater. This gave around 10000 compounds.

2.6.2 Preparing Keap1 crystal structures for in-silico analysis.

The appropriate Keap1 DC domain crystal structure was downloaded from Protein Data Bank. Any co-crystallised ligands were exported to individual sd files.

The *prepare protein* protocol available in Discovery Studio (DS) 3.5 was then used with default settings to prepare the protein for further processing; this protocol standardizes atoms names, inserts and minimises missing loops, calculates pKa and protonates the protein.

2.6.3 Preparing pharmacophore models

The prepared protein (2.6.2) was loaded into DS 3.5 along with the appropriate exported ligand. The *Receptor-Ligand Pharmacophore Generation* protocol within DS 3.5 was used to find essential interactions in the binding of the ligand with the Keap1 protein DC domain. The protocol works through selecting pharmacophore models using a Genetic Function Approximation (GFA) technique. The pharmacophore model created

is based on interactions between the receptor and the ligand. Features such as hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), hydrophobic (HY), negative and positive ionisable (NI, PI), and ring aromatic (RA) interactions are created as interaction spheres (an area for which and interaction must be contained within). All default parameters were used. All features identified were selected for use in screening.

2.6.4 Screening database against pharmacophore

The *Screen Library…* protocol within DS 3.5 was used to screen *in-silico* libraries against any pharmacophore model created in 5.6.3. This protocol works by trying to match functional groups on the ligand that are capable of making interactions with the protein (defined by the pharmacophore). For example, a HBD from the ligand was matched with a HBA on the protein. The minimum number of features that must be matched between the protein and ligand was set at 4, and the maximum set at 6. Parallel processing was utilised with 6 processor cores, and a batch size of 500. All default parameters were used.

2.6.5 Preparing ligands for docking

The *Prepare Ligands…* protocol in DS 3.5 was used to prepare compounds for the correct format needed for docking. The protocol can generate ionisation states, tautomers, and isomers of compounds as well as fix valence errors and generate 3D coordinates from 2D structural data. 'Change ionisation' was set to *true*, with a pH range of between 6.5 and 8.0. The 'generate tautomers', and 'generate isomers' options were set as *false*. 'Fix bad valencies' was set to *true* and '3D coordinate generation' was also activated. All other properties were used as default.

2.6.6 Docking

2.6.6.1 Docking using GOLD

The appropriate crystal structure of DC domain Keap1 was loaded into GOLD in pdb format. Hydrogen atoms were added, and all water molecules and ligands were deleted from the structure. The docking area was selected using DS 3.5 using the *Define and Edit Binding Site* tools to create a binding sphere around the key amino acids covering the binding area between ligands and the receptor. The docking area was defined as an 11.24 Å radius sphere around the coordinates x=4.320, y=7.121, z=1.277. Table 6 outlines the settings used in the GOLD docking for the initial experiment, the second screen used the same settings as below with Arg415 selected as a flexible residue using all available rotamers for Arginine from the library described by Lovell et. al.[241] Other settings were left as default.

Table 6: GOLD settings used during docking of Keap1 with each ligand in the defined binding pocket

Global Option	Parameter	Setting
Ligand Flexibility	Flip amide bonds	Selected
	Detect internal H-bonds	Off
	Flip ring corners	Selected
	Match template conformations	Off
	Flip all planar R-NR 1R2	Selected
	Ring-NHR	Selected
	Ring –NR 1R2	Selected
	Flip protonated carboxylic acids	Selected
	Use torsion angle distributions	Selected
	Postprocess rotatable bonds	Selected
Fitness & Search Options	Rescore	Off
	Allow early termination	On (If top 3 solutions within 1.5 Å)
	Generate diverse solutions	Off
	Use internal ligand energy offset	Off
GA Settings	Automatic	Selected
	Search Efficiency	100 %

2.6.6.2 Docking using LibDock

The *Dock-Ligands (Libdock)…* tool in DS 3.5 was used to screen compounds that passed the pharmacophore screen (2.6.4). The prepared protein structure (2.6.2) was loaded into DS 3.5. The docking area was selected using DS 3.5 using the *Define and Edit Binding Site* tools to create a binding sphere around the key amino acids covering the binding area between ligands and the receptor. The docking area was defined as an 11.24 Å radius sphere around the coordinates x=4.320, y=7.121, z=1.277. Table 7 outlines the settings used in the LibDock docking. Other settings were left as default.

Table 7: LibDock settings used during docking of Keap1 with each ligand in the defined binding pocket

Global Option	Parameter	Setting
	Number of Hotspots	100
	Docking Tolerance	0.25
	Docking Preferences	High Quality
	Conformation Method	BEST
	Minimization Algorithm	Adopted Basis NR
Advanced	Verbose	0
	Sp2-sp2 rotation	True
	Grid Scoring	True
Parallel Processing		True
	Batch Size	125
	Server Processes	6
	Preserve Order	True

2.6.6.3 Docking using CDocker

The *Dock-Ligands (CDOCKER)...* tool in DS 3.5 was used to screen compounds selected from the pharmacophore screen. The prepared protein structure (5.2.2) was loaded into DS 3.5. The docking area was selected using DS 3.5 using the *Define and Edit Binding Site* tools to create a binding sphere around the key amino acids covering the binding area between ligands and the receptor. The docking area was defined as an 11.24 Å radius sphere around the coordinates x=4.320, y=7.121, z=1.277. Table 8 outlines the settings used in the CDocker docking. Other settings were left as default.

Table 8: CDocker settings used during docking of Keap1 with each ligand in the defined binding pocket

Global Option	Parameter	Setting
	Top Hits	25
	Random Conformations	50
	Stimulated Annealing	True
	• Heating Steps	2000
	• Heating Target Temp.	700
	• Cooling Steps	5000
	• Cooling Target Temp.	300
Advanced	Forcefield	CHARMm
	Final Minimization	Full Potential
Parallel Processing		True
	Batch Size	5
	Server Processes	4
	Preserve Order	True

2.6.7 Predictive ADME

The *ADMET descriptors...* tool in DS 3.5 was used in default settings to generate ADME prediction data. Compounds were selected using the software wizard and loaded in sd format. The protocol uses QSAR models to estimate a range of ADMET related

properties for small molecules. Including: aqueous solubility, blood-brain barrier (BBB) penetration, CYP2D6 binding, hepatotoxicity, intestinal absorption and plasma protein binding (PPB).

2.6.8 Predictive Toxicity

The *Toxicity prediction…* tool in DS 3.5 was used as default settings to generate predictive toxicity data using all available predictive properties. Compounds were loaded in using the software wizard in sd format. The Toxicity Prediction (TOPKAT) protocol uses Quantitative Structure Toxicity Relationships (QSTR) models to assess the toxic potential of molecules based solely on their molecular structure. Toxicity parameters that can be computed include; Ames Mutagenicity, NTP Rodent Carcinogenicity, FDA Rodent Carcinogenicity, Weight of Evidence Carcinogenicity, Carcinogenic Potency TD50, Developmental Toxicity Potential, Rat Oral LD_{50}, Rat Maximum Tolerated Dose, Rat Inhalational LC_{50}, Rat Chronic LOAEL (lowest observed adverse effect level), Skin Irritancy, Skin Sensitization, Ocular Irritancy, Aerobic Biodegradability, Fathead Minnow LC_{50}, Daphnia EC_{50}.

2.6.9 Data mining

All data mining and analysis of docking results was performed using Pipeline Pilot, with poses visualised in Discovery Studio with protein ligand clashes, intermolecular hydrogen bonding and pi interaction visualisation enabled. Interaction surface maps were also used to help visualise hydrophobic/hydrophilic interactions and hydrogen bonding interactions with the protein. Custom nodes were written to analyse and find molecules suitable for purchase and testing *in-vitro* using the supplier information supplied within the downloaded ZINC database to find compounds available for purchase.

2.6.10 Pipeline pilot analysis of HTS screening data

Data files from the PherastarFS platereader (BMG-Labtech) were read in using input nodes for excel spreadsheets, following input the data was pipelined through a variety of nodes to determine compounds that met hit criteria. Briefly, each plate was assigned a plate number and processed independently from each other. The controls from each plate were used to determine the Z', signal window, average fluorescence intensity of controls, maximal and minimal average fluorescence polarisation values and the active cut offs for each plate with statistical analysis of each parameter analysed. The data from the screening wells for each plate were then filtered to remove compounds that had fluorescent intensities significantly different from controls (> 3 SD above or below the control means). Compounds which passed this filter were then analysed for activity,

with well data being filtered to keep those compounds that decreased the maximal fluorescence polarisation signal by >3 or 4 SD of the control means. A list was created for the two different cut offs respectively. Results for all plates were then output to an excel file.

2.6.11 Homology modelling

PDB files for Keap1, SPOP, and KLHL11 were downloaded and imported into D.S. 3.5. The sequences for all loaded proteins were then aligned against a full length sequence of human Keap1. The "Create Homology Models…" tool was then used to generate 20 models of Keap1. All settings were maintained as default apart from the number of models to be generated which was changed to 20. The produced models were analysed using visual analysis, statistical based potential and effective force field energy calculations.

3. Development of an HTS *in-vitro* screening assay

3.1 Introduction

3.1.1 Introduction to HTS

In order to effectively screen libraries to find hit compounds a suitable high throughput screening (HTS) *in-vitro* assay was needed to test selected molecules for inhibition of the interaction between Keap11 and Nrf2.

HTS assays are commonly used within drug discovery projects. Many assay formats have been developed for use in HTS, these have different advantages and disadvantages and can include affinity, cellular and enzymatic assays. A homogenous competition based affinity assay was chosen as the primary screening assay in this project to assess the interaction between Keap1 and Nrf2, along with a downstream cellular assay which is used to assess activation of the antioxidant response element (ARE).

3.1.2 Fluorescent HTS techniques

Fluorescence techniques are widely used within drug discovery because of their relative ease of adaption to HTS formats where homologous assays are preferred. In fluorescent light emission an electron within a fluorophore first absorbs light of a specific wavelength. This provides the electron with energy causing it to be raised to an excited state. The energy is then released as a fluorescence photon as the electron spontaneously relaxes back to its ground state; this process is illustrated in Figure 19.

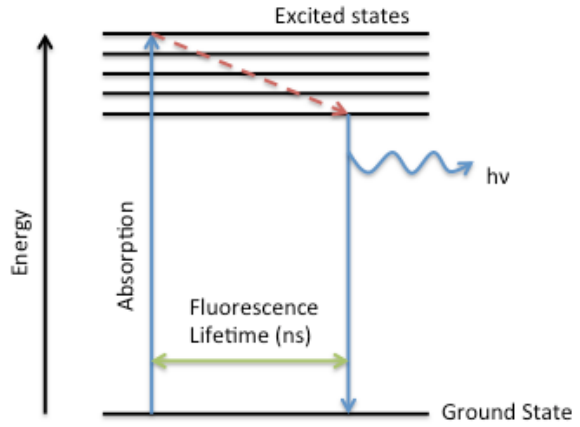


Figure 19: Jablonski diagram to illustrate fluorescence. The diagram illustrates the electronic states of the molecule and the transitions between them. As the molecule absorbs energy it moves to a higher energy excited state. Following this the molecule can undergo non-radiative transfer of energy amongst the high energy states (such as internal conversion and vibrational relaxation). Finally, the molecule returns to the ground state, where it releases energy as fluorescence.

There are currently a wide variety of fluorescent techniques available for HTS including fluorescent polarisation, intensity, lifetime and energy transfer etc., each measuring a different fluorescent property of the molecule.

Pope et al. published an outline of the techniques available for HTS in 1999.[242] Since then more applications and varieties of these techniques have emerged, but the basic fluorescent principles remain the same to this day. There are several advantages and disadvantages to the different techniques; these are highlighted in Table 9.

Table 9: Advantages and disadvantages of different fluorescent techniques for use in HTS (Reproduced from Pope et al.[242]).

Technique	Advantages	Disadvantages
Fluorescence Intensity (FLINT)	-Simple -Easily miniaturised -Suitable for fluorogenic assays	-Little information for quality control -Suffers from inner-filter effects such as reabsorption -Prone to autofluorescence interference.
Fluorescence Anisotropy/Polarisation (FA/FP)	-Simple -Immune to inner-filter effects -Ratiometric -Good for small (<15kDa) ligands -Only the ligand needs to be labelled	-Propeller effects -Prone to autofluorescence interference -Suitability limited by ligand size, difference in molecular weight, lifetime of dye -Dynamic range limited -Generally requires high concentration of receptor
Fluorescent Resonance Energy Transfer (FRET)	-Simple -Suitable for short inter or intramolecular distances (<5 nm) -Large range of donors and acceptors	-Labelling donor or acceptor can be troublesome -Suffers from inner-filter effects -Prone to autofluorescence interference -Limited to short distances for energy transfer -Most dyes only monitor donor quenching
Time Resolved FRET (TR-FRET)	-Predictable and robust -Suitable for long distances (5-10 nm) -Sensitive -Easily minaturisable	-Labelling donor or acceptor can be troublesome -Limited choice of donors and acceptors -Nonspecific energy transfer limits background

As demonstrated in the table, there are well known limitations to fluorescence detection methods that need to be accounted for when running high throughput screens. For example, autofluorescence, the natural fluorescence of a molecule, can interfere and mask the signal from the desired fluorophore in the assay. To avoid this, a counter screen for natural fluorescence can be implemented on the molecules before or after the assay. The molecules can then either be discounted, or the fluorescence background signal for that molecule deducted from the total fluorescence reading. Another common problem is quenching (also known as the inner-filter effect), this is where a component of the assay reduces the fluorescence from the fluorophore, usually due to the component absorbing light at the wavelength of emission from the fluorophore.

3.1.3 Current fluorescence based inhibition assays for the Keap1-Nrf2 interaction

There are two main fluorescence based assay types that have been explored for studying the interaction between Nrf2 and Keap1. The first of these is fluorescence resonance energy transfer assays.

3.1.3.1 Fluorescence resonance energy transfer (FRET)-based assays

FRET, an abbreviation for Förster resonance energy transfer, is named after the German scientist Theodor Förster who discovered the phenomenon in 1946.[243] FRET is the process of non-radiative transfer of excitation energy between a fluorescent donor to an acceptor (Figure 20). For FRET to occur, donor and acceptor molecules need to be between 10 to 100 angstroms from each other. In addition to the distance restriction the emission of the donor must also overlap with the absorption spectra of the acceptor.

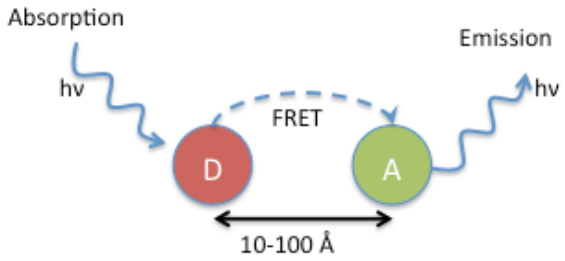


Figure 20: Showing fluorescence resonance energy transfer between a donor molecule (D) and an acceptor molecule (A).

In 2013 Schapp et al. described a steady-state FRET based assay for the identification of inhibitors of the Keap1-Nrf2 protein-protein interaction.[244] In the paper the authors used yellow fluorescent protein (YFP)-conjugated Keap1 kelch domain as the FRET acceptor and a cyan fluorescent protein (CFP)-conjugated Nrf2-derived 16-mer peptide containing the ETGE motif as the FRET donor in their assay.

Additionally, Baird et al. have described using a cell based FRET assay to deconvolute the nature of the interaction between Keap1 and Nrf2 using techniques based on multi-photon fluorescence lifetime imaging microscopy.[245] Using mCherry-Keap1 and EGF-Nrf2 fusion proteins, the group were able to identify a cycle of interaction that occurs between Keap1 and Nrf2 (Figure 21). A previous paper by Tong et al. suggested that under induced conditions the DLG motif becomes unbound from Keap1 in a hinge and latch model.[197,246] This suggests that the weak DLG interaction is interrupted first followed by the interruption of the stronger ETGE interaction. However, Baird et al. show that under induced conditions the DLG motif does not dissociate from Keap1 as a FRET signal is maintained, thus showing the motif may remain bound (Figure 21).

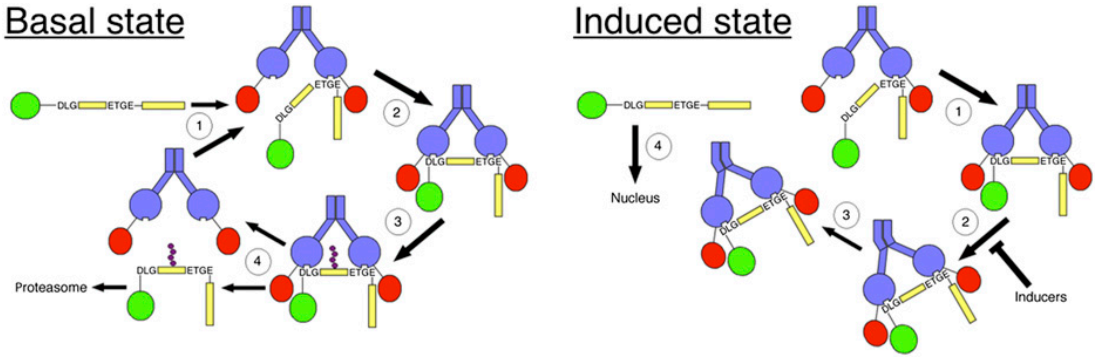


Figure 21: Keap1 FRET assay. Reproduced with permission from Baird et al, (2013). In the basal state, newly synthesised Nrf2 (yellow) binds sequentially to a free Keap1 dimer (blue), first through the ETGE (1) and then through the DLG motif (2). Once Nrf2 is bound it can then be ubiquitinated by the Keap1-dependent E3-ubiquitin ligase (3). Ubiquitinated Nrf2 is then released from Keap1 becomes degraded by the proteasome. In the induced state, Nrf2 is unable to be ubiquitinated so as a consequence; Nrf2 is not released from Keap1. This blocks the Keap1 ubiquitination complex allowing newly synthesised Nrf2 to accumulate, translocate into the nucleus and bind antioxidant response elements (ARE) of multiple genes.[245]

3.1.3.2 Fluorescence Polarisation/Anisotropy (FP/FA)-based assays

Fluorescence polarisation (FP) measures the change in orientation of the plane of polarised light in solution that is used to excite fluorescently-tagged or fluorescent molecules. During the time between excitation of a fluorophore with polarised light and emission of a photon via florescence (fluorescence lifetime) the plane of polarised light is depolarised. This is due to the rapid tumbling of small molecules in solution. If the fluorophore becomes bound to a vastly larger molecule, the speed of tumbling in solution is slowed, and hence the change in depolarisation of plane polarised light is less. This means that when the unbound fluorophore absorbs polarised light little of the fluorescence is emitted in the same polarised plane. By contrast, the receptor-bound fluorophore will emit a significantly higher proportion of its fluorescence in the same

plane of polarisation as the absorbed light. This enables the extent of depolarised fluorescence emission to be directly related to the level of binding (Figure 22).

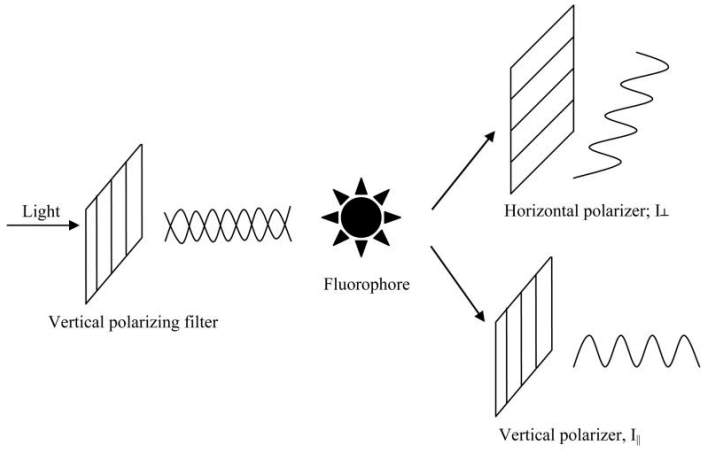


Figure 22: Basics of fluorescence polarisation. The fluorophore is excited with linearly polarised light. The intensity of polarised light is then measured through an emission polariser in both the vertical ($I_{\parallel}$) and horizontal planes ($I_{\perp}$). These values are then used to determine the FP signal (Equation 1).

From equation 1 it can be seen how the fluorescence polarisation value is calculated. The FP value is independent of the fluorophore concentration, because it is not dependent on the absolute intensity of light at either orientation, only the change in intensity between them.[247]

Equation 1: Where I=Intensity of emitted light, ⊥=perpendicular to excitation and ∥=parallel to excitation.

$$FP = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}$$

Fluorescence polarisation has both advantages and disadvantages. FP has a simple mix and read protocol, and does not use any radioisotopes, filtration or separation steps and only the ligand requires labelling. It is also fairly inexpensive; this makes the assay format ideal for HTS. Plates can also be re-measured, as FP detection does not destroy samples. However, several disadvantages or limitations need to be taken into account when designing experiments. The largest limitation is the lifetime of the dye, ligand size and molecular weight change. Short lifetime dyes such as rhodamine and fluorescein, (with a $T_F \sim 4$ ns) are usually unsuitable for monitoring protein-protein interactions, as they approach a limiting value at a protein mass of around 50 kDa. Further limitations include a lack of signal amplification between the free and bound species; this can mean that an assay will require a large fractional binding to be robust. Other factors include auto-fluorescence, and a property known as the "propeller effect". The "propeller effect" is where the motion due to the flexibility of the fluorophore, even in the bound state,

decreases the observed polarisation value. This effect can be decreased through the use of reactive dyes without aliphatic linkers between the fluorophore and the reactive group.[242]

In 2012 Hancock et al. and Inoyama et al. published the first fluorescence polarisation assays for identifying peptide inhibitors of the Nrf2-Keap1 interaction.[248,249] Different fluorescently labelled Nrf2 peptides that contained the Nrf2 ETGE motif were used to detect the direct peptide inhibitors of Keap1–Nrf2 interaction. The authors found that the FITC-labelled Nrf2 9mer peptide amide was able to show the largest assay window. This was thought to be mainly due to its reduced "propeller effect" in addition to maintaining a strong binding affinity similar to that of the longer 16mer Nrf2 peptide. The authors showed that the assay has a large tolerance toward DMSO and gave Z'-factors (a measure of statistical size, used to quantify how effective a particular assay would be in HTS) as high as 0.9 in HTS formats.

Whilst developing our assay several more papers appeared in the literature utilising fluorescence polarisation assays to identify small molecule inhibitors. A summary of the assays is shown in Table 10 below.

Table 10: Comparison of published FP assays.

Peptide	Keap1 Protein	Well Format	Volume	Z'	Reference
FITC-LDEETGEEFL-OH 10 nM	100 nM Bacterial expressed Kelch domain	384	40 µL	0.9	Inoyama et al. 2012[248] and Hu et al. 2013[211]
FITC-β-DEETGEF-OH 1 nM	200 nM Bacterial expressed Kelch domain	96	100 µL	0.66	Hancock et al. 2012[249]
FITC- βAla-DEETGEF-OH 10 nM	400 nM Bacterial expressed Kelch domain	96	80 µL	-	Zhuang et al. 2014[207]
FITC-LDEETGEEFL-OH 4 nM	12 nM Bacterial expressed Kelch domain	384	40 µL	-	Sun et al. 2014[208]

3.1.4 A pharmacologically relevant fluorescent polarisation assay

Upon first looking at the field in 2012 we believed that when developing an assay for screening of inhibitors it should ideally have several properties. It would need to be easily translatable to a HTS format, in addition it would need to be homogenous for ease of application and finally it would need to be as close to the real *in-vivo* interaction as possible. It is generally thought within the field that Keap1 forms a homodimer.[181] It was therefore thought that rather than using just the Nrf2 binding domain (kelch or DC domain) that the human full-length protein would provide a more physiological assay, since it should form the dimeric structure seen *in-vivo*. This would also facilitate study of the redox-sensing sulfhydryl groups on Keap1 (in the BTB and IVR domains) with a view to seeing how they affected binding.

3.1.5 Previous work on the project

Prior to starting on the project some initial work had been carried out by Dr Richard Mead to establish a choice of peptide and protein for the assay. A summary of this work is shown below.

3.1.5.1 Selection of peptide

Either the DLG or the ETGE Nrf2 motif could have been chosen to label for use in the assay. However, it was thought that the strongest FP signal would be seen with the higher affinity ETGE motif, which has roughly a 500 times increase in binding affinity when compared with the DLG motif ($K_D = 5 \times 10^{-9}$ vs. 1.0×10^{-6} M).[185] In addition, using the higher affinity peptide should require less target protein for the assay.[250] A 9-mer (LDEETGEFL) peptide was chosen as it was thought that a shorter peptide where the fluorescent dye is close to the peptide binding site will avoid the "propeller effect". This effect is caused by the flexibility of the fluorescent dye on the peptide, this depolarises the fluorescence signal of the bound complex.[251,252] The 9-mer ETGE peptide had also shown promising binding data to Keap1 in ITC experiments and also in x-ray crystal structures of the protein.[185,226] It was unknown as to whether labelling the peptide at the N or C terminus would have any effect on fluorescence polarisation signal, so both FITC-LDEETGEFL-OH (L1) and H2N-LDEETGEFL-FITC (L2) were trialled initially.

3.1.5.2 Selection of protein

Two possible suppliers of recombinant protein had been found, one from Origene and one from Novus biological. The Origene supplied protein contained a C-terminal MYC/DDK tag and was made in HEK293 cells, whilst the Novus supplied protein was expressed in an *in vitro* wheat germ expression system and was tagged with a GST tag. The protein molecular weights were therefore 69.5 and 94.75 kDa respectively.

3.1.5.3 Selection of Buffer

Various different buffer compositions have been identified for use in fluorescence polarisation assays. It has been seen that higher salt concentrations can have a significant effect on the binding affinity in FP assays.[249,252] Lynch et al. report that increasing the salt concentration had an adverse effect on binding. However, Lo et al. predict that there is a strong electrostatic component to the binding of Nrf2 to Keap1, in addition they reported that charge screening by electrolytes could play an important role in the ETGE motif binding to Keap1.[229] This suggests that having an increased salt concentration in the region of 100-150 mM would be beneficial to binding. Jadhav et al. reported that the inclusion of a detergent to the buffer in order to reduce artefacts caused by aggregation prone small molecules can have a dramatic effect on quality of the data in a FP assay.[253] The final composition of buffer shown in Table 11 was influenced by the conditions chosen by Wu et al. with higher salt concentration, and the inclusion of 0.1% NP-40 as a detergent.

Table 11: Buffer Composition

Component	Buffer B
NaCl	150 mM
HEPES	10 mM
NP-40	0.1 %
pH	7.6

3.1.5.4 Determination of peptide concentration

To begin with, the two different peptides were titrated to test what the detection limit was for the plate reader and what concentration would be sensible for use within the assay. Plate and volume conditions were selected from Lokesk et al. as they utilised a smaller assay volume of 20 µL in a 384 well plate format which is more amenable to HTS.[254] From Figure 23 it can be seen that the both peptides are stable down to about 0.5 nM. As expected, as the fluorophore reaches the detection limit of the plate reader, the FP signal becomes variable and the apparent polarisation signal increases. A concentration of 3 nM was initially taken forwards for testing with interaction with the protein to ensure a reliable signal when placed with the protein. No obvious difference could be seen between the N and C terminal labelled peptides at this point.

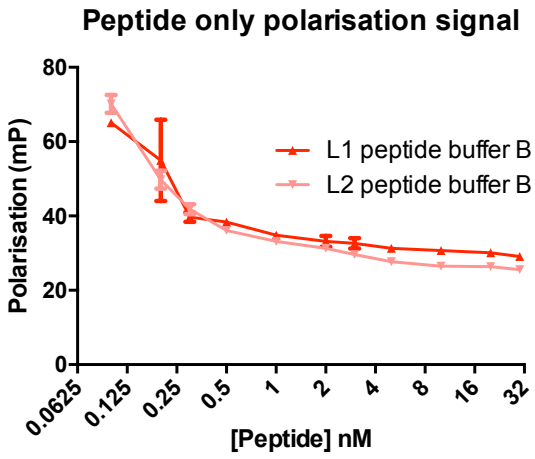


Figure 23: Optimisation of peptide concentration and buffer compositions. Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus peptide concentration in nM (N=3). L1 peptide represents FITC-LDEETGEFL-OH and L2 peptide represents H2N-LDEETGEFL-FITC (L2)

3.1.5.5 Determination of Protein assay concentration

Next work moved towards looking at what concentration of either of the proteins would give a sensible assay window for any competition-binding event to become apparent. A good assay will have a signal window (SW) >2; the signal window is defined by Equation 2.

Equation 2: Formula for calculating the signal window where AVG_{max} is the average maximal signal (FP signal gained from 0 % inhibition) and AVG_{min} is the average minimal signal (FP signal gained from 100 % inhibition).

$$SW = AVG_{max}/AVG_{min}$$

Therefore, to have a signal window greater than 2 an assay with a minimum signal of 35 mP needs to have a maximal signal of at least 70 mP.

To find the signal range of the assay, the two proteins were titrated against a concentration of 3 nM peptide (Figure 24).

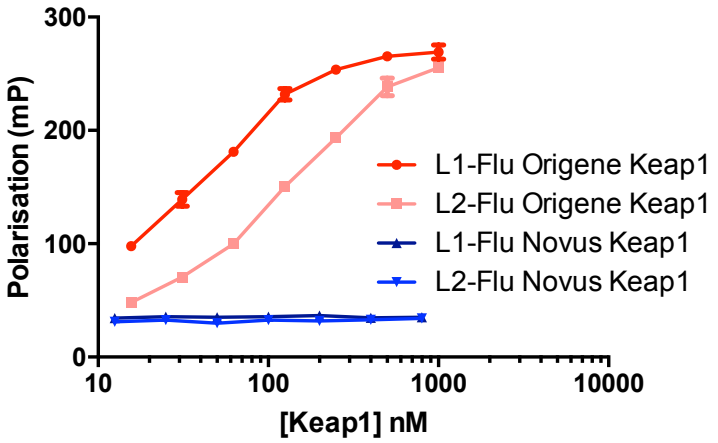


Figure 24: Titration of two commercially available Keap1 protein samples, a wheat germ expressed GST tagged protein (Novus Keap1) and a mammalian expressed sample (Origene Keap1). The protein was titrated against a 3 nM concentration of the C and N terminal labelled peptides (L1-Flu and L2-Flu) to determine optimal assay concentration of protein. Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus Keap1 concentration in nM (N=3).

From the assay it could be easily seen that the wheat germ expressed GST tagged protein from Novus did not function within the assay. There are several reasons the protein could be non-functional in the assay. Firstly, it is possible that the GST tag could be obscuring the peptide-binding site or simply causing the protein to misfold in a conformation unfavourable for binding. However, more likely is that the cell-free expression system did not contain the chaperones and the posttranslational modifications that would make the protein functional. The Origene MYC/DDK tagged protein gave a good signal within the assay for both the C and N-terminal tagged peptides. For a signal of ≥ 120 mP (giving a SW ≥ 3) a protein concentration of ~40 nM was sufficient for the N-terminal labelled peptide, whilst a 120 nM concentration of protein was required for the C-terminal labelled peptide.

3.2 Results and Discussion

3.2.1 Recombinant Expression of Keap1

In order to implement a high throughput screen, large quantities of protein needed to be produced. Purchase of the amount of protein needed for development and screening of the assay was economically unfeasible, therefore a method of producing recombinant Keap1 protein on a scale suitable for the screen was needed in-house. Whilst previous published work had shown expression of mammalian Keap1 kelch domain in bacteria, as well as the close homologue of full length mus musclus Keap1 (for use in single particle electron microscopy) no evidence could be found in the literature for synthesis or use of full length mammalian Keap1.[181,228]

Full-length Keap1 protein was the desired choice for the assay, mainly because of its increased propensity for physiological activity, because of the inclusion of the BTB domain which mediates dimerisation of the protein.[183,255,256] This dimerisation of the protein would also make the protein more suitable for a fluorescence polarisation assay in particular. As fluorescence polarisation assays rely on the binding of a fast tumbling small fluorescent probe to a large slow tumbling protein, the larger the protein, the greater the change in polarisation signal should be.[247] Additionally, the expression of the protein in a mammalian system was thought to be important due to additional posttranslational modifications that create the functional protein. The lack of these modifications can be seen in the section above (3.1.5.5) where the mammalian protein expressed in a wheat germ expression system is tested in the fluorescence polarisation assay and does not show any binding to the Nrf2 peptide (Figure 24).

3.2.1.1 Selection and expression of a Vector

A vector is DNA molecular vehicle that is used to introduce foreign DNA into a host cell. Once inside the cell the DNA can be replicated and/or expressed. There are four main types of vectors: plasmids, viral vectors, cosmids and artificial chromosomes. The most commonly used vector is a plasmid, and this was selected for introduction of the Keap1 DNA into the host system. Plasmids are double-stranded circular DNA sequences that can automatically replicate within a host cell. Many plasmids contain a variety of features that make manipulation of the plasmids possible. Most plasmids will contain a multiple cloning site, containing multiple restriction enzyme sites, which allow insertion of a transgene cut from another vector or PCR product with compatible enzymes.[257]

In order to create enough plasmid for transfection into the host cell system, bacteria are used to amplify the plasmid. Bacteria containing the plasmid can generate millions of copies of the vector in a few hours.

The commercially available expression vector available from Origene was selected for the project. The vector (Figure 25) contains a kanamycin bacterial selection construct. This ensures that only bacteria containing the kanamycin resistance gene and hence the desired construct will survive. It also contained sequences for inclusion of cMyc and FLAG-DDK tags on the C-terminal end of Keap1 to aid in identification and purification. The plasmid was successfully transformed, selected, and expressed in cultures of up to 2 litres (Chapter 2) using a competent DH5-Alpha e-coli cell line.

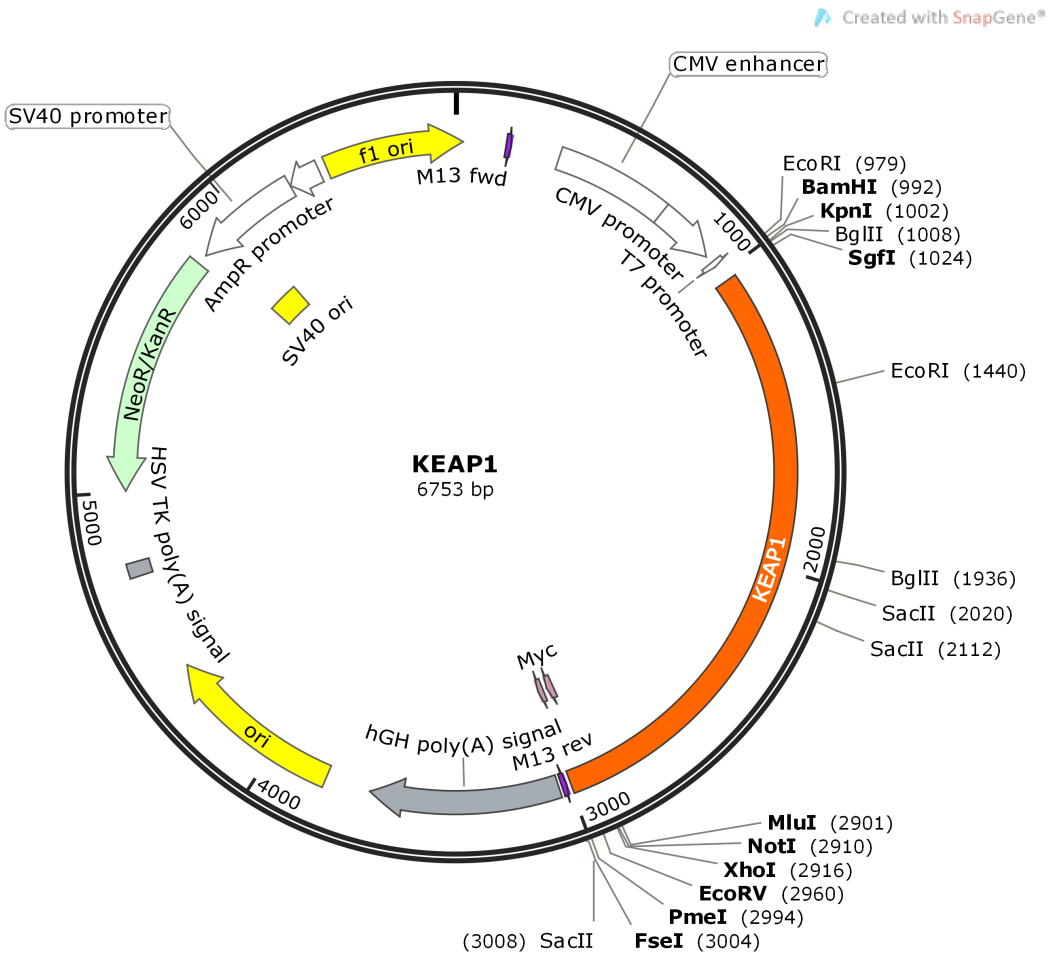


Figure 25: Plasmid map of the Keap1 cDNA showing locations of key genetic sequences and digestion sites within the plasmid.

After growth and purification of the plasmid, the next step was to validate that the isolated plasmid was the correct sequence, and no mutations or undesired replication had occurred during expression. The plasmid was sequenced, giving the desired sequence of the Keap1 transgene (Appendix 1).

3.2.1.2 Selection of expression host

Many different host systems are available including bacteria, yeast, plants, filamentous fungi, mammalian or insect cells grown in culture as well as transgenic animals and plants.[258–264] There are advantages and disadvantages to all hosts and these are important to consider when selecting a host. The selection of a host can also influence the method of purification. The main points to consider when selecting a suitable host are the amount of protein required, the degree of purity, maintained biological functionality, and any potential toxicity with the host.[265]

For the project an expression system capable of expressing the same mammalian functional protein that had previously been shown to work well in the assay was paramount. Therefore, HEK293T cells were selected as the expression system. HEK293 cells are a cell line derived from human embryonic kidney cells, and are widely used in cell biology due to their ease of growth and transfection. The T variant of the cell line is an important variant due to the inclusion of the SV40 large T-antigen. This antigen is able to bind to the SV40 promoter within transfected expression plasmids (such as that in Figure 25) to increase protein production.[266] Being a mammalian derived cell-line, they also contain the machinery needed for post-translational modifications.[265] This makes HEK293T cells the ideal for growth and expression of Keap1.

3.2.1.3 Optimisation of transfection

Transfection is a common technique for incorporating nucleic acids into cells. Many different methods for transfection exist including: chemical, non-chemical, particle based and viral (also known as transduction) methods. The most common method is chemical transfection; this involves the mixing of the DNA with a chemical solution that can be introduced to the cells that aids the uptake of the DNA. Chemical based transfection can be split depending on the type of chemical reagent used. These classes are phosphates, liposomes, and cationic polymers.[267]

The most well known method is that first discovered by F. L. Graham and A. J. van der Eb in 1973, the creators of the HEK293 cell line.[266] This method uses calcium phosphate to create a solution of precipitate with a charged surface able to bind the DNA, a proportion of which when added to cells is able to be taken up by the cells.[266]

Another efficient method uses liposomes. Liposomes are spherical vessels made up of lipids that are able to encapsulate molecules of DNA. These liposomes are able to fuse with lipophilic cell membranes and release their payload into the cell. A popular reagent that uses this technique is known as Lipofectamine®. However, whilst popular for

smaller scale transfections, the use of this reagent for large scale transfections is unsuitable due to the cost associated, as well as the toxicity the reagent exhibits to cells when used over a longer period of time.[267–269]

The final method is the use of cationic polymers such as DEAE-dextran or polyethylenimine (PEI). These positively charged cationic species bind the negatively charged DNA and the polymer-DNA complex can then be taken up into the cell via endocytosis.[267] These reagents are often much more cost effective than the liposome type transfection reagents that are advertised by manufacturers so are more suitable for larger scale transfections.

Polyethylenime (PEI) was chosen as the transfection reagent of choice. A variety of different ratios of transfection reagent to DNA as well as the length the cells were left to grow after initial transfection were explored on a small 6 well plate scale to attempt to find a suitable combination for scaling up to the larger 150 mm plates.

Analysis of cell lysates by western blotting, with quantification of amounts of Keap1 and the control, tubulin by densitometry (Figure 26) showed that a ratio of 1:4 of DNA to transfection reagent, combined with a time point of 72 hours post transfection was found to give the largest amount of protein.

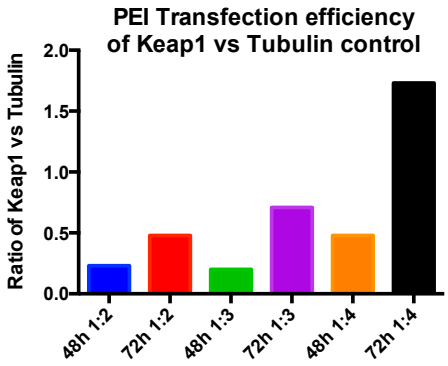


Figure 26: Transfection optimisation. Graph showing the ratio of Keap1 protein to endogenous tubulin control for a range of transfection ratios and times in HEK293T cells. Bars are labelled with the time post-transfection and the ratio of amount of DNA to transfection reagent. Bars represent the ration of Keap1 to tubulin measured by pixel density from western blotting.

3.2.1.4 Purification of expressed protein

The inclusion of a FLAG® (DDK), and cMYC tag on the C-terminal end of the protein, meant that purification of the protein from the cellular lysate could be performed using affinity chromatography, with a bead-bound ANTI-FLAG® M2 antibody. The purification was performed as per the manufacturer's instructions using a 3x molar ratio of FLAG® peptide to column FLAG® peptide concentration to elute the protein from the beads.

In order to determine the purification efficiency (measured through the efficiency of binding to the beads) the amount of Keap1 protein detected in the input and output lysates were monitored after each purification run. Analysis by densitometry showed that binding to the beads was around 50% efficient (Figure 27). Which was the maximum achieved through this method, increasing the length of binding did not appear to change the amount of protein bound.

Figure 27 shows a western blot of the input and output fractions, loaded in equal volume as well as densitometry quantification in the quantitative amounts of Keap1 in each lane. From the analysis it can be seen that the amount of Keap1 present in the sample after binding to the beads is performed is approximately half that before binding, illustrating a successful extraction of protein from the lysate.

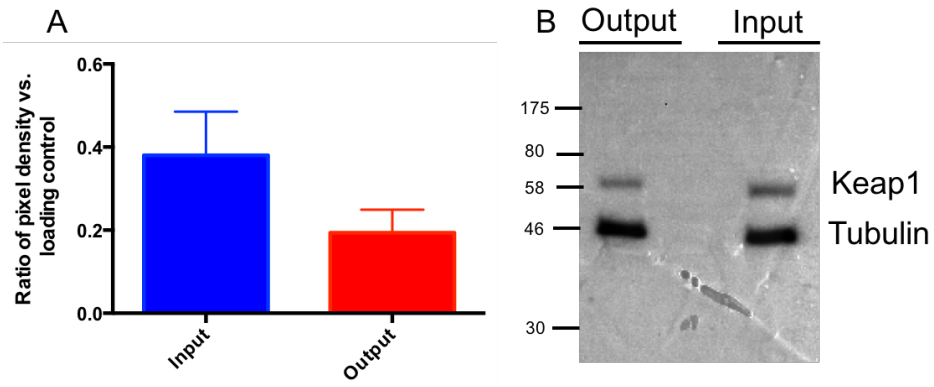


Figure 27: Purity quantification of purified Keap1. A) Bar chart illustrating average +/- SD of the densitometry data acquired from replicate western blots of Keap1 purification columns (N=3) of 'Output' lysate containing unbound proteins, and the 'Input' lysate containing unpurified protein lysate. The radio of pixel density of the Keap1 band vs. the tubulin loading control is given on the y-axis. The output bar (red) illustrates a reduction of ~50 % in Keap1 signal vs. input lysate (blue). B) Representative western blot of an SDS-PAGE gel loaded with with 10 µL fractions of each sample, and probed with Anti-Keap1 polyclonal antibody (top) and anti-alpha tubulin antibody (bottom). A reduction in Keap1 signal from input lysate when compared to output lysate is shown.

Following the affinity-bead purification, the protein was then dialysed into the assay buffer in order to remove any remaining FLAG peptide.

3.2.1.5 Western blotting to confirm protein

Western blotting with an ANTI-FLAG antibody was used to check the identity of the purified protein band. From Figure 28 it can be seen that FLAG-tagged protein can be detected at the correct molecular weight for Keap1.

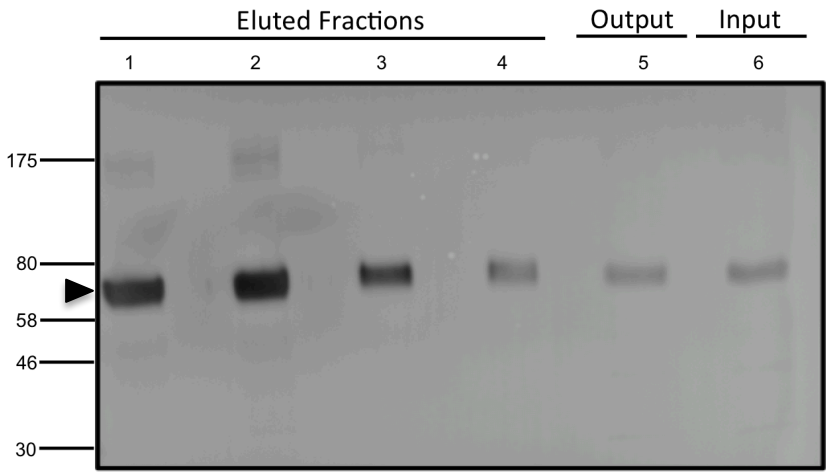


Figure 28: Western blot of an SDS-PAGE gel showing purified Keap1. Gel loaded with with 10 µL fractions of each sample, and probed with ANTI-FLAG antibody. Eluted fractions of Keap1 (1-4), Output lysate from column (5), and Input lysate to column (6) all gave a band of the expected molecular weight (arrow ~70 kDa MW). Some higher molecular weight species are visible particularly in fraction 2 which may represent Dimer (~140 kDa MW)

3.2.1.6 Purity of protein by Coomassie analysis

A 15 µL volume of each of the fractions (300 µl each) eluted from the purification column, as well as the original lysate (input) and the unbound flow through (output), was run on a 4-15% SDS-PAGE gel, total protein visualised by coomassie staining, and analysed by densitometry. This was achieved by analysing the total pixel density of the well and quantifying the Keap1 band as a proportion of total well density. The results show that the protein was produced successfully in high purities of >80 % of total protein in the well (Figure 29).

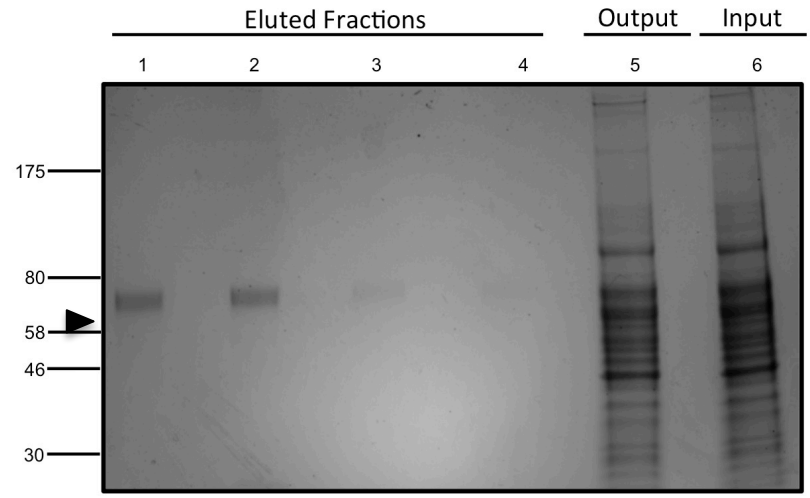


Figure 29: Representative coomassie stained SDS-PAGE gel showing eluted fractions from the Keap1 affinity purification step (1-4), 'Output' lysate containing unbound proteins (5), and the 'Input' lysate (6). 15 µL loaded per lane. The arrowhead indicates the expected molecular weight of Keap1 (~70 kDa). The Keap1 protein was eluted in the first two fractions and the Keap1 band represents >80% of total coomassie staining in that lane indicating a high purity.

3.2.1.7 Determination of protein concentration by BCA assay

The concentrations of each fraction was then determined by BCA assay. Due to the presence of NP-40 in the buffer, conventional analysis by Bradford assay could not be performed due to interference of Bradford reagent with the detergent. As BCA assays use the reduction of copper, rather than creation of a dye-protein complex to deduce protein concentration. Therefore, they are not as affected by detergents or other chemicals that commonly interfere with the dye-based Bradford assay.

Protein concentration was generally around 0.5 mg/ml within the first fraction and fell to 0.1 mg/ml in later fractions. This generally gave around 250 µg of protein per preparation (25 x 150cm dishes of transfected HEK293T cells).

3.2.1.8 Verification of protein activity by fluorescence polarisation assay

To verify protein was functional, the activity in the fluorescence polarisation assay was checked. This ensures that the synthesised protein can successfully bind the Nrf2 peptide and therefore gives an indication that the protein is suitable for use in the assay. Figure 30 shows the titration of the synthesised protein and purchased protein against a fixed 4 nM concentration of labelled peptide. The protein produced in-house shows a similar activity to that of the commercially available protein, and therefore should be suitable for use within the assay.

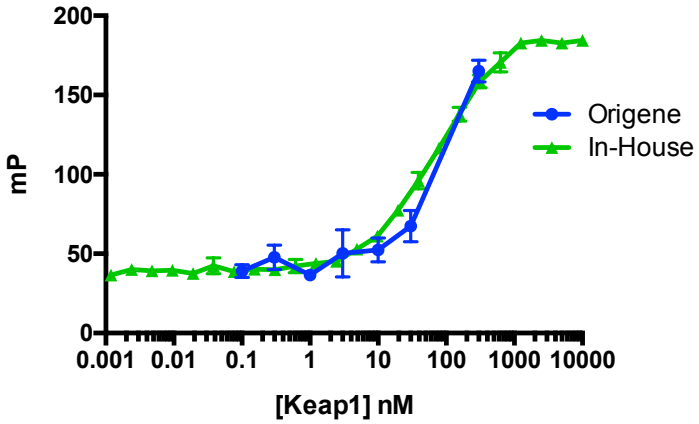


Figure 30: Activity comparison of purchased and synthesised protein. Protein titration curves showing a titration of Keap1 protein produced either using the in-house protocol (green) or purchased from supplier* (Blue). Protein was titrated against a fixed 4 nM concentration of labelled peptide (L2-Flu). Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus the protein concentration in nM (N=3). The protein produced in-house shows a similar profile to that of purchased protein. *Limited points for purchased protein due to limited availability of protein and supplied concentration.

3.2.2 Miniaturisation of HTS assay

After the initial groundwork by Dr. Mead (3.1.5) in establishing whether the assay was feasible for use with this project, and obtaining protein in quantities needed for assay development and screening the next step was to develop the assay so it was suitable for HTS.

To develop the assay into an HTS format it would need to be miniaturised, to enable screening in 384 or 1536-well plates and also to reduce the protein requirements of the assay at current volumes making it too expensive to screen thousands of compounds using purchased protein.

3.2.2.1 HTS Peptide Optimisation

First, it had to be shown that the peptides could be detected at a reduced assay volume. The peptides were therefore again titrated as in section 3.1.5.4 to obtain a usable peptide concentration.

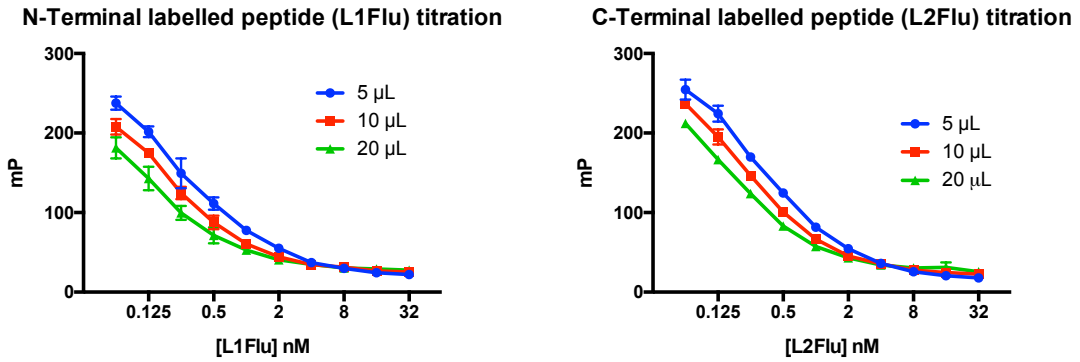


Figure 31: Peptide titration curves at various assay volumes. Left-titration curve of N-terminal labelled peptide. Right-titration curve of C-terminal labelled peptide. Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus peptide concentration in nM (N=3). For both peptides the FP signal is stable down to 4 nM even at a 5 µL assay volume.

From Figure 31 it can be seen that, even at 5 µL assay volumes, the peptide signal is still stable at low concentrations of peptide. At concentrations lower than 4 nM the detection limit of the platereader is reached and results in the artefact of increasing mP signal. A 4 nM concentration of peptide was chosen for both the C-terminal labelled peptide and the N-terminal labelled peptide.

3.2.2.2 HTS Protein Optimisation

The next stage was to test the lower assay volume with a Keap1 protein titration at the selected fixed peptide concentration to check that a good signal could be obtained at the lower assay volume.

From the data (Figure 32) a signal of ≥ 120 mP suitable for use in the screening assay can be obtained with both peptides. For the N terminal labelled peptide, a protein concentration of around 15 nM is needed to obtain the desired signal of 120 mP (which should give a desired signal window of ~4). In contrast the C terminal labelled peptide required a protein concentration of around 90 nM to obtain the same signal. However, it should be noted that the maximal signal that could be reached by the N terminal labelled peptide in the assay (~150 mP) was significantly lower than that of the C terminal labelled peptide (~200 mP).

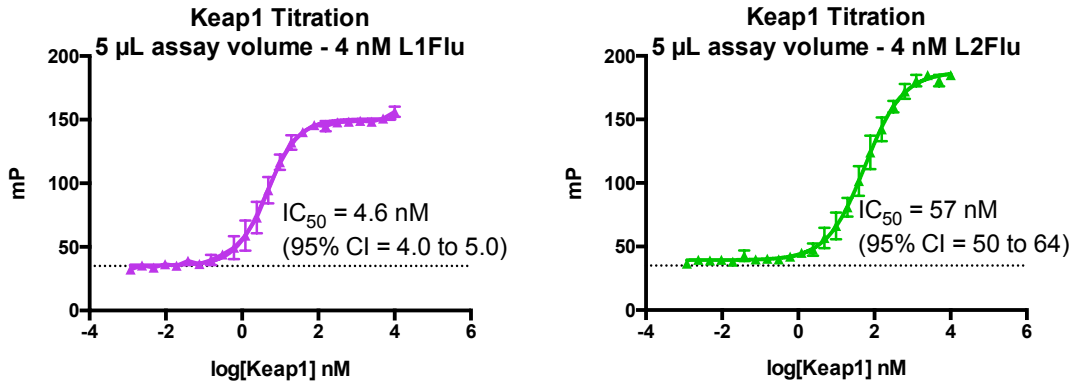


Figure 32: Protein titration curves. Both graphs show a titration of Keap1 protein against a fixed 4 nM concentration of labelled peptide. Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus the log of protein concentration in nM (N=3). IC_{50} values are indicated with their 95 % confidence intervals (95% CI) Left graph illustrates the N-terminal labelled peptide (L1Flu) whilst the right graph indicates the C-terminal labelled peptide (L2Flu).

3.2.2.3 Proof of concept in a competition binding assay

In order for the assay to be validated, it needed to be able to show that a competition-binding event could occur. This was done using the unlabelled 9-mer ETGE peptide (H_2N-LDEETGEFL-$CONH_2$). In addition, in order to utilise the assay to determine a dissociation constant (K_D) for the labelled and unlabelled peptides, the protocol detailed by Rossi et al. was followed.[270] The assay was performed at an IC_{90} concentration of Keap1 to ensure maximal fraction bound peptide.

At this point it was decided that the C-terminal labelled peptide might be better for use in competition binding studies. This was after analysing several co-crystal structures of the ETGE peptide, in particular the one published by Lo et al. in 2006.[229] Analysis of the co-crystal structures showed that the N-terminal end of the peptide is located within the active site and has the potential to interfere with binding (Figure 33). However, the C-terminal end protrudes away from the active site, out into solvent making it a more sensible location for attachment of the fluorophore. This decision was further backed up from experimental evidence with the K_D's of the two-labelled peptides determined from

the assay using Calculation 1. This showed that the K_D of the N-terminal labelled peptide was higher (2.58 nM) than the C-terminal labelled peptide (54.52 nM), suggesting that having the FITC tag closer to the active site could be increasing the affinity of the peptide. This would make it more difficult for any competing molecule to remove the peptide, potentially reducing the number and variety of molecules that could be detected in a HTS. In addition, it also introduces the possibility of increased 'false-positive' hits in the assay as compounds that favourably interfere with the fluorophore binding over the peptide could be increased.

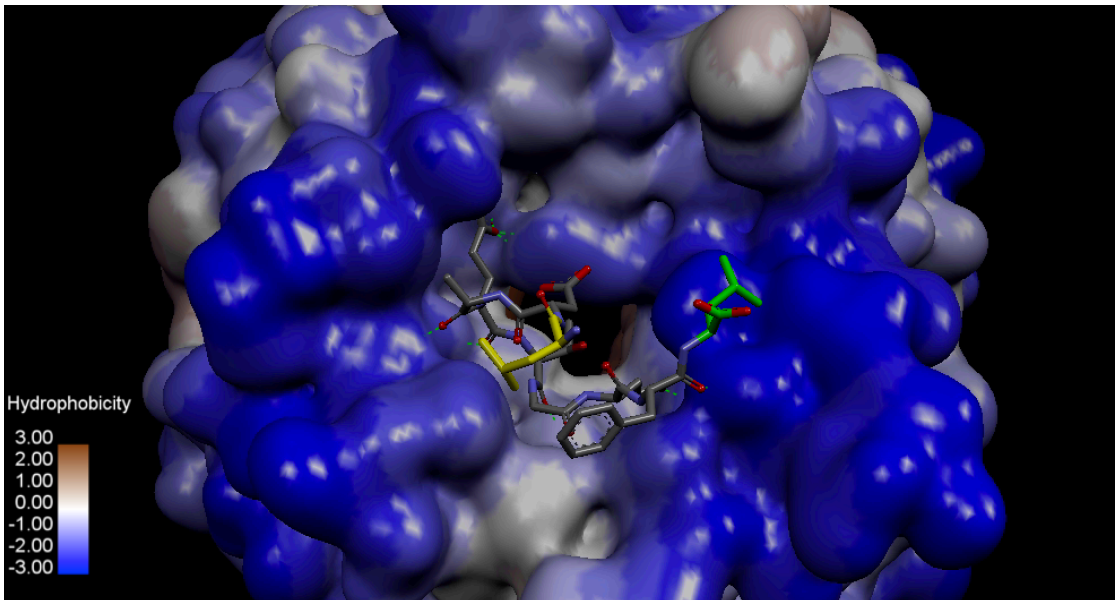


Figure 33: X-ray crystal structure of Keap1 and ETGE peptide showing the locations of the N-terminal peptide residue (yellow), and C-terminal residue (green). Protein surface illustrates hydrophobicity with blue residues indicating hydrophilic areas of the protein, brown showing hydrophobic areas with white as an intermediate between the two. Model generated from structural alignment of PDB structures 2DYH and 2FLU with their contained peptide ligands using Discovery Studio (Accelrys).

From Figure 34 it can be seen that the unlabelled peptide is able to disrupt the bound Keap-L2Flu complex (275 mP) and bind to Keap1 causing L2Flu to dissociate to its unbound state to give the baseline reading of 35 mP.

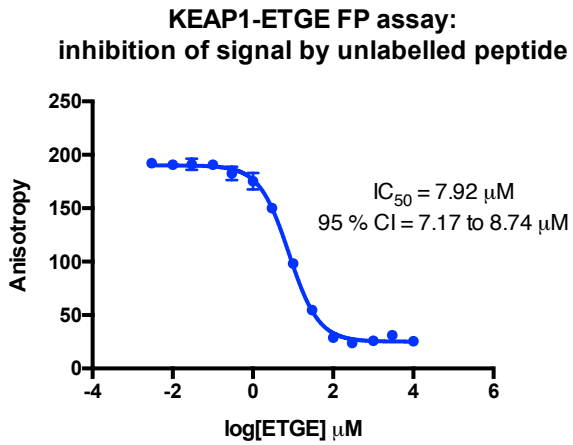


Figure 34: FP competition binding assay using C-terminal fluorescein labelled ETGE peptide (4 nM), Keap1 (1 μM), and the indicated concentrations of unlabelled ETGE peptide. Graph illustrates the data plotted as the average +/- SD of Anisotropy values versus the log of unlabelled peptide (ETGE) concentration in nM (N=3). IC_{50} values are indicated in micro molar (μM) with their 95 % confidence intervals (95% CI).

Following the protocol in Rossi et al., the binding constant for the labelled and unlabelled peptides could then be determined using Calculation 1.[270]

Prior to the calculation the fluorescence polarisation signal associated with non-specific binding at each concentration of protein was required to be determined. If a signal for non-specific binding occurred any concentration of protein this could then be factored into the calculation. Keap1 protein was therefore titrated against a fixed concentration of fluorescence ligand (4 nM) and a saturating concentration of unlabelled peptide to ensure maximal depletion of signal. Any binding occurring due to non-specific binding of the labelled peptide would therefore be seen through an increased FP signal when compared to a peptide and buffer control.

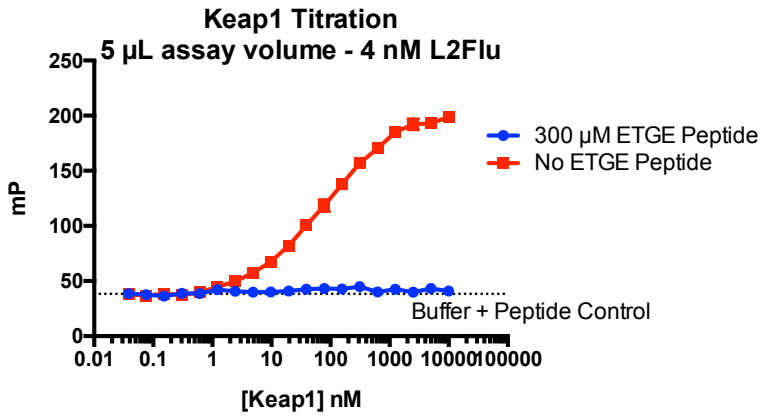


Figure 35: Protein titration curves to calculate non-specific binding. Graph shows a titration of Keap1 protein against a fixed 4 nM concentration of labelled peptide, with either a fixed 300 µM concentration of unlabelled ETGE peptide (blue), or a null-peptide control (red). The null protein control (buffer + labelled peptide only) is shown as a dotted line. Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus the protein concentration in nM (N=3).

Figure 35 shows that no observable signal due to non-specific binding is observed above the baseline control. The non-specific baseline correction can therefore be excluded from the calculation.

Calculation 1: Determination of K_D values for labelled and unlabelled ETGE peptides using fluorescence polarisation data.

$$K_D^I = \frac{[D^*R]_{50} IC_{50} K_D^{D^*}}{(D^*{}_T R_T) + [D^*R]_{50}(R_T - D^*{}_T + [D^*R]_{50} - K_D^{D^*})}$$

Where:

K_D^I =	Binding constant of inhibitor
D^* =	Donor (L2Flu)
R =	Receptor (Keap1)
$K_D^{D^*}$ =	Binding constant of Donor (L2Flu)
IC_{50} =	Free concentration of competing ligand (ETGE) at 50% displacement of specifically bound D^*
$D^*{}_T$ =	Total donor concentration
R_T =	Total receptor concentration
$[D^*R]_{50}$ =	Concentration of D^*R at I_{50}
I_{50} =	Total concentration of competing ligand (ETGE) at 50% displacement of specifically bound D^*
R_T^{50} =	Total receptor concentration required to cause 50% of D^* to be bound.

From the data generated the K_D^I can therefore be calculated for the unlabelled peptide:

First K_D for labelled peptides:

N-Terminal Labelled:

$$K_D^{D^*} = R_T^{50} - \frac{D^*{}_T}{2} = 4.58 - \frac{4}{2} = \mathbf{2.58\ nM}$$

C-Terminal Labelled:

$$K_D^{D^*} = R_T^{50} - \frac{D^*{}_T}{2} = 56.52 - \frac{4}{2} = \mathbf{54.52\ nM}$$

Followed by IC_{50} determination of the inhibitor and $[D^*R]_{50}$ leads to determination of K_D^I

$$IC_{50} = I_{50} - \frac{R^T}{2} = 7917 - \frac{1000}{2} = \mathbf{7417\ nM}$$

$D^*{}_T$ = 4 nM

R_T = 1000 nM

$$[D^*R]_{50} = D^*{}_T \frac{A_S^{50} - A_{D^*}}{A_{D^*R} - A_{D^*}} = 4\frac{103-25.14}{190.1-25.14} = \mathbf{1.89\ nM}$$

I_{50} = 7917 nM

$$K_D^I = \frac{1.89 \times 7417 \times 54.52}{(4 \times 1000) + 1.89(1000 - 4 + 1.89 - 54.52)} = \frac{764268.45}{5782.97} = \mathbf{132\ nM}$$

$$K_D^I = 132\ nM$$

From the calculation it can be seen that the binding constant for the labelled peptide (132 nM) is about 2.5 x that of the unlabelled peptide (55 nM).

3.2.2.4 HTS Assay Validation

In order for the assay to be suitable for HTS, the robustness of the assay would need to be tested. This can be done through the use of several statistical tests. For the purpose of this work, the tests chosen were taken from the chapter on HTS assay development in the Assay Guidance Manual by Sittampalam et al.[271]

The first test used is the Z' factor (Equation 3). This is a measurement of signal separation, plate uniformity and signal variability. The Z' factor can be defined by the formula below. A good Z' factor is determined by the manual as a value ≥0.4.

Equation 3: Formula for calculating Z'. Where AVG_{max} is the average maximal signal (FP signal gained from 0 % inhibition), AVG_{min} is the average minimal signal (FP signal gained from 100 % inhibition). SD_{max} is the standard deviation obtained from the AVG_{max} and SD_{min} is the standard deviation obtained from the AVG_{min} and n indicates the number of replicates in the final screening assay.

$$Z' = \frac{\left(AVG_{max} - \frac{3SD_{max}}{\sqrt{n}}\right) - \left(AVG_{min} - \frac{3SD_{min}}{\sqrt{n}}\right)}{AVG_{max} - AVG_{min}}$$

In this assay the *"max"* signal is the FP signal measured with an EC_{80} concentration of a standard agonist (inhibitor). In this case a published compound (Figure 36) from the Broad institute (PubChem BioAssay ID: 504523, 504540) was used. The *"min"* signal is designed to measure the background signal of the assay so is the FP signal measured with an EC_{10} concentration of standard agonist.

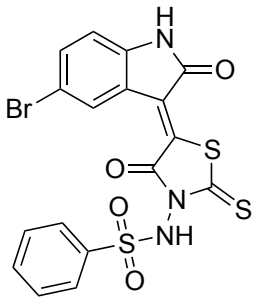


Figure 36: Chemical structure of the test compound used as a standard agonist, selected from one of several hit compounds in the Broad institute screen (PubChem BioAssay ID: 504523, 504540).

To test plate uniformity, the assay was designed with an interleaved-signal format (Figure 37). The plate has a combination of wells producing "maximum" and "minimum" signals on the plate. This format of plate design enables HTS artefacts such as plate effects to be detected; examples of these are signal drift (where the signal changes from left to right, or top to bottom across a plate) and edge effects (where wells at the edge of the plate have a different signal, usually due to evaporation happening more quickly in edge wells).

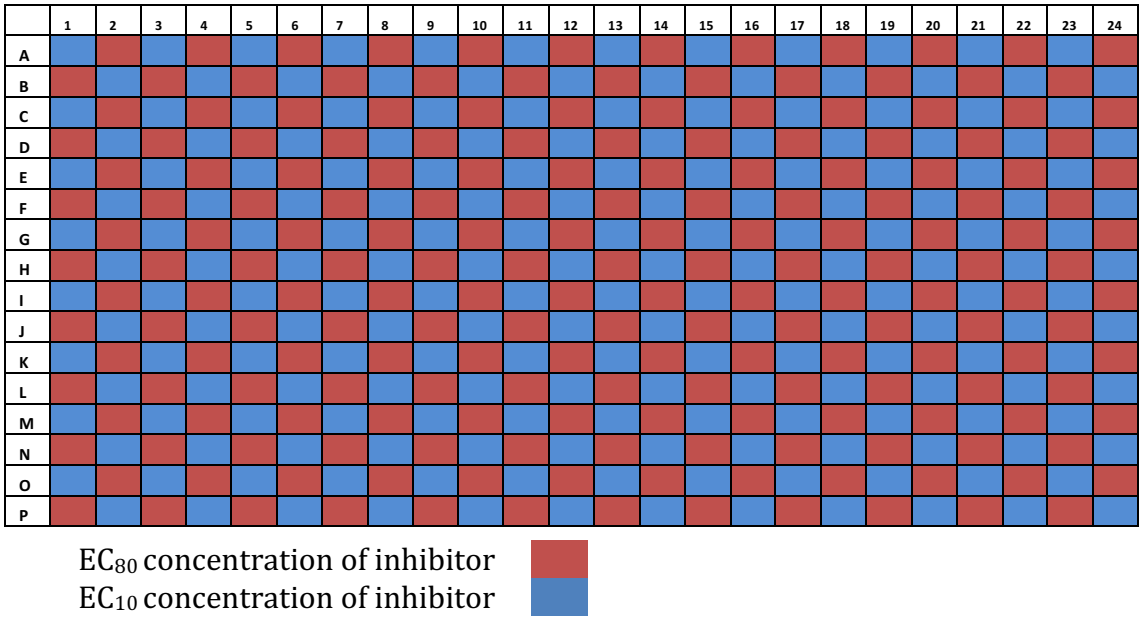


Figure 37: Plate layout for Z' and SW determination

From the results (Figure 38, and Calculation 2) a Z' factor of 0.694 was obtained, which is greater than the criterion of 0.4 for a suitable HTS assay. A signal window (SW) of 2.7 was also calculated from the plate, greater than the suitability criterion of 2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	107.4	41.6	111	40.1	108.4	41.8	107.3	39.1	107.8	40.6	103.9	45.9	104	37.8	107.7	41.3	105.8	36.2	107.5	37.3	106.6	40.1	109.2	40.4
B	37.4	98.8	38.9	108.6	38.7	104.8	40.4	104.3	39.8	106.7	38.2	104.4	37.2	104.4	37	105	38.7	106.2	36.6	105.8	40.9	103	39.7	106.9
C	109.5	40.5	107	40	109.8	40.2	107.3	40.3	111.6	39.2	105.9	45	108.3	39	107.3	38.7	107	36.1	108.6	36.5	104.4	41.8	106.8	43.1
D	43.5	84.9	38.6	111.1	40	105	39.4	106.2	39.7	107	39.4	110.1	38.5	107.7	39.5	106.3	38.1	108.5	40.1	104.4	38.7	112.8	38.5	105.8
E	106.7	40.7	110.3	37.5	105.6	39.7	111.2	38	110.4	37.6	107.4	47.9	109.2	38.5	106	40.8	106.9	38.9	107	38.7	105.3	37.6	105.7	41
F	39.7	76.7	38.6	113.1	38.7	109	38.9	104.3	40.8	106.1	37.3	108.5	38.3	107.8	39.5	104.8	38.7	106.7	36.7	103.4	36.7	105.8	36.8	104.5
G	108.1	39.9	119.7	40.4	108.5	41.3	111.4	39.8	109.6	40.2	107.8	46.9	106	39	106.9	36.3	106.8	37.6	103.1	37.9	104	41	102.2	36.2
H	40.4	72.2	38.4	108.5	37.1	105.4	38.3	105.1	40.1	105	38.7	106.1	38.3	105.6	39.6	104.8	38.3	108.2	38.3	106.7	37	105.9	38.8	102.3
I	113	39.6	109.3	45.1	106.7	40.7	107.5	38.3	107.5	37.8	106.7	43.4	106.8	37	103.9	36.7	105.3	40.8	106.5	37.4	106	36.5	107.7	37.7
J	40.9	96.4	40.7	109.6	36.8	105.7	39.3	106.4	36.2	104.8	38.7	108.2	37.2	105.6	39.6	102.5	36.4	107.9	36.3	107.7	38.2	105.1	37.5	102.7
K	109	39.7	109.7	41.7	107	40	107.1	41.1	106.1	37	104.8	42.6	105.2	40.1	101.8	39.4	102.6	38	105.3	37.6	104.3	36.8	104.1	38.4
L	37.1	98.5	39.4	106.2	38.3	107.8	38.8	106.5	40.4	104.9	39.1	105.8	38	105.6	37	104.1	37.6	105.1	36.4	101	35.8	101.6	35.4	103.3
M	109.6	35.1	111.1	40.1	105.7	39.9	106.1	35.2	107	37.4	102.9	45.2	106.5	38.1	102.7	36.6	105.3	35.9	106.2	37.3	103.1	37.7	109.7	36
N	36.5	105.4	38.5	107.1	38	104.1	37.9	107.5	38.3	104.9	35.8	107.8	38.4	103.8	36.1	104.2	37.5	106.3	37.3	107.4	37.9	102.8	38.2	100.6
O	106.4	39.7	106.9	40.3	104	37.1	108.2	40.2	105.6	37.8	105.6	43.4	102	36.7	107.9	35.4	104.2	35.5	104.9	38.6	103.5	38.2	102.4	36.5
P	41.2	92.2	44	105.5	38.5	109.2	37.1	104.5	41.1	105.9	38.3	104.3	37.3	102	38.3	105	36.6	100.9	33.8	101.7	38.5	102.7	41	106.5

Figure 38: Results from Z' and SW assay. Max signal shown in green, and min signal shown in red, a colour gradient is applied to highlight marginal wells. Values indicate fluorescence polarisation values in milli polarisation units (mP).

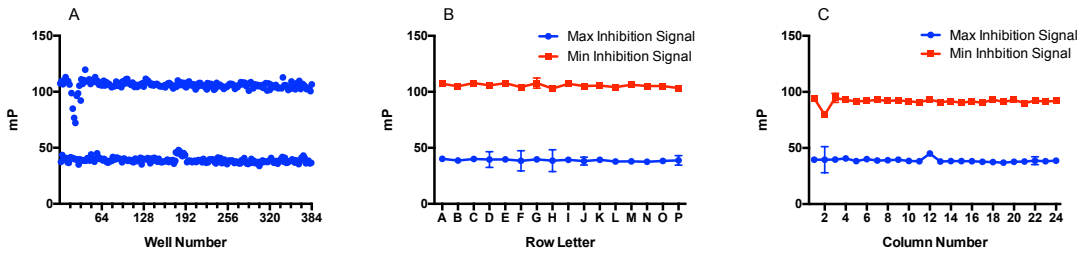


Figure 39: Graphs illustrating well data (A), row averages (B) and column averages (C) in order to illustrate the effect of drift and edge effects in the Z' assay. Graph A shows the raw fluorescence polarisation values on the Y-axis, measured in *milli polarisation units (mP)* for each individual well number (plotted on the X-axis). Graph B shows the row averages +/- SD of fluorescence polarisation on the Y-axis for both the maximal inhibition signal (Blue) and the minimal inhibition signal (Red). The corresponding row values are shown on the X-axis. Graph C shows the column averages +/- SD of fluorescence polarisation on the Y-axis for both the maximal inhibition signal (Blue) and the minimal inhibition signal (Red). The corresponding column values are shown on the X-axis.

Calculation 2: Determination of Z' and the signal window (SW) from fluorescence polarisation data.

$$Z' = \frac{\left(105.7 - \frac{3 \times 4.6}{\sqrt{1}}\right) - \left(38.9 - \frac{3\times 2.2}{\sqrt{1}}\right)}{105.7 - 38.9} = \mathbf{0.694}$$

$$SW = \frac{105.7}{38.9} = \mathbf{2.7}$$

No observable edge or drift effects were seen (Figure 39), although two column singletons were seen (one in column 2 and one in column 12). This was likely due to errors in pipetting of single columns, which could be due to a number of reasons (small bubbles in pipette tips, random machine pipetting error leading to either too much or too little protein dispensed etc). However, despite the two row errors, the plate still achieved the suitable criteria for a HTS assay, giving assurances that the assay is robust and can tolerate small errors.

3.2.3 Further assay exploration and development

3.2.3.1 DLG peptide

Whilst a successful assay was created using the ETGE Nrf2 peptide, having an assay that could use the weaker binding DLG peptide would be advantageous over the ETGE as in the literature it is believed that only the DLG Nrf2 motif would need to be disrupted for the ubiquitnation of Nrf2 to be prevented, this is known as the hinge and latch model.[246]

The assay development was trialled with the N-terminal labelled DLG 12-mer peptide (FluoNH-DILWRQDIDLGV-CONH2) using the same optimisation method as above (Figure 40). A peptide concentration of 4 nM was again seen to be suitable. However, no

activity in the assay could be seen at concentrations of protein up to 6 μM. This was hypothesised to be mainly due to the lower affinity of the DLG motif resulting in the assay not being able to detect the binding. It was thought that if the protein concentration could be increased then a signal may have been seen, however increasing the protein concentration above ~1 μM would make running the assay impractical due to the large amounts of protein required to perform the assay.

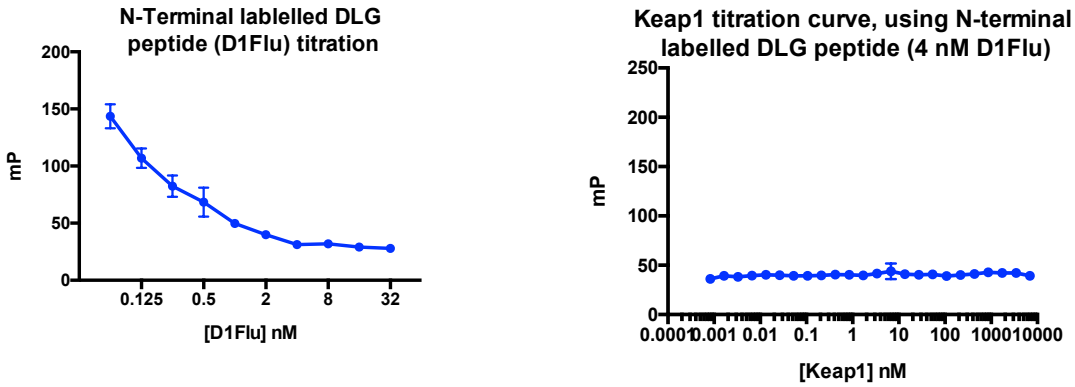


Figure 40: Optimisation of DLG peptide. Left: Peptide titration curve of N-terminal labelled DLG Nrf2 peptide (D1Flu). Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus the peptide concentration in nM (N=3). Right: Keap1 protein titration curve using 4 nM N-terminal labelled DLG Nrf2 peptide (D1Flu). Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus Keap1 protein concentration in nM (N=3).

Another approach would be to find a higher affinity DLG peptide more representative of the Nrf2 motif as a shorter peptide may not have as higher affinity as a larger more "protein-like" structure that can form more complex secondary structures. In a recent paper Fukutomi et al were unsuccessful in getting 12-mer, 14-mer or 16-mer DLG peptides to bind in ITC and DSF experiments with the kelch domain of Keap1, as had been seen in previous literature.[249,272] However, they found that by extending the number of amino acids in the peptide chain to 35-mer peptide known as DLGex that an affinity for the Keap1 closer related to that of the Nrf2 DLG domain (K_D ~ 10^6) can be seen.[272]

We decided to test to see if we were able to detect this interaction in our assay, so therefore tested the 35-mer DLGex peptide with a N-terminal FITC tag (FluoNH-MDLIDILWRQDIDLGVSREVFDFSQRRKEYELEKQ-CONH2).

Figure 41: Partial amino acid sequence of Nrf2 showing the residues of the 12-mer DLG peptide (pink) and the extra residues encompassed in the 35-mer DLGex peptide (green).

When titrating the peptide, it was observed that the increase in FP signal observed as the concentration of peptide is reduced was more gradual compared to the slope previously seen with the shorter ETGE (L1 and L2) and DLG (D1) peptides (Figure 42). The increase is normally indicative of the fluorescence levels becoming too low for the platereader to detect, leading to erroneous increases in the signal. However, this observation is not seen with the DLGex labelled peptide where only a very small increase in FP signal is seen even at much lower concentrations that the other peptides. Due to the difficulties selecting a clear concentration of peptide, a 4 nM concentration of peptide was selected to trial for interaction with Keap1, as it would provide a good comparison to the other peptides that were tested at the same concentration.

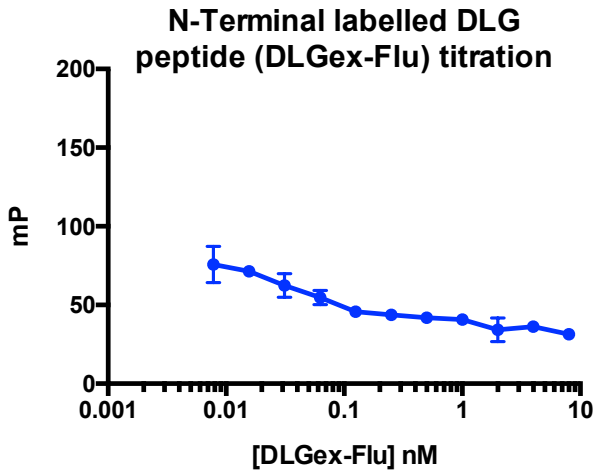


Figure 42: Peptide titration curve of N-terminal labelled DLGex Nrf2 peptide. Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus the peptide concentration in nM (N=3).

The peptide was then titrated against the protein up to a concentration of ~ 6μM. Whilst there was no change in the signal at concentrations of protein up to ~200 nM, above this the signal decreased until at the highest concentration of protein there was a large negative shift in the polarisation (Figure 43). This was hypothesised to be due to the propeller effect caused by the increased length of the peptide chain. This effect occurs due to the fluorescent tag being further extended into solvent away from the binding site because of an increased number of non-bonding residues. When the peptide chain is bound, this decreases the movement of any bound resides, however the tag, extended into solvent does not experience this effect and can end up rotating quicker than in free solution leading to a decrease in the fluorescence polarisation signal.[273]

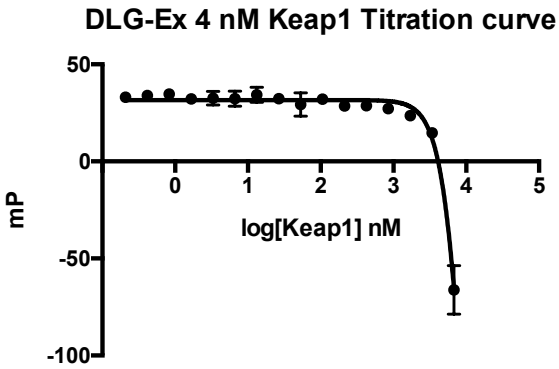


Figure 43: Keap1 protein titration curve using an extended sequence length 4 nM N-terminal labelled DLG Nrf2 peptide (DLG-Ex). Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus Keap1 protein concentration in nM (N=3).

3.2.3.2 Effect of a reducing reagent on the assay

Due to the "redox sensitive" cysteine residues on Keap1 we wanted to make sure that the change in redox state of the sulfhydryl groups did not affect the ability of the assay to detect potential inhibitors. We therefore tested a reference compound from the literature for which we had an established IC_{50} value. The compound we chose named in the literature as "Hit1" (Figure 44) and is known to have a K_D of 1.9 μM.[211]

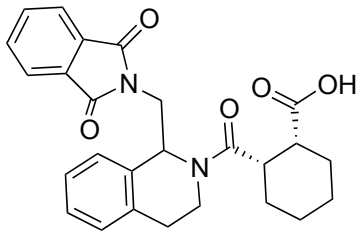
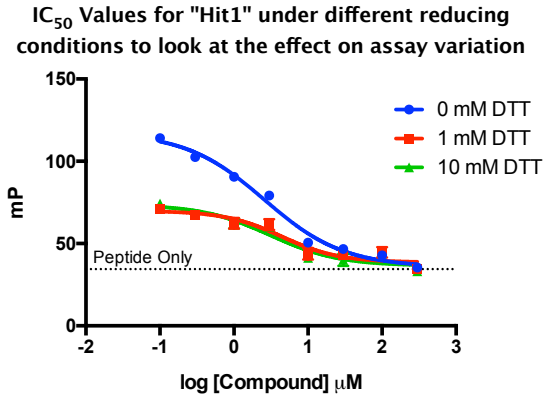


Figure 44: Chemical structure of "Hit1".[211]

We tested the IC_{50} of the compound in the assay at 0, 1 and 10 mM concentrations of DTT in the regular assay buffer (Figure 45). We found that, whilst the signal of the FP assay decreased with the addition of DTT, the IC_{50} of the compound remained similar and within in the 95% confidence intervals for all measurements made. This suggests that the reaction of the cysteine residues on Keap1 has no effect on the affinity of the ligand bound, and that the ligand is able to bind in any mono- or multi-meric form of the protein. However, the decrease in FP signal does suggest that the reducing agent is able to have an effect on the assay through a lowering of the maximum signal obtained with the same concentration of protein. This is likely to be due to the ability of reducing agents to disrupt disufide bonds in order reduce Keap1 from a larger (multimeric) form into a smaller (monomeric) form (see Chapter 7).



	0 mM DTT	1 mM DTT	10 mM DTT	Average	% Variation	SD
IC_{50} (μM)	2.56	4.43	3.12	3.38	28	0.95
95% CI	1.77 to 3.77	2.35 to 8.34	1.51 to 6.42			

Figure 45: Comparison of IC_{50} values in the presence and absence of the reducing agent DTT using the control compound "Hit1". Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus the log of the concentration of Hit1 in nM (N=3). Varying concentrations of DTT used in the assay are shown, with 0 mM in blue, 1 mM in red and 10 mM in green.

Fluorescence polarisation relies on the binding of a small fluorescent ligand to a larger molecule for signal (3.1.3.2). It can therefore be concluded that an absence of reducing agent would be beneficial for the assay due to an amplification in signal at lower protein concentrations, with no adverse effects on binding. This enables lower concentrations of protein to be utilised, reducing costs and timescales for production of protein.

3.2.3.3 Effect of a classical Nrf2 activator on the assay

As previously discussed (1.4.5.1) most classical Nrf2 activators work through covalent reaction with sulfhydryl groups within cysteine residues on Keap1. It was important that the assay only identified non-reactive inhibitors of the Keap1-Nrf2 interaction. Therefore, the effect of reactive Nrf2 activators had to be tested in the assay to check whether they were able to disrupt the Nrf2-Keap1 interaction.

Previous literature has suggested that classical Nrf2 activators should not disrupt the Nrf2 ETGE motif from binding. The hinge and latch model, which is most prominent in the literature, suggests that only the DLG motif (the latch) is released upon binding of classical Nrf2 activators with Keap1.[197,246] Therefore, as the ETGE peptide motif is utilised in the fluorescence polarisation assay, no effect should be seen in the assay on addition of a covalent Keap1 binder.

Three known classical activators were tested in the assay (Figure 46); Bardoxolone-Methyl (CDDO-Me), Sulforaphane (SFN), and Dimethyl-fumarate (DMF, BG-12).

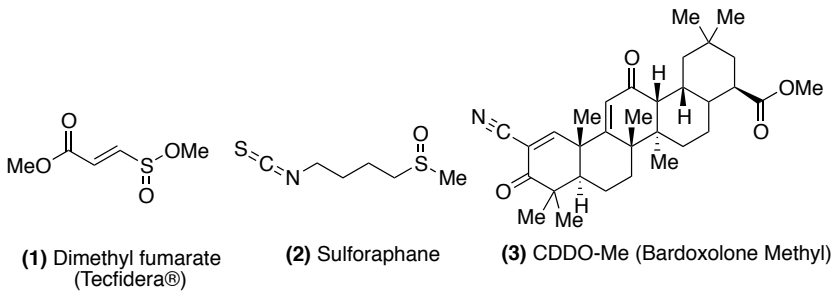


Figure 46: Activators of Nrf2 that work through a classical mechanism of covalent modification of Keap1.

From the results (Figure 47) it can be seen that no effect on the binding of the ETGE peptide can be seen, with the fluorescence polarisation signal unchanged. This suggests that the assay should be very robust in filtering out reactive activators, selecting for the desired non-reactive activators of Nrf2.

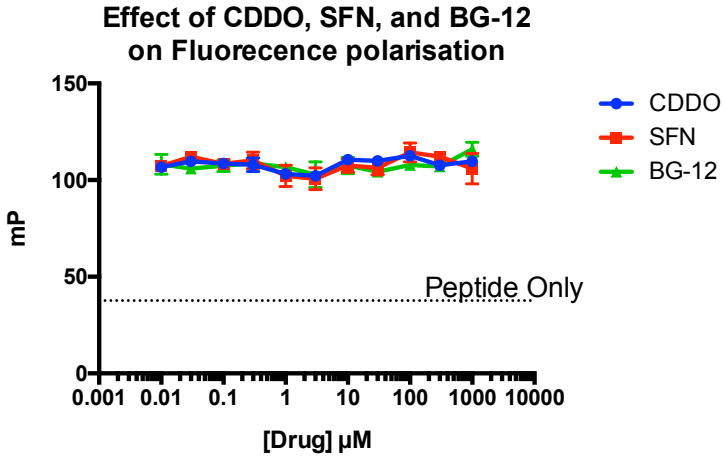


Figure 47: Graph showing how varying concentration of classical activators fails to have any effect on reduction of fluorescence polarisation signal, and hence disrupt the Nrf2 peptide from binding to Keap1. Average +/- SD of fluorescence polarisation measured in milli polarisation units (mP) versus the compound concentration in nM (N=3). Bardoxolone-Methyl (CDDO-Me) is represented in blue, sulforaphane (SFN) in red and Dimethyl-fumarate (BG-12) in green.

3.3 Conclusions

To conclude a HTS assay was developed to probe the protein-protein interaction between Keap1 and Nrf2, using mammalian full length recombinant Keap1 protein and different fluorescein labelled Nrf2 based peptide motifs. The assay was successfully miniaturised to a 5 μL, 384 well format and using 90 nM protein and 4 nM peptide concentrations the Z' factor and signal windows determined. The Z' factor was found to be 0.694 and the signal window 2.7, values in excess of those required for a robust assay. In addition, the assay was used to determine the binding constant of the ETGE peptide, and was found to be 132 nM vs. a binding constant of 55 nM for the labelled peptide.

The assay can now be used to test a library of compounds for inhibitors of the interaction between Keap1 and Nrf2.

The assay was also tested for interactions with the weaker affinity DLG motif on Nrf2, as successfully establishing an assay with the weaker DLG peptide may lead to discovery of a more diverse set of compounds being able to disrupt either one or both the ETGE and DLG motifs. Both a shorter 9-mer peptide and a longer 35-mer peptide were tested for affinity, with no detectable binding being observed. Furthermore, the assay was tested for its selectivity towards non-reactive inhibitors through the testing of reactive inhibitors in the assay. The assay showed the ability to distinguish between reactive and non-reactive inhibitors. This gives confidence that any inhibitors identified within the assay will be reacting through direct inhibition of Nrf2 binding to Keap1, rather than through reaction with the redox sensitive cysteines on Keap1.

Additionally, a methodology for synthesising recombinant full-length protein in-house was developed enabling a significant reduction in the cost associated with running of the assay. Furthermore, producing the protein in-house will enable greater control over supply and quality of the protein potentially reducing assay variance.

4. High throughput screening to find inhibitors of the Keap1-Nrf2 interaction.

4.1 Introduction

Following successful establishment of a protein synthesis method capable of producing enough protein for a HTS and development of an assay capable of screening large drug libraries the next stage was to choose and implement a screen of large diverse libraries of compounds.

Two libraries were screened as part of the project, internally a focused library was screened initially utilising the protein prepared in-house (3.2.1). This was then followed by second screen with a much larger library which was provided and screened by an external source under our direct supervision where larger libraries were able to be handled more efficiently through utilising dedicated screening centres.

4.2 Results and Discussion

4.2.1 Screening of the MRCT Protein-Protein inhibitor (PPI) Library

The initial library screened was kindly donated by the Medical Research Council Technology (MRCT), a not for profit spin out from the MRC that collaborates with academics to progress small molecule drug targets. The library contained 13372 compounds pre-selected by the MRCT as compounds likely to inhibit protein-protein interactions based on their ability to mimic alpha-helices. The compounds were provided without the knowledge of the structures.

The library was screened using the assay as described previously (Chapter 2) at a single point concentration of 30 μM.

Starting with the 13372 compounds one compound was identified as a weak hit through a triage process.

4.2.1.1 Plate layout

The plate layout shown below (Figure 48) was used to screen the MRCT library. Inclusion of negative and positive controls on every plate ensured the robustness of the assay, as the Z-score and assay window could be calculated for each plate. Calculating these parameters for each plate enables the signal for each individual compound to be interpreted appropriately, and each plate could be directly compared, even if one plate was run at a different time to another in the library. Having these controls on the plate also enables quality control across the screen with any plate failing to meet the minimum requirements on either Z-factor (the measurement of reproducibility of the screen as defined in 3.2.2.4 HTS Assay Validation as needing to be >0.4) or signal window (>2 as define in 3.2.2.4) to be rejected and re-screened.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	P	P	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	P	P
B	N	N	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30	X31	X32	X33	X34	X35	X36	X37	X38	X39	X40	P	P
C	P	P	X41	X42	X43	X44	X45	X46	X47	X48	X49	X50	X51	X52	X53	X54	X55	X56	X57	X58	X59	X60	N	N
D	P	P	X61	X62	X63	X64	X65	X66	X67	X68	X69	X70	X71	X72	X73	X74	X75	X76	X77	X78	X79	X80	N	N
E	N	N	X81	X82	X83	X84	X85	X86	X87	X88	X89	X90	X91	X92	X93	X94	X95	X96	X97	X98	X99	X100	P	P
F	N	N	X101	X102	X103	X104	X105	X106	X107	X108	X109	X110	X111	X112	X113	X114	X115	X116	X117	X118	X119	X120	P	P
G	P	P	X121	X122	X123	X124	X125	X126	X127	X128	X129	X130	X131	X132	X133	X134	X135	X136	X137	X138	X139	X140	N	N
H	P	P	X141	X142	X143	X144	X145	X146	X147	X148	X149	X150	X151	X152	X153	X154	X155	X156	X157	X158	X159	X160	N	N
I	N	N	X161	X162	X163	X164	X165	X166	X167	X168	X169	X170	X171	X172	X173	X174	X175	X176	X177	X178	X179	X180	P	P
J	N	N	X181	X182	X183	X184	X185	X186	X187	X188	X189	X190	X191	X192	X193	X194	X195	X196	X197	X198	X199	X200	P	P
K	P	P	X201	X202	X203	X204	X205	X206	X207	X208	X209	X210	X211	X212	X213	X214	X215	X216	X217	X218	X219	X220	N	N
L	P	P	X221	X222	X223	X224	X225	X226	X227	X228	X229	X230	X231	X232	X233	X234	X235	X236	X237	X238	X239	X240	N	N
M	N	N	X241	X242	X243	X244	X245	X246	X247	X248	X249	X250	X251	X252	X253	X254	X255	X256	X257	X258	X259	X260	P	P
N	N	N	X261	X262	X263	X264	X265	X266	X267	X268	X269	X270	X271	X272	X273	X274	X275	X276	X277	X278	X279	X280	P	P
O	P	P	X281	X282	X283	X284	X285	X286	X287	X288	X289	X290	X291	X292	X293	X294	X295	X296	X297	X298	X299	X300	N	N
P	P	P	X301	X302	X303	X304	X305	X306	X307	X308	X309	X310	X311	X312	X313	X314	X315	X316	X317	X318	X319	X320	P	P

P	Positive Contol (Low mP)
N	Negative Control (High mP)
X	Sample

Figure 48: Screening plate layout based on a 384 well plate format. Sample wells are shown in grey, positive control wells are shown in green and negative control wells are shown in red. All wells contained 5 µL of 90 nM Keap1 protein with 4 nM labelled peptide. Positive control wells also contained 30 µM concentration of control compound MS16. Negative control wells contained DMSO vehicle control.

4.2.1.2 Screening

In order to screen the assay as quickly as possible, automated dispensing was used. This included use of the Matrix platemate, which enabled quick dispensing of protein from reservoirs into the assay plates. Following this the plates were manually moved to the ECHO550 acoustic dispenser where all screening compounds were added first. Acoustic dispensing has several advantages over traditional dispensing, firstly tip-less transfer results in less loss of compound due to molecule adsorption to plastics. It also cuts down on compound waste and carryover. This results in more reliable and reproducible assay results.[274,275] Addition of the protein to the compound first enables any potential hits to bind with the protein before peptide is added. Following the addition of the screening compounds the control compound MS16 was added to the positive control wells, followed by the fluorescent ETGE peptide, which is added to all wells. The plates were then incubated at room temperature for 10-minutes and read on the platereader. The basic results of the screen with all compounds before any data filtration can be seen in Figure 49.

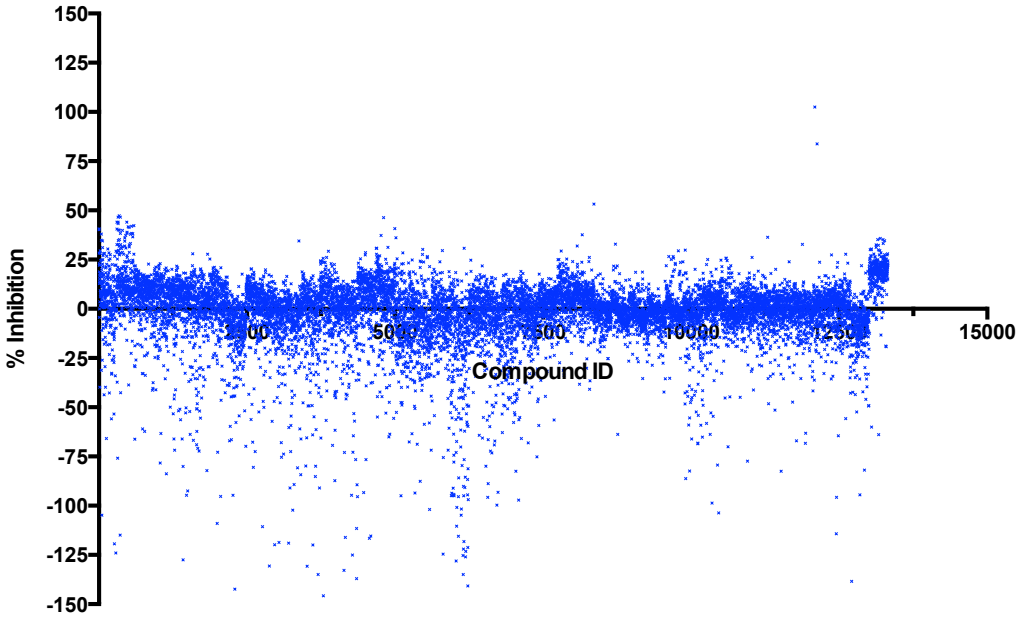


Figure 49: Graph illustrating the percentage inhibition of each compound in the MRCT library. Activity for every compound has been converted to a percentage inhibition based on positive and negative control data for the corresponding plate. Hit compounds are shown in the upper half of the graph, with the most active having a larger positive percentage inhibition.

4.2.1.3 Data analysis

Due to the large amount of data that needed to be processed a data pipelining protocol was designed to handle the analysis of the data (Figure 50). The goal of the protocol was to make it as simple as possible to exclude compounds that interfered with the assay through fluorescence and find compounds that met the criteria of the assay, with an FP value 3 or 4 standard deviations greater than the mean of the controls. Pipeline pilot works through the programming of individual 'nodes' that perform a single operation to the data being input. These nodes can be generally summed up in categories such as read, write and Boolean filters.

The data were read from the excel format produced by the platereader and assigned a plate number. The values for polarisation and fluorescent intensity were then calculated for each well in the plate using the standard formulas as described in Equation 4 and Equation 5.[247,270]

Equation 4: Equation for calculating fluorescence polarisation (FP). Where I=Intensity, ⊥=perpendicular and ‖=parallel.

$$FP = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}$$

Equation 5: Equation for calculating fluorescent intensity from parallel and perpendicular intensity values. Where I=Intensity, ⊥=perpendicular and ‖=parallel.

$$I = I_{\parallel} + 2I_{\perp}$$

From this the average intensity, and polarisations of the positive and negative control wells can be calculated along with the standard deviations. The samples are then filtered based on whether they reduce the average signal of the negative controls by greater or equal to 3 or 4 times the standard deviation of the negative controls. These equate to Z scores of >3 and >4 respectively. Following this, the samples are further triaged based on fluorescence intensity. Any compound that has fluorescence intensity that either exceeds or reduces the mean intensity of the controls by +/- 3SD of the mean is filtered out as a false positive within the assay. Within the protocol no data are ever deleted when filtered but stored in a separate file from the final results.

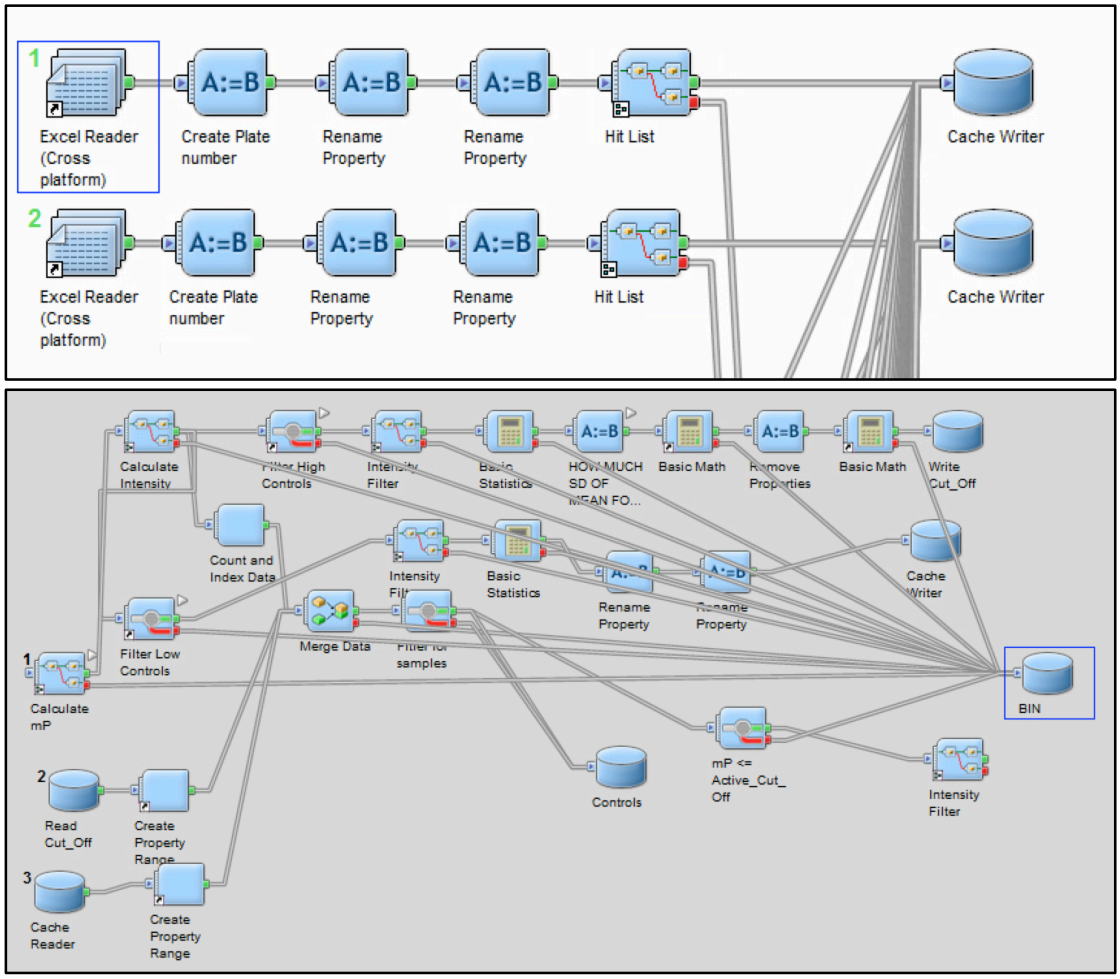


Figure 50: Pipeline protocol written to identify active compounds from the screening data. Top: A section of the top level of the protocol inducing the node 'Hit List' for identifying hits from the assay. Bottom: Subprotocol level within the 'Hit List' generation node showing all nodes used to identify hits. Data flows from left to right along interconnecting grey lines. Square nodes represent data processes and cylinders represent data storage caches.

Following implementation of the protocol, two lists were generated giving 44 compounds identified at 3 SD, of those 44 compounds 23 hits also made the 4 SD cut off. This gave the assay a 0.33 % hit rate at 3 SD and 0.17 % at 4 SD (Appendix 2).

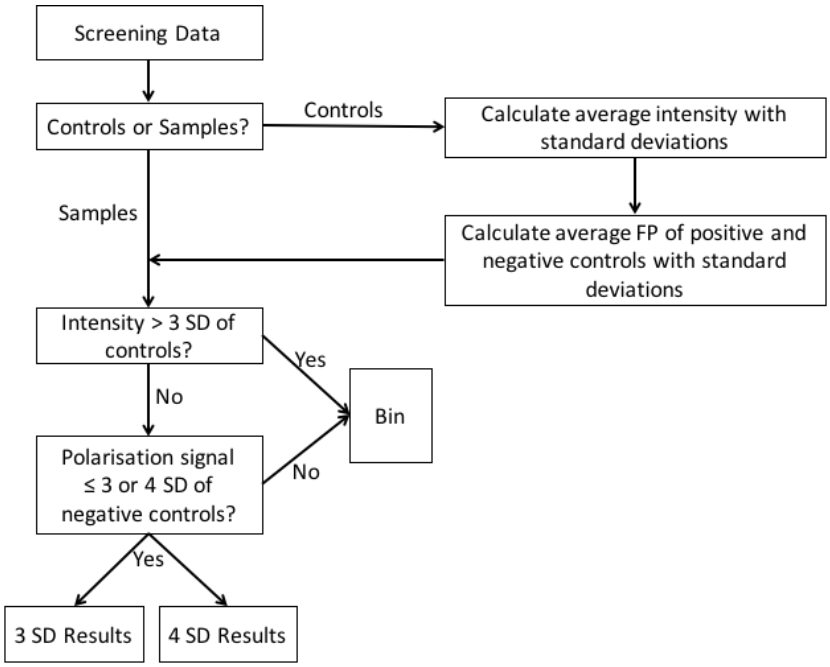


Figure 51: Flowchart to show a simplified view of how screening data is processed.

Due to the intellectual property relating to the chemical structure being owned by MRCT, the results from the screen were sent to collaborators within the MRCT so hit structures could be analysed. Ten compounds were selected and resupplied by the MRCT, along with chemical structure information. Selection was based on structural diversity of the hits, and also availability of the compounds.

4.3.1.4 Conformation screen, dose response fluorescence polarisation assay.

The 10 compounds selected for resupply were then tested in a multiple point dose response assay. This was to check that compounds selected within the screen were true hits, results relating to assay error were eliminated, and IC_{50}'s of the hits could be assessed.

Only one compound out of the 10 resupplied was seen to be active in the assay, MRT00160026. Whilst concentrations could not be reached to produce a full dose response, the IC_{50} could be calculated though constraining the curve to the FP value of the peptide only. This compound was therefore calculated to have an IC_{50} of 180 μM (Figure 52). The compound was resynthesised and found to remain active.

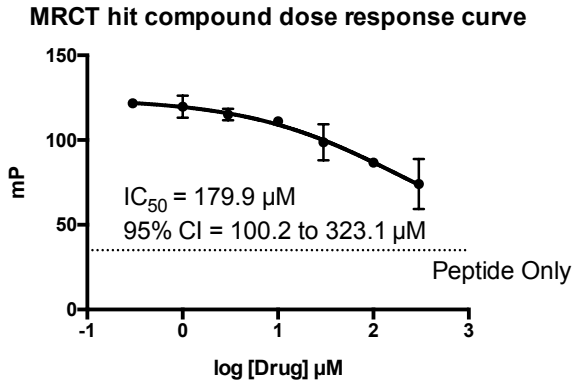


Figure 52: Dose response fluorescence polarisation assay for MRCT00160026. Average +/- SD of fluorescence polarisation measured in millipolarisation units (mP) is shown on the y-axis versus the log of the compound concentration in µM on the x-axis (N=3). IC_{50} values with 95 % confidence intervals are indicated on the graph along with the fluorescence polarisation signal for the peptide only control wells.

4.3.1.5 Testing of structurally similar analogues of MRT00160026

A search for commercially available structural analogues of MRT00160026 was performed using the online tool Scifinder and a number of analogues were purchased for testing to see if any structure activity relationship could be gained around the compound. (Appendix 3)

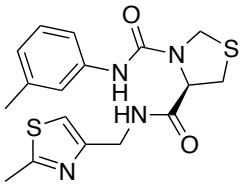


Figure 53: Chemical structure of MRT00160026

These 14 analogues were tested and all were found to have no activity in the fluorescence polarisation assay (Figure 54).

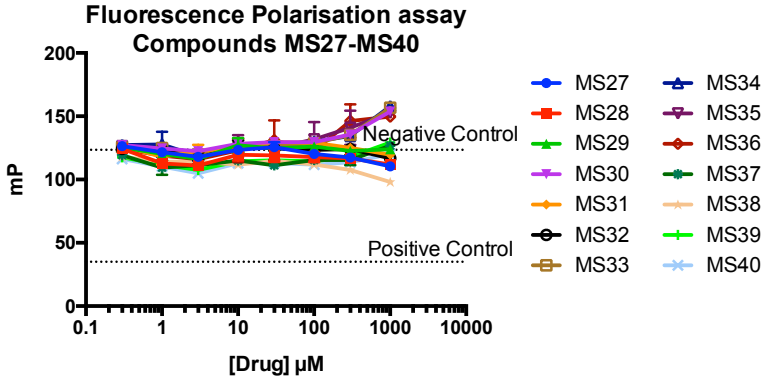


Figure 54: Fluorescence polarisation assay to test MRCT analogues for potential inhibition of fluorescently labelled Nrf2 peptide to Keap1. Y-axis shows polarisation in mP with error in standard deviation (N=3). Negative (DMSO), and positive (30 µM Hit1) controls are shown as indicated.

4.3.1.6 ARE reporter cell line assay

The ARE assay is made up of two constructs stably transfected into a Chinese Hamster Ovary (CHO) or a C6 (rat astrocyte) cell lines. The two constructs both contain a TK-eGFP reporter construct that consists of a 123bp thymidine kinase promoter inserted in the multiple cloning site of peGFP (Clontech). The control cell line contains this construct alone (TK-eGFP) whereas the Nrf2 reporter construct (ARE-TK-eGFP) also contains four repeats of a 41bp GST ARE motif (TAGCTTGGAAATGACATTGCTAATCGTGACAAAGCAACTTT) 3' to the TK promoter. These cell lines were created in Sheffield prior to the project for use in a screen for CNS penetrant activators of Nrf2.[218]

All of the molecules were tested in the C6 cell line ARE reporter assay for a dose dependent response with BG12 (DMF/Tecfidera®), a known reactive Nrf2 inducer being included as a positive control. Any molecules with a significant increase in fluorescent intensity at the final concentration would then be tested alongside the ARE negative cellular line (TK-eGFP) to assess whether any effect seen was ARE dependent. Figure 55 shows that none of the compounds apart from the control were able to induce a meaningful ARE response.

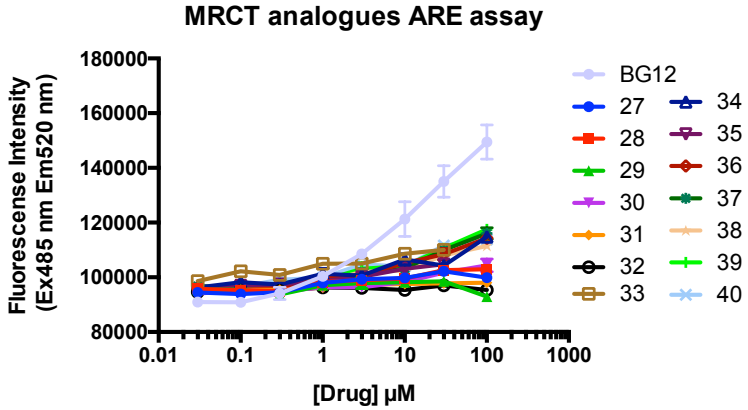


Figure 55: Cellular ARE assay data for the MRCT compound structural analogues. A positive assay response is shown by BG12. Y-axis indicates fluorescent intensity (Ex485 nm, Em520 nm) error plotted as +/- standard deviation (N=3). X-axis indicates compound concentration in µM.

4.3.1.7 MRCT screen conclusions

In conclusion, a 13372 compound library predicted to inhibit protein-protein interactions was screened. A data processing pipeline was developed to quickly process screening data, and can be adapted to any future HTS to enable rapid analysis of data.

Using the data pipelining program compounds with an inhibition signal of 3SD below the controls were selected with assay interfering compounds filtered out prior to

selection. From this 44 compounds were identified and from these 10 compounds were selected and resupplied for testing in a full dose response curve by MRCT.

From these compounds, one showed the ability to inhibit binding of Nrf2 to Keap1 but at a low IC_{50} of 180 μM. Structural similar analogues were then purchased and tested in the fluorescence polarisation assay and the cellular reporter ARE assay to establish if any basic structure activity relationship could be established around the compound. These compounds showed no activity in the FP assay suggesting that the identified compound may be a singleton in its structural class, with very tight SAR. No activation was seen with any molecule in the cellular reporter assay either. However, as no activity was seen in the fluorescence polarisation assay at concentrations normally required to see a response in the cellular assay. A drop off in activity is traditionally seen when using cellular assays due to factors such as drug metabolism and cellular penetration, so it is therefore not unexpected that compounds with little or no activity in biochemical assays show no activity in the cellular assay. The mild affinity of the single compound for the target, with no ability for medicinal chemistry to develop the compound further, makes further development of the compound very difficult. Therefore, additional, larger libraries should be screened to facilitate the identification of structural class with a greater ability for medicinal chemistry development.

The small hit conformation rate achieved by this screen is disappointing, with only one compound out of the 44 that met the 3SD cut off showing activity. However, such high throughput screens always have unavoidable noise within the assay, and whilst avoided through using assay validation techniques such as measuring Z factor, ultimately there are always some compounds that do not confirm activity seen in the HTS. In the case of this assay a higher rate may have been achieved if more than 10 out of the 44 compounds could have been tested in the confirmation assay. Additionally, repeats of the single point screen could have further reduced the false positive rate within the screen. However, this is not practical due to the time, effort and cost associated with even a single high throughput screen and such screens are often only done in singleton even in large pharmaceutical companies with large amounts of capital.

4.2.2 Screening of European Lead Factory Library

Following on from the MRCT screen where no promising hits were identified we were able to screen a second larger library of 318132 compounds under the European Lead Factory (ELF) program. However, due to the size of the library it was not feasible to perform the screen in house. Therefore, the screen was performed externally at the European screening centre (Oss, The Netherlands). The assay was transferred and

optimised for the equipment in Oss using the same method as described in Chapter 3 under my direct instruction from Sheffield. Staff at the European screening centre executed the screen with data analysis performed under my direct supervision.

Protein synthesis for the screen was performed by the ELF at the structural genomics consortium (SGC) in Oxford.

Resynthesis of compounds, running of the MST and redox assays were performed and carried out by the ELF at the European Screening Centre in Newhouse, Scotland.

All other work within the screen including dose response fluorescence polarisation assays, cellular assays, ITC assays, *in-silico* work and analysis were performed by myself at Sheffield.

4.2.2.1 Peptide

A fresh batch of ETGE fluorescein C-terminal labelled peptide was synthesised and purchased for the screen to ensure there was enough of the same batch to perform the whole screen and subsequent analysis with a single batch. This peptide was titrated on the uHTS screening equipment and compared to the previous batch used previously in-house. The two batches were found to have similar profiles, and concentrations of 3 and 5 nM were selected for testing with protein. Interestingly with the new platereader, increases in fluorescence polarisation were observed at increased concentrations. However, this was believed to be due to the narrower florescence window on the different platereader as when the fluorescence signal begins to saturate the detector, artifactual increases in FP signal can be observed, which can also be seen at low concentrations as the detector struggles to pick up the fluorescence signal.

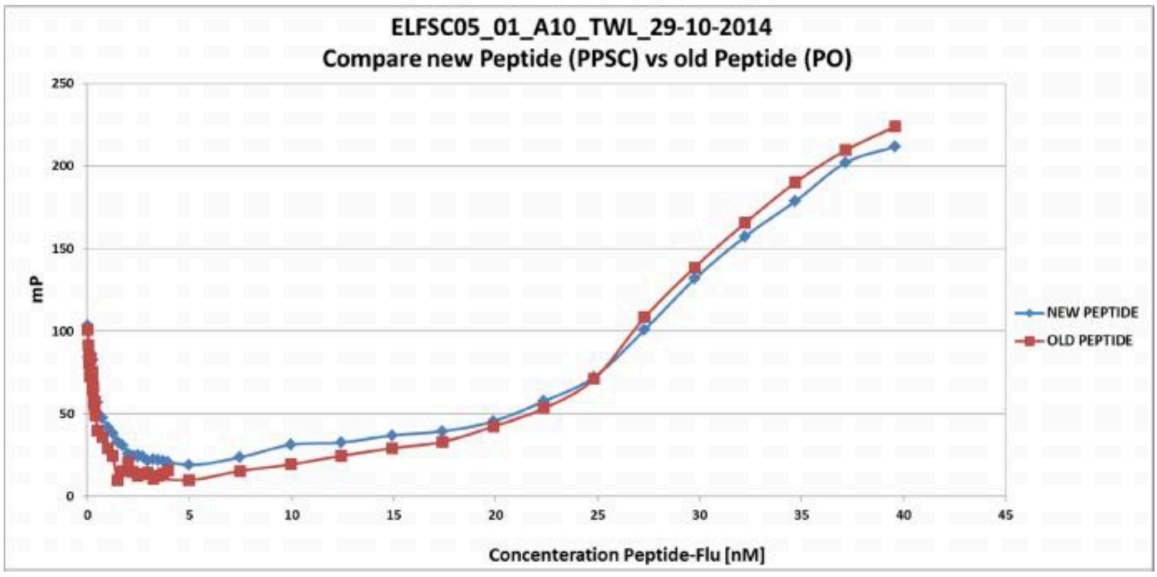


Figure 56: Comparison between the new batch of peptide and the old batch of peptide used in the MRCT screen.

4.2.2.2 Protein

Due to the increased demand for protein needed within the screening and follow up experiments (which was calculated to be around 15 mg) Baculovirus expressed protein was established and produced by the SGC in Oxford on behalf of the project. A variety of different constructs were trialled in baculovirus and mammalian expression vectors to evaluate which gave the optimal expression level. The following constructs were evaluated with all constructs containing a TEV cleavable 10xHIS-FLAG tag (full sequences can be found in Appendix 4):

- A1 – Full length
- A2 – 6 amino acid C-terminal truncation (all domains intact)
- A3 – 47 amino acid N-terminal truncation (all domains intact)
- A4 – 47 amino acids at the N-terminus and 6 amino acids at the C terminus (all domains intact)
- A5 – 177 amino acid N-terminal truncation (no BTB domain)
- A6 – 177 amino acid N-terminal and 6 amino acid C-terminal truncations (no BTB domain)
- A7-A12 same constructs as above in mammalian expression vectors

The expression was evaluated *in-baculo* and compared to mammalian cell expression (Figure 57). The expression constructs A2, A3 and A4 were most expressed in both systems. The A2 construct was selected due to the large increase in protein expression seen for a minor truncation of only 6 amino acids. The baculovirus expression system

was selected for use due to each expression and purification run able to produce more protein which would cut down on lead times for expression. Baculovirus expression systems as mentioned earlier contain post-translational modification machinery meaning the desired structural form of the protein should be maintained.

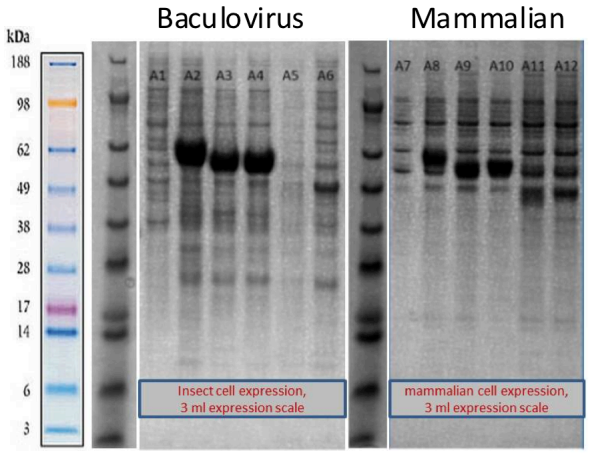


Figure 57: Coomassie stained SDS-PAGE gel showing protein expression of lysates of different expression vectors for Keap1 protein. Molecular weights markers indicated on the left of each gel. Left panel represents protein expressed in a baculovirus expression system and the right panel represents protein expressed in a mammalian expression system. (Data provided from the SGC in Oxford).

Once synthesised, this new protein was then tested for activity in the FP assay to check it had comparable activity to the mammalian synthesised batches used in previous screens. This was done first on the screening equipment at Sheffield which used the BMG Labtech Pherastar FS plate reader and then in Oss on the uHTS equipment which used the Perkin Elmer EnVision plate reader. The protein was found to be slightly less active than mammalian expressed protein with an EC_{50} about double that seen in 3.2.2.2 but was deemed acceptable. Similar EC_{50}'s could be calculated between different platereaders, even with slightly differing concentrations of peptide showing good continuity of data between sites (Figure 58).

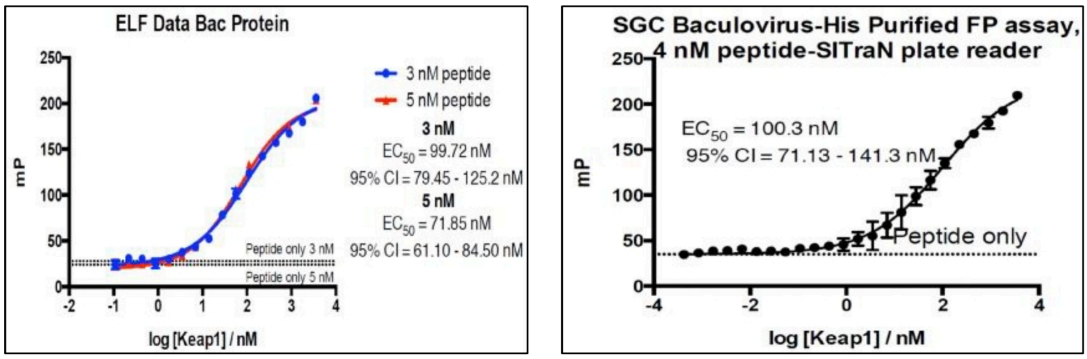


Figure 58: Comparison of protein titration curves performed with the screening batch between Oss (left) and Sheffield (right). Left panel shows titration of baculovirus expressed Keap1 protein in the presence of 3 nM (blue) or 5 nM (red) peptide, EC_{50} values with 95 % confidence intervals shown on the right. Right panel shows the same titration of protein in the presence of 4 nM peptide, with EC_{50} value and 95 % confidence intervals shown above the curve.

4.2.2.3 Optimal plate type, Z prime and assay range determination

Three different plates were trialled; the 1536 variant of the Greiner 384 well plate used in the internal MRCT screen, and two 1536 well plates from Corning, a non-binding variant and a standard variant.

Z prime data showed that all plates gave good Z factors of >0.6 but the Corning plates gave a slightly better value than the Greiner plate (0.78/0.76 vs. 0.71). However, the two varieties of Corning plate showed very similar Z factors (non-binding = 0.78 vs. standard = 0.76).

The assay range (the difference between the positive and negative controls) was also determined for each plate and again the Corning plates showed an increased assay window compared with the Greiner plates (113/112 vs. 99 mP). The more cost effective standard plate was therefore selected for use within the screen.

4.2.2.4 uHTS Screen

Compounds were passed through a triage of screens consisting of a primary screen, multiple hit conformation screens, deselection screens and analytical analysis. The triaging process is detailed in Figure 59.

The primary screen of the 318132-member compound library was successfully screened in two days. The primary screen consisted of a single point screen at a compound concentration of 10 µM. The screen yielded a similar hit rate to the previous screen with ~1000 compounds identified at 3SD giving a 0.3 % hit rate. From this a single point conformation screen was then performed at 10 µM where 377 compounds achieved the set criteria of a Z score of >= 4 and a percentage inhibition of FP signal of > 20 %. These compounds were progressed to full dose response screen and de-selection

assays. The dose response assay was performed in duplicate with the same conditions as the primary fluorescence polarisation screen using a 7-point log dose response curve using 20 μM to 20 nM compound concentrations. From this 318 compounds were able to meet the criteria of a $pEC_{50} \geq 5$ with a Hill slope of between 0.5 and 2.

The de-selection screen, as in the previous MRCT screen, was then performed to remove compounds that were seen as interfering with the assay readout through auto fluorescence. Any compound with a fluorescence intensity of greater than 3 standard deviations away from the controls at 10 μM concentration was not progressed further.

A second deselection screen was also performed on the 318 compounds to remove any compounds that may react with nucleophiles such as the cysteine sulfydryls in Keap1. These compounds were undesirable, as compounds with redox liability are highly likely to demonstrate non-specific inhibition of susceptible targets such as proteins with catalytic cysteine residues or essential redox functionality i.e. NAD/NADH. In addition, this screen had specifically set out to look for non-reactive activators of Nrf2 through inhibition of the Keap1-Nrf2 interaction therefore it was important to deselect for these molecules as they could also activate Nrf2-directed transcription via reaction with sulfhydryl groups on Keap1 (1.4.5.1 Indirect Nrf2 inducers). This was done through the use of a redox assay wich was set up and performed by the ELF team in Newhouse, Scotland. In the assay the compound is first incubated with DTT, any reactive molecules such as those containing Michael acceptors should react with the sulfhydryls in the DTT in a redox reaction. Resaurin is then added. Resaurin is a redox indicator and upon reduction turns from light blue to the highly red fluorescent Resorufin. Any compound that reacts with DTT reduces the concentration of free DTT and will decreases amount of Resaurin reduced to Resorufin, therefore giving a reduced fluorescent signal. Compound 9,10- Phenanthrenequinone was used as a positive control in the assay.

There were 20 compounds which met the selection criteria of a pIC_{50} in the redox assay of < 4.7 and showed acceptable fluorescent intensities as described above. These compounds were then subjected to an analytical purity assessment whereby only compounds with a purity of >70 % by LCMS analysis were resynthesised. The structures of the final 8 compounds were then disclosed.

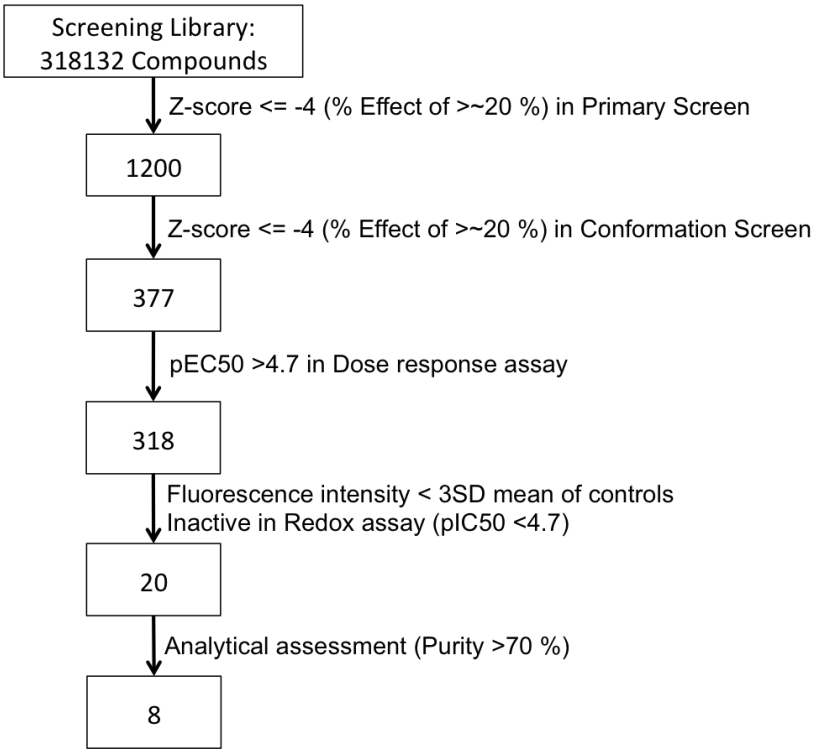


Figure 59: Flow chart detailing the triage of compounds through the screening cascade.

4.2.2.5 Analysis of Hits

From the structures (Figure 60) it can be seen that two compounds ELF1 and ELF2 have high structural similarity and therefore can be clustered together. All other compounds are structurally diverse from each other so can be considered as singletons.

The structural properties of each compound can also be analysed to look at whether any of the compounds possess the desired properties to be a brain penetrant compound, such as a low polar surface area of less than 80 and a logP of between 1.5 and 2.5. These properties and others can also be used to analyse how drug-like the compounds are.

Table 12: Illustrates the molecular properties of the hits, compound number is listed along the top with various molecular properties described down the side. Where MWt = molecular weight of the compound, AlogP = atomic partition coefficient, clogD = calculated distribution coefficient, HBA = number of hydrogen bond acceptors, HBD = number of hydrogen bond donors, tPSA = topological polar surface area, QED = quantitative estimate of drug likeness.

	ELF1	ELF2	ELF3	ELF4	ELF5	ELF6	ELF7	ELF8
MWt	461	475	229	354	353	383	349	334
AlogP	2.81	2.01	3.37	3.59	3.46	2.28	4.63	3.21
clogD	1.81	2.01	2.75	2.31	3.46	2.28	3.17	2.51
HBA	7	8	3	5	5	6	4	5
HBD	1	1	0	0	1	0	1	1
tPSA	81.3	98.4	19.4	25.9	92.4	114.2	57.6	72.8
QED	0.69	0.66	0.79	0.79	0.76	0.74	0.79	0.85

From Table 12 it can be seen that a few of the molecules have properties that, whilst are not perfectly within the criteria specified for a brain penetrant compound, do contain some of the key properties, showing potential for development. Many of the molecules whilst not containing all the properties for a brain penetrant molecule do contain many of the physiochemical properties that are desired for a lead drug candidate. All of the compounds fit within Lipinski's traditional rule of 5, with molecular weights of less than 500 Da, logP's of less than 5, number of hydrogen bond donors being less than 5 and number of hydrogen bond acceptors being less than 10. Additionally, many of the molecules score reasonably highly on the QED measure, an estimate of drug likeness as proposed by Bickerton et. al. in 2012 (with a compound that a 'perfect drug' having a score of 1).[276]

Furthermore, many of the molecules also lack the problematic carboxylates seen in previously published inhibitors that look to contribute to their lack of cellular potency, this may mean the molecules may be easier to optimise. Whilst very rarely does a hit compound have all the properties properties for a lead candidate the molecules are now able to be tuned using medicinal chemistry approaches to explore the SAR around the compounds and press towards a lead candidate.

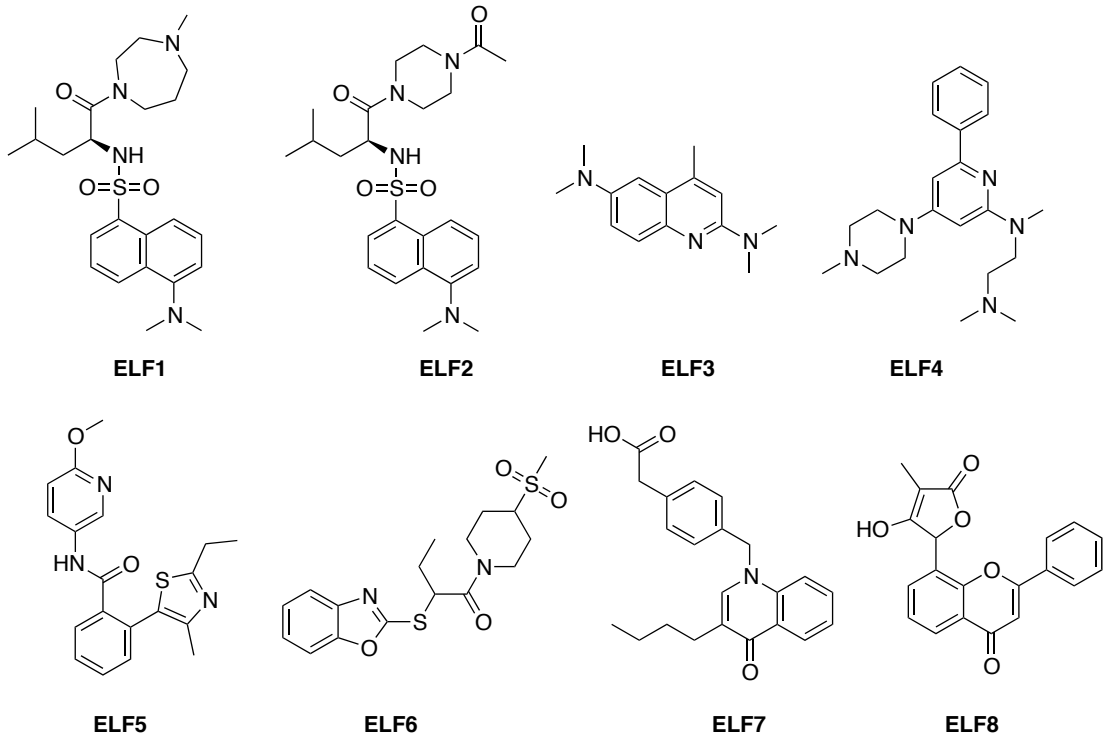


Figure 60: Structures of hits from the ELF screen.

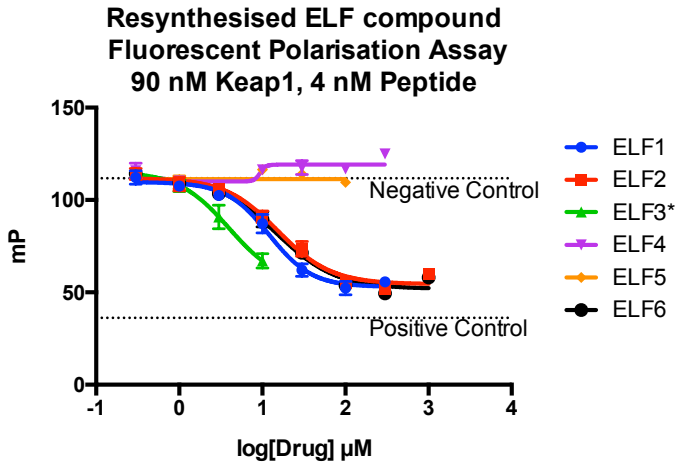
4.2.2.6 Analysis of resynthesised molecules

Due to synthesis of ELF7 and ELF8 requiring multi-step synthesis only compounds 1-6 were successfully re-synthesised by the medicinal chemistry team at the University of Dundee. The compounds, once synthesised, were then sent to Sheffield for analysis in various assays.

4.2.2.6.1 Fluorescence Polarisation Assay

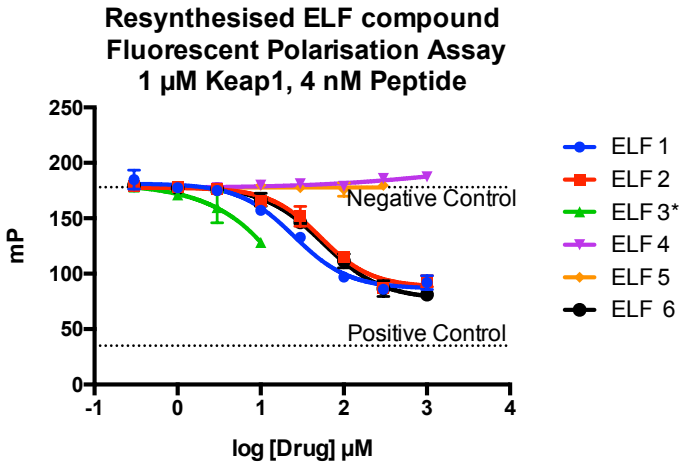
The resynthesised compounds were tested at Sheffield in the fluorescence polarisation assay to confirm any activity seen from the compound DMSO stock in the screening plate. The compounds were tested at the standard assay protein concentration of 90 nM and also at a protein concentration of 1 μM in order to determine an estimate for the K_D of the compounds as described in Chapter 3.

Figure 61 shows how compounds ELF1, ELF2 and ELF6 are able to successfully compete off the Nrf2 ETGE peptide for binding to Keap1. All show reasonably good IC_{50}'s for compounds selected from a HTS. ELF1 has an IC_{50} of 12 μM, with ELF 2 slightly lower at 15.5 μM. Both compounds within the same structural series showing activity is encouraging as it suggests that there is space for chemical development of this series through a hit to lead process. ELF 6 was also able to successfully compete for binding of ETGE peptide in the assay with an IC_{50} value of 14.5 μM.



	ELF1	ELF2	ELF3	ELF4	ELF5	ELF6
IC50 / µM	12.02	15.53	n.d	n.d	n.d	14.49
95 % CI / µM	9.778 to 14.78	11.98 to 20.13	n.d	n.d	n.d	11.14 to 18.84

*Compound Fluorescent



	ELF1	ELF2	ELF3	ELF4	ELF5	ELF6
IC50 / µM	25.11	52.74	n.d	n.d	n.d	55.62
95 % CI / µM	19.40 to 32.49	41.87 to 66.44	n.d	n.d	n.d	45.94 to 67.34

*Compound Fluorescent

Figure 61: Graphs illustrating dose response curves for the 6 resynthesised ELF compounds. Left graph illustrates the dose response curve at a Keap1 concentration of 90 nM (screening concentration). Y-axis shows polarisation response in mP with +/- standard deviation shown (N=3), and X-axis show the concentration of compound (µM) in a log scale. IC_{50} and 95 % CI intervals are shown below the graph for compounds where an accurate curve could be fitted to the data. Right graph illustrates the dose response curve at a Keap1 concentration of 1 µM (for K_D determination). Y-axis shows polarisation response in mP with standard deviation shown (n=3), and X-axis show the concentration of compound (µM) in a log scale. IC_{50} and 95 % CI intervals are shown below the graph for compounds where an accurate curve could be fitted to the data.

K_D's were then calculated from the data using Keap1 at 1 μM concentration and methods described in Chapter 3.

ELF1 = 1240 nM

ELF2 = 2660 nM

ELF6 = 2810 nM

The K_D's calculated indicate that ELF 2 and ELF 6 have K_D values roughly twice that of ELF 1.

4.2.2.6.2 Cellular ARE reporter assay

All 6 of the resynthesized compounds were tested in both the CHO and C6 cellular reporter lines for ARE expression.

In the CHO cell line shown in Figure 62 ELF2, ELF3, ELF5 and ELF6 show an increase in fluorescence in the ARE-TK-eGFP cells. ELF3 however also shows an increase in fluorescence in the TK-eGFP control cell line showing that the compound is fluorescent at the wavelengths used as seen in the FP assay (4.2.2.6.1) where it interfered with the assay. ELF6 shows the largest increase in fluorescence from those tested. As ELF1 is structurally similar to ELF2 it was unusual to observe that there was no activity shown with ELF1. However, some toxicity was beginning to be observed at the top concentration of ELF1 which can lead to a reduction in the total amount of fluorescence observed due to decreased cell number. Additional properties that can affect cellular activity include cellular penetration, solubility, stability and metabolism of the compound.

The same assay replicated in C6 cells (Figure 62) showed a slightly different picture. ELF1 was able to show a small amount of activity in the C6 cells with ELF2 and ELF6 showing a smaller increase at the top concentration. ELF3 also showed its fluorescent properties in this assay.

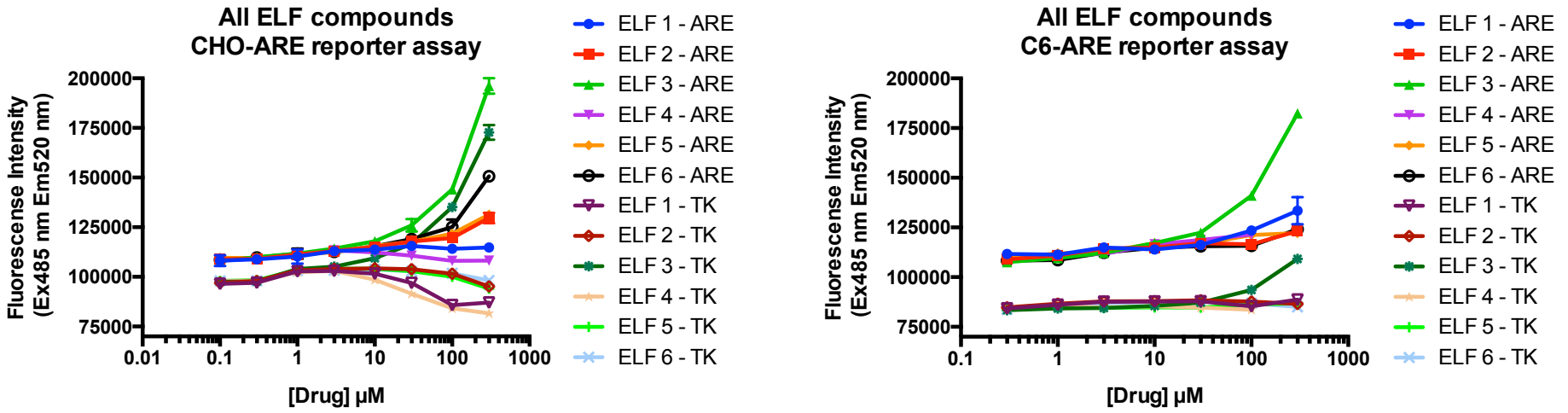
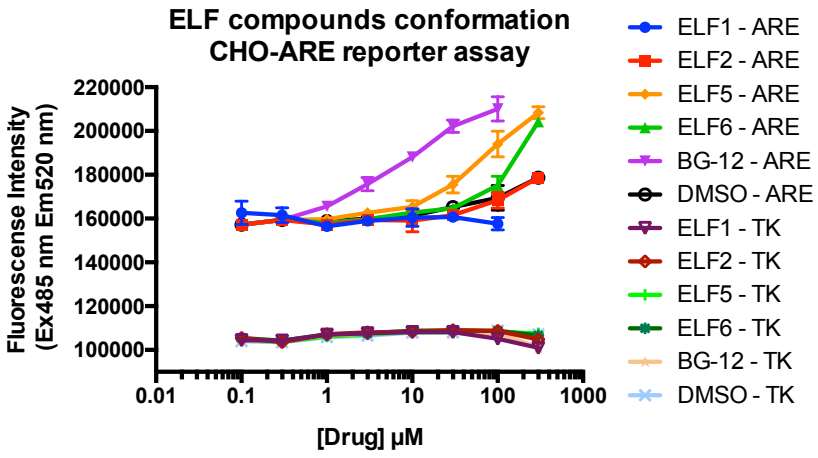
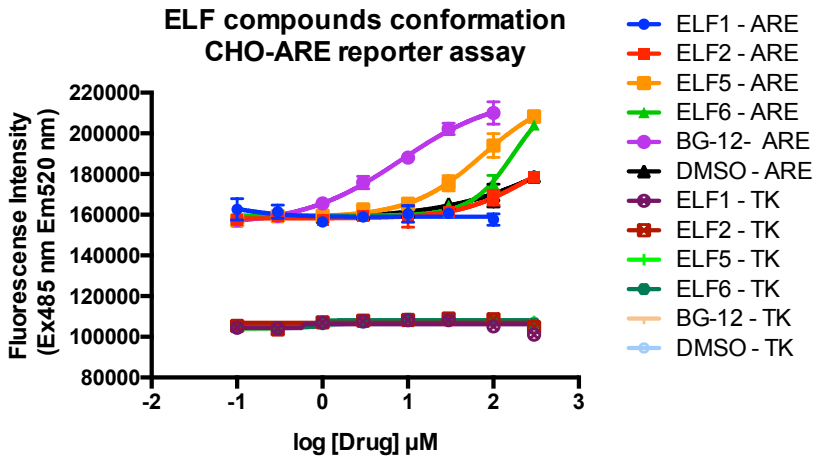


Figure 62: Graphs illustrating the results of the ELF compounds in the ARE reporter assay in CHO cells (left), and C6 cells (right). Legend on right of each graph indicates the compound number and cell line "ARE" is 4xARE-TK-eGFP cell line for ARE-Nrf2 reporter activity. "TK" is TK-eGFP control cell line with no AREs for Nrf2 to bind the promoter. Y-axis indicates fluorescent intensity (Ex485 nm, Em520 nm) error plotted as standard deviation (N=3). X-axis indicates compound concentration.





	ELF1	ELF	ELF3	ELF4	ELF5	ELF6	BG12
EC50 / µM	n.d	228.8	n.d	n.d	75.2	171.5	8.249
95 % CI / µM	n.d	9.660 to 5420	n.d	n.d	40.92 to 138.2	64.88 to 453.5	5.050 to 13.47

Figure 63: Hit compounds activate Nrf2-ARE directed transcription. Graphs illustrate two repeats of hit compounds in the assay. Y-axis indicates fluorescent intensity (Ex485 nm, Em520 nm) error plotted as standard deviation (N=3). X-axis indicates compound concentration. Hit compounds shown are the BG12 control, ELF6, ELF5, and marginally ELF2. EC_{50} values are shown in the table below.

The compounds active in either the CHO or C6 assay were repeated in the CHO assay (shown in Figure 63, n=3) as this assay saw the greatest response from the compounds and would therefore make determination of an EC_{50} value possible. A DMSO titration control was included due to the high concentrations of compound being used (because of a lack of observed activity in these compounds) leading to new concentrations of DMSO previously not tested for in the assay. In addition, BG12 was included in the assay as a positive control.

The EC_{50} values were determined for all compounds that showed an effect. ELF2 had an EC_{50} value of 229 μM, ELF6 a value of 171.5 μM, and ELF5 a value of 75.2 μM. The control compound BG12 showed an EC_{50} value of 8.2 μM within the compound's normally observed EC_{50} of ~10 μM.

Interestingly the DMSO showed an increase at the final concentration similar to that of ELF2 suggesting that the DMSO was able to have an effect on ARE activation, an effect that was not seen in the TK control cell line. This also leads to the suggestion that the effect seen for ELF2 was simply the effect of the DMSO and may not be due to the compound itself giving it parity to its structural partner ELF1, which showed no activity.

ELF5 also was fairly interesting as it showed an activation within the cellular assay but failed to show activity in the FP assay where some fluorescence interference being seen at higher concentrations in the FP assay lead to wells being excluded from the dataset. However, no fluorescence interference was seen within the cellular assay. It will therefore be interesting to see whether this compound will show binding to Keap1 in a biophysical assay. Potentially it could be working in the cellular assay via a non-Keap1 mediated activation of Nrf2 through the GSK-3β pathway which would also be picked up using this assay.

4.2.2.6.3 Microscale Thermophoresis (MST)

Microscale thermophoresis (MST) uses a laser to locally heat a capillary containing a 4 μl sample of binding partner molecules (e.g. target protein and compound) in aqueous solution (Figure 64). These molecules will move differently though the temperature gradient depending of their charge, size and hydration shell. Upon binding of a molecule to its receptor the receptor-ligand complex is formed, this has a different movement velocity compared to free unbound molecules. By monitoring the difference in velocities the K_D of the interaction between the binding partners can be determined.

Keap1 contains 5 tryptophan residues, which have intrinsic fluorescence. These residues can be utilised for label free MST experiments using the excitation/emission wavelengths Ex 280 nm, Em 350 nM.

All 6 ELF compounds were tested by the ELF at the screening centre in Newhouse, alongside the control compound Hit1 (3.2.3.2) which had been shown in the literature to have a K_D value of 1.9 μM.[211]

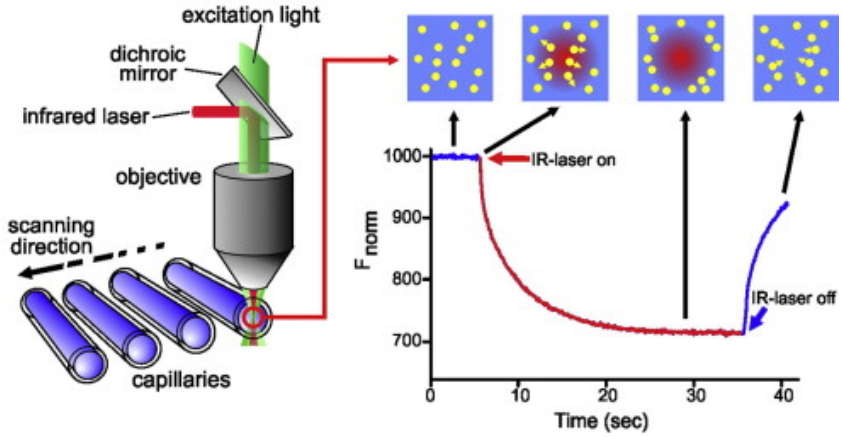


Figure 64: How a MST machine functions. Left image illustrates the instrumental workings of a MST machine. A focused IR-Laser is used to locally heat a 4 µL volume of sample. The fluorescence within the capillary is excited and emission of the sample is detected through the thermophoresis of the fluorescent molecules through the temperature gradient. The right image represents the output of a typical MST experiment. An initial fluorescence value is obtained from when the molecules are homogeneously distributed (Time = 0 sec). Then the IR laser is activated (Time = 5 sec) and thermophoretic movement of the fluorescent molecules is observed, leading to a decrease in molecules in the locality of the laser. Once the laser is deactivated (Time = 35 sec) reverse diffusion of the molecules occurs as the area locally heated by the laser cools and a homogeneous distribution resumes. Molecules bound in complexes produce different movement velocities when compared with unbound/free molecules. Reproduced with permission from Jerabek-Willemsen et. al. (2014).[277]

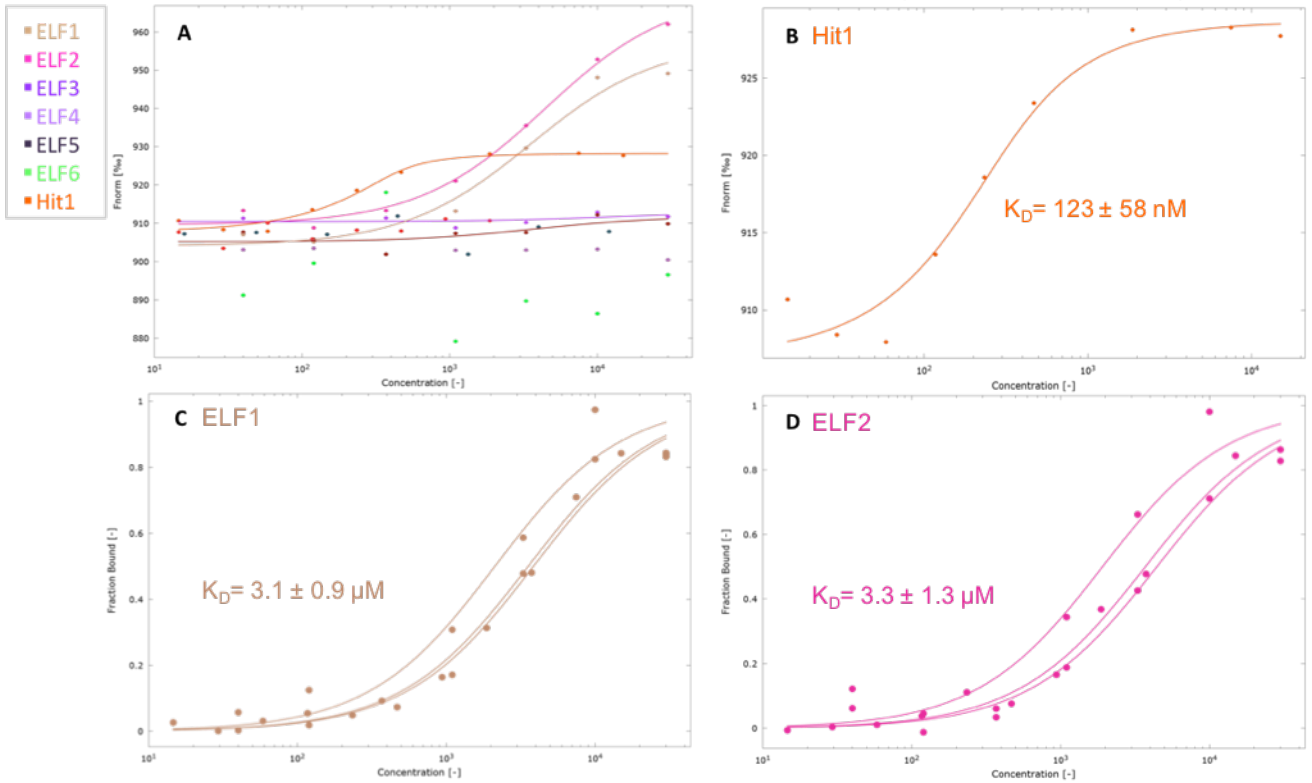


Figure 65: Illustrates MST titration curves for various compounds against a 200 nM concentration of Keap1 protein in 50mM HEPES, 300mM NaCl, pH 7.5. A: Illustrates MST titration curves for all compounds ELF1-6 as well as the control compound Hit1, legend is shown to the left of the panel. X-axis shows concentration of ligand in molar, Y-axis indicates the normalised fluorescence (F_{norm} = F_{hot} (fluorescence before laser is activated)/F_{cold} (fluorescence after laser is activated)). B: Shows a representative titration curve for Hit1 K_D = 123 nM ± 58 nM (N=5). X-axis shows concentration of ligand in molar, Y-axis indicates the normalised fluorescence (F_{norm} = F_{hot}/F_{cold}). C: Illustrates 3 replicate titration curves for compound ELF1. X-axis shows concentration of ligand in molar, Y-axis indicates the amount of fraction bound. K_D = 3.1 ± 0.9 µM D: Illustrates 3 replicate titration curves of for compound ELF2. X-axis shows concentration of ligand in molar, Y-axis indicates the amount of fraction bound. K_D = 3.3 ± 1.3 µM.

From Figure 65 it can be seen that compounds ELF1, ELF2 and the control Hit1 show binding to Keap1, whilst ELF3, ELF4, ELF5 and ELF 6 do not show any binding to Keap1. Interestingly, ELF6 shows an unusual profile with a greater scatter of points when compared to other bound or unbound compounds, this is a known characteristic of aggregation and may be why ELF6 shows a greater scatter of points.

The K_D values for ELF1, ELF2 and Hit1 were then determined. Hit1 showed a K_D of 123 ± 58 nM. This value is around 15 fold less than the literature value of 1.9 μM which was measured using surface plasmon resonance (SPR) and just the isolated kelch domain, which could account for the differences in observed K_D. ELF1 and ELF2 had largely similar K_D values of 3.1 ± 0.9 μM and 3.3 ± 1.3 μM respectively. This gives good parity when compared to the values determined by the fluorescence polarisation assay (4.2.2.6.1).

After establishing whether the compounds could successfully bind to Keap1, further experiments were performed to establish if the compounds ELF1 and ELF2 could compete for binding with each other. Additionally, in order to validate that the compounds were binding the desired kelch domain of Keap1 and not an allosteric site, the ability of ELF2 to compete for binding with the literature compound Hit1 as well as the ETGE peptide was explored.

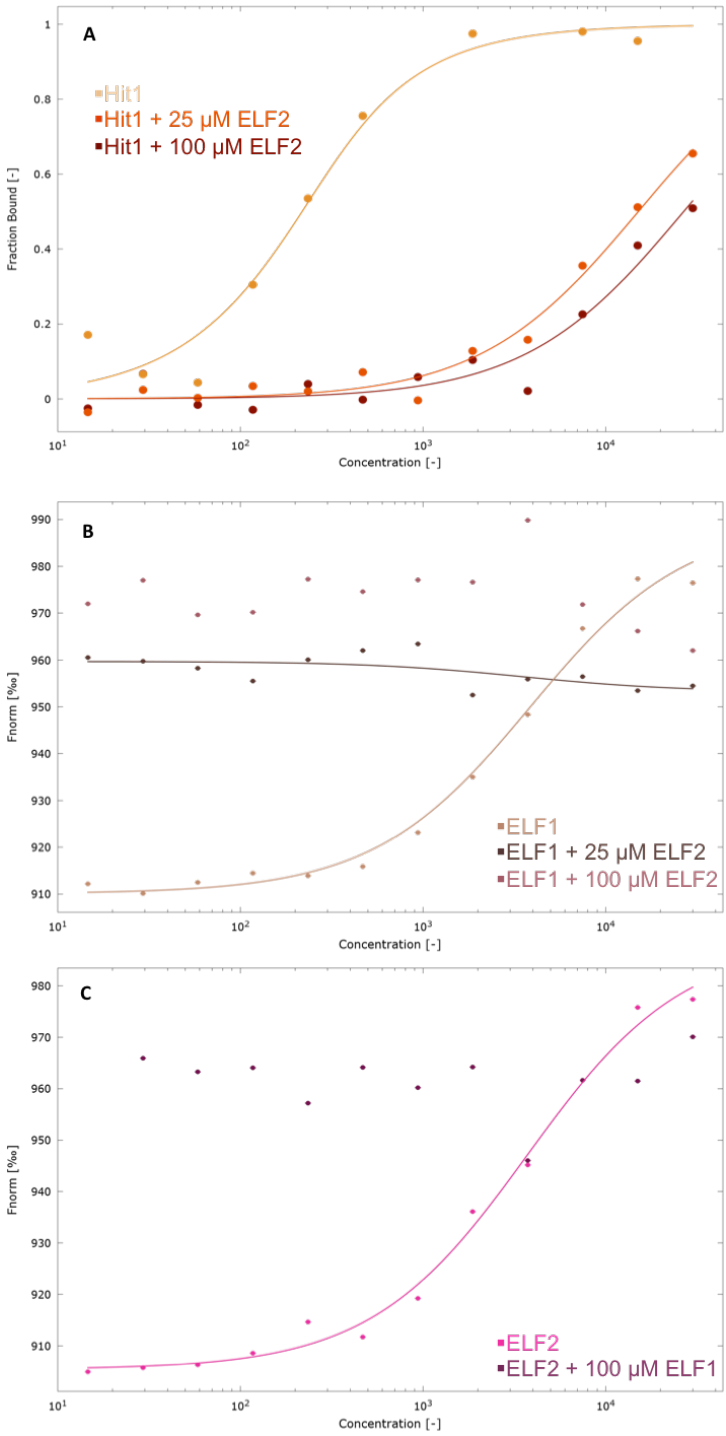


Figure 66: Use of MST to show ELF compounds participate in competitive inhibition of Keap1. A: Illustrates titration dose response curves for Hit1 alone or Hit1 with either a 25 µM or 100 µM concentration of ELF2 added to each point. Curves shifted further to the right indicate an increase in the value for K_D. X-axis shows concentration of ligand in molar, Y-axis indicates the amount of fraction bound. B: Shows the effect of addition of either 25 µM or 100 µM concentrations of ELF2 to a dose response titration of ELF1. X-axis shows concentration of ligand in molar, Y-axis indicates the normalised fluorescence ($F_{norm} = F_{hot}/F_{cold}$). C: Illustrates the effect of addition of a 100 µM concentration of ELF1 to a titration dose response curve of ELF2. X-axis shows concentration of ligand in molar, Y-axis indicates the normalised fluorescence ($F_{norm} = F_{hot}/F_{cold}$).

Figure 66 panel A shows how ELF2 is able to shift the binding curves for Hit1. This indicates that ELF2 is able to compete for binding at the desired binding site within the kelch domain. Furthermore, panel B and C indicates that ELF1 and ELF2 disrupt each others binding curve indicated they are able to compete for the same binding site on the protein.

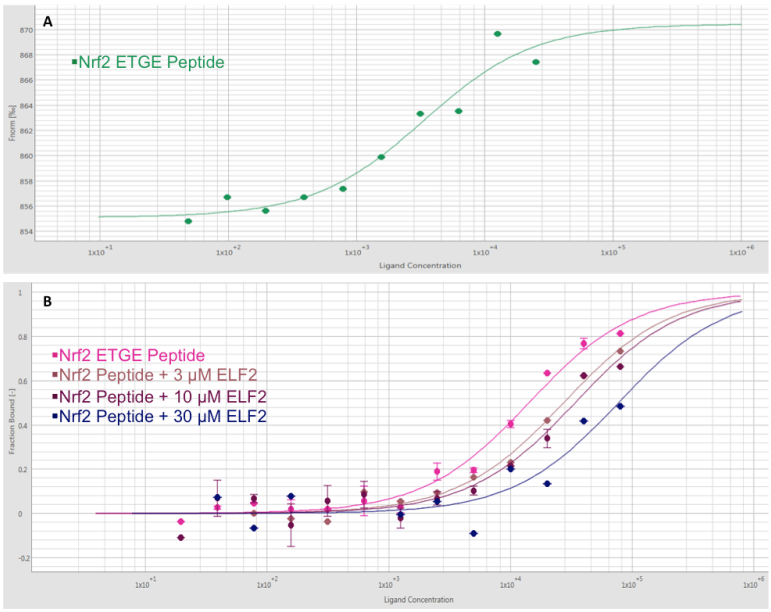


Figure 67: MST assay curves showing binding of the ETGE peptide to Keap1. A: Indicates the titration dose response curve for the unlabelled Nrf2 ETGE peptide H-LDEETGEFL-NH2. X-axis shows concentration of ligand in molar, Y-axis indicates the normalised fluorescence ($F_{norm} = F_{hot}/F_{cold}$). K_D = 3.25 µM B: Shows the effect of addition of 3, 10 and 30 µM concentrations of ELF2 to the titration dose response curve of the Nrf2 ETGE peptide. Curves shifted further to the right indicate an increase in the value for K_D. X-axis shows concentration of ligand in molar, Y-axis indicates the amount of fraction bound.

The ability of the compounds to disrupt binding of the unlabelled peptide and not just the labelled peptide is important, as it indicates that their activity in the fluorescence polarisation assay isn't driven by some artefact related to labelling of the peptide with FITC. Figure 67 shows how compound ELF2 is able to shift the binding curves for the Nrf2 peptide rightwards, giving an increase in the value for K_D. This shows that ELF2 is able to compete for binding of the Nrf2 peptide sequence that it was screened to inhibit. Interestingly the K_D value for the peptide of 3.25 µM is higher than that obtained in the fluorescent polarisation assay of 132 nM and the value shown in the literature of 352 nM.[278] This could be down to differences in technologies, or down to other factors such as buffer compositions.

4.2.2.6.4 Isothermal titration calorimetry (ITC)

Briefly, Isothermal titration calorimetry (ITC) is a technique used to measure the heat exchanges associated with molecular interactions at a constant temperature. It provides a complete thermodynamic profile in a single experiment, including enthalpy changes (ΔH), binding affinity/association constant (K_a) and stoichiometry (n). ITC works though two sample cells, one containing a reference solution and one containing the sample solution (Figure 68). The reference cell is kept at a constant power, whereas the sample cell is kept at a constant temperature with the power required to maintain the constant temperature monitored. Upon injection of a ligand solution to a receptor solution if a binding event occurs an exothermic or endothermic temperature change will occur depending on the type of binding. The amount of power required to maintain a steady state temperature is then monitored.[279]

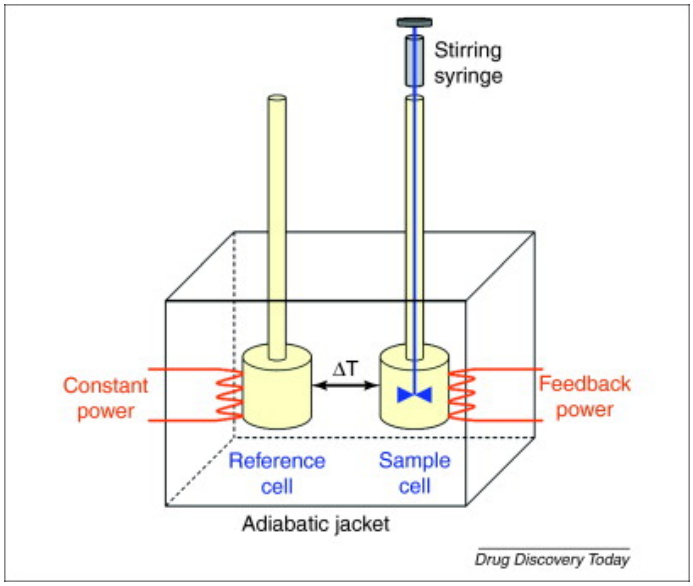


Figure 68: Diagram illustrating the basic layout of an ITC machine. Two cells, a sample cell and a reference cell are located within an adiabatic chamber. Both cells should contain identical buffer systems with the sample cell containing the receptor of interest. Constant power is applied to the reference cell to maintain a constant temperature; the power applied to the sample cell is variable and monitored to maintain a constant temperature. When the ligand is titrated into the sample cell the heat produced (exothermic) or absorbed (endothermic) due to a binding event occurring is monitored through the amount of power required to maintain the steady state temperature. Figure reproduced with permission from Núñez et. al 2012.[279]

Only a limited amount of protein was available for ITC due to the large quantities required for ITC. Therefore, the ETGE peptide was first used for validation of the technique followed by ELF2 as it was the compound which had shown activity in all the assays performed at the time of the experiment. A 1 mM solution of ETGE peptide or ELF2 was titrated into a solution of 59 μM Keap1.

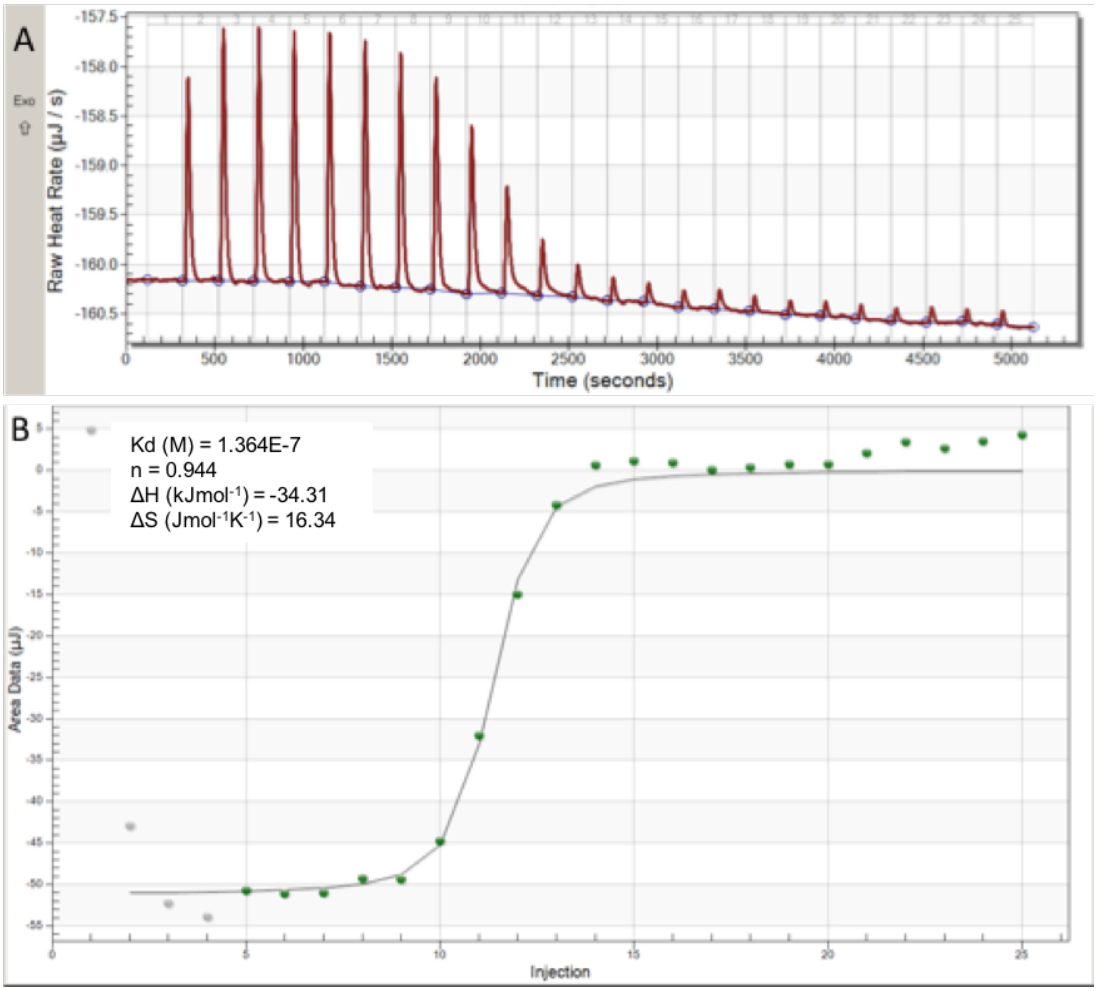


Figure 69: Isothermal titration calorimetry traces for ETGE peptide. 25 1.49 µL injections of 1 mM unlabelled ETGE peptide was titrated into 290 µL of Keap1 protein at 59 µM. Panel A shows the raw blank corrected titration data. X-axis shows the time in seconds, and the Y-axis shows the raw heat rates in µJ/s. Panel B shows the integrated areas of the peaks showing the integrated titration curve calculated K_D, stoichiometry (n), ΔH, and ΔS can be seen in the box to the top left of the image. Injection number is labelled along the X-axis with the area in µJ on the Y-axis.

Figure 69 shows a titration curve observed as the peptide is added to the protein. This shows that binding of the peptide to the protein is occurring and validates the technique can be used to observe binding events. The K_D can be calculated from the curve as 136 nM this is very close to the K_D observed by fluorescence polarisation showing of 132 nM showing excellent parity between the two techniques in this instance. Furthermore, the stoichiometry can be calculated, this can be seen as 0.944 which indicates a 1:1 binding of one Keap1 molecule with one molecule of ETGE peptide, as would be expected. The entropic and enthalpic contributions to binding can also be determined from the curve. The enthalpic contribution can be calculated to be -34.31 $kJmol^{-1}$ and the entropic contribution as 16.34 $Jmol^{-1}K^{-1}$.

However for ELF2, as Figure 70 shows only baseline heats of titration were observed, and no binding event was recorded. This is probably due to the concentration of protein required for a heat change to be observed was in excess of the maximal protein concentration that could be achieved with Keap1 before precipitation of protein was observed.

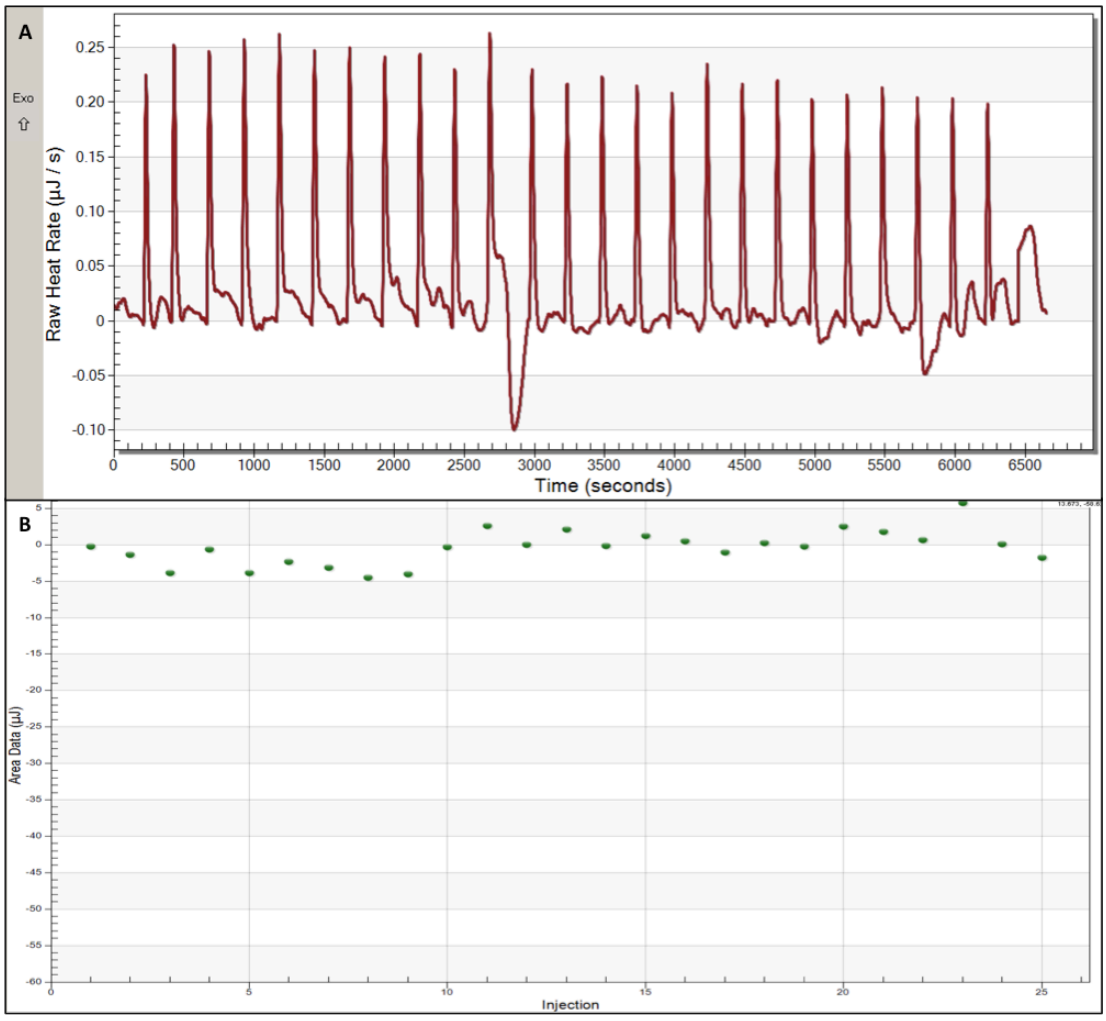


Figure 70: Isothermal titration calorimetry traces for ELF2. 25 1.49 µL injections of 1 mM ELF2 compound was titrated into 290 µL of Keap1 protein at 59 µM. Panel A shows the raw blank corrected titration data. X-axis shows the time in seconds, and the Y-axis shows the raw heat rates in µJ/s. Panel B shows the integrated areas of the peaks showing that no visible titration curve can be seen. Injection number is labelled along the X-axis with the area in µJ on the Y-axis.

4.2.2.6 *in-silico* analysis of hits

Docking was performed on all of the ELF hits as well using compound 16 (1.4.5.2.2) as a control to establish potential binding poses to Keap1. The same protocol was followed as that established in 6.3.2.5 with a flexible docking algorithm used within the GOLD program. The docking poses were scored using the ChemPLP scoring function and compared against the binding poses of compound 16 to establish if any of the same interactions could be made. Interactions of particular interest were those to Arg415, which is thought to be a key hinge binding interaction. Arg483, Ser508, Phe478 and Gly462 are also key to look at as Jiang et. al. have previously showed that this region contributed the most effective binding energy compared to other areas of the binding pocket.[209]

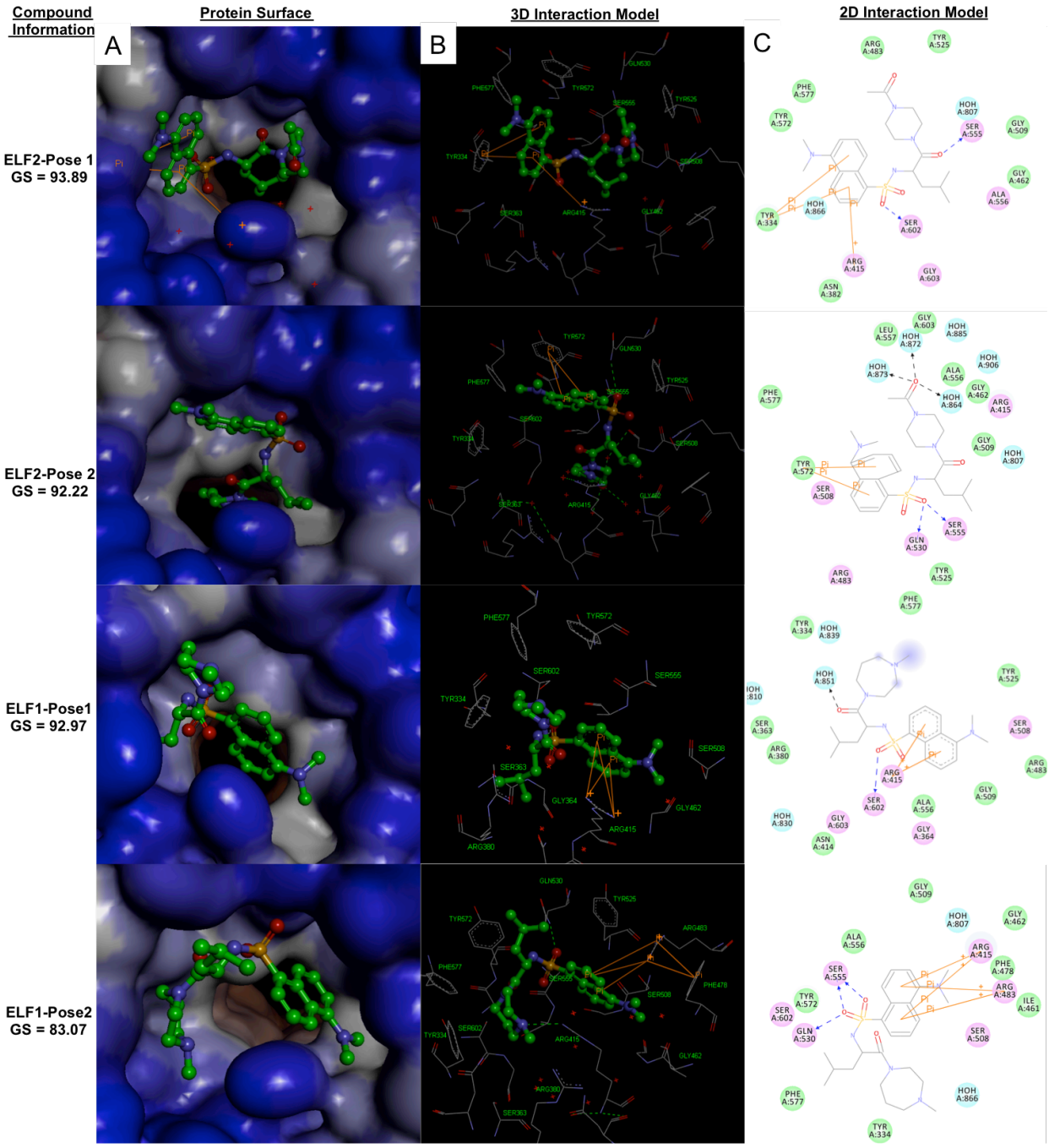


Figure 71: Top two docking poses achieved for ELF1 and ELF2. Column A shows docking poses with the hydrophobic surface of Keap1 a blue to brown gradient shows hydrophobicity with brown being most hydrophobic, and blue most hydrophilic. Column B shows the 3D interaction map of the compounds with key amino acids of Keap1. Hydrogen bond interactions are shown with dashed green lines, π interactions are shown with solid orange lines, and amino acid residues are labelled in green. Column C shows a 2D interaction map of the compound with key amino acids and interactions shown, again hydrogen bond interactions are shown with dashed green lines, π interactions are shown with solid orange lines. Compound ID, pose number and docking score (GS) are also shown.

Figure 71 shows the top two scoring docking poses for each ELF1 and ELF2 respectively. These compounds have shown to be of the most interest to the project, having shown the most promise in the majority of assays tested, and having been the highest scoring docking compounds out of the set, with the exception of ELF7, which was unable to be retested in any of the biological assays due to difficulty in re-synthesis of the compound. The structural series around ELF1 and ELF2 was therefore selected as the lead hit series and were explored in further detail using docking analysis. This enables the exploration of approaches which would enable further development of these compounds.

The top-scoring pose from all of the ELF compounds is for compound ELF2 and is shown in the first row of Figure 71. This pose achieved a docking score of 93.89 compared with the score of the control compound 16 of 92.12. In this pose the isobutane moiety sits within the hydrophobic core of the binding site, where the napthaline moiety sits in compound 16 (Figure 72-A). Using the five different binding areas in the binding site as defined in 6.3.1.8 it can be seen that the isobutane moiety is sitting in the central cavity, with the dansyl group sitting in pocket 3 where it is able to form several π-cation interactions with Tyr334, and Arg 415, this area is the same as that occupied by one of the methoxybenzene groups in compound 16 (Figure 72). The sulphonamide is also able to act as a hydrogen bond acceptor for the hydroxyl of Ser602. Additional hydrogen bonding interactions are formed by the amide carbonyl acting as another HBA with Ser555, where the piperizine moiety occupies pocket 4 of the binding site.

If x-ray crystallography is able to validate this binding pose there are a variety of alterations that could potentially be made to the structure to enhance its potency through additional interactions and displacement of favourable waters.

For example, pocket 2 is not occupied by this docking pose and leaves potential for substitution off either the aromatic ring system or the sulphonamide amine towards the pocket. In particular, an additional hydrogen bonding interactions could be made with Arg380 through the addition of a hydrogen bond acceptor in the vicinity. Water molecules occupying both pockets 1 and 2 also have the potential to be displaced and could give a favourable increase in binding energy. Further interactions can be made through occupation of pocket 1, this could be accessed through alteration of the piperizine ring to extend a hydrogen bond acceptor into the pocket gaining interactions with Arg483. Alteration of the piperizine ring to an aromatic ring may also have potential benefits, as it opens up the potential for π-π interactions with Tyr525 similar to those achieved with the other methoxybenzene ring on compound 16.

The second highest scoring pose for ELF2 shown in the second row of Figure 71 achieved a score of 92.22. However, this docking pose shown isn't as appealing as the top scoring pose due to its lack of interactions with the key Arg415 residue. This lack of the hinge binding interaction gives concern that this pose is not a representative binding pose, although molecules such as that of published compound 15 shown in Figure 17 have been shown to bind weakly without the interaction to Arg415. There are however several interesting features of the pose. In this pose, the piperazine moiety penetrates into the centre of the structure deeper than aromatic molecules are able to achieve. In the pose the terminal amide carbonyl is able to form three salt bridges through three water molecules. However, this docking pose is most likely unachievable in real world situations, as the polar piperazine side chain would have to penetrate through the hydrophobic cavity entrance before achieving any interactions with buried water molecules. Figure 72-B also shows an overlay of the pose with the control compound 16 showing how the carbonyl moiety extends beyond the reach of the naphthalene ring of published compound 16.

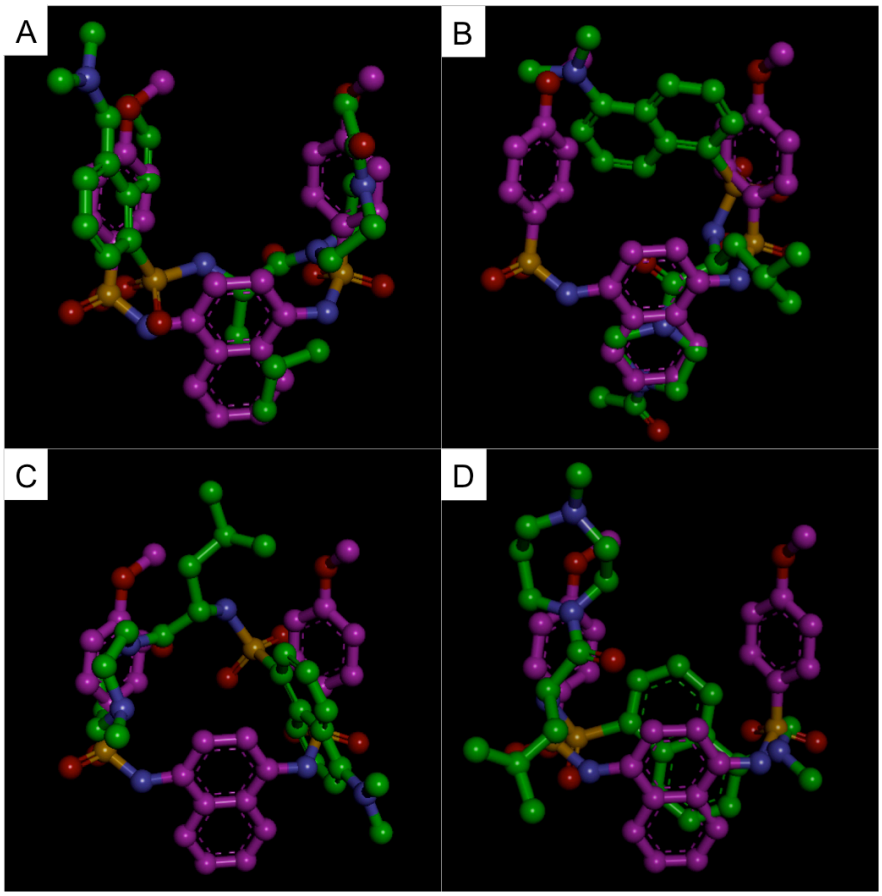


Figure 72: Docking poses superimposed onto the co-crystal structure of compound 16 from PDB ID: 4IQY. A = ELF2 Pose 1, B= ELF2 Pose 2, C = ELF1 Pose 1, D = ELF1 Pose 2. Compound 16 is always illustrated in pink, and the docking pose of interest in green.

The two highest scoring poses for ELF1 are shown in the third and fourth rows of Figure 71. Interestingly these poses have similar binding conformations with slight differences. In both the conformations the dansyl aromatic ring is positioned in the centre of the Keap1 binding pocket with the dimethylamine group pointing towards the hydrophobic pocket 1. However, whereas in pose 1 the aromatic ring is more towards the centre of the binding site, nearer where the napthaline is in compound 16 (Figure 72-D) in pose 2 the ring is slightly askew from the centre and is closer to pocket 1. In pose 2 the orientation of the aromatic ring system enables it to form bridging π-cation interactions between Arg415 and Arg483, whereas in pose 1 π-cation interactions can only be formed with Arg415, more similar to those formed by compound 16.

A single hydrogen bond interaction is made between the sulphonamide and the hydroxyl on Ser602 in pose 1, with an additional interaction shown forming a salt bridge with a water molecule. Pose 2 is able to form 3 hydrogen-bonding interactions between the sulphonamide moiety, which forms two hydrogen bonds to Ser555, and a single one to Gln530.

The potential opportunity for enhancement of activity, should these poses be experimentally determined as the correct orientation, is more limited. This is because the poses already occupy more of the available binding surface. Pose 1 has the most potential for development out of the two poses as the molecule occupies the central pocket more than in pose 2 leaving a larger amount of the outer pockets free. Pose 1 is able to occupy pocket 3 with the piperizine-like moiety and the aliphatic chain extends into pocket 2. Additional interactions could be achieved through adding a hydrogen bond acceptor to the terminus of the aliphatic chain for interactions with Arg380. Further extension from the amine group on the napthaline could lead to molecules able to better extend into pocket 1 more than the methyl group, dipropylamino and dimethyl amino groups could be trialled. Furthermore, alterations to the aromatic ring with the introduction of heteroatoms within the ring could bring additional activity through the alteration of the ring electrostatics, for example ring systems such as quinolone could be trialled, as well as 5,6 systems such as indole, where the nitrogen of the indole located in a similar locale to the dimethylamine group.

As previously stated pose 2 occupies a higher amount of the outer binding surface where additional interactions can be obtained than pose 1. Pose 2 is able to occupy pocket 1 with the aromatic ring system forming two π-cation interactions between two arginine residues (415 and 483) leaving chemical development in this direction difficult

as any extension of the molecule from the amine is likely to form a steric clash with the protein. However, small changes could be made such as altering the heteroatom for example to oxygen. This would give a methyl ether with a greater ability as a hydrogen bond acceptor. Other changes could be explored, such as whether primary or secondary amines are tolerated. However, due to the charged, hydrophobic nature of the pocket this would be predicted to be disadvantageous. The part of the molecule with the largest scope for medicinal chemistry is likely to be the piperizine moiety that occupies part of pocket 2. This could be aromatised leading to additional π-π or π-cation interactions with neighbouring Tyr334, Arg380 and even additional interactions with Arg415.

Details of a future potential medicinal chemistry plan for initial hit exploration around the molecule ELF2 is summarised below (Figure 73). Although changes are layed out around ELF2, they could also be potentially beneficial for ELF1 which due to a structurally similar core ultimately be shown to have a similar binding mode to ELF2.

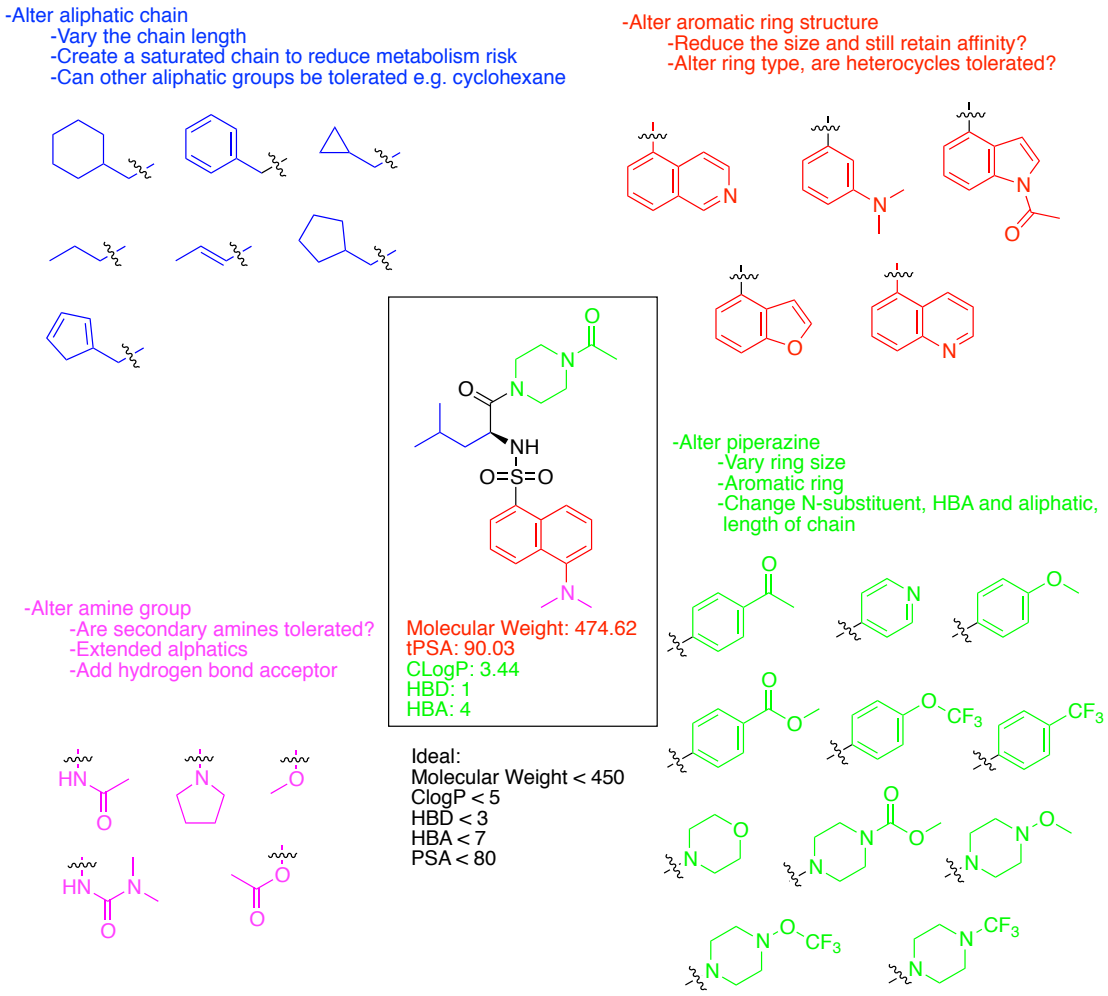


Figure 73: Illustration of potential modifications that could be made to the hit molecular series. Potential molecular substitutions, that by *in-silico* modelling have the potential to be beneficial are shown in blue, red, magenta and green depending on the substitution position on the central molecules. The molecular properties for ELF2 are shown with those that match the ideal CNS penetrant properties shown in green and those outside the criteria shown in red.

4.3 Conclusion

In conclusion two different chemical libraries were screened against the target Keap1 using the fluorescence polarisation assay. The first screen with a library provided by MRCT contained compounds designed to mimic alpha-helices and inhibit protein-protein interactions. The library was successfully screened, and several *in-silico* data analysis tools were developed for rapid statistical and iterative analysis of the assay data, (including tools for removing assay interference compounds) which can be easily adapted for use with future high throughput screens. A hit rate for the assay of 0.3 % was determined at a cut off three standard deviations below the mean of the negative controls which gave 44 compounds classified as hits. MRCT chemists' manually selected ten compounds for resupply, screening out frequent hitters and selecting molecules for structural diversity. The selected molecules were then screened in subsequent conformation and deselection assays, including a full dose response curve and a cellular ARE reporter assay. Only one compound was able to confirm the activity seen in the primary screen and gave an IC_{50} of 180 μM, which did not meet the criteria for hit selection. Structurally similar analogues around this hit were purchased and tested in both the screening and cellular assays, with no activity seen in either assay.

The second screen was performed on a much larger library of 318132 molecules with a larger structural diversity of compounds. The screen was performed externally at the European Lead Factory screening centre and was successfully re-optimised for use with equipment at the facility with Z prime values achieved in excess of 0.7. The primary screen of the library gave a similar hit rate to the MRCT screen of 0.3 %. These compounds were screened in confirmation assays at both single point and dose response. Deselection assays were performed to determine the redox potential of the molecules with any molecules showing redox reactivity screened out. A single concentration fluorescence screen was also undertaken with the data used in parallel with fluorescence data collected from the primary screens to remove compounds shown to interfere with the assay. The contents of the screening wells containing active molecules were then subject to purity test by HPLC-mass spectrometry with impure samples being discarded. The remaining 8 compounds were then sent for resynthesis where 6 were successfully resynthesised. The molecules were then validated in a full dose response curve fluorescence polarisation assay, and tested in the cellular ARE reporter assay. Three molecules ELF1, ELF2 and ELF6 were validated in the FP assay and able to show full dose response curves. ELF6 and ELF5 showed significant activity in the cellular assay. ELF2 showed activity but was not significantly above the DMSO

control, ELF1 was toxic at the top concentration so activity could not be determined to the same concentration as the other candidates.

The compounds were also tested in biophysical assays, with ELF1 and ELF2 showing binding to the target. Both molecules K_D's were determined as around 3 μM in the MST assay. ELF2 was able to compete for binding with the ETGE peptide and the control compound Hit1 evidenced by a shift the K_D values. ELF2 was also tested for Keap1 binding by ITC, but no binding was observed at the maximum concentration of protein that was practical to use in the assay.

All compounds were also subjected to a detailed *in-silico* analysis where ELF1 and ELF2 were able to demonstrate poses that achieved good docking scores when compared to a known binding control. Using these docking poses a detailed medicinal chemistry plan was proposed based on the structural information from *in*-silico experiments to illustrate where additional activity or desirable molecular properties could be gained through structural development of the molecules leading to an increase in binding energy.

These compounds are promising enough to start pursuing chemical development of these molecules in the form of an initial hit to lead program. The program should ideally aim to be informed by an X-ray co-crystal structure of the molecules in the Keap1 binding site to confirm binding poses. In the absence of X-ray crystallography, chemical input should first focus on making molecular alterations that enable elucidation of the correct binding pose as predicted by *in-silico* work. Once the binding mode is established and the development of the molecules have a path forward, a full hit to lead program can then be established.

5. In-silico techniques for finding potential ligands for protein binding

5.1 Introduction

5.1.1 Drug-Receptor interactions

The interactions that are involved in drug-receptor complexes are the same as those usually involved in interactions between organic molecules and can include ionic interactions, ion-dipole, dipole-dipole, covalent bonding, hydrogen bonding, hydrophobic interactions, cation-π, halogen bonding, and van der Walls interactions. Normally the types of bonds formed are weak non-covalent interactions, however as can be seen from previous chapters with the approval of Tecfidera® (1.4.5.1) covalent drug receptor interactions are becoming more prevalent and in some cases such as chemotherapeutic agents, it can even be desirable as a covalent bond ensures the drug can exert its toxic effect on the cancerous cell for as long as possible.

5.1.1.1 Covalent Bonds

Covalent bonds are the strongest type of bond and as previously stated are rarely formed by drug-receptor interactions. However, when desirable they can be extremely effective at forming interactions with the receptor as the bond can add between -40 to -110 kcal/mol in stability to the complex.[280]

5.1.1.2 Ionic/Electrostatic interactions

Proteins contain many different amino acid side chains, and at the physiological pH of around 7.4, the basic side chains of arginine, lysine and sometimes histidine are protonated forming cationic groups. Conversely, the acidic groups such as the carboxylic acid groups on aspartic acid and glutamic acid become deprotonated, forming anionic groups. This gives opportunities for drug molecules with opposing charges to the receptor group to be attracted. These interactions are able to work over longer distances compared to other interactions and are generally worth around -5 kcal/mol, however, they can be worth more if a neighbouring interaction is able to reinforce the interaction.[280]

5.1.1.3 Ion-Dipole & Dipole-Dipole Interactions

These interactions occur due to the phenomena of electronegativity of atoms. Those atoms with a greater electronegativity (such as Cl, Br, O and S) can create an electronic dipole between it and its neighbouring atoms. This electronic dipole can then attract other dipoles (dipole-dipole) or ions (ion-dipole). These interactions can be worth between -1 and -7 kcal/mol.[280]

5.1.1.4 Hydrogen Bonding

Hydrogen bonds are a special type of dipole-dipole interaction, formed between protons of an electronegative atom (e.g. H-O), and one or more nearby electronegative atoms containing a nonbonding pair of electrons. Hydrogen bonds form between electronegative atoms such as oxygen, nitrogen and fluorine. Hydrogen bonding works through the hydrogen bound electronegative atom removing electron density from its hydrogen, creating a partial positive charge on the hydrogen. This positive charge is then strongly attracted to the nonbonded pair of electrons. Bonds can either be described as neutral or charge assisted, where an ionic interaction enhances the hydrogen bond (Figure 74). On average hydroxyl to carbonyl (O-H --- O=C) bonds are around 2.75 Å, whilst the equivalent between an amine and a carbonyl is around 2.90 Å.[280] Hydrogen bonding is optimal when the donor and acceptor molecules are orientated at 180 degrees to each other, and are usually worth around -3 to -5 kcal/mol. Interestingly for each hydrogen bond formed, the binding affinity of a ligand to its receptor usually increases by an order of magnitude, they are therefore very important interactions in receptor-ligand binding.[281]

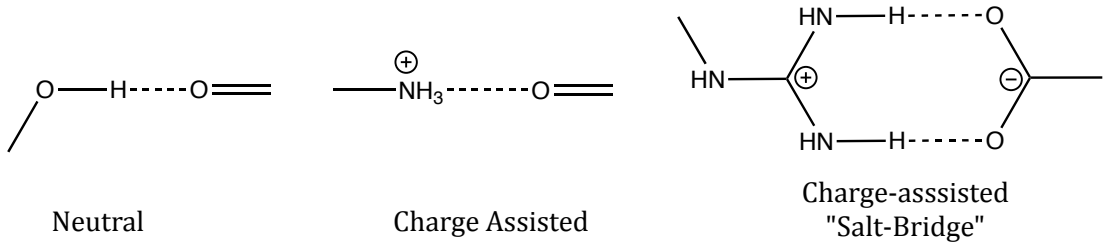


Figure 74: Different types of hydrogen bond commonly seen in receptor-ligand complexes

5.1.1.5 π-π and cation-π interactions

π-π and π-cation interactions were first described in 1990 and have since been identified as important contributing interactions to the binding affinity of ligand-receptor complexes.[282] These type of electrostatic interactions form due to the increased negative electrostatic potential at the centre of an aromatic ring creating a gradient of electrostatic potential from positive, at the edge of the ring, to negative at the centre. Cations are attracted to the anionic like charge at the centre of the aromatic ring creating a π-cation interaction. In π-π interactions the positive edges of other aromatic π-systems are attracted to the negative centre of aromatic π-systems. These interactions are estimated to account for around -2 to -4 kcal/mol.[281,282]

5.1.1.6 Importance of these interactions in Drug-Receptor complexes

As previously stated, most receptor-ligand interactions are noncovalent and the formation of bonds between the receptor and ligand will only occur if the change in free energy (ΔG) is negative. This binding energy ΔG is made up of the sum of the enthalpic contribution (ΔH) and the entropic contribution (-TΔS) giving Equation 6.

Equation 6: Equation for calculating the Gibbs free energy, where ΔH is energy due to enthalpy, ΔS is energy due to entropy, and T = Temperature.

$$\Delta G = \Delta H - T\Delta S$$

The binding energy can be related to the equilibrium constant between the receptor ligand complex K_{eq} using Equation 7. This relationship means that a change in free energy of a few kcal/mol can have a massive effect on the binding equilibrium, especially if K_{eq} is small.

Equation 7: Equation showing relationship of the Gibbs free energy ΔG to the equilibrium constant K_{eq}, where R = molar gas constant, and T = Temperature.

$$\Delta G^0 = -RT\ lnK_{eq}$$

Recent advances in computer aided drug design (CADD) have enabled predictions in binding energies to be calculated, utilising the knowledge in strength of different bonds and when they can form. This enables the software to be able to predict the likelihood that a ligand will bind to a receptor.

5.1.2 CADD

Discovery of a drug candidate is a difficult task (Figure 75) with only a small number of candidates able to meet the strict criteria. Computers have become an ever more essential tool in modern medicinal chemistry, as well as in the world as a whole. Computer-Aided Drug Design (CADD) has become an important part of this. This is because the economic viability of more traditional experimental drug development has decreased whilst the practicality, speed and accuracy of CADD has increased exponentially. Large numbers of virtual compounds can be screened rapidly with an increasingly degree of accuracy.[283] This allows the fast identification of *hit* compounds. Hit compounds are compounds that show initial activity against a therapeutic target and the systematic modification of these 'hits' could give rise to lead compounds which eventually become modified to final drug candidates. The application of molecular models allows the estimation of the activity of a molecule towards its potential target; predicted *pharmacodynamics,* as well as its interactions with the rest of the human body; *pharmacokinetic* properties.

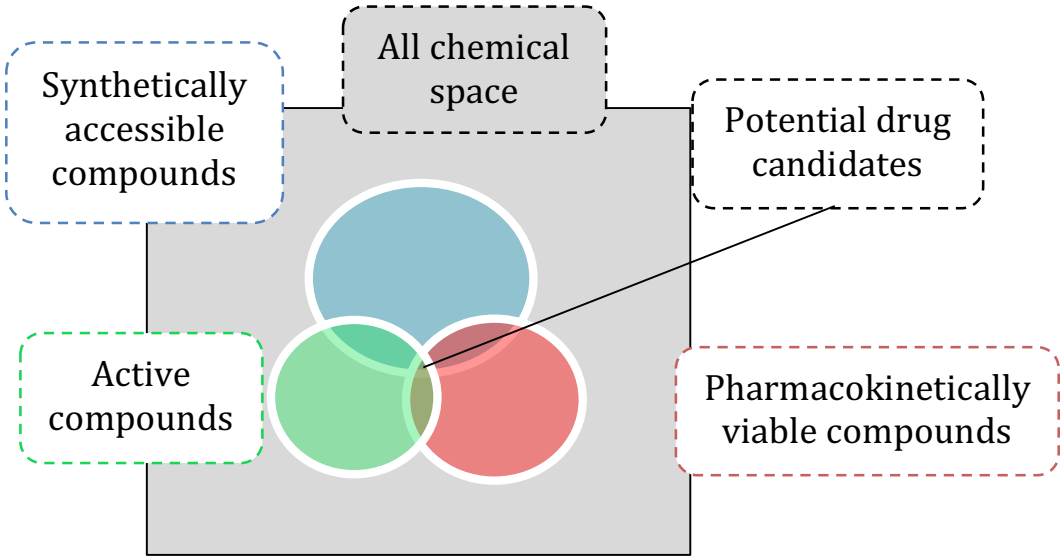


Figure 75: Venn diagram illustrating the difficulty in finding a potential drug candidate.

CADD occupies an increasingly important position in drug development, and spans an ever-widening array of methodologies. Hence, only a selection of relevant techniques that are appropriate within the context of the project will be discussed here.

A classic iterative process of structure-aided drug design (SADD) similar to that used in this project can be seen below (Figure 76). Classically in target based drug discovery a druggable target is selected. Normally this will be a protein, but could also be a nucleic acid. These protein targets can be split into 6 superfamilies. G-protein-coupled receptors (which are currently the target for of ~50% of all drugs), enzymes (especially protein kinases, proteases, esterases, and phosphatases), ion channels, nuclear hormone receptors, structural proteins and membrane transport proteins.

Once the target is selected, in order to perform CADD ideally there needs to be a crystal structure of the target. This can either be created through homology modelling or through synthesising and crystallising the protein. Homology modelling can be used where a crystal structure for a similar protein family member may already be available. This can then be modified, usually through a series of mutations and/or insertions of amino acids to those within the desired target protein, followed by energy minimisation refinement to altered areas of the structure.

The structure of the protein can then be either visually or computationally (using programs such as fpocket or Qsite finder) analysed to look for possible small molecule binding sites. These binding sites are probed for inhibitors using docking, docked molecules can then be analysed and scored for interaction with the binding sites. Molecules can then be selected and tested for binding in biochemical assays. Any

molecules which successfully bind a target can be further optimised using classical medicinal chemistry combined with the knowledge gained from docking poses of the molecule.

If a co-crystal structure of a protein with a known inhibitor is available, then the interactions the inhibitor makes with the protein can be analysed to find further molecules that are able to make the same interactions, this is known as pharmacophore modelling. Pharmacophore modelling can also be utilised where there is an absence of crystal structures but where a variety of ligands are available. These ligands can be "feature mapped" creating a negative image of the binding site based on the interactions from the known binders. This is known as a ligand-based pharmacophore.

CADD can be utilised throughout the drug development process, enabling alterations to molecules (substations and additions) in order to achieve the desired potency and ADMET properties. CADD is therefore a key part of the drug discovery process.

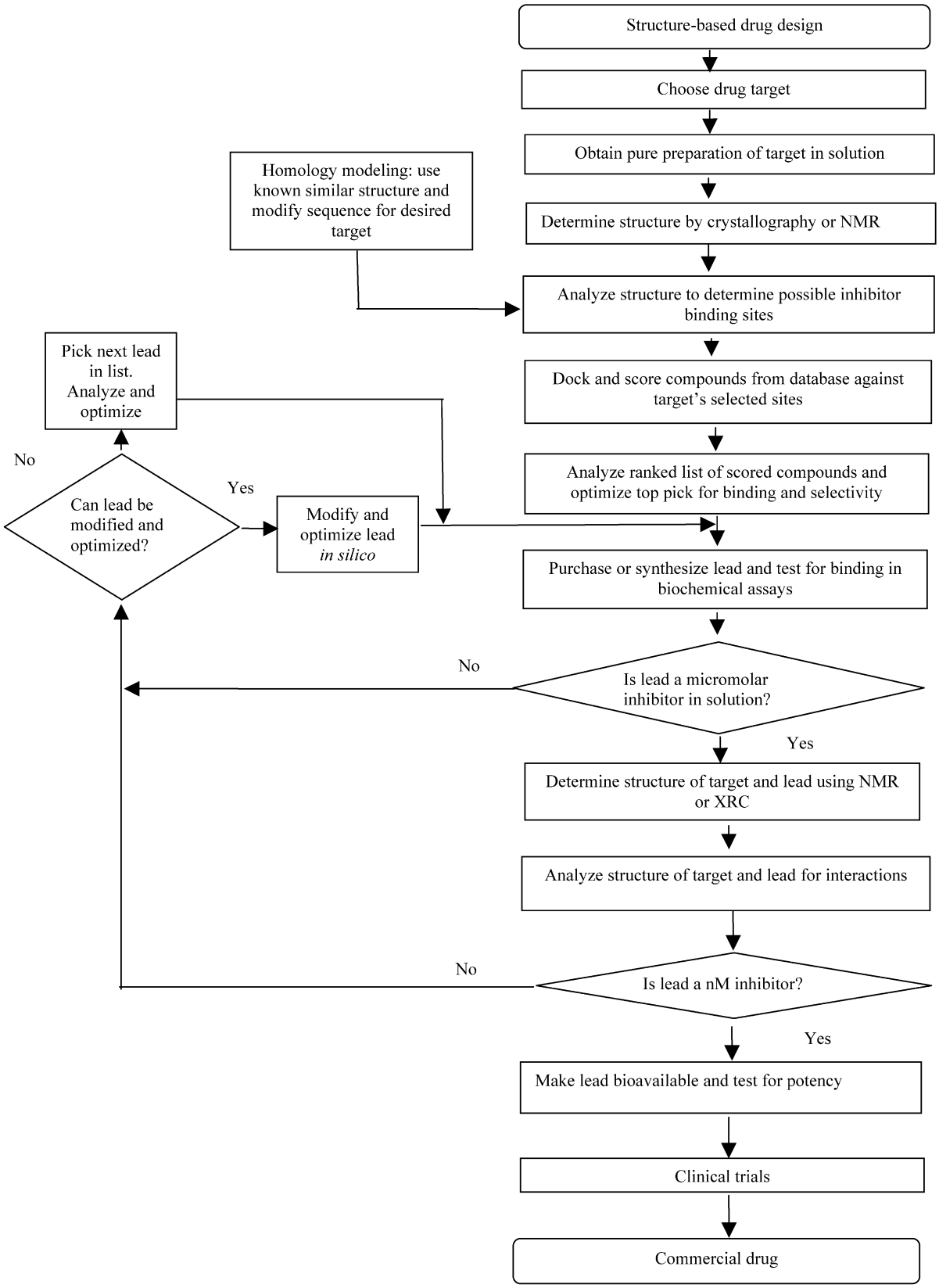


Figure 76: A flow chart representing a classical approach to structure-aided drug design (reproduced with permission from Anderson, 2003).[284]

5.1.3 CADD Techniques

There are two main strategies that can be taken within CADD. These strategies are known as “direct” or “indirect” design. In direct design a crystal structure of the target is usually known, this is gained from either X-ray or NMR crystallography or through homology modelling. This enables direct modelling, and characterisation of receptor-ligand interactions. In indirect design however the target structure is unknown. The design therefore is based around structural features of compounds known to be either active or inactive against the desired target, using information gained from biological assays.[285–287]

5.1.3.1 Pharmacophore mapping

Pharmacophore mapping is traditionally an indirect technique, used when the structure of the receptor is not known. Traditionally the 2D or 3D computer generated structures of active compounds are mapped together to explore any interactions they may share. This has a major disadvantage as it fails to take into account that unfavourable conformations of compounds may occur in receptor-ligand binding where a more energetically favourable interaction can be found that can overcome the energy barrier preventing the unfavourable conformation. However, when a co-crystal structure of a protein-ligand complex is available, pharmacophore mapping can be adapted into a direct technique. Using the structure of the co-crystallised ligand bound into the receptor a 3D pharmacophore can be generated. This contains additional information about the location of the interactions needed for binding of the ligand with the receptor. This approach enables non-receptor binding features on the ligand to be ignored, meaning molecules that match the pharmacophore are more likely to bind the target, as they should satisfy spatial and structural conditions for binding.

A pharmacophore model, according to the IUPAC definition, is "an ensemble of steric and electronic features that is necessary to ensure the optimal intermolecular interactions with a specific biological target and to trigger (or block) its biological response".[288] In the form used in this project, a 3D pharmacophore was defined as the essential features or chemical substructures and their corresponding 3D locations that are responsible for the similar biological activities of a set of compounds. Typically, pharmacophore features include hydrophobic, aromatic, hydrogen bond acceptor, hydrogen bond donor, positive ionizable and negative ionisable features.[289]

The following figure shows a typical 3D pharmacophore model with three types of chemical features and their 3D location constraints:

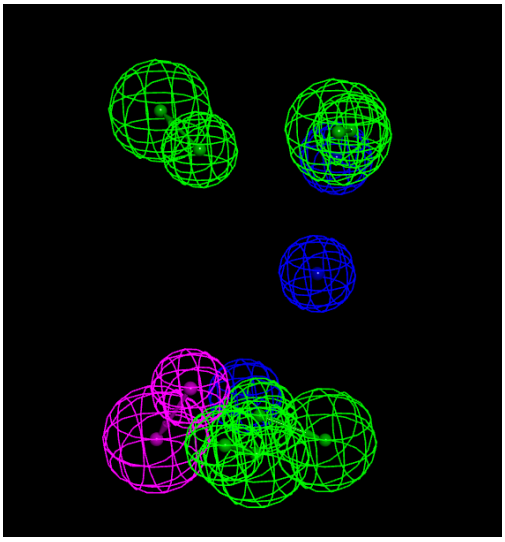


Figure 77: A Typical Pharmacophore model. The pharmacophore model consists of hydrogen bond acceptors (HBA, in green), hydrogen bond donors (HBD, in pink), and negative ionizable centres (in dark blue). Since a hydrogen bond is directional, each HBA feature is specified with two points: one is the location of the associated heavy atom, and the other is the projection point, which defines the direction for forming a hydrogen bond with a target. Each sphere defines the location tolerance of a particular feature point.

5.1.3.2 Docking

Molecular modelling can be used to fit or "dock" a molecule into a binding site. A docking program will position a molecule into a binding site based on defined ideal bonding distances for potential interactions (hydrogen bonds, Van der Waals forces etc).[289] Once positioned in the binding site, the predicted interactions of the docked ligand with the receptor are evaluated and scored. The exact scoring method depends on the scoring function used, but the result is a quantitative comparison of the strength and stability of the predicted protein-ligand interactions for each dock of a simulation. The docking score is an overall prediction of the combination of steric and electronic interactions between the ligand and the surface of the specified binding site. Multiple conformations of the considered molecules are docked with lower energy conformations being favoured over high-energy conformations. The strength of the interactions of each dock with the target binding site is quantified, relative to the other docks of that simulation.[290]

However, it is well documented that it is difficult to select a scoring function that reliably ranks ligands with experimentally determined results, such as binding energies and co-crystal structures where the binding pose is known.[291–293] This often means that more than one scoring function is used in docking to attempt to reduce the number of false positives produced by the process.

There are three general categories that the scoring functions can be loosely associated with: force-field based, knowledge-based and empirical scoring functions. Force-field based functions estimate the intermolecular interactions between all the atoms in the ligand and the protein to define a score. Knowledge-based functions use published data from protein ligand co-crystal structures to create statistically driven rules such as how likely interactions are to happen depending on the frequency an atom pair occurs in the literature. Lastly empirical scoring functions count the number of various different types of interactions between a protein-ligand pair, such as the number of hydrogen bonds, hydrophobic interactions etc.[294]

In addition to how the docking program scores protein-ligand interactions, there are several other important other considerations to take into account. Proteins and ligands are not rigid structures in solution or *in-vivo* environments. They can therefore undergo conformational changes when a ligand-protein binding event takes place. It is therefore important to consider including flexibility of the protein within docking. However, including protein flexibility within the calculation increases the complexity of the docking calculation to be performed by the machine. This has an impact on the time to run a simulation and therefore the cost.

There are several approaches to tackle molecular flexibility within docking: rigid-body docking, where both the receptor and ligand are kept rigid, rigid-body flexible-ligand docking, where the receptor is kept rigid and the ligand allowed to flex, and fully flexible docking, where both the receptor and the ligand are allowed to flex.

In addition, there are methods that allow partial flexibility of the receptor, such as allowing the rotation of side chains of specific amino acids known to have flexibility in the region of the binding site. Comparing several crystal structures of the chosen receptor will reveal flexible residues, such as those within *apo* (where the protein is not bound to its substrate) and *holo* (where the protein is bound to its substrate) forms of proteins. Structure B-factors will also give an indication on how flexible the protein is, as the B-factor is a measure of the fluctuation of atoms about their average positions.

Finally, it is important to take into consideration the quality of a given crystal structure. This can be determined through looking at the resolution, R values, R-free values, and B-factors. The R value also known as R-work is a measure of the agreement between the experimentally observed x-ray data and the crystallographic model. Mathematically, it is the difference between the experimentally observed values and the ideal calculated values. The R-free is a similar measure, but is used to calculate over modelling of the

data based on a small sample of the data that is never used in refinement of the structure.

5.1.3.3 Predictive ADMET

Drug discovery is an expensive business, with each new medicine from discovery lab to pharmacy shelf costing over $800 million.[295] Much of this expense comes within the research and development (R&D) stage where over $200 million per drug is spent. This has led to a search for cost savings within R&D. A large amount of time within the drug discovery process is taken up optimising a hit compound into a potential clinical candidate, or lead. This stage is known as hit to lead optimisation. If this process could be quickened through selecting out compounds with non-drug-like properties, then the cost of optimisation could be reduced. This is because the starting molecules will have properties closer to that of a final drug.

A classic example of this theory is *Lipinski's Rules* of 5.[240] These are a set of rules based on known approved drugs. Lipinski focuses on the number of hydrogen bond donors (HBD), and acceptors (HBA), the molecular weight, and the octanol-water partition coefficient (logP). However, as more drugs have been approved, these criteria have been further refined to include additional properties. An example of these properties is shown in Figure 78. Additional properties that have been included include molecular polar surface area (mPSA), and the number of rotatable bonds. The mPSA can be defined as the sum of surfaces of polar atoms (usually oxygens, nitrogens and attached hydrogens) in a molecule. It is a very useful property to use at it has been shown to correlate well with human intestinal absorption, blood-brain barrier (BBB) penetration, and Caco-2 monolayer permeability. Palm el al. showed that molecules with a mPSA of greater than 140 $Å^2$ had poor cell permeability, and that molecules with mPSA of less than 60 $Å^2$ had the best chance of passing the BBB.[236] The number of rotatable bonds is also important; Veber et al. (2002) found that greater than 10 rotatable bonds had a detrimental effect on membrane permeability. Additionally, Veber et al. found that a reduced polar surface area correlates better with increased permeation rate than lipophilicity (ClogP).[296]

```
Num_atoms is between 12 AND 70
 AND Molecular_PolarSurfaceArea <= 120
 AND Molecular_Weight is between 192 AND 500
 AND N_Count >=1
 AND O_Count >=1
 AND (Cl_Count + Br_Count + I_count + F_Count) <=7
 AND Num_H_Donors_Lipinski <=5
 AND Num_H_Acceptors_Lipinski is between 2 AND 9
 AND Num_RotatableBonds is between 2 AND 8
```

Figure 78: Filter for High Drug Likeness derived from the collective works of Murcko, Ghose, Mozziconacci, Oprea, Lipinksi and Palm.[236–240]

In addition to using more traditional pharmacodynamics properties, predicting pharmacokinetic properties of a drug has become of equal importance, after a study in the 1990s indicated that poor pharmacokinetics and toxicity were important causes for costly late-stage failures in drug development.[297] Through recent advances prediction of a range of absorption, distribution, metabolism, excretion and toxicity (ADMET) properties can now be performed. Examples of the more important factors which can be predicted are intestinal absorption,[298,299] aqueous solubility,[300] BBB penetration,[299] cytochrome P450 (2D6 isoform) enzyme inhibition, hepatoxicity,[301,302] and plasma protein binding.[302–304] Such criteria are factored in to the development of specific compounds as well as the selection of potential drug candidates. Minimising the time to find a compound with the desired 'drug-like' properties could potentially save large amounts on R&D budgets, and reduce the timeframe for bringing a drug to market.

5.1.4 Software packages

A variety of different software packages are available to cater for the different functions required by the computational chemist, these are too numerous to list but they can be broken down by the function that they perform. With regards to this project, software will be needed to perform library curation, pharmacophore generation, pharmacophore screening, docking, ADMET prediction and data handling. A number of software suites now exist that combine many of these functions into one program to aid with workflow and data handling. The most popular of these include MOE (http://www.chemcomp.com), Schrödinger's Small-molecule drug discovery suite (http://www.schrodinger.com), and Accelrys's Discovery Studio (http://accelrys.com). All three packages whilst differing in specific algorithms for the programs, are capable of performing the tasks required by the project.

5.1.4.1 Discovery Studio

Discovery Studio is the suite of software that was available to this project, a screenshot of the user interface can be seen in Figure 79. Discovery Studio includes applications covering the following areas:

Ligand Design:

- Tools for enumerating molecular libraries and library optimisation

Pharmacophore modeling:

- Tools for creation, validation and virtual screening using pharmacophore models.

Structure-based Design:

- Tools for fragment-based docking and refinement, receptor-ligand docking and pose refinement, as well as de-novo design.

Macromolecule design, validation and engineering:

- Protein-protein docking
- Antibody design and optimisation
- Tools for membrane-bound proteins, including GPCRs

QSAR:

- Multiple linear regression, partial least squares, recursive partitioning, Genetic Function approximation and 3D field-based QSAR.

Predictive ADME and toxicity:

- Including TOPKAT modeler for prediction of ADME toxicological properties of molecules.

Simulations:

- Molecular Mechanics (including implicit and explicit-based solvent and membrane models), Molecular Dynamics, and Quantum Mechanics (including hybrid QM/MM calculations).

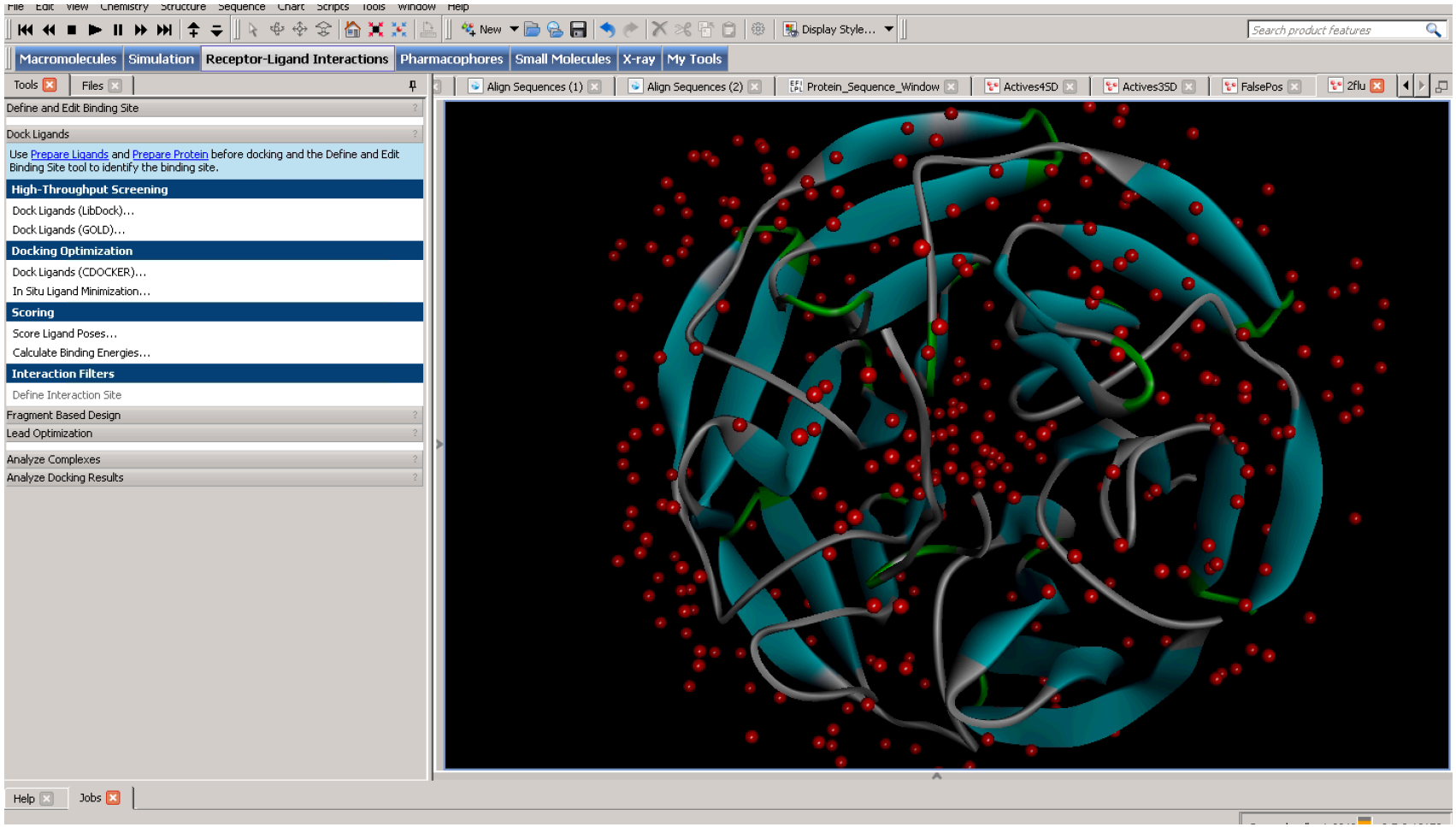


Figure 79: Screenshot of the Discovery Studio interface showing a protein x-ray crystal structure loaded into the visualisation window on the right, with a subset of the tools that can be used for analysing receptor-ligand interactions shown on the left.

5.1.4.2 Pipeline Pilot

Pipeline pilot is a data flow and visualisation program developed by Accelrys. The Discovery Studio interface detailed above is built on the foundations of pipeline pilot. It is used to channel large amounts of data through processing sections called "components", or "nodes". Each component performs a separate task on the data, for example filtering out all non-active compounds, or doing a mathematical calculation. Users can either choose from a selection of pre-designed components or can choose to design their own using the software's scripting language "pilot-script". In a typical protocol data records are read in from the left using a reading component and pipelined (to the right) for analysis, processing and reporting. An example protocol can be seen in the screenshot in Figure 80.

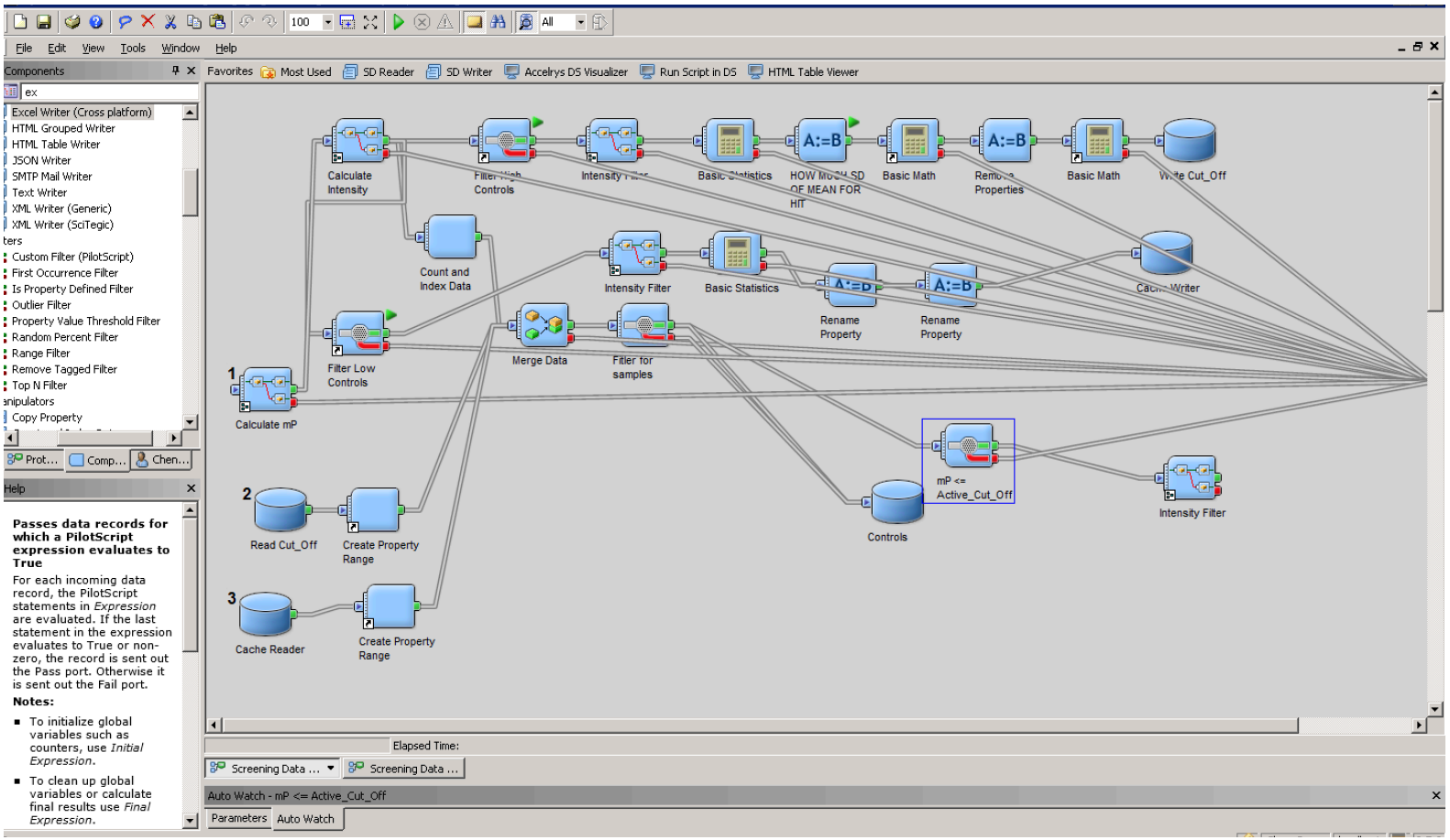


Figure 80: Screenshot of the pipeline pilot user interface showing a multitude of data processing components in the main window. New components can be selected from the boxes to the left of the screen and dragged onto the workspace to connect with the existing pipeline.

5.2 Results and Discussion

5.2.1 Initial Screening

5.2.1.1 Selection of a screening library

In order to perform *in-silico* virtual screening of Keap1, a screening library had to first be selected. Whilst there are many virtual screening libraries available the largest non-commercial database of compounds is the ZINC database.[235] ZINC is comprised of around 21 million compounds in ready-to-dock 3D formats. It was this database that was selected for use within the project.

However, the ZINC database is comprised of molecules with a large range of molecular properties. A subset of the database that contained only those compounds that have drug-like properties was compiled specifically for this project. This would favour molecules that were more likely to penetrate the BBB so as to reach the therapeutic area of interest, and have properties suited for oral administration for preferred delivery of the drug. To do this a drug-likeness filtering program was written in Pipeline Pilot 8.5 (Figure 81) combining the works of Murcko, Ghose, Mozziconacci, Oprea, Lipinksi and Palm.[236–240] Table 13 below shows a comparison between the criteria selected and other traditional criteria. The chosen criteria were selected as an up-to-date version of the drug-like criteria for an orally absorbed drug, taking into account further molecular properties that have been investigated since Lipinski's original paper over a decade ago.

Table 13: Comparison of properties for drug-and lead-like properties.

Suggested	Lipinski Drug-Like[238]	Lead Like[305]
≤ 5 HBD	≤ 5 HBD	≤ 3 HBD
2-9 HBA	≤ 10 HBA	≤ 3 HBA
Molecular Weight 192-500	Molecular Weight< 500	Molecular Weight< 300
n/a	logP<5	logP<3
2-8 Rotatable Bonds		≤ 3 Rotatable bonds
mPSA <=120		
Halogen atoms <=7		
N atoms >=1		
O atoms >=1		
Total atoms 12-70		

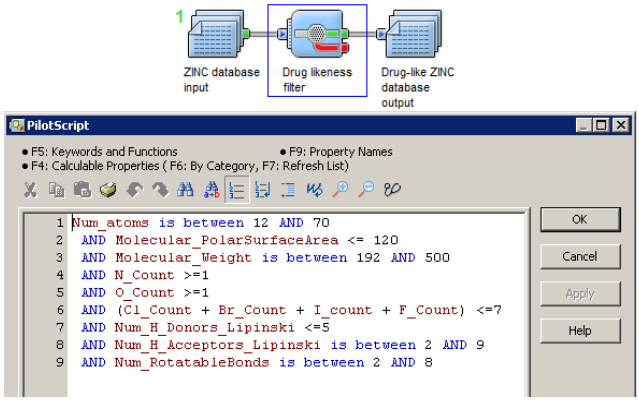


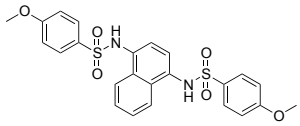
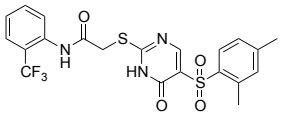
Figure 81: Pipeline Pilot Protocol for creating a drug-like subset of the ZINC database.

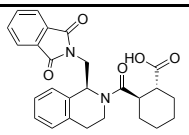
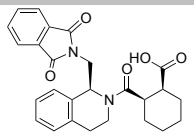
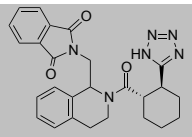
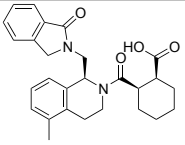
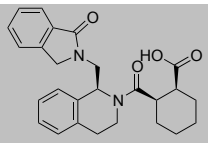
The screen resulted in a final virtual screening library size of 3.9 million compounds.

5.2.1.2 Selection and analysis of a KEAP1 crystal structures

A search of the PDB database revelled that no full length structure of Keap1 was available. Several suitable, high-resolution crystal structures of the DC domain of Keap1 were chosen and downloaded. These are detailed in the table below. Mutant and non-human isoforms were discounted when selecting a structure to use. This was to try and obtain a crystal structure folded as closely to the active human structure as possible to ensure maximum pharmacological relevance.

Table 14: Review of available crystal structures. *shows structures that were published at time of structure selection in 2012.

PDB ID	Domain of Keap1	Co-crystallised ligand	Residues of Keap1	Species	Resolution (Å)	R-Value	R-Free
1ZGK*	DC	n/a	321-609	Homo sapiens	1.35	0.112	0.133
2FLU*	DC	Nrf2 ETGE 16-mer	325-609	Homo sapiens	1.50	0.180	0.199
1U6D*	DC	n/a	321-609	Homo sapiens	1.85	0.149	0.182
4IQK	DC	Cpd 16 	321-609	Homo sapiens	1.97	0.155	0.179
3VNG	DC	FUU (crystal)	321-609	Homo sapiens	2.10	0.196	0.255
3VNH	DC	FUU (soaked)	321-609	Homo sapiens	2.10	0.201	0.211
3ZGC	DC	Nrf2 ETGE 7-mer	321-609	Homo sapiens	2.20	0.173	0.188
4IN4	DC	Cpd 15 	321-609	Homo sapiens	2.59	0.194	0.250
2DYH*	DC	Nrf2 DLG 15-mer	309-624	Mus Musculus	1.90	0.174	0.211
2Z32*	DC	Prothymocin alpha (39-54)	309-607	Mus Musculus	2.00	0.169	0.223
3ADE*	DC	Sequestosome-1 (346-359)	309-624	Mus Musculus	2.80	0.178	0.232
3WZD	DC	Phosphorylated Sequestosome-1 (346-359)	321-609	Mus Musculus	2.60	0.194	0.255
4IFL	DC	Nrf2 ETGE 16-mer	321-609	Homo sapiens	1.80	0.193	0.235

4IFN	DC	 12O	321-609	Homo sapiens	2.40	0.278	0.352
3WN7	DC	Nrf2 DLGex 34-mer (17-51)	321-609	Mus Musculus	1.57	0.189	0.223
4L7B	DC	 1VV	321-609 (R354D mutation)	Homo sapiens	2.41	0.229	0.246
4L7C	DC	 1VW	321-609 (R354D mutation)	Homo sapiens	2.40	0.266	0.281
4L7D	DC	 1VX	321-609 (R354D mutation)	Homo sapiens	2.25	0.256	0.276
4N1B	DC	 2FS	321-611 (R354D mutation)	Homo sapiens	2.55	0.246	0.287

2FLU (Figure 82) was selected for the pharmacophore study, due to its high resolution, and presence of a co-crystallised Nrf2 ligand. The availability of the co-crystallised ETGE peptide in the active site enables the creation of a pharmacophore of interactions between the peptide and the ligand.

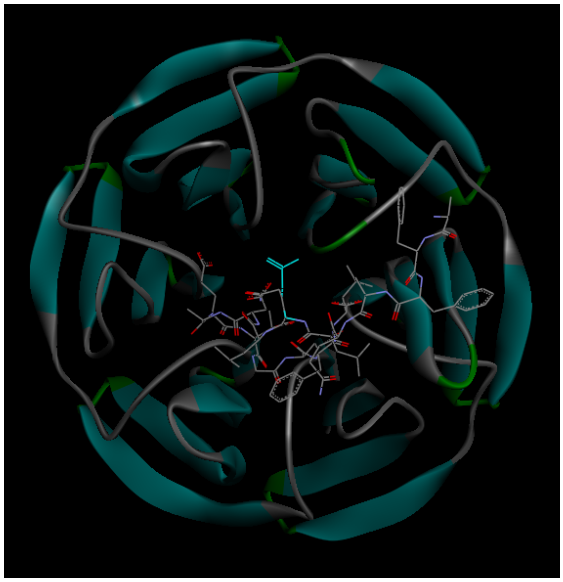


Figure 82: The crystal structure of KEAP1 DC domain (PDB ID: 2FLU)[229] with the Nrf2-ETGE 16mer peptide co-crystallised in the active site.

5.2.1.3 Developing a high throughput virtual screen using pharmacophore modelling

Using Discovery Studio 3.5 the interactions between the peptide and protein were modelled to create a 3D pharmacophore (Figure 83). The model creates binding spheres around identified interactions. These interactions include HDB, HBA and ionic interactions. In addition, the program also creates exclusion spheres, these are areas where if a molecule were to interact with these spheres it would have a negative effect on a binding interaction, usually due to a steric clash with the protein.

The model enables screening of a large number of molecules in a relatively short period of time when compared with standard docking analysis as, unlike docking, the binding interactions are already pre-defined and fixed. This means that the program does not have to work out all possible interactions between the protein and the molecule. The program is simply able to orientate the molecule within the space encompassed by the pharmacophore to see if a given molecule can orientate itself to fit two or more of the defined interactions.

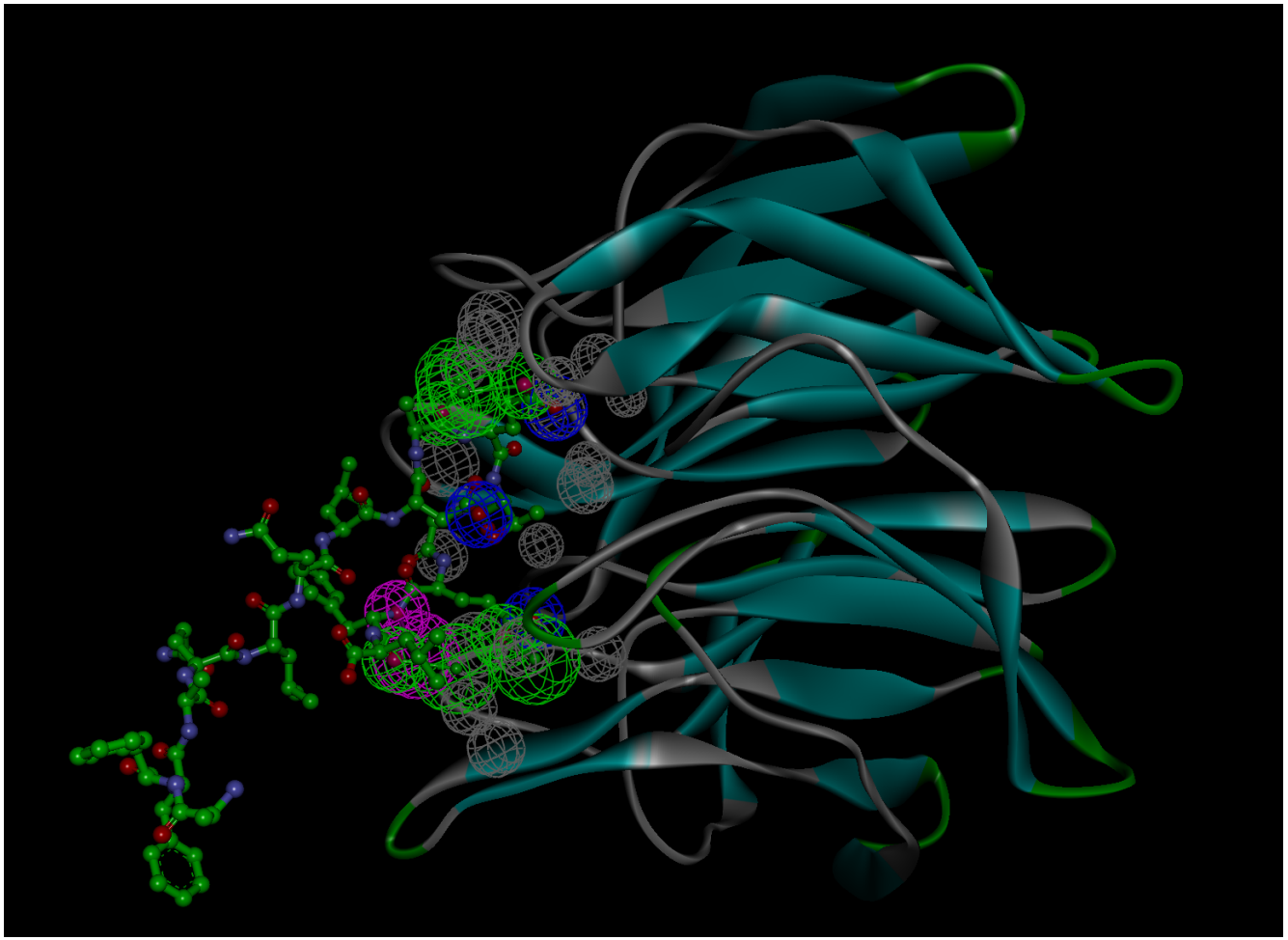


Figure 83: Pharmacophore model of interactions between Keap1 DC domain and Nrf2 peptide. Interactions shown in spheres are hydrogen bond donors (green), hydrogen bond acceptors (pink), ionic interactions (blue) and exclusion spheres (grey).

The drug-like library was then screened against the pharmacophore. This resulted in about 3000 molecules (686 after ionisation and tautomeric forms were filtered out) from the initial set of 3.9 million that were able to match the features set out by the pharmacophore, with varying scores.

These molecules then needed further refinement to a number economically feasible to test.

5.2.1.4 Refinement of molecules using docking

Discovery studio contains two complimentary docking programs, Libdock, and GOLD. Docking was performed using both programs, and the results were analysed. Through analysis of the high scoring docking poses from each program and comparisons with the scores achieved in the pharamacophore screening, a selection of molecules was chosen for progression.

Although no correlation was seen between the docking scores and the scores in the pharmacophore screen (Figure 84), good correlation was seen between the two docking programs (R^2 of 0.41) showing that molecules that conform well to the protein surface within the pocket also seem to make the most interactions with the protein.

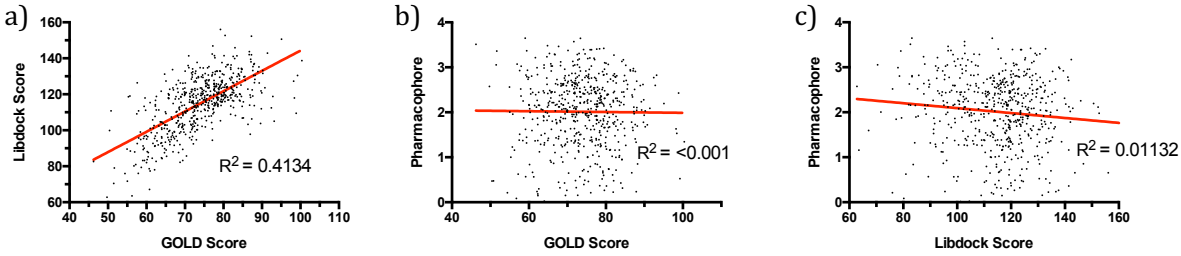


Figure 84: Comparison of scoring functions to look for correlation in the data set of 686 molecules selected for docking to the Keap1 crystal structure. a) Docking comparison: ChemPLP (GOLD) score vs. Libdock score. b) Pharmacophore comparison: Pharmacophore score vs ChemPLP (GOLD) score. c) Pharmacophore comparison: Pharmacophore score vs Libdock.

5.2.1.5 Analysis of docking results

A group of 20 molecules was selected based on scores from docking and the pharmacophore screening. The top five scoring molecules that were tractable synthetically or commercially were selected from each of the 3 virtual screening experiments, giving a total of 15 molecules. The remaining 5 molecules were selected from molecules that scored well in all three processes, but which had not already been selected as top 5 in each experiment. This was achieved through normalising the scores for each of the screens between 0 and 1 and then summing the normalised scores to give

a total accumulative score out of 3. Molecules were then purchased and tested in biological assays.

5.3.1.6 Testing of analogues in fluorescence polarisation assay

Available analogues were tested in the fluorescence polarisation assay and were found to have no effect (Figure 85).

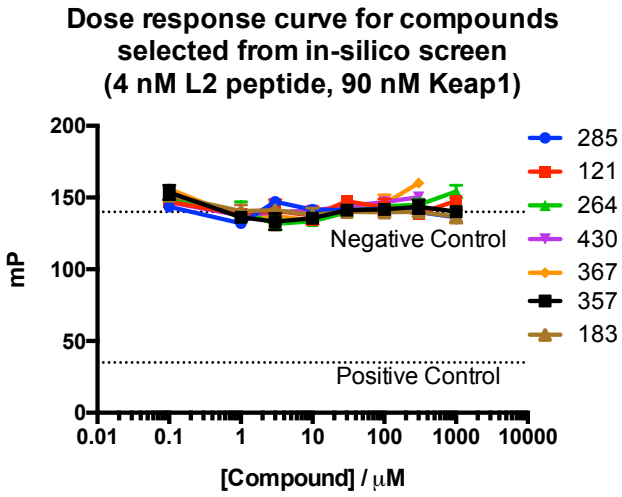


Figure 85: Dose response fluorescence polarisation assay testing available analogues from in-silico screening program. Y-axis shows polarisation in mP with error in standard deviation (n=3). Negative (DMSO), and positive (30 μM Hit1) controls are shown as indicated.

5.2.1.7 Discussion and conclusions

The lack of activity seen in the assay was thought to be down to a number of possibilities, the first being the low number of molecules that could be tested from those selected in the screen. Normally from an *in-silico* screen >100 molecules would be tested due to the large number of variables in biological systems that are unable to be accounted for using an *in-silico* simulation, which can lead to a large number of false positive compounds appearing in the *in-silico* screen. However, due to the availability, synthetic tractability and cost of the molecules selected from the *in-silico* data, only a few molecules from the set could be tested. This is a common problem in computational screening methods as the libraries available for *in-silico* work are larger than the chemical space which is easily accessible using from current synthetic methods, and much larger than the commercially available chemical space. Additionally, commercial availability of compounds can change rapidly leading to molecules that were once available becoming unobtainable within a reasonable timeframe. All of these factors lead to molecules being selected that cannot be tested *in vitro.* To overcome this the *in-silico* library can be filtered down to only include molecules that have commercial

suppliers, and therefore minimising the rate of attrition when selecting compounds for *in-vitro* testing, but at the potential expense of chemical diversity.

Another reason could be that the model created was not representative enough of how a small molecule might bind the active site. This can happen due to the limitations of the X-ray crystal structures used in the studies. X-ray crystal structures show proteins and their constituent amino acids as a fixed moiety, without the flexibility that naturally enables them to flex and move when solubilised in aqueous solutions. In this study a protein crystal structure of Keap1 with the large biological peptide co-crystal structure was utilised. Since this section of work was completed a protein crystal structure with a small molecule co-crystallised in the active site has been made available in the PDB.[210] When comparing the surface of the protein between 2FLU (peptide) and 4IQK (small molecule) (Figure 86) it can be seen that the structure of the protein has altered slightly. In particular there is a larger amount of available space in the upper right and upper left regions of the binding pocket. This is mostly due to Arg415, which due to its flexible aliphatic tail and charged guanidinium head group, is able to flex changing the available protein surface of the binding site.

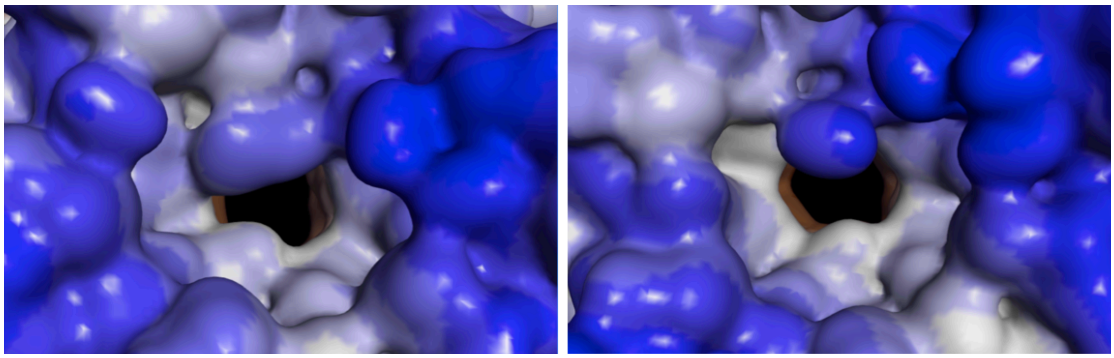


Figure 86: Comparison of hydrophobic protein surfaces to analyse available binding pockets between PDB ID's 2FLU (left), and 4IQK (right). Hydrophobicity is shown as a gradient from Brown (most hydrophobic) to Blue (least hydrophobic).

To compensate for this change in binding structure one or both of two *in-silico* techniques can be applied. The first is to use the crystal structure of Keap1 with the small molecule co-crystallised making the assumption that all small molecules would make use of this form of the binding site. The second is to predict the flexibility of the arginine using a library of known rotamers/conformers of the amino acid. This would enable the software to use the rotamer library in docking studies so that molecules being docked in the active site have the opportunity to see all of the potential different binding surfaces available to them depending on the conformation of the arginine.

Whilst this method has obvious advantages, the largest disadvantage is the increase in computational resource needed for this type of study.

5.2.1.8 Analysis of recently published data

Due to the lack of hits from the previous screen, it was decided to repeat and re-optimise the *in-silico* screen in light of the possible issues arising from the previous screen, and new data being published enabling a different approach.

The structural data that had been made available was first visually analysed to see if any insight could be obtained as to the differences or similarities between peptide and small molecule binding in the active site. At the time of analysis three small molecule co-crystal structures had been published (Figure 87), in addition to the peptide co-crystal structures of the Nrf2 ETGE and DLG peptides (PDB ID: 2FLU and 2DYH, for ETGE and DLG respectively). As discussed in chapter 1 (1.4.5.2.2), the published molecules known as 16 (Compound 2, Figure 87) and 15 (Compound 3, Figure 87) were discovered through a high throughput screen.[210] Compound 15 (PDB ID: 4IN4) had a weak affinity towards the protein (IC_{50} = 118 μM), whilst compound 16 (PDB ID: 4IQK) showed a better binding affinity (IC_{50} = 2.7 μM). In addition, a small molecule coded as FUU, (Compound 1, Figure 87) has recently been released into the Protein Data Bank without an accompanying publication, (PDB ID: 3VNH).

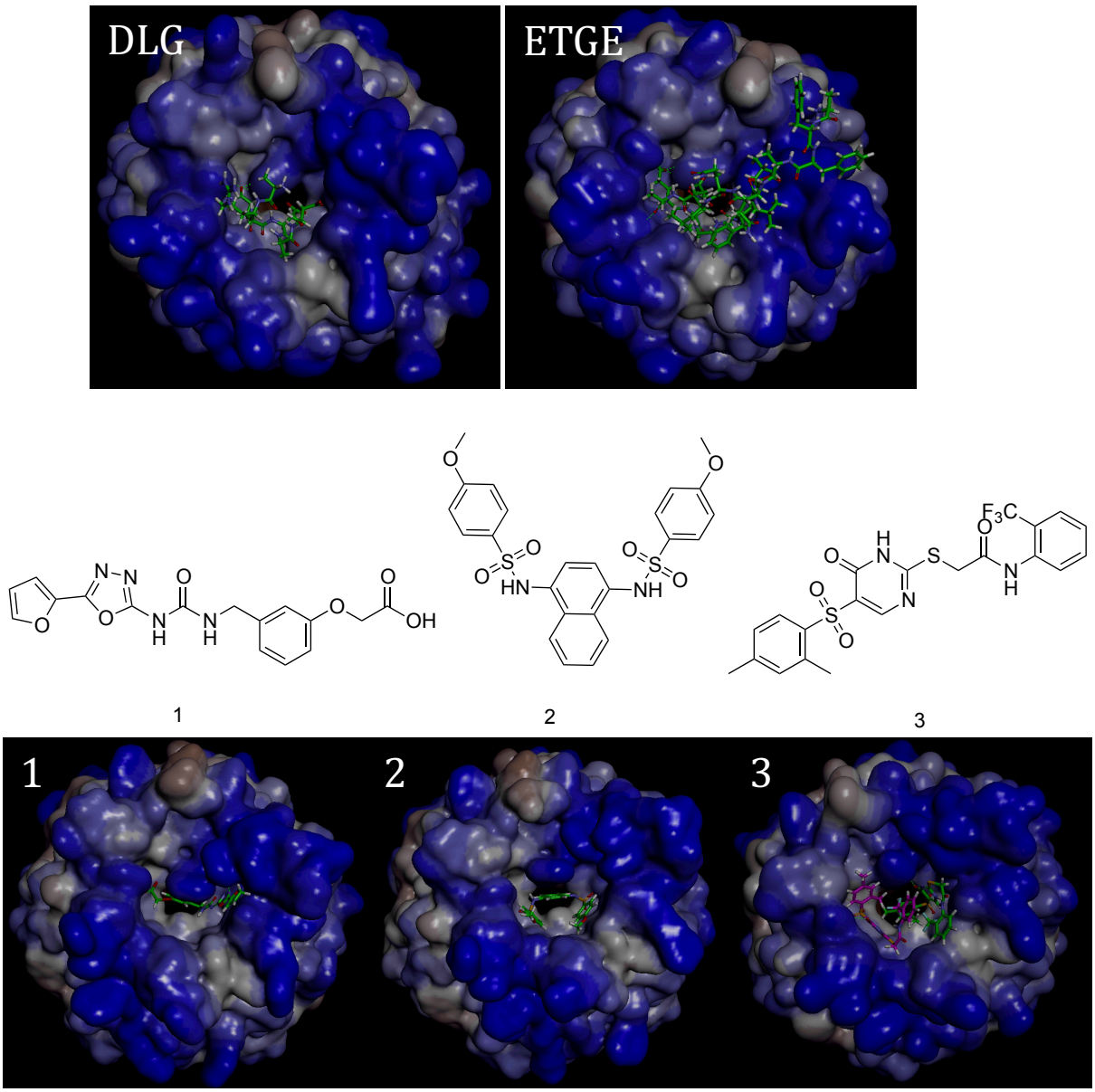


Figure 87: *in-silico* analysis of the binding modes of Keap1 to peptides and small molecules. Top: Binding modes of Nrf2 peptides to Keap1, shown with a hydrophobic surface. Middle: Recently published small molecule inhibitors of the Keap1-Nrf2 interaction (compounds 1, 2 and 3). Bottom: Binding modes of the three small molecule inhibitors with Keap1.

Analysis of the published data revealed that a few of the small molecules shared a similar binding orientation. These molecules contained an aromatic backbone that would orientate itself into the centre of the beta-propeller structure of Keap1. As previously discussed (1.4.5.2.2) Jiang et. al. illustrated the binding mode of compound 16 with Keap1 and showed that the surface of the pocket contained four smaller cavities in which different molecular groups can sit forming interactions with the protein, these are shown in Figure 88.[209] Several key amino acids were identified which were involved in binding to the small molecules and peptides. In particular, Arg415 was found to form key hydrogen bonding or π-cation interactions with a number of molecules. It also works to bridge the gap between pockets 1 and 2, and can be seen in a variety of rotamers in different crystal structures, showing it is flexible and not held rigid in the protein structure. Arg483, Ser508, and Ser363 also appear to be important to binding, making key hydrogen bonding interactions in many docking poses.

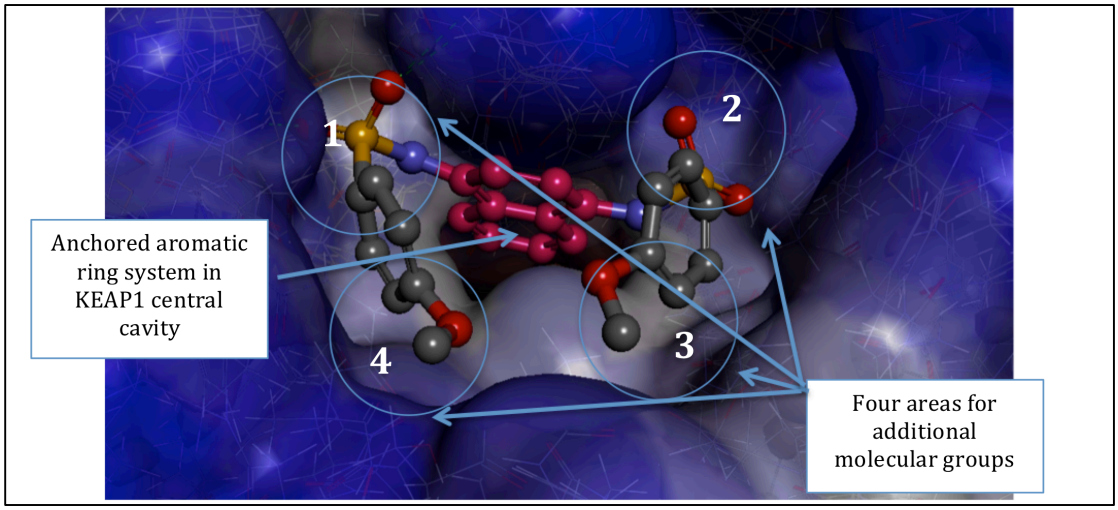


Figure 88: Further analysis of published co-crystal structures showed molecules bound in the binding site with an aromatic ring into the central cavity of the protein, this anchors the molecule enabling it to make further interactions with the four smaller pockets highlighted 1-4

From analysis of the other small molecules recently published it was clear that the formation of this active site conformation was favourable when fused rings were incorporated into the molecules as it enables the flexible arginine residue to form a key backbone cation-π interaction with the molecules.

It was therefore decided future computational studies should use the 4IQK crystal structure using algorithms to gain a flexible Arg415 where possible.

5.2.2 Second *in-silico* screen

5.2.2.1 Selection of Library

As in the previous screen the ZINC library was utilised. However, due to the trouble in sourcing compounds identified by the previous screen, the commercially available subset of ZINC containing compounds that were known to be in stock was used. This selection of the library was comprised of around 12.5 million compounds at the time of access.

5.2.2.2 Filtering of Library

Following a similar line to the previous screen, a filter was used to identify compounds primarily likely to be brain penetrant. In order to maintain as wide a structural diversity as possible the stricter drug-like elements of the filter were not used with a view to implementing later if necessary. In addition, the compounds were already reduced to a small enough number to be manageable in screening.

The filter selected compounds with a mPSA of less than 80, a molecular weight of between 350 and 800 and a CLogP of between 1.5 and 2.5. This resulted in a subset of around 300000 compounds from the initial 12.5 million.

The predictive ADMET tool in discovery studio was used to calculate the brain penetrance, and the molecules were then further filtered to keep molecules predicted to have a high brain penetrance. This filter then reduced the number of compounds to the final library of around 10000 compounds.

5.2.2.3 Selection and Validation of a pharmacophore model.

Three pharmacophores were generated and tested against a list of known Keap1 binders to see if any single pharmacophore could identify all known binders of Keap1.

The pharmacophores were generated from three different receptor-ligand complexes; the first pharmacophore was generated from the ETGE peptide in complex with Keap1 using the PDB entry 2FLU as used previously in 5.2.1.3. This pharmacophore shown in Figure 89 was shown to be able to identify 15 out of 35 known actives (a 43 % hit rate), however the score of fit to the pharmacophore did not match published activities, and none of one particular set of compounds was identified (compounds from the Broad screen) as active.[211,212]

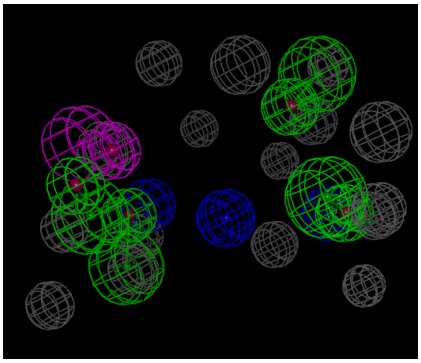


Figure 89: Pharmacophore generated from interactions between the ETGE peptide and Keap1 using PDB structure 2FLU. Showing; HBA (Green), HBD (Pink), Negative Ionisable centre (Blue), and exclusion spheres (grey).

The second pharmacophore (Figure 90) was generated from the PDB entry 41QK that contains the Evotec "compound 16" (1.4.5.2.2). This pharmacophore was able to identify 31 out of 35 compounds, a 89 % hit rate. However, again the score of fit for the molecules towards the pharmacophore did not match published activities. The pharmacophore in general scored compounds most similar to "compound 16" with the highest fit, which, as the pharmacophore is generated from this molecule is to be expected. However, the software was unable to distinguish analogues of "compound 16" by activity. In conclusion, whilst this phamacophore with its high hit rate should be suitable for selecting comopunds from a large library for further screening, the scoring function is unsuitable for predicting the activity of any selected molecule.

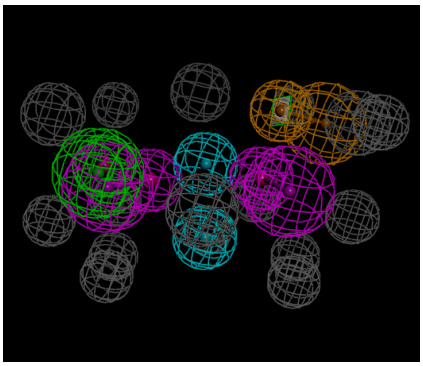


Figure 90: Pharmacophore generated from interactions between the compound "compound 16" and Keap1 using PDB structure 4IQK. Showing; HBA (Green), HBD (Pink), Aromatic π-π interactions (orange), Hydrophobic centres (light blue) and Exclusion spheres (grey).

The final pharmacophore (Figure 91) was generated from the PDB entry 4L7D that resulted from the work of Jnoff et. al. (1.4.5.2.2) with the Broad institutes compound "Hit 1". The pharmacophore was able to identify 32 out of 35 compounds, a 91 % hit rate. However, again the score of fit for the molecules towards the pharmacophore did not match published activities even within compounds of the same chemical family.

In conclusion only pharmacophore 3 and pharmacophore 2 seem to be equally suitable for selection of molecules for further study. However, no pharmacophore model is sophisticated enough to have a corrolation between the fit of the pharmacophore model and that of published activities.

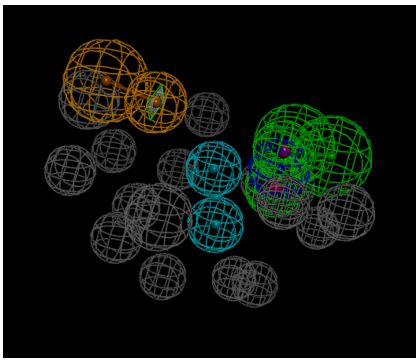


Figure 91: Pharmacophore generated from interactions between the compound "Hit 1" and Keap1 using PDB structure 4L7D. Showing; HBA (Green), Aromatic π-π interactions (orange), Negative Ionisable centre (Blue), Hydrophobic centres (light blue) and Exclusion spheres (grey).

5.2.2.4 Selection and Validation of a docking model.

A similar approach was taken with respect to the docking model. From previous discussion (5.2.1.7) 4IQK was thought to be more suitable for docking studies aimed at finding small molecules when compared to those used previously. This was mainly due to the flexibility of Arg415 in Keap1 and to a lesser extent other amino acids, this gives a conformational flexibility to the available binding surface for potential molecules. As no small molecule tool compounds were available for the previous screen, validation of that model was therefore difficult, therefore with tool compounds now available, validation of the new model could be easily performed. With this in mind two docking algorithms were trialled. The CDocker algorithm and the GOLD algorithm. CDocker is a more computationally intensive version of the Libdock protocol as it utilises molecular mechanics (CHARMM) to energy minimise molecules in the active site in their most favourable energetic conformation, and enables flexible protein residues to be chosen. The second algorithm is the GOLD algorithm discussed above, this algorithm can also cope with conformational flexibility so is suitable for adding the flexibility seen within the binding site.

CDocker was unable to dock the test molecule "compound 16" back within the pocket it had been removed from. From Figure 92 it can be seen that the overlay of the docking poses is very poor and hence CDocker was regarded as unlikely to be a suitable algorithm for docking small molecules in this binding pocket.

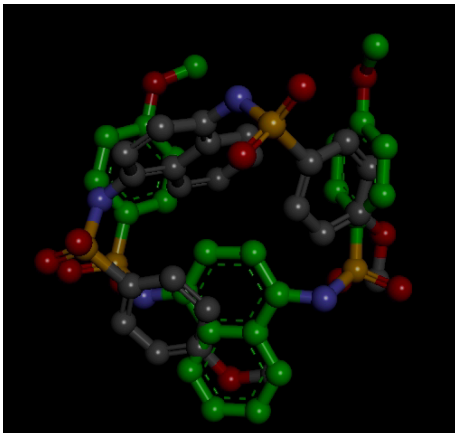


Figure 92: Comparison of the docking pose of compound 16 (grey) with the experimental data from the crystal structure (green).

In the GOLD results the software was much more successful and was able to create docking poses very similar to that of the experimental data, validating the model. The top three scoring docking poses all had a good RMSD of 1.21 Å to the binding conformation seen in the crystal structure (Figure 93).

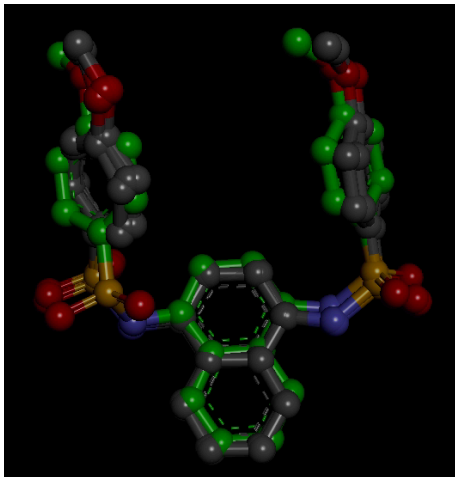


Figure 93: Comparison of the top three scoring docking poses of compound 16 (grey) with the experimental data from the crystal structure (green).

The ability of the GOLD algorithm to successfully predict the binding pose of the test molecule shows that it is suitable for use in further screening experiments, where due to the genetic algorithm utilised by GOLD compound 16 will continue be used as validation in all experiments, to ensure continued accuracy.

5.2.2.5 Screening of the new *in-silico* library

The 10000 molecules were then screened using both pharmacophore models. Molecules that were able to match both pharmacophores (2686 molecules) were then progressed through to docking studies using the GOLD algorithm.

Results from the screen were then analysed by cluster analysis with the top 1000 molecules split into 100 cluster centres using the FCFP_6 fragment collection. Molecules were then selected for purchase based on the largest diversity (by number of cluster centres) that could be gained from the fewest number of suppliers in order to lower costs. A set of 90 ligands was then purchased for testing *in-vitro.*

5.2.2.6 Fluorescent polarisation assay screening

All 90 ligands were screened in the fluorescence polarisation assay to determine their ability to bind Keap1 and inhibit Nrf2 from binding. This was performed as previously described using a $\log_{10}$ based eight-point dose response curve starting at 1 mM concentration. From the results seen in Figure 94, most compounds appeared to increase the fluorescence polarisation signal rather than to have the desired effect of decreasing the signal. This is mainly due to autofluorescence of the compounds in the excitation emission wavelengths of the assay. As previously discussed in chapter 3 this can cause erroneous readings in fluorescence polarisation signals and can cause increases as well as decreases in signal. An example of this can be seen in Figure 95 where two compounds that have dose dependent autofluorescence can be seen to have opposing, interfering effects on the fluorescence polarisation signal.

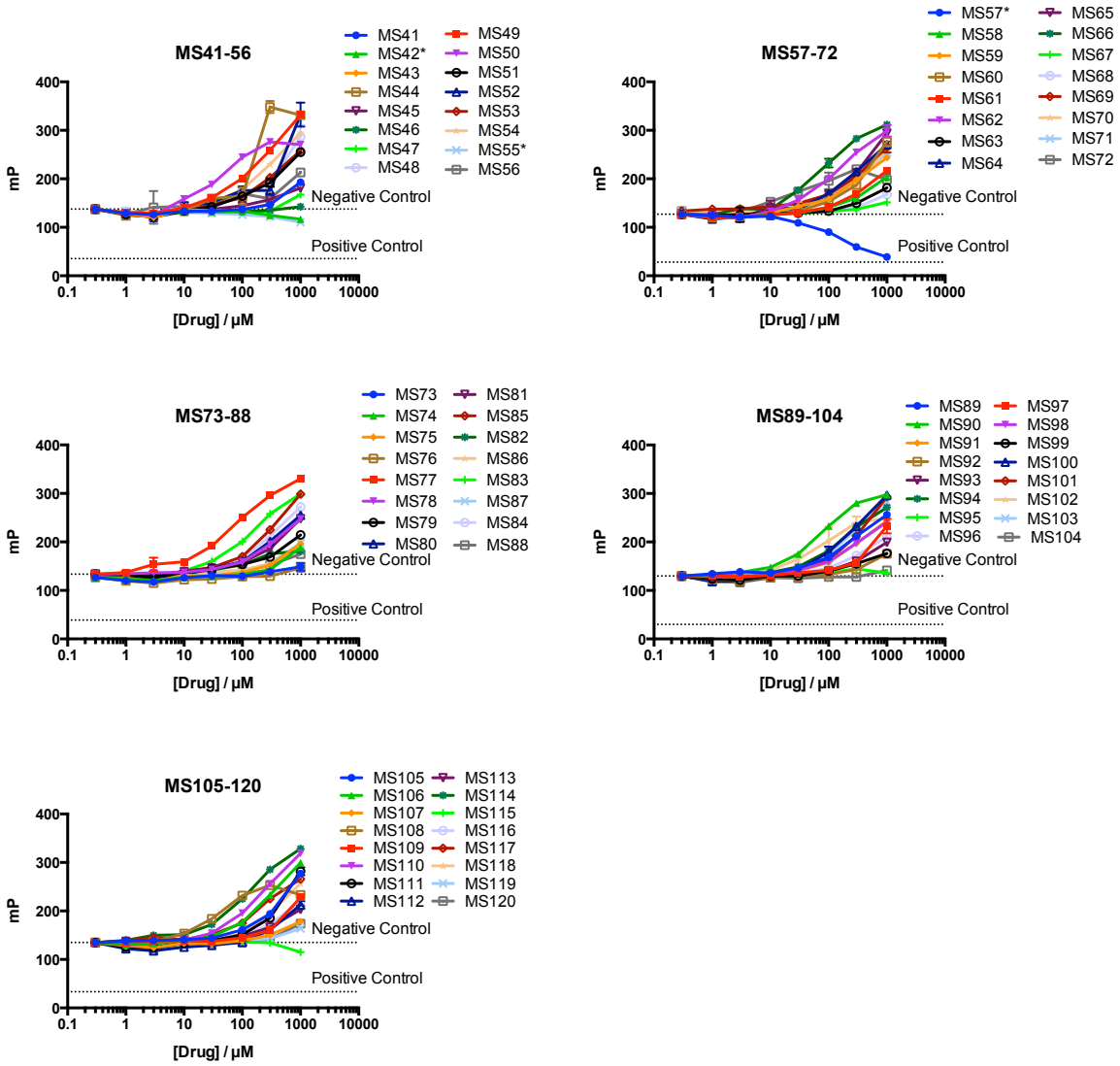


Figure 94: Dose response fluorescent polarisation assay of compounds selected from *in-silico* screen (n=3). Average of fluorescent polarisation values plotted with errors shown as standard deviation. Positive compounds are those that significantly reduce the fluorescence polarisation signal below the level of the negative control (shown as the uppermost dotted line). Compounds that increase the fluorescence polarisation signal above the level of the negative control do not represent assay hits and have been found to be autofluorescent in the wavelengths of the assay and hence interfere with the signal.

Whilst most compounds increased the signal due to autofluorescent interference, a few compounds did decrease the signal at the highest concentration of 1 mM. Compounds MS42, MS55, and MS115 were shown to decrease the signal slightly, however none of these compounds are significantly active enough to take forward as all have IC_{50}'s > 1 mM which would not meet the criteria for a hit compound. Whilst MS57 did appear to show a dose response, further interrogation of the data revealed this to be a fluorescence-based effect, as could be seen from Figure 95.

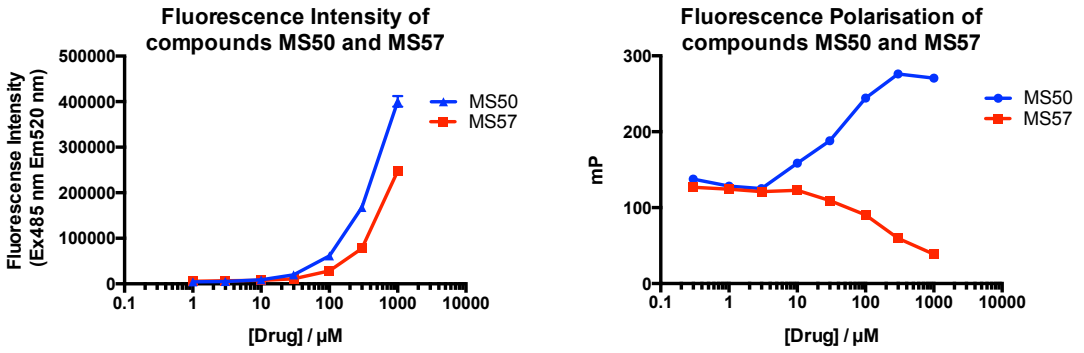
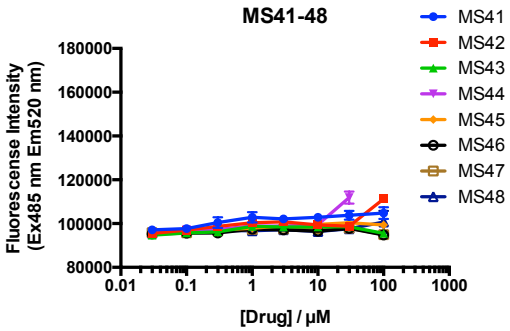


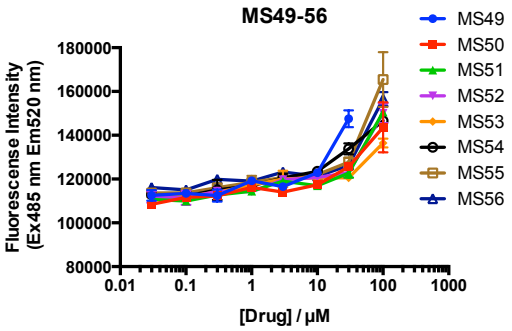
Figure 95: Dose response graphs of MS50 and MS57 showing how autofluorescence can cause either an increase or decrease in fluorescence polarisation signal. Left panel shows the how the variation in concentration of MS50 (blue), and MS57 (red) affects fluorescence intensity with error plotted as standard deviation (n=3) (y-axis). The right panel shows the effect of a range of concentrations of compounds MS50 (blue) and MS57 (red) on fluorescence polarisation with error plotted as standard deviation (n=3) (y-axis).

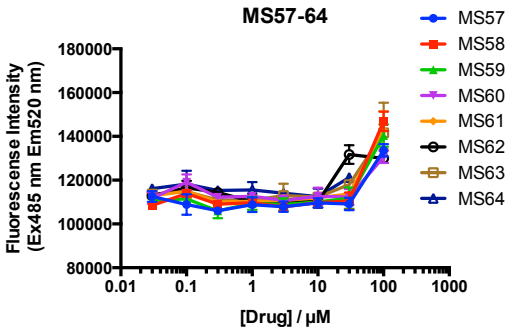
5.2.2.7 Cellular Screening – C6-ARE Reporter Assay

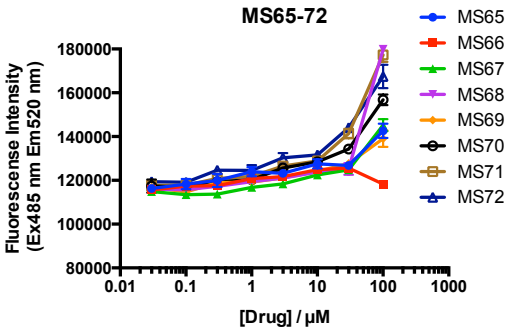
Even though no compound met the criteria for hit selection, all the compounds were tested out of interest in the C6 cellular reporter assay to assess their ability to activate ARE transcription (Figure 96). As previously described the assay uses a 4xARE-TK-eGFP promoter to give an eGFP signal upon binding of Nrf2 to the ARE sequence.

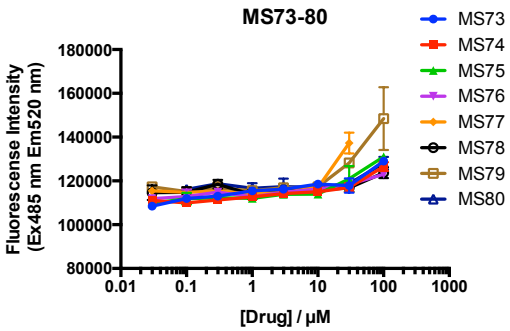
From Figure 96 it can be seen that the majority of compounds fail to produce a dose dependent response in the assay. Many compounds however appear to have activity at only the top concentration (or the top concentration before toxicity mediated cell death is observed), with only MS70, MS71, MS72, and MS79 showing an increased response at two different concentrations.

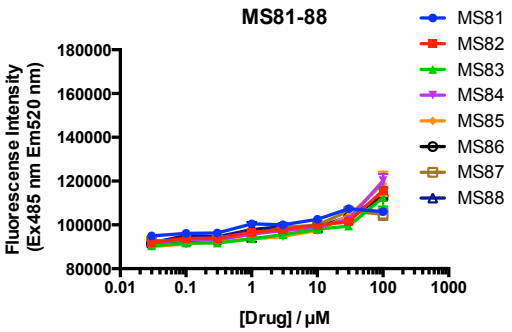

MS41-48
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS41
MS42
MS43
MS44
MS45
MS46
MS47
MS48

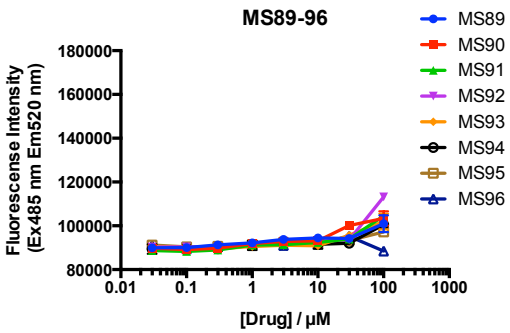

MS49-56
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS49
MS50
MS51
MS52
MS53
MS54
MS55
MS56

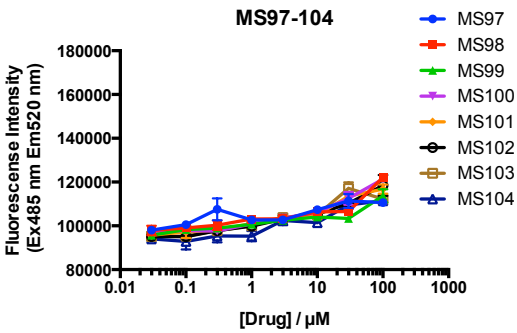

MS57-64
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS57
MS58
MS59
MS60
MS61
MS62
MS63
MS64


MS65-72
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS65
MS66
MS67
MS68
MS69
MS70
MS71
MS72


MS73-80
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS73
MS74
MS75
MS76
MS77
MS78
MS79
MS80


MS81-88
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS81
MS82
MS83
MS84
MS85
MS86
MS87
MS88


MS89-96
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS89
MS90
MS91
MS92
MS93
MS94
MS95
MS96


MS97-104
Fluorescense Intensity (Ex485 nm Em520 nm)
180000
160000
140000
120000
100000
80000
0.01
0.1
1
10
100
1000
[Drug] / µM
MS97
MS98
MS99
MS100
MS101
MS102
MS103
MS104

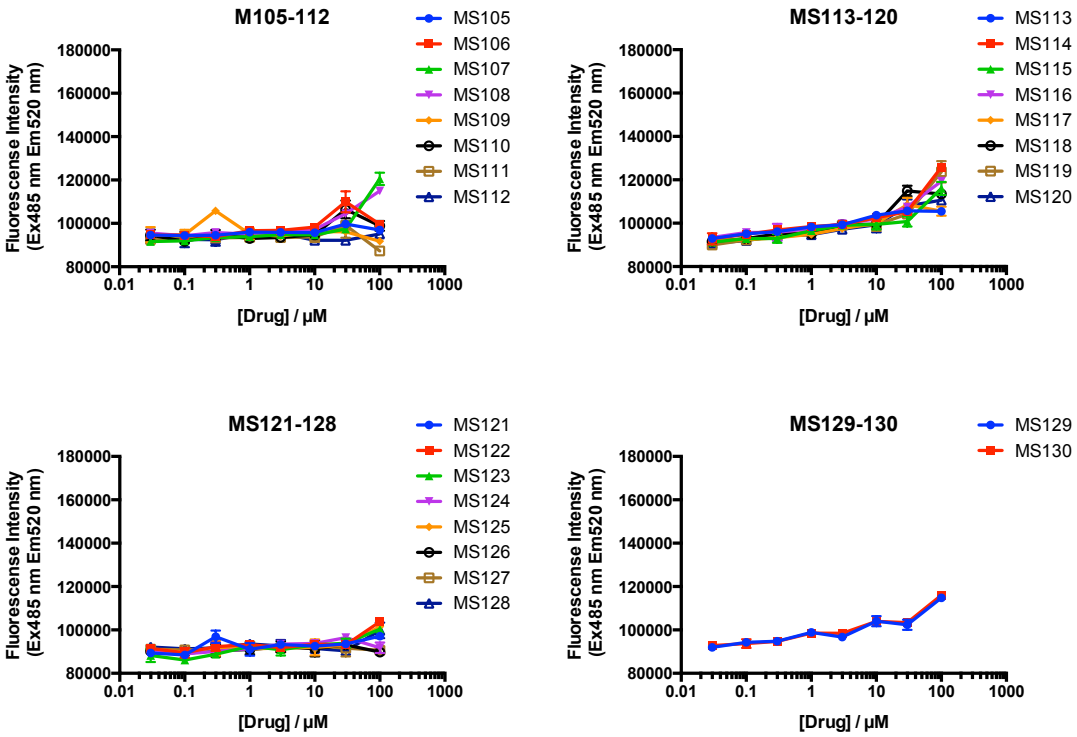


Figure 96: ARE reporter cell line assay. "ARE" is 4xARE-TK-eGFP cell line for ARE-Nrf2 reporter activity. Legend on right of each graph indicates the compound number. Each panel shows results of individual cell assays with a maximum of 8 compounds run in each assay. Positive responses in ARE activation are indicated by increases in fluorescence intensity. Each compound is tested 8 concentrations ranging from 100 µM to 0.03 µM (x-axis), and their relevant fluorescent intensities (Ex485 nm, Em520 nm) corresponding to eGFP induction with error plotted as standard deviation (n=3) are shown.

A potential explanation for positives could be the autofluorescent nature of the compounds seen from the FP assay. Therefore compounds with a dose dependent response and those with a significant increase in fluorescent intensity at the final concentration were tested alongside the ARE negative cellular line (TK-eGFP) to assess whether any effect seen was ARE dependent.

Figure 97 shows the tested compounds, with BG12 a known reactive Nrf2 activator included as a control. From the assay it could be said that no compound was able to achieve a response similar to BG12. Compounds MS56, MS68, MS70, MS71, and possibly MS72 looked like they were able to mildly activate ARE at the highest concentration of 100 μM. However, none of the compounds that showed mild activity in the fluorescence polarisation assay showed activity in the cellular assay, which suggests that either any effect these compounds may be having, is likely to be independent of direct inhibition of the Nrf2-Keap1 interaction or a much higher concentration of compound (> 1 mM) would be needed before an effect could be seen.

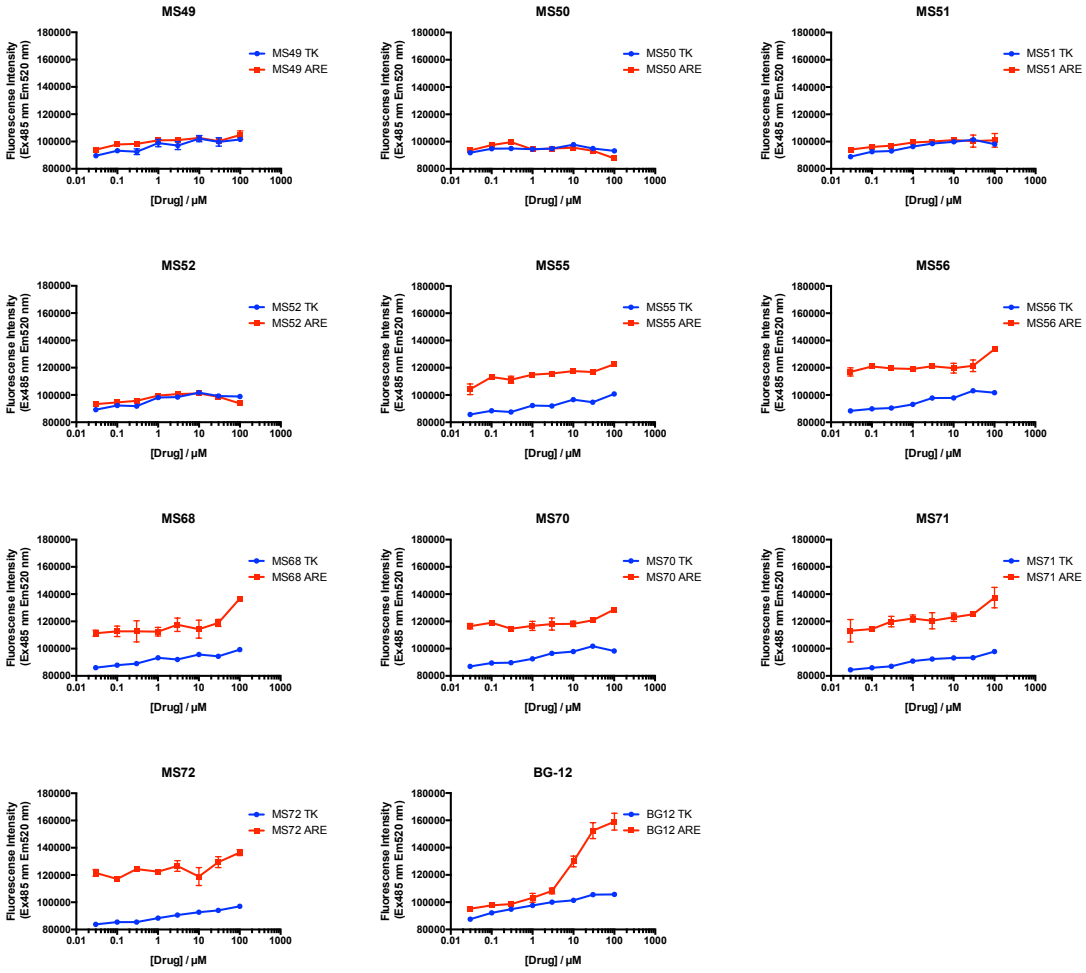


Figure 97: Selected *in-silico* compounds tested in an 8-point dose response in both cell lines. C6-4xARE-TK-eGFP shown in red and C6-TK-eGFP shown in blue. Legend on right of each graph indicates the compound number and cell line. Y-axis indicates fluorescent intensity (Ex485 nm, Em520 nm) error plotted as standard deviation (n=3). X-axis indicates compound concentration.

5.3 Conclusions and Future Work

Two virtual screening approaches were developed to try and identify novel chemical starting points for an Nrf2-Keap1 inhibitor. The initial screen developed virtual screening libraries of brain penetrant molecules that were used in *in-silico* pharmacophore and docking screens derived from the binding of the co-crystalised Nrf2 peptide containing the ETGE motif to Keap1. These screens were used to triage the molecules, which were selected for testing in the fluorescence polarisation assay. No compound was able to disrupt the interaction between the Nrf2 peptide and Keap1.

The second virtual screen was able to utilise more information, based on findings from the initial screen and also from work published in the literature since the initial screen. The *in-silico* screening procedures were rigorously optimised in each case against published Keap1 binders. A newly optimised virtual screening library was developed containing predicted brain penetrant compounds that could be easily sourced after screening, enabling a higher number of selected compounds to be available for *in-vitro* testing after screening.

This new library was submitted to the *in-silico* screening pipeline and 90 molecules were selected for *in-vitro* testing. These molecules were tested first in the fluorescence polarisation assay where many were found to interfere with the signal in the assay due to autofluorescence of the compounds. Some molecules were found to have a minimal effect in the assay, but only at the highest concentration tested of 1 mM. All molecules were then tested in the cellular reporter assay where only a few molecules were able to minimally activate an ARE response. No molecule was shown to be active in both the FP assay and the cellular assay.

The two virtual screening libraries developed are now available for screening in any additional screens and the techniques used in the development of the library can be further used to create additional virtual screening libraries that can be optimised towards any additional drug target.

The virtual screening tools developed here will also be able to look at potential binding orientations of any molecules found as a result of high throughput screening in order to assess potential modifications to hit molecules that may lead to an increase in binding affinity or to improve the pharmacokinetic properties of the molecule whilst retaining potency.

6. Characterisation of mammalian Keap1 protein.

6.1 Introduction

As previously discussed in Chapter 1 (1.4) the Keap1 mediated regulation of Nrf2 is an important pharmacological mechanism for the reduction of oxidative stress and the mechanism by which this is achieved has been widely researched. The regulation mechanism described in 1.4.4 being the most widely accepted mechanism. Since establishment of this mechanism there has been a wide interest in the structural biology of Keap1 and how this affects its ability to regulate Nrf2.

Keap1 is a cysteine rich 69-kDa protein containing 627 amino acids, of which 27 are cysteines. These cysteine residues are distributed throughout the protein and can undergo oxidation or electrophilic modification. Different electrophilic compounds have different propensities to react with individual cysteine residues.[306–308] Specific cysteines can also vary in function depending on the prevailing cellular environment, Cys273 and Cys288 act under both basal and stressed conditions to control levels of Nrf2, whereas Cys151 is primarily only activated under stressed conditions.[186,187,202,309]

Keap1 is the substrate adaptor for the ubiquitin ligase complex assembly comprising of Cul3 and Rbx1. These self-assemble to form a Cul3-RING ligase, which is responsible for ubiquitination of Nrf2.[181–183] Keap1 is a member of the BTB-Kelch family of proteins and is comprised of five domains. The N-terminal domain, the BTB domain, the central IVR domain, the DGR domain, and the C-terminal domain. The BTB domain mediates homodimerisation of the protein and is also important for forming interactions with the H2 and H5 helixes of the first cullin repeat for binding to Cul3.[256] The central IVR domain contains a 3-box motif that is partially involved in Cul3 binding where a small hydrophobic grove binds the N-terminal extension of Cul3.[200] The IVR domain also contains the majority of the key reactive cysteine residues that act as sensors towards oxidation as well as the nuclear export signal (NES) sequence.[181,184] The final domain is the DRG or Kelch domain, which binds the ETGE or DLG motifs of Nrf2.[226,246,272,310] The DRG domain consists of 6 Kelch repeats, and contains some of the reactive cysteine residues believed to be involved in redox sensing.[185] Of these cysteine residues seven of them are known to be highly reactive towards ROS and electrophiles. These cysteine residues are thought to act as redox sensors and are very important in the regulation of Nrf2.[186,187]

Whilst a full length crystal structure has not been solved single particle electron microscopy has shown Keap1 to be a cherry-bob-like structure.[181]

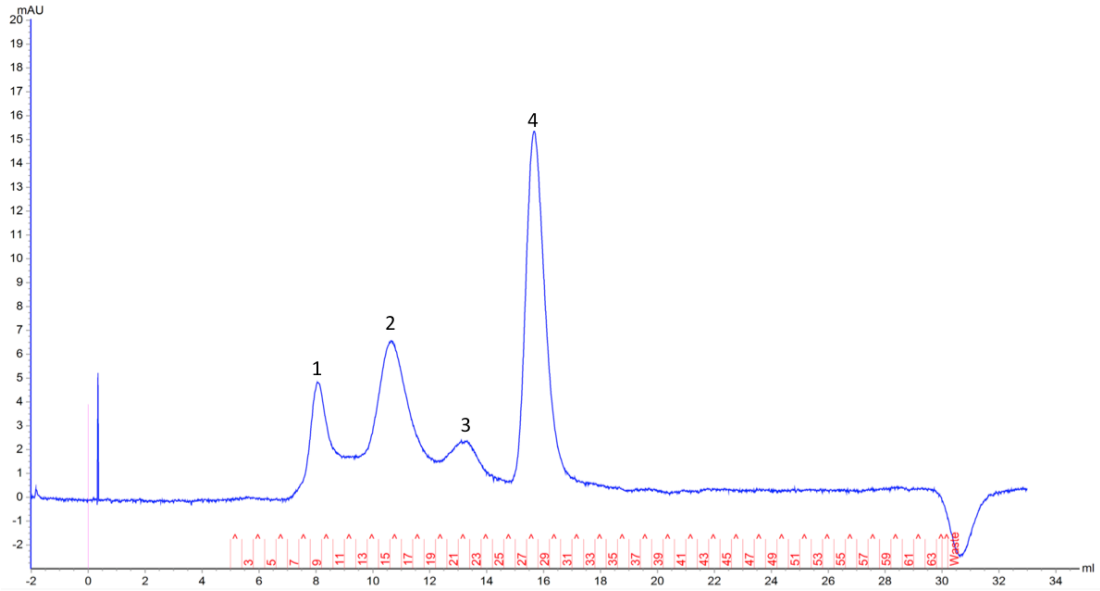
The majority of structural and biochemical work in the literature has focussed on the isolated Kelch domain expressed in bacteria and the assumption has been made that full length Keap1 exists as a dimer as it is well known for other BTB containing proteins.[255,311] The only literature support for the dimerisation of Keap1 comes from the work on full length mouse Keap1. Crosslinking studies performed by Dinkova-Kostova et al. showed dimeric forms of the protein, whilst electron microscopy studies of the same protein showed a cherry-bob shaped dimeric structure.[181,234] No characterisation has been performed on the full-length human mammalian expressed form of the protein. Characterisation of this form may help to elucidate the function of the protein in humans and whether it differs from that found for bacterially expressed isolated Kelch domain, or that of bacterially expressed full length mouse protein. To this end we set out to establish whether full length human Keap1 expressed in mammalian cells existed as a dimer. However, due to protein constraints, additional further work was conducted on the baculovirus expressed human Keap1 truncated at the C-terminus by 6 residues (as used in 4.2.2).

6.2 Results and Discussion

6.2.1 Size Exclusion Chromatography

6.2.1.1 SEC under physiological conditions

The first aspect of the purified protein we wanted to look at was its behaviour on gel filtration size exclusion chromatography (SEC). The technique is able to separate proteins based on their size, and therefore is able to give a good estimate of the molecular weight of different protein forms. Monomeric and dimeric species were predicted to be the most common observed structural forms seen under physiological conditions.



Peak	Average Elution Volume (n=4)	Calculated Molecular Weight (KDa)	Number of monomeric units
1	7.95	1472.8	21
2	10.49	279.9	3.8
3	13.005	54.1	n/a
4	15.43	11.1	n/a

Figure 98: Representative gel filtration trace of 100 µg Keap1 injection with 1 mM DTT. UV trace at 280 nm shown in blue, fractions collected shown in red. Table shows elution volumes, estimated globular molecular weights from protein standards, and approximate number of monomeric units of Keap1 by weight.

Under physiologically similar conditions (20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.5) four peaks were seen (Figure 98), and after western blotting of all fractions only peaks 1 and 2 were seen to contain a band at the correct molecular weight for Keap1 (Figure 99). The remaining peaks showed no band via western blot analysis.

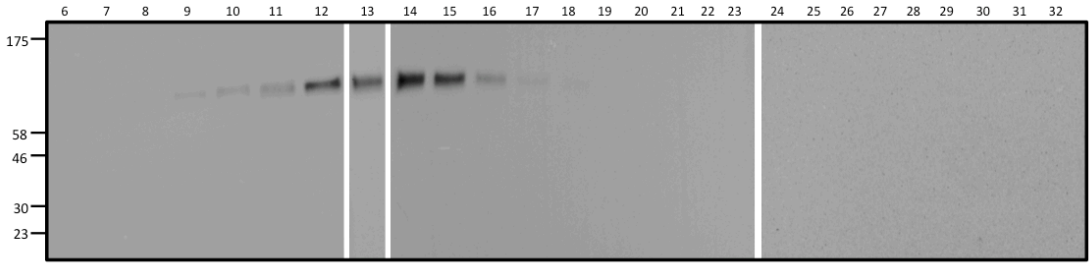


Figure 99: Western blot images of SDS-PAGE gels loaded with 15 µL of each sample from the reducing gel filtration fractions and probed with Anti-Keap1 antibody. Each fraction was loaded with a volume of 15 µL of sample with 5 µL 4x laemelli buffer. Fraction numbers are indicated along the top of the image and molecular weight standard weights in kDa are indicated to the left of the image.

Peak 1 was estimated to have a molecular weight of ~1473 kDa. This should be considered a crude estimate as there was no calibration standard available at this molecular weight, and it is also very close to the void volume of the column. This made it very likely that this was a solubilised aggregated form of the protein. Previous literature has not mentioned the propensity of the protein to form solubilised aggregates despite the presence of multiple cysteine residues in Keap1, which could theoretically cause intermolecular disulphide bonds and lead to aggregation in concentrated protein preparations. The presence of soluble aggregates under reducing conditions is therefore interesting as this suggests that another mechanism may be causing aggregation of the protein.

Peak 2 corresponded to a molecular weight of around 286 kDa which would indicate a tetrameric form of Keap1 may exist in the human form of the protein, contrary to the literature precedence for the dimeric form of the protein in *mus-musculus*.[181]

Peak 3 corresponds to a potential degradation product which can also be seen on the input coomassie (Figure 100) at around 50 kDa, and does not appear by western analysis.

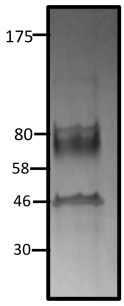


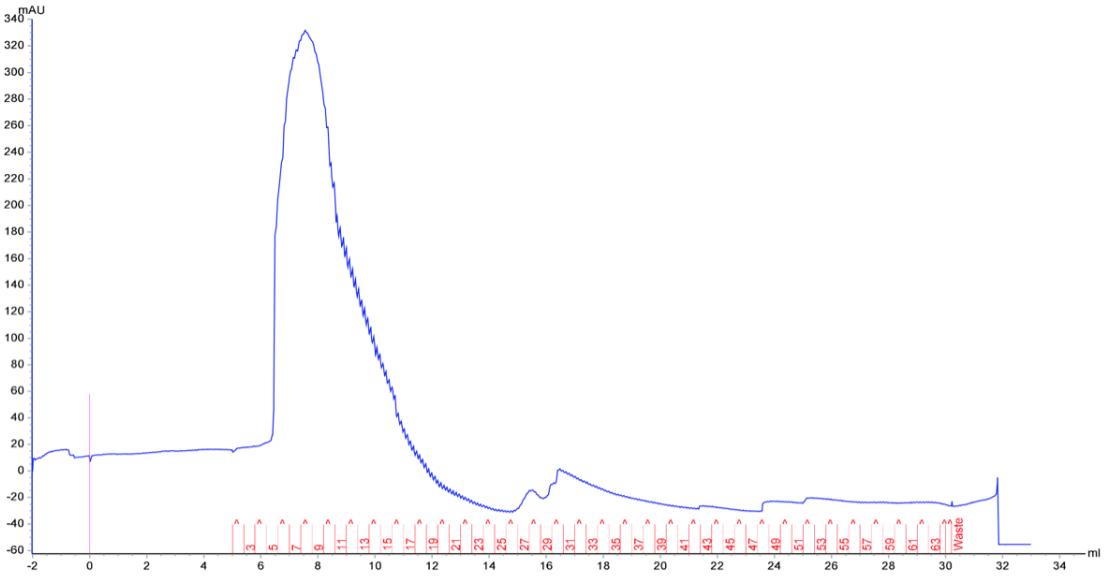
Figure 100: Coomassie stained SDS-PAGE gel of the Keap1 batch used for Gel Filtration column chromatography. Molecular weight standard weights in kDa are indicated to the left of the image.

Peak 4 at a molecular weight of around 10 kDa is also visible when the storage buffer of the protein was run on the column as a control (data shown in Appendix 6). The most likely constituent of the buffer to cause the peak was thought to be residual surfactant NP-40 as surfactants are able to form large micelles that can often have large micelle molecular weights and globular size.

The most informative peaks are therefore peaks 1 and 2, as these are the only peaks which correspond with Keap1 by western analysis. These peaks correspond to globular molecular weights of aggregated (1) and tetrameric (2) protein forms of Keap1.

6.2.1.2 SEC under non-reducing conditions

Previous work (3.2.3.2) had shown that the maximal obtained signal in the FP assay was reduced upon the addition of DTT, with no effect on the IC_{50}'s observed for known small molecule binders. This seemed to suggest that DTT might have an effect on reducing the amount of larger aggregated forms of the protein, and thus reducing the FP signal itself. In order to test this, the protein was analysed by size exclusion chromatography without the addition of DTT (Figure 101).



Peak	Average Elution Volume (n=4)	Calculated Molecular Weight (KDa)	Number of monomeric units
1	8.05	1379.6	19.6

Figure 101: Representative gel filtration trace of 100 µg Keap1 injection without DTT. UV trace at 280 nm shown in blue, fractions collected shown in red. Table shows elution volumes, estimated globular molecular weights from protein standards, and approximate number of monomeric units of Keap1 by weight.

The results show a single broad peak by gel filtration SEC at an elution volume of 8 mL matching that of the first peak in the reduced sample, this was hypothesised to be an aggregated form of the protein. The peak is also so broad it may mask any peaks seen at elution volumes that come before ~14 mL. Therefore, it cannot be said for definite whether any other forms of the protein are present under non-reducing conditions. Western blot analysis (Figure 102) showed that Keap1 is present. The intensity of the bands by western matched the intensity of the gel filtration peak by UV. Interestingly, the UV adsorption levels were significantly higher for the same sample injection volume when compared to the reduced sample. This can be accounted for by the increased propensity of larger molecular structures to be able to absorb increased amounts of UV due to the larger number of aromatic rings and π-bonds within the structure able to absorb UV light at 280 nm. Due to this phenomenon it may be true that the amounts of aggregated protein in the reduced sample as suggested by UV intensity could be misleading and could therefore appear to be a greater amount than actually exists.

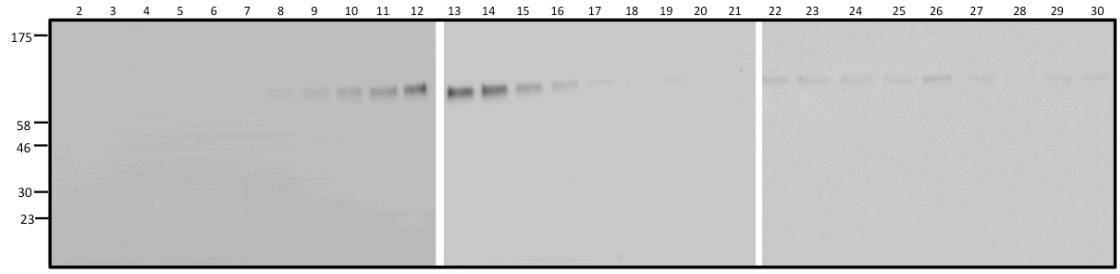


Figure 102: Western blot of non-reducing gel filtration fractions of equal volume loading imaged with Anti-Keap1 antibody. Fraction numbers are indicated along the top of the image and molecular weight standard weights in kDa are indicated to the left of the image.

6.2.2 Could Keap1 form tetramers and aggregates? A literature and *in-silico* exploration

Results from gel filtration suggested that under physiological conditions the major form of the protein is a tetramer. The hypothesis that homo-tetrameric forms of Keap1 could be formed was then investigated further.

6.2.2.1 Literature precedence

It is well known in the literature that dimeric forms of the protein exist.[181,183,186] With the most compelling evidence coming from the electron microscopy images collected by Ogura et. al. In this paper the authors used single particle electron microscopy to create a 3D representation of mus-musculus Keap1 (Figure 103). What is interesting to note is that this paper uses a gel filtration methodology to isolate the dimeric species with their SEC showing only one peak, contrary to the profile seen in 6.2.1. Detailed methodology and buffer compositions are not given, so a direct comparison with the methodology used here cannot be made.

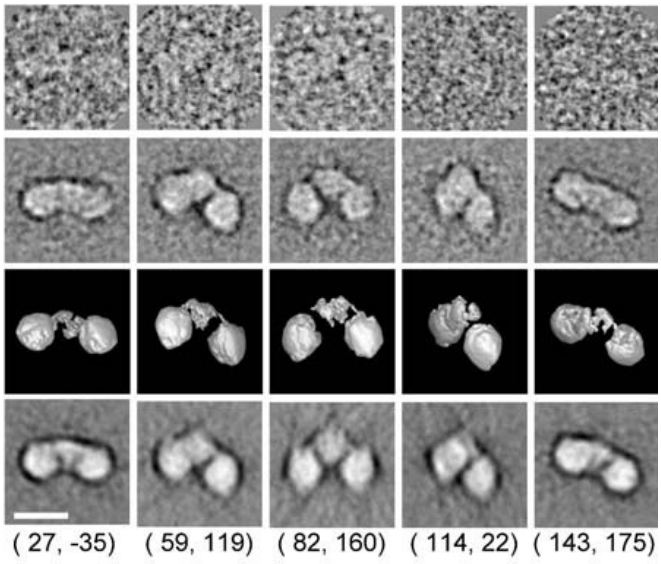


Figure 103: Raw electron micrograph images with different Euler angles (Row 1). Row 2 shows two-dimensional averages, row 3 shows the three-dimensional surface reconstructions of the protein. Row 4 shows the re-projections of the three-dimensional images showing consistency of image through reconstruction. The Euler angle (β, γ) is indicated below columns. Protein is displayed in bright shades. Scale bar shown is 100 Å. Reproduced with permission from Ogora et. al. (2010).[181]

Keap1 is part of a family of BTB/POZ domain containing proteins. Poxvirus and Zinc Finger (POZ) domain is an alternate name for the previously described BTB domain in proteins which contain a zinc finger.[312] BTB domains are well known as dimerization domains.[312–314] However, the literature suggests there is precedence for BTB/POZ domain containing proteins to form higher order oligomers.[255] In fact several families of BTB proteins are able to form species larger than dimers. For example the T1 family, which is usually contained within ion channels promotes tetramerization.[315] In addition members within the family of BTB containing proteins with Zinc finger, Kelch, MATH, and Rho domains have also been shown to form dimers, oligomers and complexes with other proteins. An example of this is the drosophila protein GAGA, which has been shown to form homo-oligomers. These oligomers are able to interact with a higher affinity than monomers to their DNA target.[316] Additionally, a BTB-MATH protein SPOP (a cullin-3 E3 ubiquitin ligase substrate adaptor similar to Keap1) is able to form higher order oligomers.[317,318] The formation of these oligomers has interestingly been shown to increase the catalytic efficiency of the E3 ligase complex, and attenuation of this activity can be caused through addition of a self-assembly inhibitor.[317]

The structural basis for SPOP oligomerisation has been determined by X-ray crystallography.[318] This study showed that the protein was able to form oligomers through a second dimerization domain on the C-terminal end of the protein (Figure 104). Further analysis through chemical cross-linking, and mutagenesis showed that it was the C-terminal domain that was responsible for the oligomeric forms of the protein, with Tyr353 playing an important role.[318]

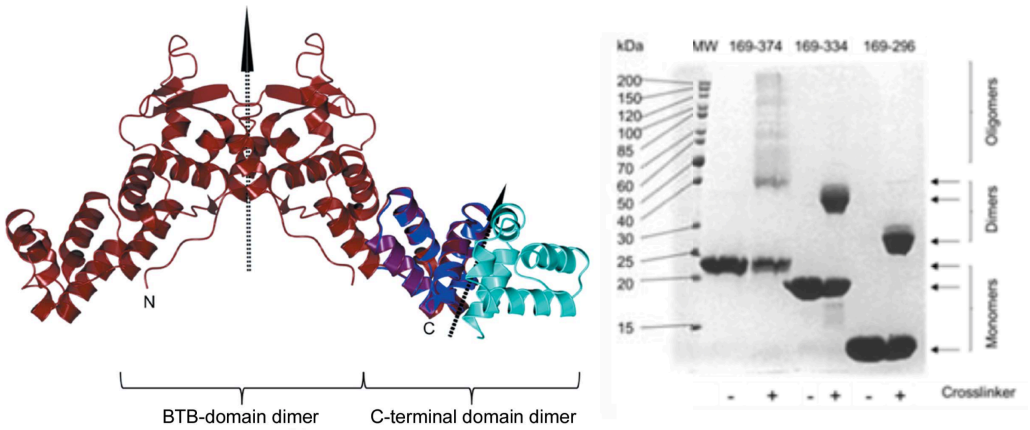


Figure 104: SPOP dimerisation. Left: Superposition of the X-ray crystal structure of SPOP C-terminal domain dimer (blue and cyan) with the X-ray crystal structure of SPOP$_{169-374}$ Y353E dimer (brown). Right: SDS-PAGE gel showing cross-linking of different truncated forms of SPOP indicating that only when residues 334-374 (including Tyr353) are incorporated does oligomerisation occur. Reproduced with permission from van Geersdaele et. al. (2013).[318]

6.2.2.2 *in-silico* comparisons between Keap1 and SPOP

With our results indicating oligomeric forms of Keap1, and the literature on SPOP suggesting that an additional homodimerisation interaction could occur through the C-terminal end of the BTB domain, the potential for Keap1 to form similar interactions was evaluated *in-silico*. Firstly, the sequence similarity of the hypothesised interaction domain between the two proteins was evaluated. This showed that the two regions had some sequence similarity (31.8 %), but not to a significant extent (Figure 105).

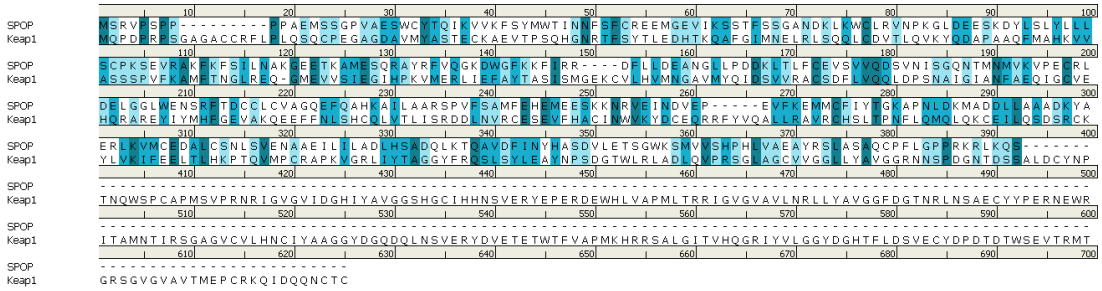


Figure 105: Sequence alignment of SPOP and Keap1 proteins showing some sequence homology in the BTB domains. The residue number is indicated along the top of the sequence. Sequence similarity is indicated with darker shades of turquoise indicating a higher sequence similarity, with the darkest shade showing exact sequence matches.

The structural similarity was evaluated next. The X-ray crystal structures of the BTB domain of Keap1 were superimposed with the BTB domain of SPOP and were shown to have a reasonable secondary structural similarity within the BTB domain with an RMSD of 4.46 Å (Figure 106). The SPOP crystal structure included the amino acids in the oligomerisation domain and this can be seen in the overlaid structure in Figure 106 indicated in pink. It can be seen that this part of the protein forms an alpha helical structure continuing on from the C-terminal end of the main BTB domain. It is therefore unfortunate that the crystal structure of the BTB domain of Keap1 stops just short of the amino acids which would be required to form the proposed second dimerisation domain (2DD). The C-terminal oligomerisation domain of SPOP contains five alpha helixes with oligomerisation interface contained within the final three helixes.

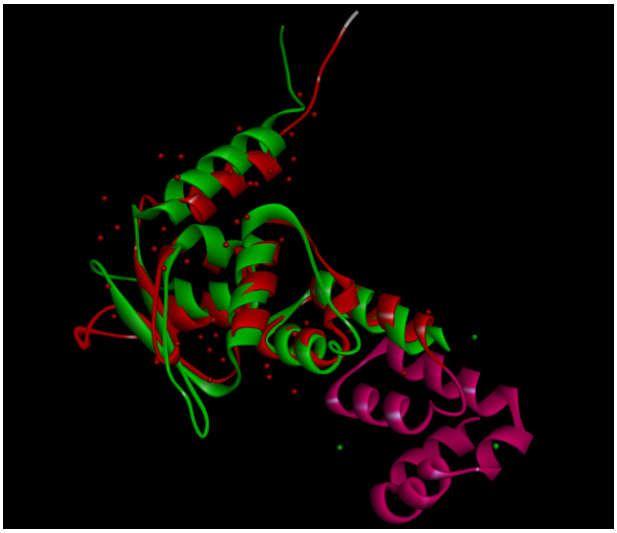


Figure 106: Structural superposition of the BTB domains of SPOP (PDB ID: 4J8Z) and Keap1 (PDB ID: 4CXI). The BTB domain of Keap1 is illustrated in green, the BTB domain of SPOP is illustrated in red, with the C-terminal oligomerisation domain illustrated in pink.

The secondary structure of this section of Keap1 was therefore predicted *in-silico* using a web tool called Protein Predict.[319] The results showed that the amino acid sequence related to this region was predicted to have an alpha helical structure (Figure 107), similar to that in seen in SPOP, with several regions of predicted protein-protein interaction activity.

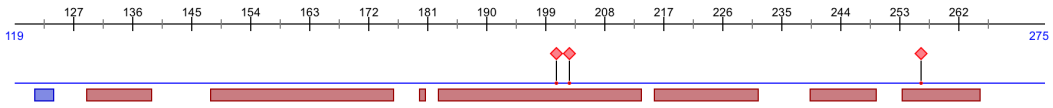


Figure 107: Secondary structure prediction for residues contained within the suggested interaction domain region. Beta-sheet regions are shown as blue bars, alpha-helical regions are shown as red bars and predicted protein-protein interaction regions are shown with red diamonds.

Next a homology model of Keap1 was created. The model was created to help visualise how the helical region of Keap1 may fold based on a more similar structure from the same protein family (KLHL11). The crystal structure of the BTB domain of Keap1 and KLHL11 (a Kelch family member with a BTB domain containing a helical region C terminal to the BTB domain) were used to create the model (Figure 108).

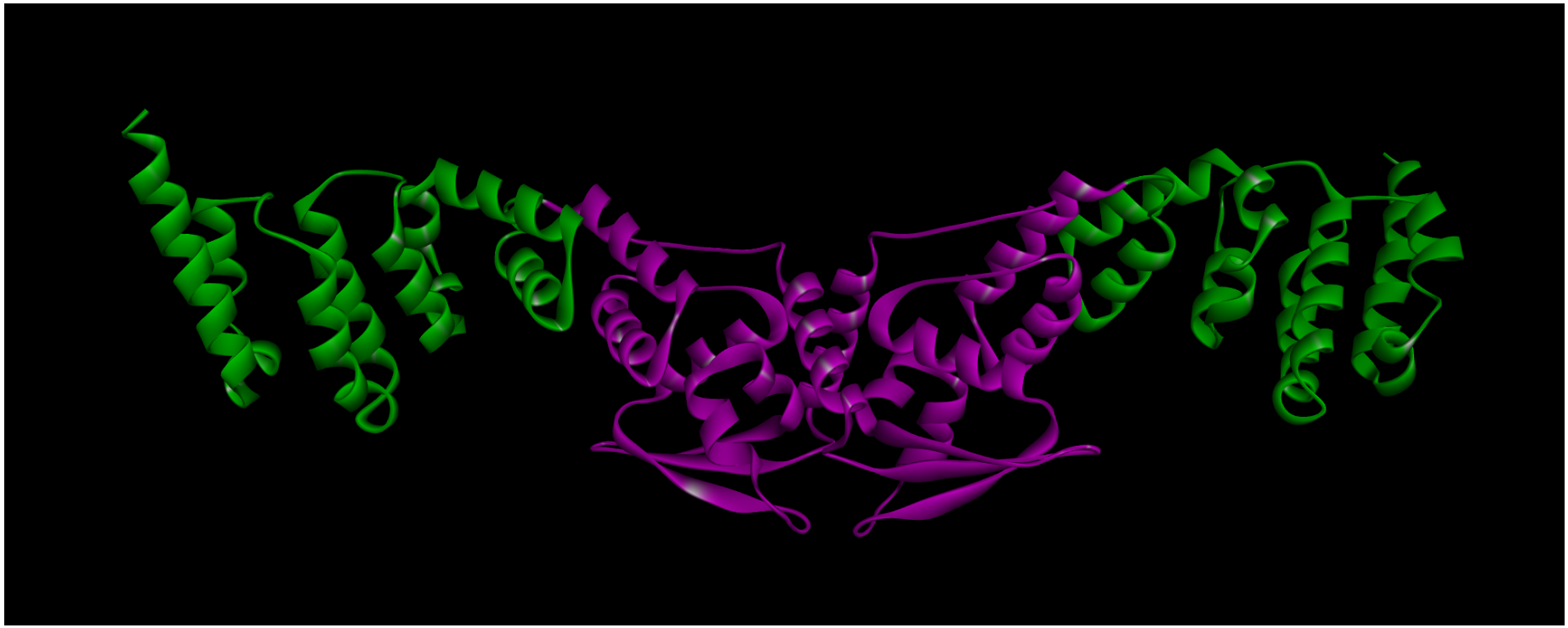


Figure 108: Homology model of Keap1 dimer (residues 50-315) showing the BTB domain (pink) and the extended alpha helical region (green) which forms part of the BACK or IVR domain of Keap1.

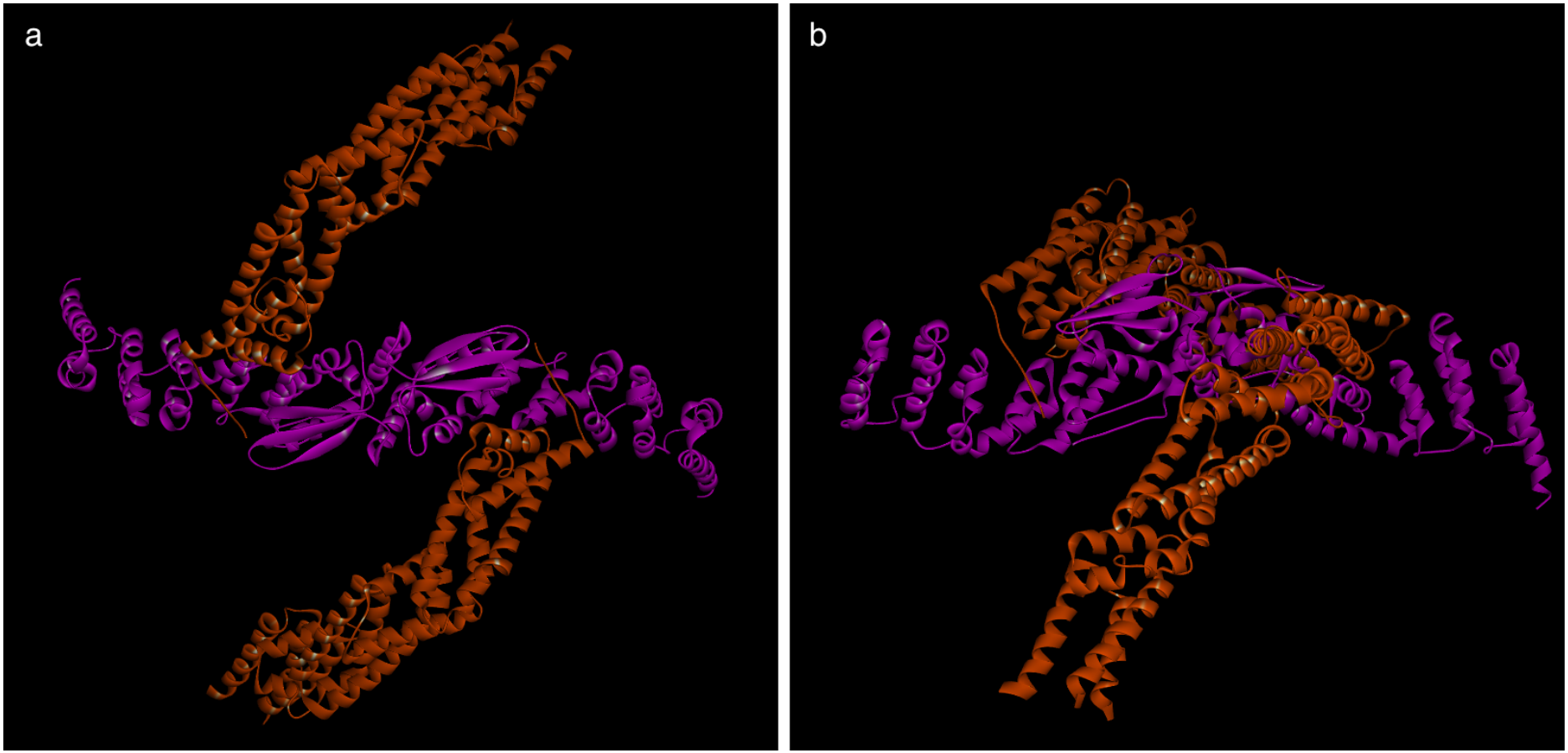


Figure 109: Dimeric model of Keap1 (residues 50-315) indicating how Cul3 (orange) may bind the Keap1 (pink) shown from two viewpoints a) shows an aerial view and b) shows a direct side on view of the Keap1 dimer similar to that shown in Figure 108.

The Cul3 binding orientation was then modelled using the KLHL11 structure as a template. This model was used to indicate how Cul3 binding may occur in Keap1 relative to the rest of the IVR and the potential interaction domain (Figure 109).

The protein surface of the homology model of Keap1 was then investigated. The extended helical domain in the model was found to be similar to other protein-protein binding domains, with exposed hydrophobic surfaces, which are well known features of protein-protein interaction binding sites. Another interesting feature was that the binding pocket containing a sulfhydryl group known to be the target for some reactive inhibitors, including CDDO-Me, is relatively nearby to the Cul3-Rbx binding site (indicated in Figure 110) and the BTB dimerisation domain. This leads to the suggestion that binding of CDDO-Me or similar sulfhydryl binder to this site may have an effect on either the binding of Cul3 to Keap1 or of Keap1 to itself, and may provide a basis for how these reactive inhibitors disrupt the ubiquitination of Nrf2, which is still a much debated topic. However, research has shown that the reaction of the cysteines does not cause a dissociation of Cul3.[204,245] This has instead led to the most popular hypothesis that reaction of the cysteines causes a conformational change in the structure of Keap1 preventing the ubiquinitation machinery from performing its task.[204] Whether the conformational change in Keap1 has the potential to affect the binding orientation of Cul3 leading to a non-favourable conformation of Cul3 for ubiquitination is not yet known. The proximity of the CDDO-Me binding site to the Cul3 binding site suggests that this remains a possibility. Any effect on the conformation of the BTB domain (potentially through the binding of CDDO-Me) may also have the ability to disrupt dimerisation or oligomerisation.

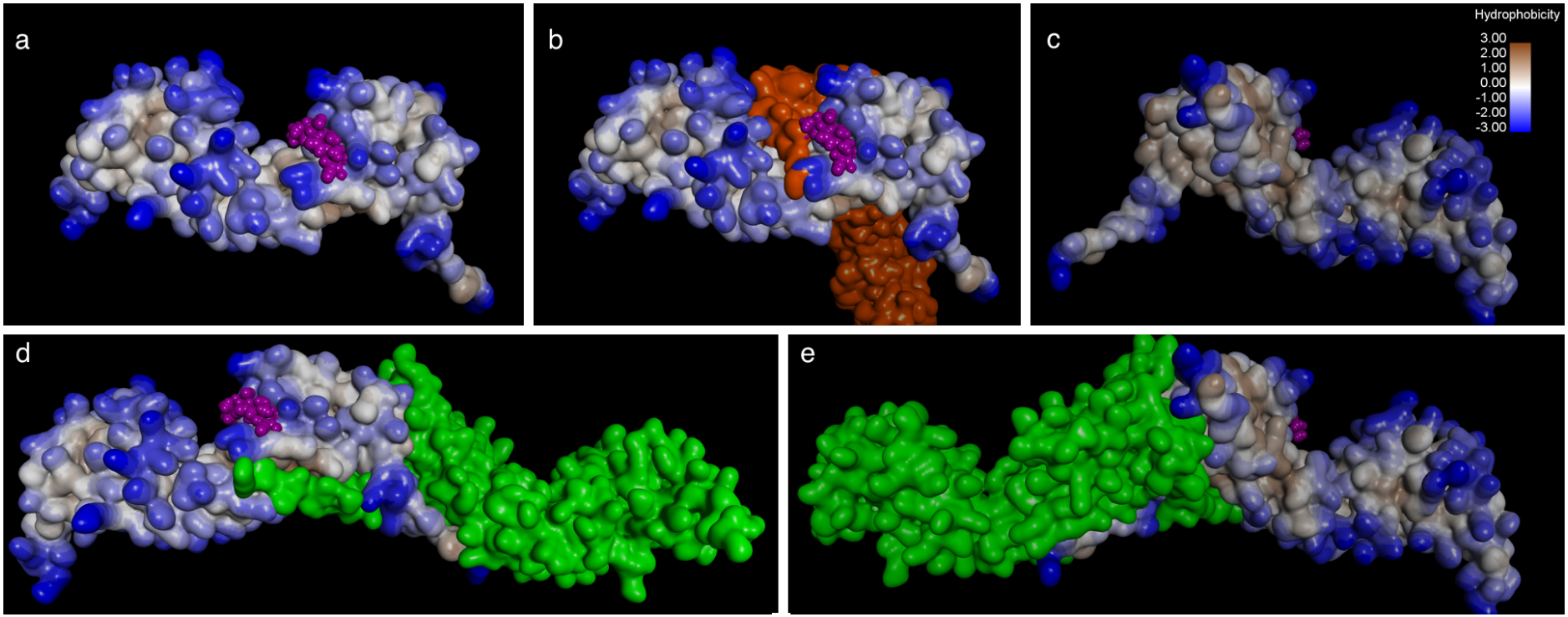


Figure 110: Various orientations of a monomeric unit of the Keap1 homology model are shown with different binding partners. a) shows the location of the CDDO-Me binding site (shown in pink space filling spheres) in relation to the rest of the model which is shown with a soft hydrophobic protein surface with blue representing the most hydrophilic areas and brown the most hydrophobic areas as indicated by the legend in c), the grove for the binding of Cul3 can be clearly seen in the centre of the molecule, and another shallower hydrophobic grove can be seen on the left of the molecule. b) shows how the protein surface of Cul3 (orange) interacts with the model in the grove seen in a) and shows its relative proximity to the CDDO-Me binding site (pink spheres). c) shows the model shown in a) rotated 180 degrees around the vertical axes, the reverse face of the protein can be seen with three notable hydrophobic faces one on the left of the molecule for the binding interface with a second molecule of Keap1, the second near the centre of the molecule for the interaction with Cul3 and a third on the right of the molecule. d) shows the binding mode for the dimeric protein with the second homomeric unit shown in green, the location of the CDDO-Me binding site (shown in pink space filling spheres) can also be seen in proximity to the dimeric binding interface. e) shows the model shown in d) rotated 180 degrees around the vertical axis.

In SPOP, the E3 activity is directly related to the formation of multimeric forms of the protein, formed through the oligomerisation domain in the C-terminal region of the protein.[317] We therefore set out to investigate whether reaction of the active site for CDDO-Me or reacting the sulfhydryl groups in general with sulforaphane lead to a decrease in the levels of multimeric forms of Keap1 which may indicate that multimeric forms of Keap1 could play a similar role.

6.2.3 Cross Linking of Keap1

In order to test the hypothesis that Keap1 is able to form oligomers, as suggested by SEC and *in-silico* studies, cross-linking was used to show the extent to which these are able to readily form. Dinkova-Kostova et al. used cross-linking to observe dimeric forms of recombinant bacterially expressed mouse Keap1 protein.[234] The authors observed only monomeric and dimeric protein, although this may have been due to the limitations in resolving power of the SDS-PAGE gels used or the fact that the protein was bacterially expressed.[234] The recombinant human baculovirus expressed Keap1 protein was cross-linked in the presence of glutaraldehyde over a period of two hours in a similar method to Dinokova-Kostova et al. However, in order to investigate higher order oligomers an altered western blotting protocol was developed to enable proteins larger than 140 kDa to be imaged. The new protocol utilised a 3-8 % Tris-Acetate gel instead of the standard Bis-Tris gels, along with a high molecular weight marker to enable proteins of up to 460 kDa to be characterised by weight.

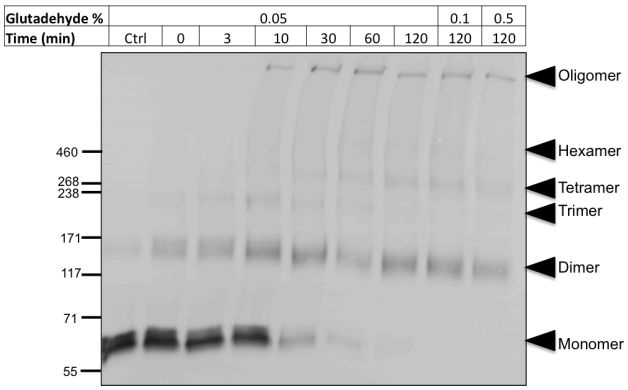


Figure 111: Western blot analysis of a SDS-PAGE gel of Keap1 after cross-linking with glutaraldehyde imaged with anti-Keap1 antibody. Molecular marker weights in kDa are indicated to the left of the image. Cross-linking of Keap1 (0.25 µM) performed with either 0.05 %, 0.1 % or 0.5 % glutaraldehyde (as indicated), was carried out in 10 mM HEPES, 150 mM NaCl, 0.01% NP40, and 2.5 mM TCEP at pH 8.0. The samples were analysed by SDS-PAGE after 0, 3, 10, 30, 60 or 120 min of incubation. Note the progressive formation of larger multimeric protein forms, and the progressive reduction of monomeric protein. Bands related to the molecular weights for monomeric, dimeric, trimeric, tetrameric, hexameric and larger oligomeric Keap1 are indicted with black arrows.

Figure 111 shows that crosslinking of Keap1 was successfully achieved with bands being observed at a variety of molecular weights. The control column with no crosslinking reagent added shows the fully reduced protein as a monomer (~70 kDa). Following addition of glutaraldehyde, dimeric (~140 kDa) bands begin to appear over time and a reduction in monomeric protein is observed. After 120 minutes the monomeric band can no longer be observed. A band at ~210 kDa (a molecular weight equivalent to trimeric protein) can be observed appearing directly after addition of crosslinking reagent with maximal intensity of the band peaking after 10 minutes of reaction, however the existence of this species seems to be short-lived as the band intensity then appears to decrease as the reaction progresses.

Bands at equivalent molecular weight to tetrameric protein (~280 kDa) can be observed after 30 minutes, and appear to be more stable than the trimeric protein bands as they can still be observed at 120 minutes and after exposure to increased concentrations of crosslinking reagent. These bands are followed by fainter bands at a slightly larger molecular weight, possibly consistent with hexameric protein, however these do not appear to form in the same amounts seen from bands at 140 kDa and 280 kDa.

Lastly the bands representing the highest molecular weight appear near the top of the gel just below the gel stack. This suggests higher order soluble oligomeric forms of the protein. These bands appear in observable concentrations after a reaction time of 10 minutes and once formed do not significantly increase or decrease in intensity over increasing time or crosslinking reagent concentration. This suggests either a limit to the amount of this form that is able to form or a limit to the amount of soluble aggregated protein that is able to penetrate into the gel, with the remainder remaining as insoluble aggregates, and not getting loaded into the gel.

Following the successful establishment of a cross-linking technique capable of observing mulitmeric forms of Keap1, the effect of two different sulfhydryl reactive compounds on the cross-linking of the protein was investigated. As previously discussed, due to the proximity of the CDDO-Me binding site, it was hypothesised that binding of a sulfhydryl inhibitor to Keap1 may prevent or alter the formation of larger species, either through altering the kinetics of the reaction or preventing some species from forming.

The experiment was therefore repeated with controls in the presence of CDDO-Me which is known to react mainly with Cys151, and Sulforaphane (SFN) which is more reactive and able to react with the majority of the 27 cysteines within Keap1.[203,320]

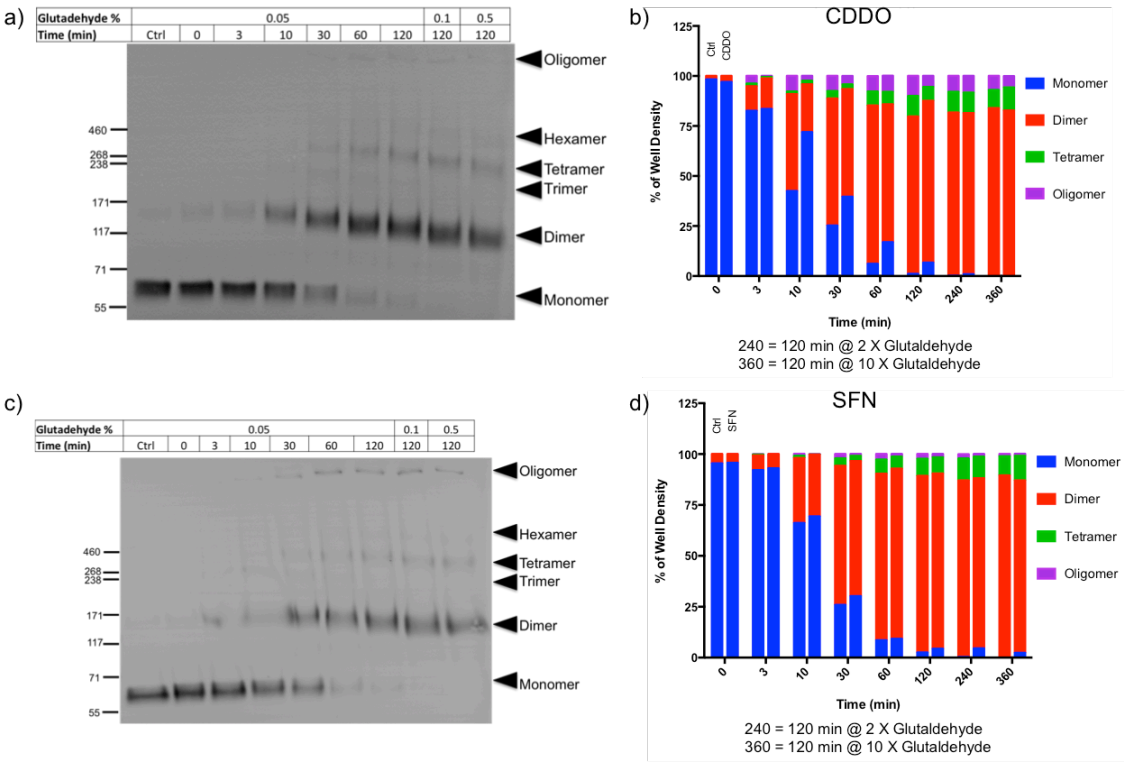


Figure 112: Crosslinking of Keap1 in the presence of reactive Keap1 inhibitors. a) Representative SDS-PAGE analysis of Keap1 in the presence of CDDO after cross-linking with glutaraldehyde imaged with anti-Keap1 antibody. Molecular weight standard weights in kDa are indicated to the left of the image. Cross-linking of Keap1 (0.25 µM) performed with either 0.05 %, 0.1 % (2X) or 0.5 % (10X) glutaraldehyde (as indicated), was carried out in 100 µM CDDO-Me, 10 mM HEPES, 150 mM NaCl, 0.01% NP40, and 2.5 mM TCEP at pH 8.0. The samples were analysed by SDS-PAGE after 0, 3, 10, 30, 60 or 120 min of incubation. Bands related to the molecular weights for monomeric, dimeric, trimeric, tetrameric, hexameric and larger oligomeric Keap1 are indicted with black arrows. b) Bar chart illustrating the normalised pixel density of the four most prominent bands (monomer = blue, dimer = red, tetramer = green, oligomer = purple) from the experiment conducted in the presence of CDDO-Me (a) and a control experiment conducted in the absence of CDDO-Me (shown in Appendix 7) in relation to the overall well density are shown over time. c) Representative SDS-PAGE analysis of Keap1 in the presence of SFN after cross-linking with glutaraldehyde imaged with anti-Keap1 antibody. Molecular weight standard weights in kDa are indicated to the left of the image. Cross-linking of Keap1 (0.25 µM) performed with either 0.05 %, 0.1 % (2X) or 0.5 % (10X) glutaraldehyde (as indicated), was carried out in 100 µM SFN, 10 mM HEPES, 150 mM NaCl, 0.01% NP40, and 2.5 mM TCEP at pH 8.0. The samples were analysed by SDS-PAGE after 0, 3, 10, 30, 60 or 120 min of incubation. Bands related to the molecular weights for monomic, dimeric, trimeric, tetrameric, hexameric and larger oligomeric Keap1 are indicted with black arrows. d) Bar chart illustrating the normalised pixel density of the four most prominent bands (monomer = blue, dimer = red, tetramer = green, oligomer = purple) from the experiment conducted in the presence of SFN (c) and a control experiment conducted in the absence of SFN (gel not shown) in relation to the overall well density are shown over time.

From Figure 112 all bands were observed in both the control experiment (Appendix 7) and experiments with either 100 μM CDDO-Me or 100 μM SFN. However, interestingly the amount of monomer and dimer in relation to the control appeared to be different at the 10-minute time point onwards. To investigate this further, the kinetics of the reaction were analysed by fitting one phase association (for formation of multimers) or decay (for decay of monomer) rate curves to the data as appropriate (Table 15). This showed that the rate of dimer formation in all cases is equal to the rate of decay of monomer. Interestingly, the rate of decay of momomeric Keap1 in the presence of CDDO-Me was reduced by around half (42 %) when compared to control, and by around 20 % in the presence of SFN. However, the rate of decay for CDDO-Me is significantly reduced as indicated by the 95 % confidence intervals. This is interesting as CDDO-Me is known to be more specific in reacting with Cys151 of Keap1 when compared to sulphoraphane, which has been shown to react with a larger variety of cysteine residues.[256,320] This may be why CDDO-Me is able to show a greater ability to slow the formation of dimer/reduction of monomer in this experiment.

When looking at the rate of formation of larger species only the rates of tetramer can be meaningfully interpreted as the rates for the oligomeric protein are affected by large errors in the rates, again suggesting that the ability larger insoluble aggregates to penetrate the gel may be affecting the result (as discussed earlier). The rate of tetramer formation is again slowed by the addition of either CDDO-Me or SFN with both results showing significance through non-overlapping 95 % confidence intervals. Both show a reduction of around 50 % in the rate of formation when compared to untreated control.

Table 15: Table showing the calculated rates of formation and decay for species of Keap1 formed through crosslinking of the protein in the presence of either CDDO-Me, SFN or a null compound control. Shown with Standard Errors and 95 % Confidence intervals.

		K / (10^{-4} s^{-1})	Standard Error / (10^{-4} s^{-1})	95% CI / (10^{-4} s^{-1})
K expressed as rate of decay	**Monomer Ctrl**	8.44	1.13	5.95-10.94
	Monomer CDDO	4.94	0.32	4.23-5.64
	Monomer SFN	6.74	0.39	5.88-7.60
K expressed as rate of formation	**Dimer Ctrl**	9.00	1.07	6.68-11.32
	Dimer CDDO	5.34	0.42	4.26-6.41
	Dimer SFN	7.80	0.84	5.64-9.97
	Tetramer Ctrl	3.18	0.39	2.35-4.00
	Tetramer CDDO	1.50	0.31	0.70-2.29
	Tetramer SFN	1.66	0.20	1.14-2.18
	Oligomer Ctrl	30.70	36.85	0.00-110.30
	Oligomer CDDO	7.47	3.84	0.00-17.35
	Oligomer SFN	6.56	2.67	0.00-13.43

In conclusion the addition of a sulfhydryl reactive compound does have an affect on the rate of formation of cross-linked species of Keap1. The effect of CDDO-Me on the rate of reaction is around twice that of sulforaphane, showing that perhaps a more specific Cys151 reactive compound is more beneficial for reducing the amounts of Keap1 dimer formed. This gives some support to the hypothesis that CDDO-Me, via reaction with Cys151 may interfere with Keap1 self-assembly and subsequently also Nrf2 ubiquitination. However, the mechanism for how this occurs is not conclusive and further investigation would be needed to clarify this as a valid mechanism of action.

6.2.4 Mass spectrometry

A sample of protein was analysed by MALDI-MS to see if any of the larger forms of the protein could be picked up under native buffer conditions. Four forms of the protein were observed up to the limit of the detector where the weak concentration of the protein sample produced too much background noise to accurately determine peak masses about ~300 kDa. Individual peaks corresponding to protein masses of Keap1 monomer, dimer, trimer and tetramer were all observed by the detector (Figure 113).

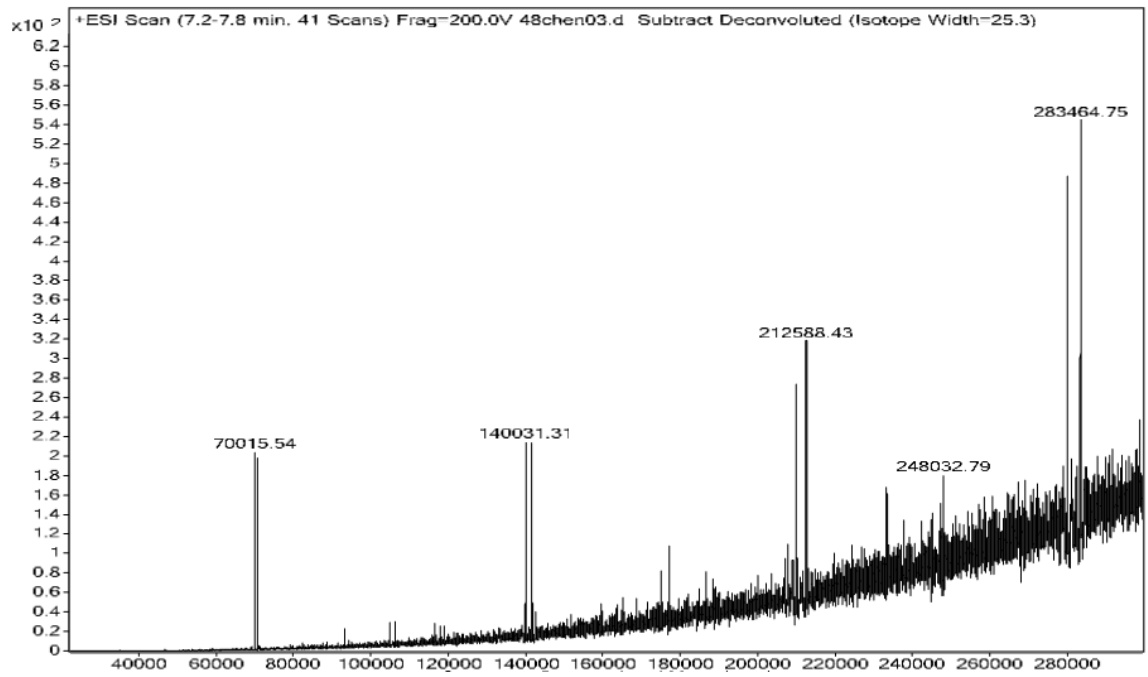


Figure 113: Mass spectrometry analysis of full length recombinant mammalian Keap1 protein under physiological buffer conditions (20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.5) at 0.5 µg/µL. Four distinct protein mass peaks can be identified corresponding to masses of Monomeric (~70 kDa), dimeric (~140 kDa), trimeric (~210 kDa), and tetrameric (~280 kDa).

This gives further evidence towards the hypothesis that Keap1 is able to form stable oligomeric forms that could influence its interaction with Nrf2.

6.2.5 Electron microscopy

To investigate the structural form of the protein, electron microscopy was used to image the protein in the presence or absence of a reducing agent. The hypothesis was that under non-reducing conditions the particles seen would represent larger aggregates and smaller oligomeric forms of the protein such as those seen in the size exclusion chromatography, whereas under reducing conditions only the predicted more stable smaller forms of the protein such as the monomeric, dimeric or tetrameric forms would be visible.

Figure 114 shows electron micrographs of reduced and non-reduced protein samples. It can be clearly seen that under non-reducing conditions the protein forms extended "worm-like" oligomeric structures and also larger globular aggregates. The extended structures measure on average over 400 angstroms in length (with some longer than 600 angstroms) and are around 60 angstroms in diameter, which is the same diameter reported for each globular domain (both the BTB and Kelch domains) of Keap1.[181] The extended length suggests these structures are likely to be high order oligomers of Keap1 protein consistent with those seen in size exclusion chromatography under non-reducing conditions. The linearity of the molecules however suggests binding between Keap1

monomeric units is most likely occurring through the N (BTB) or C (Kelch) terminal domains rather than domains within the central region of the protein which should create wider structures. As both of these domains contain cysteine residues in which di-sulphide bonds could form between molecules under non-reducing conditions from these images it is difficult to say for certain which domain the binding is occurring through, and it may be a mixture of both. At higher magnification (Figure 115) some of the extended particles can be seen in further detail with some resembling horseshoe like shapes. The larger aggregate particles can also be seen as larger clumped masses of protein.

In the reduced image there is a dramatic change in the structure of the proteins with much smaller particles observed and none of the larger aggregated masses seen. There are a higher proportion of globular particles that measure around 60 by 60 angstroms potentially representing monomeric protein viewed from a perspective of the C terminal kelch domain “bottom up” or the N terminal IVR domain “top down”. Further particles can be observed indicated by the black arrows in the figure which show linear particles similar to those observed in the literature when observing side, top, or bottom projections of Keap1 dimer (Figure 103).

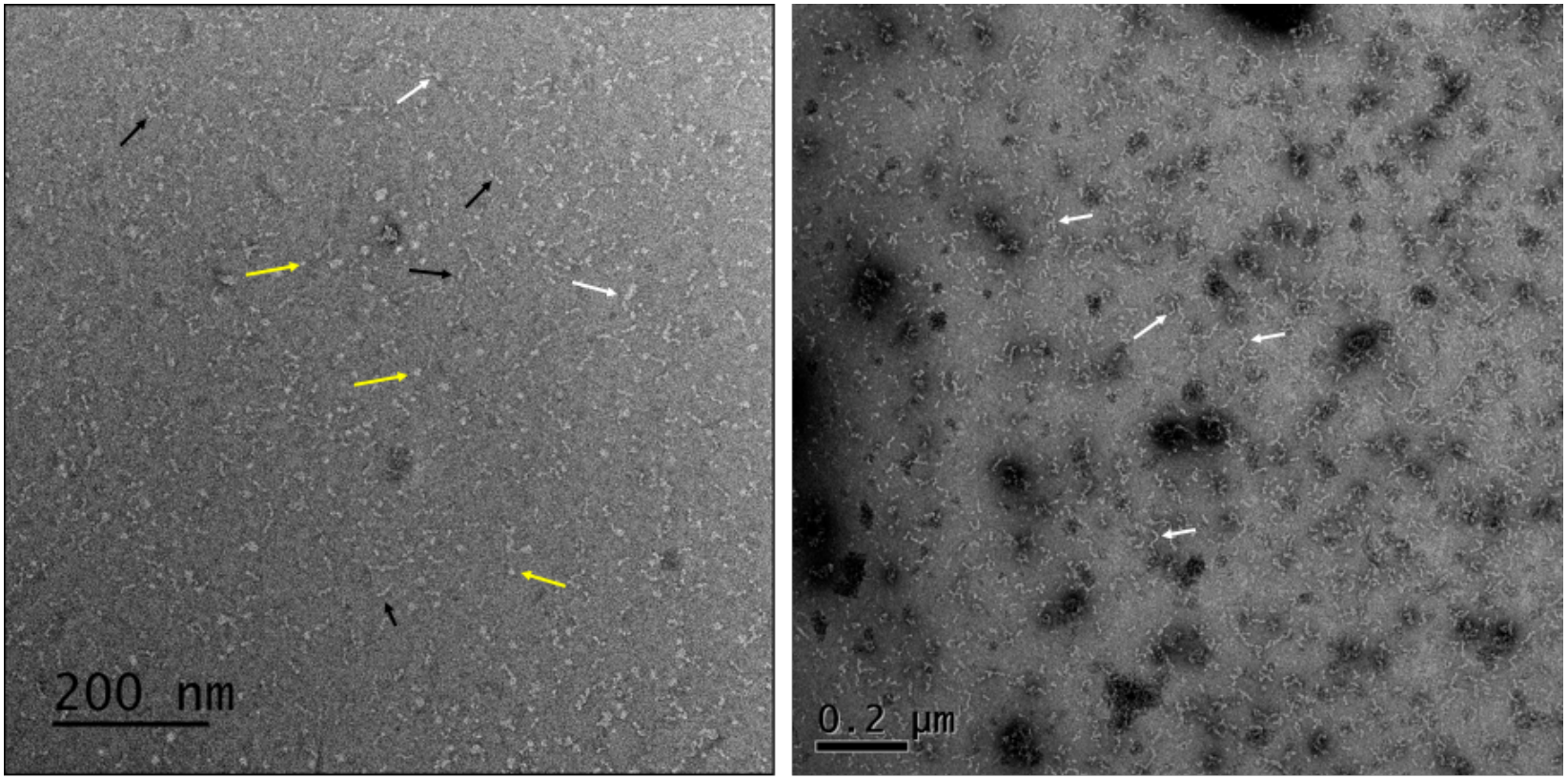


Figure 114: Representative electron micrographs images of recombinantly expressed human Keap1 protein at 21000X. Imaged using negative staining methodology and viewed using a Philips CM200 electron microscope. Left image shows protein in the presence of 1 mM TCEP scale bar is indicated in the bottom left of the image. Yellow arrows indicate circular particles measuring on average 60x60 Å, black arrows indicate longer particles measuring on average 230x60 Å, whilst white arrows show particles similar in size to those seen in the right image (>400 Å in length). Right hand image shows protein in the absence of the reducing agent TCEP. Scale bar is indicated in the bottom left corner, with white arrows showing "worm-like structures" which vary in size but are mostly all greater than 400 Å, with a width of ~60 Å. Also visible in the darker areas of the image are larger aggregated clumps of protein.

When observed in further detail at higher magnification (Figure 115 and Figure 116) the linear and spherical particles can be measured and compared to the measurements referenced in the literature for dimeric mouse Keap1 using the same methodology. The literature indicates that dimeric forms of the protein measure 160 Å across the two side-by-side kelch domains, and 103 Å along the length of an individual protein molecule from BTB to kelch domain, with each individual globular domain measuring 60x60x78 Å.[181] Measurements from the images taken (Figure 116) indicate that the spherical particles do match the measurements seen in the literature for a globular domain of Keap1, with particles measuring on average 60x60 Å. However, the linear particles varied in size somewhat, with the average particle size measuring around 200x60 Å. The linear particles tended to fit into two clusters with particles which had linear lengths of less than 180 Å and those which had lengths of greater than 210 Å, with distribution of the observed particles split evenly between the categories. The smaller cluster of these linear particles can be said to be broadly similar to the dimeric structures seen in the literature (a and b, Figure 116). However, the larger of these structures is more interesting (c and d, Figure 116). There are two possibilities for these structures, firstly they may simply be extended dimeric structures. However, for these structures one would expect to observe more separated distinct round spheres as the globular Kelch and BTB domains move slightly away from each other which is not observed clearly in this instance. Another possibility is the longer particles may be a larger oligomeric form of the protein, such as that observed of tetrameric protein observed in cross-linking, mass spectrometry, and SEC studies. If two dimeric forms of the protein were linked together via the N-terminal end of the IVR domain as previous suggested, then a 2D structure similar to Figure 117 might be expected. This structure would leave the Cul3 binding domain free for association to Cul3 and would give elongated particles (around 50% longer) when observed from above as indicated in Figure 117.

Whilst this is an interesting hypothesis, further investigation is clearly needed to elucidate the potential formation of higher order oligomers of Keap1.

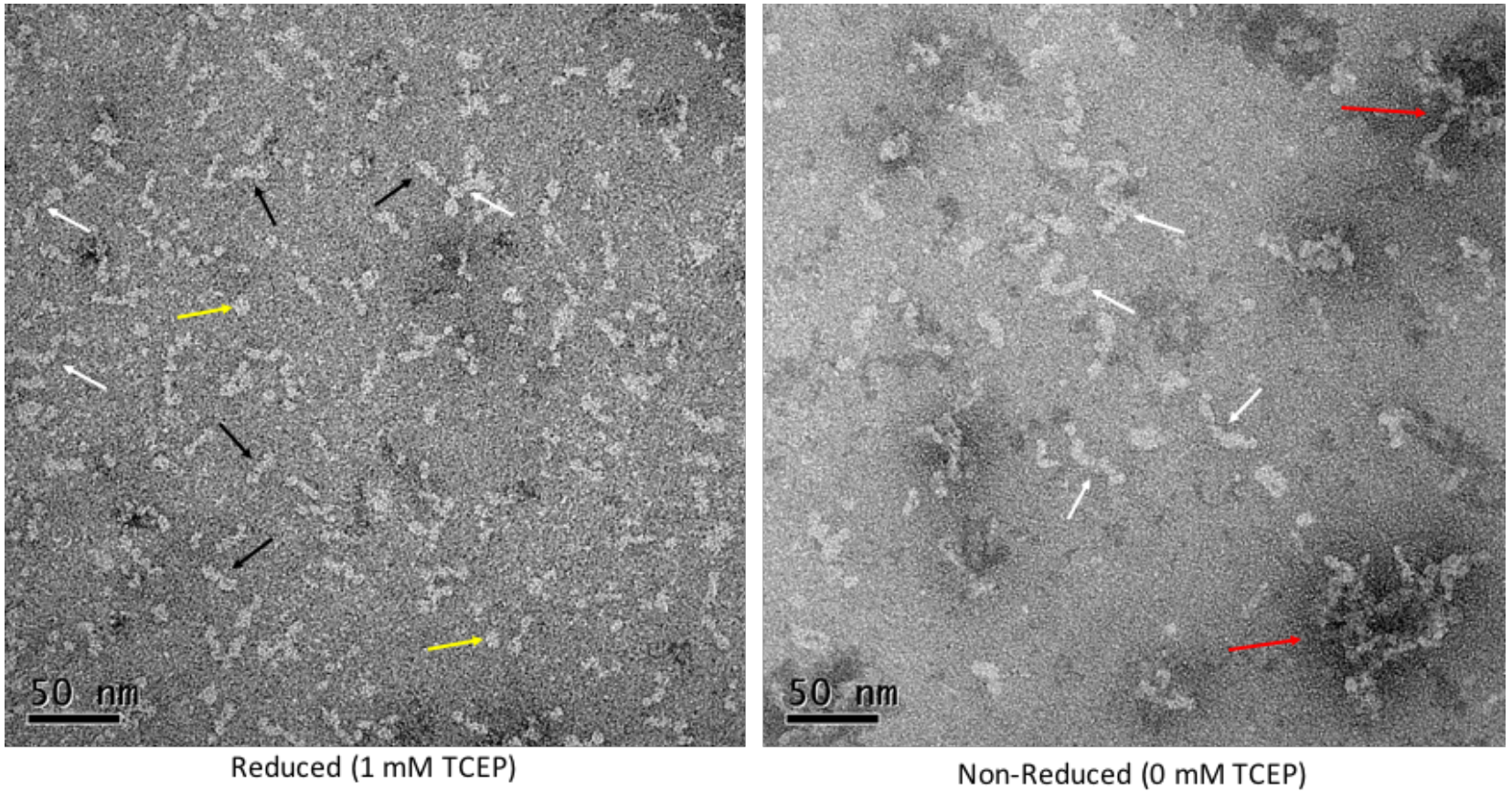


Figure 115: Representative electron micrographs images of recombinantly expressed human Keap1 protein at 52000X. Imaged using negative staining methodology and viewed using a Philips CM200 electron microscope. Left image shows protein in the presence of 1 mM TCEP scale bar is indicated in the bottom left of the image. Yellow arrows indicate circular particles measuring on average 60x60 Å, black arrows indicate longer particles measuring on average 230x60 Å, whilst white arrows show particles similar in size to those seen in the right image (>400 Å in length). Right hand image shows protein in the absence of the reducing agent TCEP. Scale bar is indicated in the bottom left corner, with white arrows showing "worm-like structures" which vary in size but are mostly all greater than 400 Å, with a width of ~60 Å. Also visible in the darker areas of the image are larger aggregated clumps of protein indicated with red arrows.

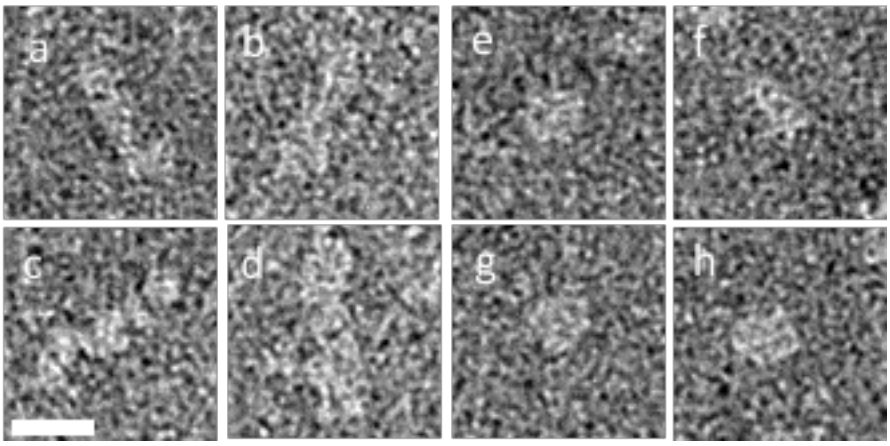


Figure 116: Selected electron microscopy images of negatively stained Keap1 particles at 52000X. Protein is shown in bright shades. Scale bar indicates 100 Å. Images a-d show examples of the linear Keap1 projections with particles measuring on average 203 Å x 60 Å, images a and b show the slightly smaller linear projections at ~180x60 Å and c and d represent the lightly longer linear projections at ~230x60 Å. Images e-h show examples of Keap1 potential cross sections with distinct round spheres being seen, these particles measure on average 60 Å x 60 Å.

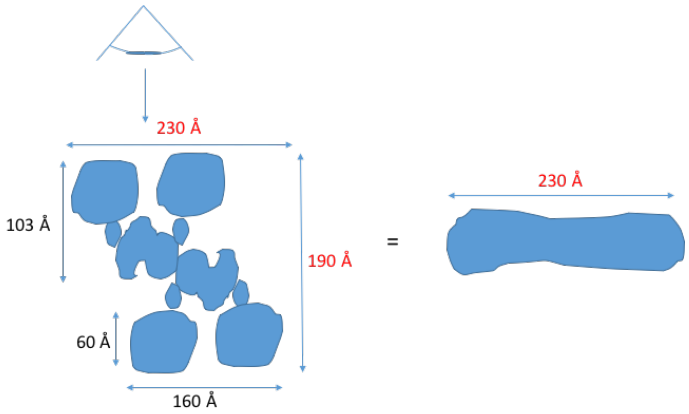


Figure 117: Hypothesised structure of tetrameric Keap1 consisting of two Keap1 dimers. Literature reported measurements of Keap1 are shown in black whilst additional measurements seen in this study using electron microscopy are shown in red against hypothesised vertices. A potential explanation of the longer Keap1 particles observed is shown with the viewpoint of the tetrameric structure observed from above as indicated leading to a linear image observed by EM with length 50 % longer than that observed for dimeric Keap1.

6.3 Discussion and Conclusion

Whilst results from the SEC, *in-silico*, mass spectrometry and crosslinking experiments performed here suggest that a larger oligomeric form of the protein (possibly in the form of a tetramer) may be achievable, more direct structural evidence will be needed to validate this hypothesis. Literature evidence shows that larger assemblies have previously been reported in the literature for BTB containing proteins. The most similar to Keap1 and most prevalent of these examples is SPOP, a member of the BTB-MATH family of Cul3 dependant-E3 ubiquitin substrate adaptors.[317,321] SPOP has structural and functionality similarity to Keap1 in several ways. SPOP contains a BTB domain that mediates its dimerisation similar to Keap1; it also contains a 3-box motif, which with the BTB domain has been shown to bind Cul3.[317,321] The way the BTB and 3-box domains bind Cul3 are shown to be structurally very similar to other BTB-Kelch proteins such as KLHL11.[200]

Interestingly, SPOP is able to form higher order oligomers through a second dimerisation domain C-terminal to the 3-box domain which enhance the ubiquitination activity of the E3 ligase.[321] A structurally similar domain to the 3-box motif resides in the BACK domain of Keap1, and has been predicted above by bioinformatics analysis to contain the alpha helical structure consistent with dimerisation domain in SPOP. One crystallised BTB-Kelch family protein KLHL11 includes a 3-box domain (although no evidence has yet been seen for KLHL11 oligomerisation), which was used to create a homology model for the 3-box domain of Keap1. This also enabled modelling of how Cul3 interacts with Keap1, in addition to modelling the hydrophobic surface of the 3-box domain. This analysis showed a hydrophobic surface that could be capable of facilitating a protein-protein interaction.

Furthermore, electron microscopy studies showed that under non-reducing conditions large aggregates of protein could be observed with some “worm-like” structures. Whereas under reducing conditions these structures broke down into smaller groups of linear structures. The smaller of the two groups may represent a dimeric form of the protein whereas the larger of the groups shows sizes in excess of that which would be expected for the dimeric protein.

In order to prove the hypothesis confirming the existence of a tetrameric form of protein, further experiments using electron microscopy represents the best chance of observing the oligomer. If a more homogeneous population of protein particles could be created through analysis of different fractions from SEC by EM this would make characterisation of the different forms easier. Furthermore, from a more homogeneous population, image classification and modelling could be used to create a 3D image of the protein.

Monoclonal antibody bound protein could also be imaged in order to analyse where specific domains are located within the particles, this would enable a greater understanding in how Keap1 molecules can bind together. This could also be used to inhibit interactions between domains. For example, a BTB domain binding antibody would be expected to inhibit dimerisation.

Further research into the effect of oligomerisation on ubiquitination of Nrf2 is also warranted. For example, as structures larger than monomer remain under the presence of reducing agent performing an experiment in which the reducing agent is titrated in the presence of Nrf2, Keap1 and the rest of the ubiquitination machinery (including Cul3 and Rbx1). The level of ubiquitation of Nrf2 can then be monitored to see whether levels of ubiquitination are lower in the presence of reducing agent.

Additional experiments into larger oligomeric forms of the protein could be investigated through the use of deletion or single point mutations to Keap1. The alpha helical region predicted to make up the dimerisation domain could be deleted and cross-linking experiments repeated to see if oligomeric protein is still observed. Alternatively, the N-terminal region of the protein responsible for BTB-dimerisation could be deleted, to see if the protein is able to form dimers through the oligomerisation domain without the use of the BTB-dimerisation region.

Overall, characterisation of mammalian Keap1 has led to some interesting hypotheses that will be of interest to future research within the field.

7. General Discussion and Conclusions

There remains an unmet need for the effective treatments of ALS. This devastating and fatal disease remains one of the disorders for which there continues to be no cure. For patients with the disease a treatment cannot come soon enough. Whilst development of a new therapy can take many years, this project aimed to start the long journey towards discovery of a new therapeutic for the treatment of ALS. The project aimed to focus on the pathogenically associated mechanism of elevated oxidative stress within ALS patents, and more specifically on the inhibition of the protein-protein interaction between Nrf2 and Keap1 which should lead to reduction of the elevated levels of oxidative stress observed in ALS patients.

This project has successfully established a reliable high throughput screening assay that is able to probe for inhibitors of the Nrf2-Keap1 interaction. In addition, a complete screening pipeline for these inhibitors was created with computational models, multiple *in-vitro* assays, as well as biophysical characterisation techniques that enable validation of the binding of any compound to Keap1. Having all of these methods is an attractive package to pharmaceutical companies whose expertise and financial involvement will be needed in the future of the project in order to develop the molecules through pre-clinical and clinical trials which are often too costly for academic institutions to pursue on their own.

The screening pipeline consisted of several key features. First the development of the high throughput screening assay which used fluorescence polarisation to probe for inhibitors between Keap1 protein and the Nrf2 peptide was developed. The assay was successfully miniaturised to a 5 µL, 384 well format and using 90 nM protein and 4 nM peptide concentrations. Key parameters for screening assays were determined such as the Z' factor and signal windows. The Z' factor was found to be 0.694 and the signal window 2.7, values in excess of those required for a robust assay. In addition, the binding constants of the unlabelled and labelled ETGE peptides was determined using the fluorescence polarisation assay. The value for the unlabelled peptide was determined as 132 nM vs. a binding constant of 55 nM for the labelled peptide.

During the project timeline a variety of other screening assays were published.[157,210–212,244,322] All of these assays used isolated kelch domain, with some using mouse protein, and others mammalian protein. This is in contrast to our assay which used full length mammalian protein. Whilst both approaches seem reasonably effective in identification of inhibitors of the interaction between Nrf2 and Keap1, the full length assay as detailed here may have some advantages for future studies. Apart from being closer to a physiological

model, the presence of some key sulfhydryl residues (absent from the kelch domain) in the BTB and IVR domains of Keap1 enabled the assay to be used to model the effect of modification of the sulfhydryls on the binding of molecules to the kelch domain. In future work the non-reactive inhibitors of the Keap1-Nrf2 interaction identified in this work can be incubated in the presence of a reactive Nrf2 activator such as CDDO-Me or Tecfidera® to judge how the binding affinity of non-reactive inhibitors may change in the presence of a reactive Nrf2 activator. Although, experiments described within Chapter 3 where reduced and non-reduced protein were shown to have no effect on the binding affinity of a known non-reactive inhibitor sourced suggests that no effect would be seen on strong binding non-reactive inhibitors.

The assay was also tested for interactions with the weaker affinity DLG motif on Nrf2, as successfully establishing an assay with the weaker DLG peptide may lead to discovery of a more chemically diverse set of compounds. This is because weaker affinity compounds that can only disrupt the DLG "hinge" but not the ETGE "latch" motif could be as efficacious as those that can disrupt both the ETGE and DLG motif. Both a shorter DLG 9-mer peptide and a longer 35-mer peptide were tested for affinity, with no detectable binding being observed. However, if a labelled DLG peptide could be found where binding was observed within the assay then the experiment suggested above with reactive Nrf2 activators could be repeated to see if the hinge and latch model is observed in this system and whether it is representative of a physiological mode of action.

Due to the presence of the full length protein the assay could be further adapted to incorporate the whole E3 complex (Cul3, Rbx1 etc). Labelled whole Nrf2 protein could potentially then be used to model the binding and inhibition of Nrf2 to the larger complex. This would open up further avenues of exploration using assays which could probe the levels of ubiquitination of Nrf2. In combination with the work presented in Chapter 6 into the oligomerisation of Keap1, this may yield further interesting results and additional therapeutic strategies in inhibition of Keap1 through modulation of the ubiquitination of Nrf2, rather than through its direct interaction with Keap1.

Utilising the screening assays developed for the project, two different chemical libraries were screened against the target Keap1. The first screen with a library provided by MRCT contained compounds with protein-protein interaction inhibition properties modelled around alpha helixes. This library provided a good starting screen to test the capabilities of the screening cascade. Following screening additional counter and validation screens were developed due to the large number of compounds which were showing fluorescent

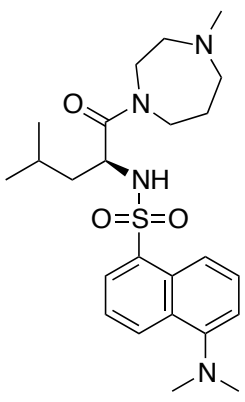
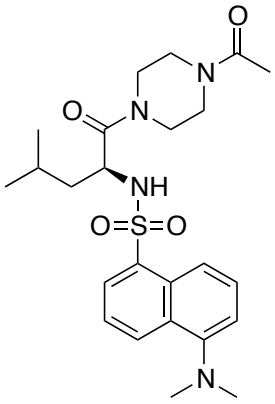
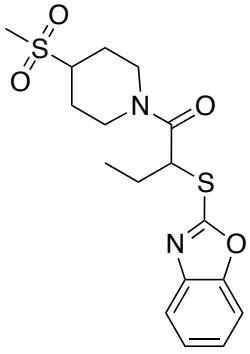
interference within the assay. This affected how the data from the screening was processed. Rapid statistical analysis techniques were developed using *in-silico* data pipelining tools to enable additional fluorescence intensity data collected during the screen to be utilised for removal of compounds which were shown to interfere with the assay. Other screens documented in the literature did not appear to take this approach and instead tended to use secondary assays, which often used costlier low throughput biophysical techniques to refine down screening hits to those which showed true binding to Keap1.[157,210–212,244,322] Using additional information from the cheaper screening assay we were therefore able to refine down compounds to a lower number before biophysical characterisation, reducing the overall cost of the screen, and developing *in-silico* protocols which could be utilised in additional screens within the department. Whilst certainly quicker and more cost effective than running all "hits" in the fluorescence assay through biophysical assays this approach may also mean that compounds that both interfere with the assay as well as bind the target may be missed as false negatives. However, these molecules could be inherently more difficult to develop, due to additional screening assays that are not susceptible to the fluorescence interference from the compounds needing to be established.

For the first screen a hit rate for the assay was determined at a cut off three standard deviations below the mean of the negative controls. This gave 44 compounds (hit rate of 0.3%) classified as hits. MRCT chemists' then manually selected ten compounds for resupply, screening out frequent hitters and selecting molecules for structural diversity. The selected molecules were then screened in subsequent confirmation and deselection assays, including a full dose response curve and a cellular ARE reporter assay. Only one compound was able to confirm the activity seen in the primary screen and gave an IC_{50} value of 180 μM, which did not meet the criteria for hit selection. Structurally similar analogues around this hit were purchased and tested in both the screening and cellular assays, with no activity seen in either assay. This suggests that this compound is a singleton with a very weak binding mode to the protein preventing any additional increases in affinity from being made.

Using lessons learnt from the initial screen, a second larger screen was performed using a library of 318132 molecules containing a larger structural diversity of compounds. The screen was performed externally at the European Lead Factory screening centre and was successfully re-optimised for use with equipment at the facility with increased Z prime values of >0.7. The primary screen of the library gave a similar hit rate to the MRCT screen of 0.3 %. These compounds were screened in confirmation assays, first at a single

concentration point and then in a concentration dose response. Deselection assays were also performed to determine the redox potential of the molecules with any molecules showing redox reactivity screened out. A single concentration fluorescence screen was also undertaken and data used in parallel with fluorescence data collected from the primary screens to remove compounds that interfere with the assay. The contents of the screening wells containing active molecules were then subject to purity test with impure samples being discarded. Six compounds out of the remaining 8 were successfully resynthesised followed by validation in a full dose response fluorescence polarisation assay, as well as being tested in the cellular ARE reporter assay. Three molecules ELF1, ELF2 and ELF6 were validated in the FP assay and able to show full dose response curves. ELF6 and ELF5 showed significant activity in the cellular assay. ELF2 showed activity but was not significantly above the DMSO control, ELF1 was toxic at the top concentration so activity could not be determined to the same concentration.

Table 16: Summary table of the three main hit compounds identified from the screen of the ELF library, more detail can be found in Chapter 4. Where MW = molecular weight of the compound, AlogP = atomic partition coefficient, HBA = number of hydrogen bond acceptors, HBD = number of hydrogen bond donors, tPSA = topological polar surface area, QED = quantitative estimate of drug likeness, N.D. = Data not determined.

Compound ID	ELF1	ELF2	ELF6
Structure			
FP IC50 (µM)	12.02	15.53	14.49
Cellular ARE Assay IC50 (µM)	N.D.	229	171.5
MST KD (µM)	3.1	3.3	N.D.
MW	461	475	383
AlogP	2.81	2.01	2.28
tPSA	81.3	98.4	114.2
HBA	7	8	6
HBD	1	1	0
QED	0.69	0.66	0.74

Whilst not all of the compounds were able to show activation of ARE directed transcription in this assay, other publications have shown activation of ARE through investigating mRNA and protein expression levels of downstream genes of Nrf2 such as HO-1 and NQO-1.[207] This could be used as an additional way of looking at whether the compounds are able to activate downstream effects of Nrf2 activation, and would provide additional evidence that the compounds were working through the desired mechanism of action. It may also be interesting to observe the levels of cytoplasmic and nuclear Nrf2, which could be done through western blotting experiments. Inhibition of Keap1 by the compounds should lead to an increase in nuclear levels of Nrf2.

The compounds were also tested in biophysical assays, with ELF1 and ELF2 showing binding to the target with both molecules showing K_D's determined as around 3 μM in the MST assay. ELF2 was also able to compete for binding with the ETGE peptide and the control compound Hit1 shown by a shift in the K_D values. ELF2 was also tested for detection by ITC but no binding was observed at the maximum concentration of protein that was practical to use in the assay.

Whilst MST worked extremely well as a technique it is also partially susceptible to compound absorption/fluorescence interference because of its reliance on endogenous protein fluorescence. It was unfortunate that greater concentrations of protein could not be achieved with full length protein before precipitation of the protein was observed. However, it would be worth re-attempting to establish an effective ITC assay for low micro-molar compounds using isolated kelch domain. Utilising the isolated kelch domain a greater solution protein concentration should be achievable compared to the full length form, and therefore the binding event may be easier to observe.

All the active compounds were also subjected to detailed *in-silico* analysis utilising models which were developed in Chapter 4. In these models ELF1 and ELF2 were able to achieve reasonably predictive binding poses with good docking scores when compared to a known binding control compound. Using these docking poses a detailed medicinal chemistry plan was proposed based on structural information to illustrate where additional activity could be gained through structural development of the molecules leading to an increase in binding energy. However, in order to validate these docking poses X-ray crystallisation studies of Keap1 in the presence of each of the ligands needs to be performed. After the correct binding pose is established a more direct approach to modification of the ligands can be used to increase binding affinity and tune physicochemical properties.

Whilst the pathway for drug development is long, these compounds represent an important step towards a new potential treatment for ALS. Once a candidate with suitable physicochemical properties and target affinity has been established, there are *in-vivo* models of ALS established at the University of Sheffield that are able to show characteristic phenotypes of the disease, such as loss of motor function and a dramatically reduced lifespan. These models have already shown to respond well to Nrf2 activators, with S[+]apomorphine (a reactive Keap1 binder) able to delay decline in motor function.[218] These models therefore represent a good pathway for future progression of any compounds that result from the "hit to lead" development of these molecules.

Computational techniques continue to become more and more widely used within the drug-discovery industry due to ever increasing accuracy of predictions, speed of calculations and cost effectiveness. The computational techniques utilised throughout all aspects of this project have enabled a more rapid progression to hit candidates, and will continue to influence the project as it progresses further. Within the project, *in-silico* techniques have enabled faster data analysis of screening results (through use of new data pipelining technology), as well as techniques such as virtual screening which have attempted to identify new chemical matter, and also being able to inform on the direction of development of hit candidates.

As the development of the compounds progresses, *in-silico* methods will be able to further influence the project. As suggested earlier, once a co-crystal structure is solved this will enable *in-silico* docking studies to be more widely utilised to influence design of the compounds. De-novo design could potentially also be utilised to help rapidly design new molecules around known fragment binders (such as around the core structures of the hit ELF compounds) in the presence of the active site to obtain functional groups that are not commonly utilised.

Whilst two virtual screening approaches were developed to try and identify novel chemical starting points for an Nrf2-Keap1 inhibitor additional to those identified by the HTS, neither of these approaches was able to replicate the activity of compounds identified through a HTS. There are many different reasons for why this might be, but the greatest influence on the success may simply be around the difficulty in finding molecules suitable for inhibition of protein-protein interactions. This may mean that much larger numbers of compounds may need to be tested from *in-silico* screens to enable identification of a single active compound. The second reason may centre around the selection of molecules for screening, this project focused on compounds which had a high probability of being able to

penetrate the BBB. This lead to filtration of compounds prior to virtual screening. However, whilst some attempt was made to relax the criteria for selection this may mean that active compounds may have been discarded due to their properties, which could potentially have been altered at a later stage during development. Counter to this however, *in-silico* screening is meant to reduce later stage development of compounds, because it is often easier to gain activity against a target than it is to change the property of the molecule to something you desire. This is because the group within the molecule that is responsible for your target activity may well also be the one responsible for the undesired property. Therefore, it is generally better to start from an ideal property viewpoint and gain activity than the other way around.

In summary, during this project a wide variety of techniques have been used to explore and address the aims of the project. It has achieved its goals of creating a cost efficient, reproducible high throughput screening assay which is able to probe the ability of small molecules to inhibit the interaction between Keap1 and Nrf2.

Several compounds were discovered as inhibitors of the protein-protein interaction between Keap1 and Nrf2, with the compounds being validated in a number of secondary screening assays. Furthermore, computational methodology was used to enhance both the screening of the compounds, using powerful data analysis tools, and the design of the inhibitors through protein-ligand modelling and *in-silico* prediction of physiochemical properties of the molecules in order to influence their design towards penetration of the blood-brain barrier.

Whilst the majority of the project focussed on the key aim of identification of hit molecule inhibitors of the Keap1-Nrf2 protein-protein interaction, unexpected results were obtained upon analysis of the full length mammalian Keap1 protein synthesised for the screening assay.

Size exclusion chromatography experiments suggested that a larger tetrameric form of the protein may exist, with further exploration through *in-silico* experiments identifying a location on the protein surface of Keap1 for a site of homo-oligomerisation. Interestingly, larger assemblies of BTB proteins have previously been reported in the literature, with the most prevalent of these examples being SPOP, a member of the BTB-MATH family of Cul3 dependant-E3 ubiquitin substrate adaptors.[317,321] SPOP is able to form higher order oligomers through a second dimerisation domain C-terminal to the 3-box domain which enhances the ubiquitination activity of the E3 ligase.[321] This leads us to question whether ubiquitination activity of the E3 ligase in the ubiquitination of Nrf2 would be enhanced

through Keap1 oligomerisation. Further research into the effect of oligomerisation on ubiquitination of Nrf2 is therefore warranted. For example, in the presence of the complete ubiquitination machinery the level of ubiquitation of Nrf2 can be monitored to assess whether levels of ubiquitination are altered under different oligomeric states of Keap1.

SPOP contains a BTB domain which mediates its dimerisation similar to Keap1; it also contains a 3-box motif, which with the BTB domain has been shown to bind Cul3.[317,321] The way the BTB and 3-box domains bind Cul3 is structurally very similar to other BTB-Kelch proteins such as KLHL11.[200] A structurally similar domain to the 3-box motif resides in the BACK domain of Keap1, and has been predicted through bioinformatics analysis in this work to contain the alpha helical structure consistent with dimerisation domain in SPOP. One crystallised BTB-Kelch family protein KLHL11 includes a 3-box domain, and was utilised to create a homology model for the BTB and 3-box domain of Keap1 with Cul3. This illustrated how Cul3 interacts with Keap1, and enabled modelling of the hydrophobic surface of the 3-box domain. The analysis showed a hydrophobic surface that could potentially mediate a protein-protein interaction.

Electron microscopy studies illustrated that under non-reducing conditions large aggregates of protein are observed with some “worm-like” structures. Whereas under reducing conditions these structures broke down into smaller groups of linear structures. The smaller of the two groups may represent a dimeric form of the protein whereas the larger of the groups shows sizes in excess of that which would be expected for the dimeric protein, with a tetrameric structure being proposed. These larger structures explain why a larger fluorescence polarisation signal could be obtained under non-reducing conditions, as the larger structures under non-reducing conditions lead to slower rotation of the peptide-protein bound structure in solution reducing the amount of light becoming depolarised when compared to smaller peptide-protein bound structures.

In order to prove the hypothesis confirming the existence of a tetrameric form of protein further experiments using electron microscopy represents the best chance of observing the oligomer. If a more homogeneous population of protein particles could be created through analysis of different fractions from SEC by EM this would make characterisation of the different forms easier. Furthermore, from a more homogeneous population, image classification and modelling could be used to create a 3D single particle image of the protein.

Additional experiments into larger oligomeric forms of the protein could be investigated through the use of deletion or single point mutations to Keap1. The alpha helical region predicted to make up the dimerisation domain could be deleted and cross-linking experiments repeated to see if oligomeric protein is still observed. Alternatively, the N-terminal region of the protein responsible for BTB-dimerisation could be deleted, to see if the protein is able to form dimers through the oligomerisation domain without the use of the BTB-dimerisation region.

Overall, the unexpected results in the characterisation of mammalian Keap1 protein in a tetrameric form was an interesting discovery, and whilst further work needs to be done to confirm the functional relevance of the tetramer, this body of work will be of interest to future research within the field.

In conclusion, whilst this project has seen success in its ultimate aims, a great deal of investigation still needs to be done into both the project and ALS as a whole. For example, unless biomarkers can be discovered to detect early onset of the disease any therapeutic administered will at most be able to halt progression of the disease. The discovery of biomarkers for the disease represents an important milestone that will need to be overcome. Any future treatment for the disease will need to combine both biomarkers to enable early detection and then an effective therapeutic than can then treat the disease once diagnosed.

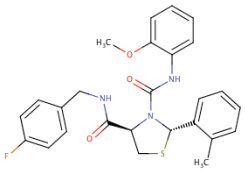
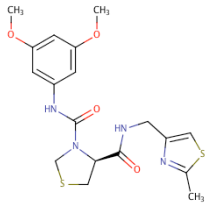
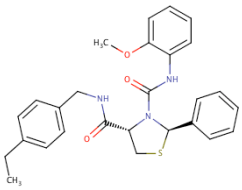
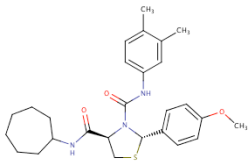
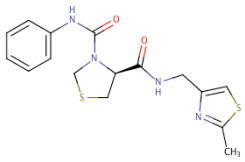
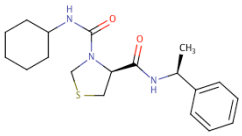
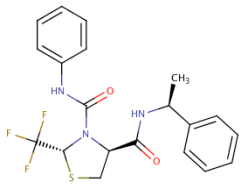
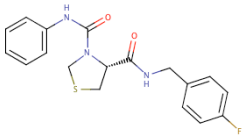
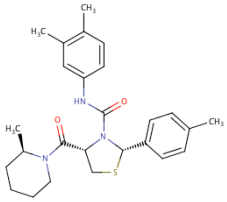
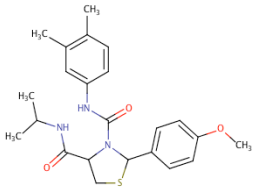
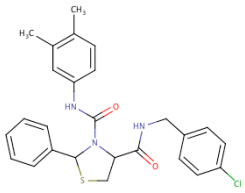
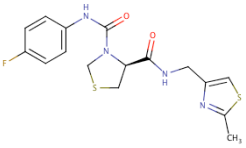
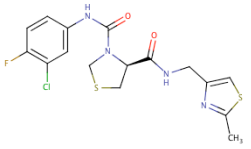
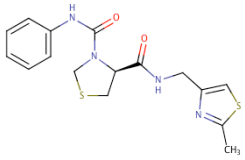
Appendix

Appendix 1: Sequencing of Keap1 cDNA plasmid. Matching residues are shown in turquoise. Literature Keap1 sequence is shown in red typeface, with the forward and reverse primer sequencing reactions shown below.

Appendix 2: Table of hits from MRCT screen with plate details.

MRCT	Percent_Inhibition	Plate	Well Row	Well Col	Parallel	Perpendicular	mP	Z'	SW	SD_Cut_Off	Active_Cut_Off	Intensity
MRT00168919	102	40	B	14	27517	25670	34.73	0.51	2.55	4	69.16	78857
MRT00168920	84	40	D	14	28422	25964	45.20	0.51	2.55	4	69.16	80350
MRT00160422	53	27	P	19	27193	23009	83.34	0.25	2.98	4	87.31	73211
MRT00160403	38	27	F	15	30808	25380	96.60	0.25	2.98	3	97.70	81568
MRT00167614	34	11	N	11	21405	18387	75.84	0.53	2.11	3	78.12	58179
MRT00161093	33	29	D	4	26006	21165	102.63	0.68	2.66	4	113.96	68336
MRT00161961	29	32	K	8	26512	22125	90.20	0.57	2.91	4	98.41	70762
MRT00160111	29	26	D	7	30602	25438	92.15	0.56	2.69	3	96.91	81478
MRT00162233	27	32	F	18	23498	19526	92.32	0.57	2.91	4	98.41	62550
MRT00161941	27	32	C	4	26873	22330	92.33	0.57	2.91	4	98.41	71533
MRT00163476	26	36	L	12	35000	29380	87.29	0.64	2.47	4	88.67	93760
MRT00161959	26	32	G	8	25030	20775	92.89	0.57	2.91	4	98.41	66580
MRT00160026	25	26	I	3	35299	29173	95.02	0.56	2.69	3	96.91	93645
MRT00161958	25	32	E	8	23959	19865	93.42	0.57	2.91	4	98.41	63689
MRT00160182	25	26	B	6	34895	28830	95.17	0.56	2.69	3	96.91	92555
MRT00160030	25	26	A	8	33490	27664	95.27	0.56	2.69	3	96.91	88818
MRT00158482	25	21	B	20	36772	30543	92.54	0.57	2.69	3	92.88	97858
MRT00160108	24	26	N	5	32338	26678	95.91	0.56	2.69	3	96.91	85694
MRT00161955	24	32	O	6	18413	15241	94.25	0.57	2.91	4	98.41	48895
MRT00161946	24	32	M	4	24722	20452	94.52	0.57	2.91	4	98.41	65626
MRT00159960	23	26	E	9	32284	26605	96.44	0.56	2.69	3	96.91	85494
MRT00162218	23	32	H	14	25685	21235	94.84	0.57	2.91	4	98.41	68155
MRT00159951	23	26	C	7	33391	27501	96.73	0.56	2.69	3	96.91	88393
MRT00162234	22	32	H	18	24546	20274	95.31	0.57	2.91	4	98.41	65094
MRT00161940	22	32	A	4	25560	21096	95.68	0.57	2.91	4	98.41	67752
MRT00161094	22	29	F	4	24527	19604	111.55	0.68	2.66	4	113.96	63735
MRT00161099	22	29	B	6	28057	22424	111.59	0.68	2.66	4	113.96	72905
MRT00161116	21	29	D	10	28303	22588	112.30	0.68	2.66	4	113.96	73479
MRT00162014	20	32	G	22	25764	21197	97.25	0.57	2.91	4	98.41	68158
MRT00161092	20	29	B	4	28977	23081	113.26	0.68	2.66	4	113.96	75139
MRT00162220	19	32	L	14	22787	18728	97.77	0.57	2.91	4	98.41	60243
MRT00161943	19	32	G	4	23299	19145	97.87	0.57	2.91	4	98.41	61589
MRT00163349	18	36	J	5	35057	29143	92.12	0.64	2.47	3	92.34	93343
MRT00162190	18	32	L	6	24803	20355	98.50	0.57	2.91	3	101.76	65513
MRT00161102	18	29	H	6	26011	20658	114.70	0.68	2.66	3	117.75	67327
MRT00160821	18	28	F	10	30354	24787	100.96	0.55	3.89	3	101.24	79928
MRT00162008	17	32	K	20	24263	19888	99.09	0.57	2.91	3	101.76	64039
MRT00161963	17	32	O	8	23950	19628	99.18	0.57	2.91	3	101.76	63206
MRT00162005	17	32	E	20	23119	18928	99.67	0.57	2.91	3	101.76	60975
MRT00162217	16	32	F	14	24038	19679	99.71	0.57	2.91	3	101.76	63396
MRT00161097	14	29	N	4	25557	20184	117.47	0.68	2.66	3	117.75	65925
MRT00162232	14	32	D	18	18569	15147	101.49	0.57	2.91	3	101.76	48863
MRT00161793	14	31	P	4	19826	15939	108.68	0.63	2.90	4	108.77	51704
MRT00161716	13	31	B	5	21742	17440	109.80	0.63	2.90	3	111.49	56622

Appendix 3: Table illustrating structures with similar structures to the MRCT hit.

 MS27	 MS28	 MS29
 MS30	 MS31	 MS32
 MS33	 MS34	 MS35
 MS36	 MS37	 MS38
 MS39	 MS40	

Appendix 4: Keap1 Sequences for Baculovirus expression

1) Full length Keap1 with C-Terminal TEV-10His-FLAG

```
  1 MQPDPRPSGA GACCRFLPLQ SQCPEGAGDA VMYASTECKA EVTPSQHGNR
 51 TFSYTLEDHT KQAFGIMNEL RLSQQLCDVT LQVKYQDAPA AQFMAHKVVL
101 ASSSPVFKAM FTNGLREQGM EVVSIEGIHP KVMERLIEFA YTASISMGEK
151 CVLHVMNGAV MYQIDSVVRA CSDFLVQQLD PSNAIGIANF AEQIGCVELH
201 QRAREYIYMH FGEVAKQEEF FNLSHCQLVT LISRDDLNVR CESEVFHACI
251 NWVKYDCEQR RFYVQALLRA VRCHSLTPNF LQMQLQKCEI LQSDSRCKDY
301 LVKIFEELTL HKPTQVMPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSDG
351 TWLRLADLQV PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT
401 NQWSPCAPMS VPRNRIGVGV IDGHIYAVGG SHGCIHHNSV ERYEPERDEW
451 HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG TNRLNSAECY YPERNEWRMI
501 TAMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE TETWTFVAPM
551 KHRRSALGIT VHQGRIYVLG GYDGHTFLDS VECYDPDTDT WSEVTRMTSG
601 RSGVGVAVTM EPCRKQIDQQ NCTCAENLYF QSHHHHHHHH HHDYKDDDDK
```

2) 6 Amino acid C-Terminal truncation with C-Terminal TEV-10His-FLAG

```
  1 MQPDPRPSGA GACCRFLPLQ SQCPEGAGDA VMYASTECKA EVTPSQHGNR
 51 TFSYTLEDHT KQAFGIMNEL RLSQQLCDVT LQVKYQDAPA AQFMAHKVVL
101 ASSSPVFKAM FTNGLREQGM EVVSIEGIHP KVMERLIEFA YTASISMGEK
151 CVLHVMNGAV MYQIDSVVRA CSDFLVQQLD PSNAIGIANF AEQIGCVELH
201 QRAREYIYMH FGEVAKQEEF FNLSHCQLVT LISRDDLNVR CESEVFHACI
251 NWVKYDCEQR RFYVQALLRA VRCHSLTPNF LQMQLQKCEI LQSDSRCKDY
301 LVKIFEELTL HKPTQVMPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSDG
351 TWLRLADLQV PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT
401 NQWSPCAPMS VPRNRIGVGV IDGHIYAVGG SHGCIHHNSV ERYEPERDEW
451 HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG TNRLNSAECY YPERNEWRMI
501 TAMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE TETWTFVAPM
551 KHRRSALGIT VHQGRIYVLG GYDGHTFLDS VECYDPDTDT WSEVTRMTSG
601 RSGVGVAVTM EPCRKQIDAE NLYFQSHHHH HHHHHHDYKD DDDK
```

3) 47 Amino acid N-Terminal truncation with C-Terminal TEV-10His-FLAG

```
  1 ---------- ---------- ---------- ---------- -------GNR
 51 TFSYTLEDHT KQAFGIMNEL RLSQQLCDVT LQVKYQDAPA AQFMAHKVVL
101 ASSSPVFKAM FTNGLREQGM EVVSIEGIHP KVMERLIEFA YTASISMGEK
151 CVLHVMNGAV MYQIDSVVRA CSDFLVQQLD PSNAIGIANF AEQIGCVELH
201 QRAREYIYMH FGEVAKQEEF FNLSHCQLVT LISRDDLNVR CESEVFHACI
251 NWVKYDCEQR RFYVQALLRA VRCHSLTPNF LQMQLQKCEI LQSDSRCKDY
301 LVKIFEELTL HKPTQVMPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSDG
351 TWLRLADLQV PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT
401 NQWSPCAPMS VPRNRIGVGV IDGHIYAVGG SHGCIHHNSV ERYEPERDEW
451 HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG TNRLNSAECY YPERNEWRMI
501 TAMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE TETWTFVAPM
551 KHRRSALGIT VHQGRIYVLG GYDGHTFLDS VECYDPDTDT WSEVTRMTSG
601 RSGVGVAVTM EPCRKQIDQQ NCTCAENLYF QSHHHHHHHH HHDYKDDDDK
```

4) 47 Amino acid N-Terminal truncation with 6 Amino acid C-Terminal truncation and TEV-10His-FLAG

```
  1 ---------- ---------- ---------- ---------- -------GNR
```

```
 51 TFSYTLEDHT KQAFGIMNEL RLSQQLCDVT LQVKYQDAPA AQFMAHKVVL
101 ASSSPVFKAM FTNGLREQGM EVVSIEGIHP KVMERLIEFA YTASISMGEK
151 CVLHVMNGAV MYQIDSVVRA CSDFLVQQLD PSNAIGIANF AEQIGCVELH
201 QRAREYIYMH FGEVAKQEEF FNLSHCQLVT LISRDDLNVR CESEVFHACI
251 NWVKYDCEQR RFYVQALLRA VRCHSLTPNF LQMQLQKCEI LQSDSRCKDY
301 LVKIFEELTL HKPTQVMPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSDG
351 TWLRLADLQV PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT
401 NQWSPCAPMS VPRNRIGVGV IDGHIYAVGG SHGCIHHNSV ERYEPERDEW
451 HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG TNRLNSAECY YPERNEWRMI
501 TAMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE TETWTFVAPM
551 KHRRSALGIT VHQGRIYVLG GYDGHTFLDS VECYDPDTDT WSEVTRMTSG
601 RSGVGVAVTM EPCRKQIDAE NLYFQSHHHH HHHHHHDYKD DDDK
```

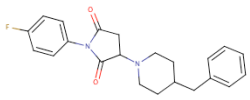
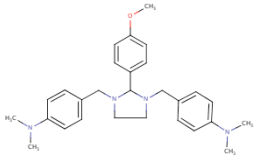
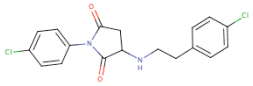
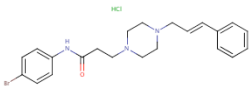
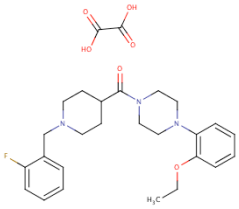
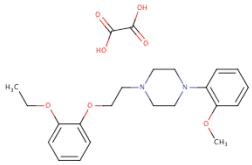
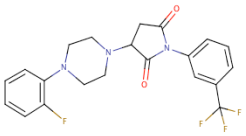
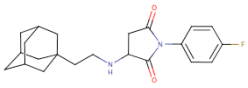
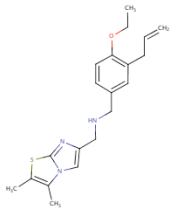
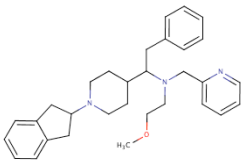
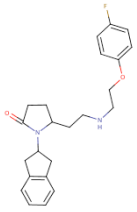
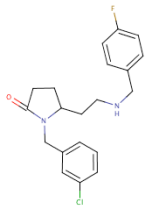
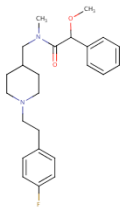
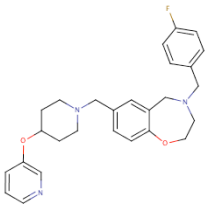
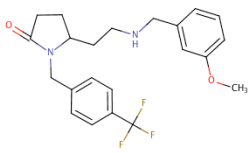
5) 177 Amino acid N-Terminal truncation with C-Terminal TEV-10His-FLAG

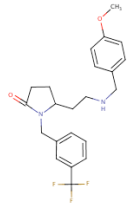
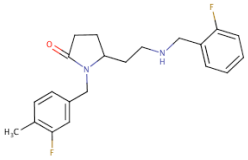
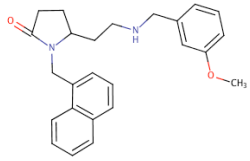
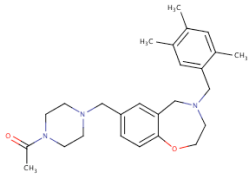
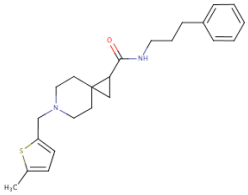
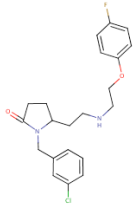
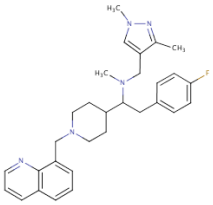
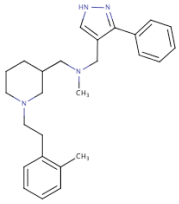
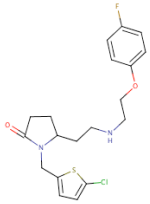
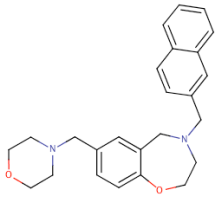
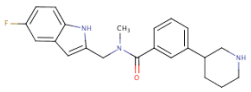
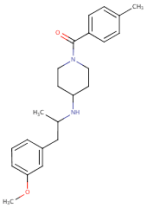
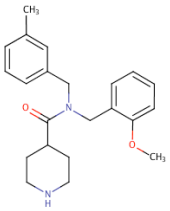
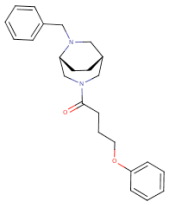
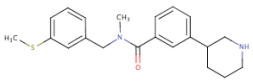
```
  1 ---------- ---------- ---------- ---------- ----------
 51 ---------- ---------- ---------- ---------- ----------
101 ---------- ---------- ---------- ---------- ----------
151 ---------- ---------- -------QLD PSNAIGIANF AEQIGCVELH
201 QRAREYIYMH FGEVAKQEEF FNLSHCQLVT LISRDDLNVR CESEVFHACI
251 NWVKYDCEQR RFYVQALLRA VRCHSLTPNF LQMQLQKCEI LQSDSRCKDY
301 LVKIFEELTL HKPTQVMPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSDG
351 TWLRLADLQV PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT
401 NQWSPCAPMS VPRNRIGVGV IDGHIYAVGG SHGCIHHNSV ERYEPERDEW
451 HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG TNRLNSAECY YPERNEWRMI
501 TAMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE TETWTFVAPM
551 KHRRSALGIT VHQGRIYVLG GYDGHTFLDS VECYDPDTDT WSEVTRMTSG
601 RSGVGVAVTM EPCRKQIDQQ NCTCAENLYF QSHHHHHHHH HHDYKDDDDK
```

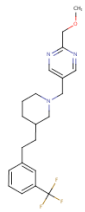
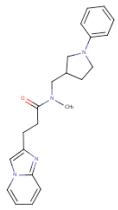
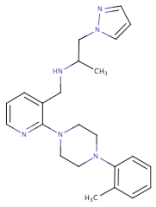
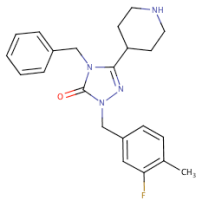
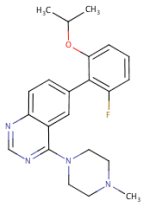
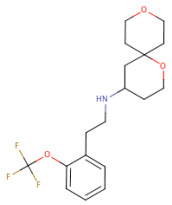
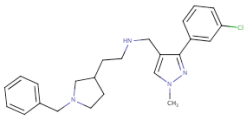
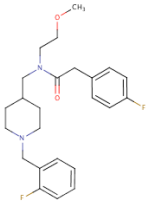
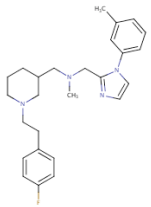
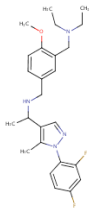
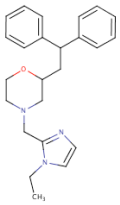
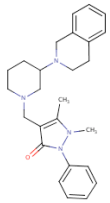
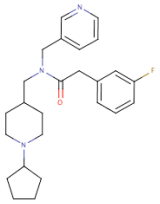
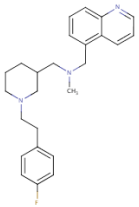
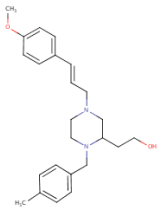
6) 177 Amino acid N-Terminal truncation with 6 Amino acid C-Terminal truncation and TEV-10His-FLAG

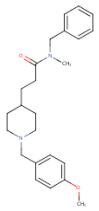
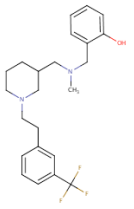
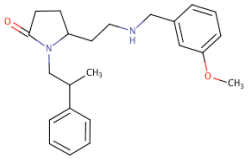
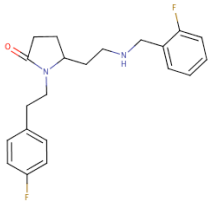
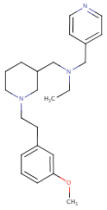
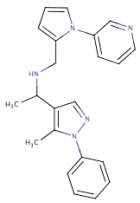
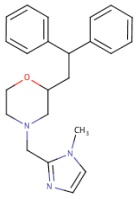
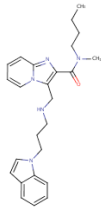
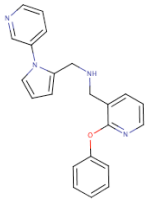
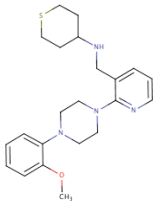
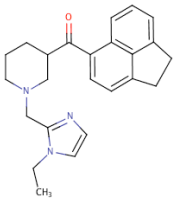
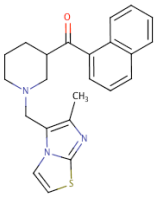
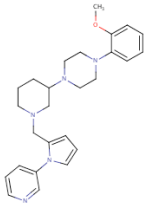
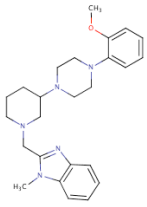
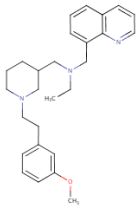
```
  1 ---------- ---------- ---------- ---------- ----------
 51 ---------- ---------- ---------- ---------- ----------
101 ---------- ---------- ---------- ---------- ----------
151 ---------- ---------- -------QLD PSNAIGIANF AEQIGCVELH
201 QRAREYIYMH FGEVAKQEEF FNLSHCQLVT LISRDDLNVR CESEVFHACI
251 NWVKYDCEQR RFYVQALLRA VRCHSLTPNF LQMQLQKCEI LQSDSRCKDY
301 LVKIFEELTL HKPTQVMPCR APKVGRLIYT AGGYFRQSLS YLEAYNPSDG
351 TWLRLADLQV PRSGLAGCVV GGLLYAVGGR NNSPDGNTDS SALDCYNPMT
401 NQWSPCAPMS VPRNRIGVGV IDGHIYAVGG SHGCIHHNSV ERYEPERDEW
451 HLVAPMLTRR IGVGVAVLNR LLYAVGGFDG TNRLNSAECY YPERNEWRMI
501 TAMNTIRSGA GVCVLHNCIY AAGGYDGQDQ LNSVERYDVE TETWTFVAPM
551 KHRRSALGIT VHQGRIYVLG GYDGHTFLDS VECYDPDTDT WSEVTRMTSG
601 RSGVGVAVTM EPCRKQIDAE NLYFQSHHHH HHHHHHDYKD DDDK
```

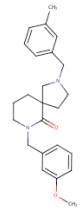
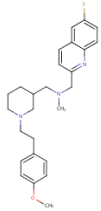
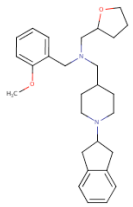
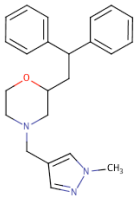
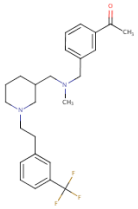
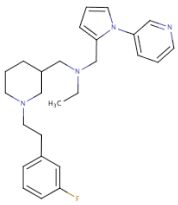
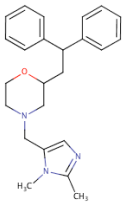
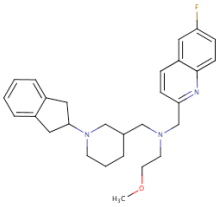
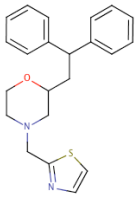
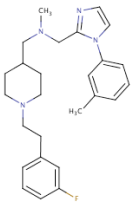
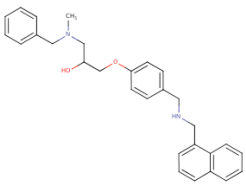
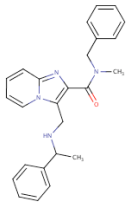
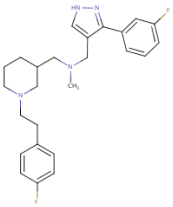
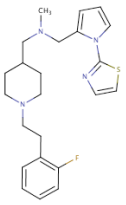
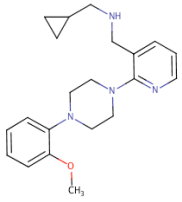
Appendix 5: Table illustrating chemical structures of compounds selected from *in-silico* screening

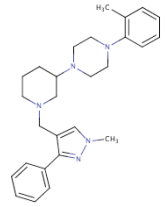
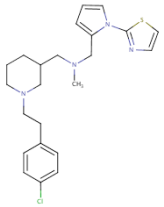
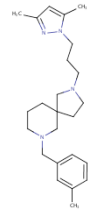
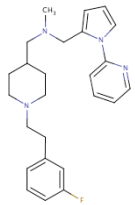
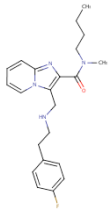
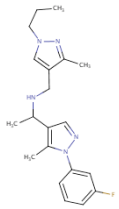
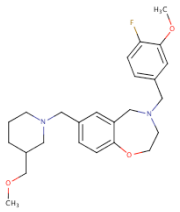
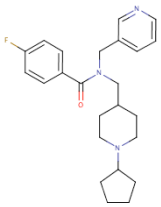
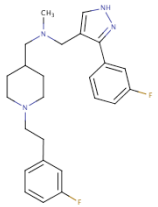
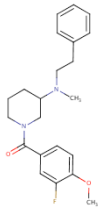
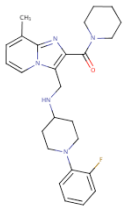
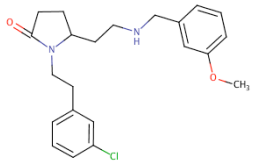
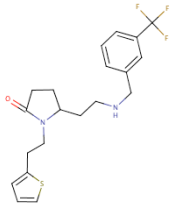
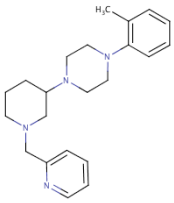
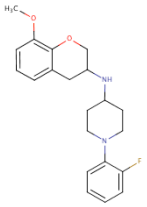
 MS41	 MS42	 MS43
 MS44	 MS45	 MS46
 MS47	 MS48	 MS49
 MS50	 MS51	 MS52
 MS53	 MS54	 MS55

 MS56	 MS57	 MS58
 MS59	 MS60	 MS61
 MS62	 MS63	 MS64
 MS65	 MS66	 MS67
 MS68	 MS69	 MS70

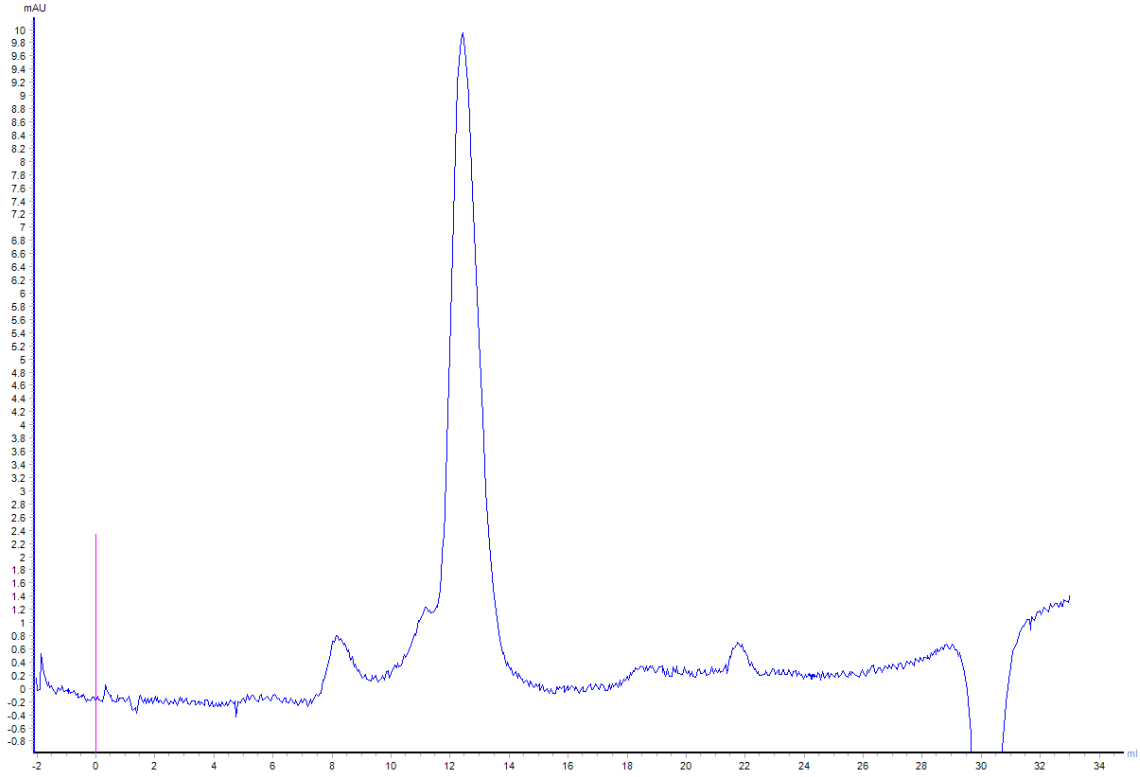
 MS71	 MS72	 MS73
 MS74	 MS75	 MS76
 MS77	 MS78	 MS79
 MS80	 MS81	 MS82
 MS83	 MS84	 MS85

 MS86	 MS87	 MS88
 MS89	 MS90	 MS91
 MS92	 MS93	 MS94
 MS95	 MS96	 MS97
 MS98	 MS99	 MS100

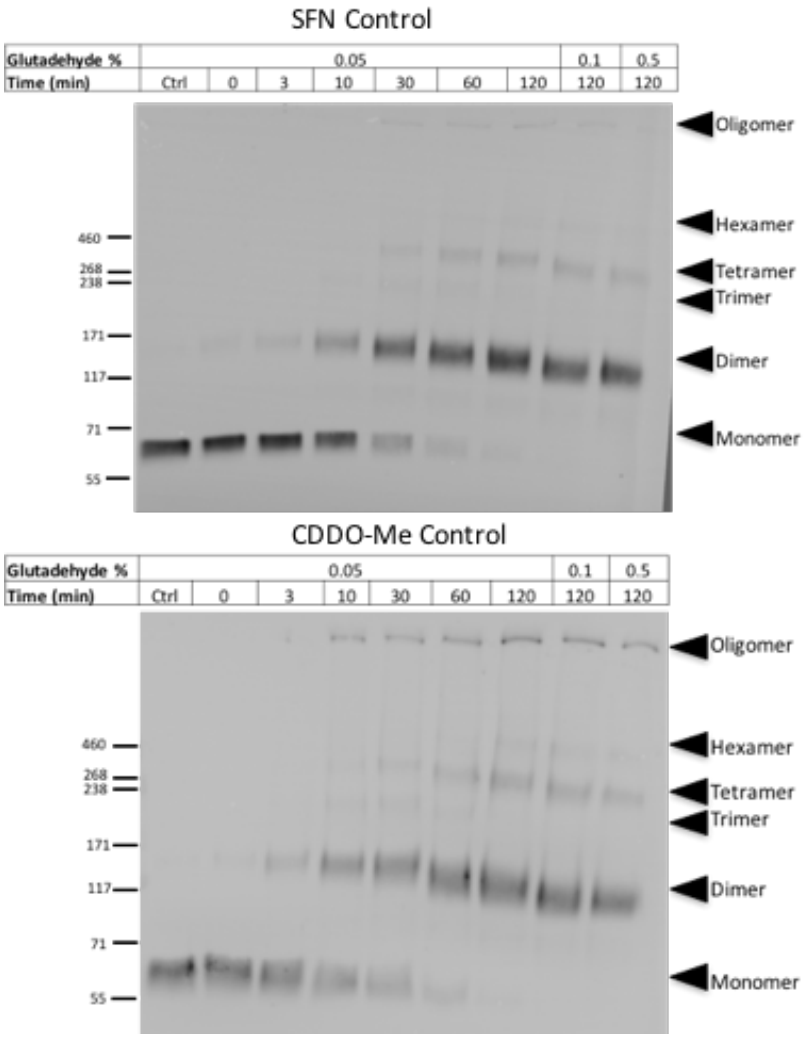
 MS101	 MS102	 MS103
 MS104	 MS105	 MS106
 MS107	 MS108	 MS109
 MS110	 MS111	 MS112
 MS113	 MS114	 MS115

 MS116	 MS117	 MS118
 MS119	 MS120	 MS121
 MS122	 MS123	 MS124
 MS125	 MS126	 MS127
 MS128	 MS129	 MS130

Appendix 6: Representative gel filtration trace of Buffer B injection with 1 mM DTT. UV trace at 280 nm shown in blue, and injection point shown in pink. Y axis shows absorption at 280 nM in mAU, X axis shown elution volume in ml.



Appendix 7: Control experiments for crosslinking experiments shown in Figure 112. Representative control SDS-PAGE analysis of Keap1 in the absence of CDDO (bottom) or Sulforaphane (top) after cross-linking with glutaraldehyde imaged with anti-Keap1 antibody. Molecular weight standard weights in kDa are indicated to the left of the image. Cross-linking of Keap1 (0.25 µM) performed with either 0.05 %, 0.1 % (2X) or 0.5 % (10X) glutaraldehyde (as indicated), was carried out in 10 mM HEPES, 150 mM NaCl, 0.01% NP40, and 2.5 mM TCEP at pH 8.0. The samples were analysed by SDS-PAGE after 0, 3, 10, 30, 60 or 120 min of incubation. Bands related to the molecular weights for monomeric, dimeric, trimeric, tetrameric, hexameric and larger oligomeric Keap1 are indicted with black arrows.



Bibliography

1. Kumar, P. & Clark, M. *Kumar and Clark Clinical Medicine*. *Russell J. Bertrand Russell Arch.* (Saunders Ltd., 2005). at <http://books.google.com/books?id=03BrAAAAMAAJ&pgis=1>
2. Rowland, L. P. How amyotrophic lateral sclerosis got its name: the clinical-pathologic genius of Jean-Martin Charcot. *Arch. Neurol.* **58,** 512–515 (2001).
3. Kaplan, P. M., Cottrell, R. C. & Ward, N. Luckiest Man: The Life and Death of Lou Gehrig. *Libr. J.* **130,** 90 (2005).
4. Goetz, C., Bonduelle, M. & Gelfand, T. *Charcot: Constructing Neurology*. (Oxford University Press, 1995).
5. McDermott, C. J. & Shaw, P. J. Diagnosis and management of motor neurone disease. *BMJ* **336,** 658–62 (2008).
6. Byrne, S. *et al.* Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J. Neurol. Neurosurg. Psychiatry* **82,** 623–7 (2011).
7. Rowland, L. P. & Shneider, N. A. Amyotrophic Lateral Sclerosis. *N Engl J Med* **344,** 1688–1700 (2001).
8. Shaw, P. J. & Strong, M. J. *Motor Neuron Disorders*. (Butterworth-Heinemann, 2003).
9. Kiernan, M. C. *et al.* Amyotrophic lateral sclerosis. *Lancet* **377,** 942–55 (2011).
10. Hudson, A. . in *Clin. Neurol. Vol. 4* 11–14 (1996).
11. Talbot, K. Motor neuron disease: The Bare Essentials. *Neurol. Pract.* **9,** 303–309 (2009).
12. Bruijn, L. I., Miller, T. M. & Cleveland, D. W. Unraveling the mechanisms involved in motor neuron degeneration in ALS. *Annu. Rev. Neurosci.* **27,** 723–49 (2004).
13. Proctor, P. H. & Reynolds, E. S. Free radicals and disease in man. *Physiol. Chem. Phys. Med. NMR* **16,** 175–195 (1984).
14. McCord, J. M. Human disease, free radicals, and the oxidant/antioxidant balance. *Clin. Biochem.* **26,** 351–7 (1993).
15. Lennon, S. V, Martin, S. J. & Cotter, T. G. Dose-dependent induction of apoptosis in human tumour cell lines by widely diverging stimuli. *Cell Prolif.* **24,** 203–14 (1991).
16. Shaw, P. J., Ince, P. G., Falkous, G. & Mantle, D. Oxidative damage to protein in sporadic motor neuron disease spinal cord. *Ann. Neurol.* **38,** 691–5 (1995).
17. Shibata, N. *et al.* Morphological evidence for lipid peroxidation and protein glycoxidation in spinal cords from sporadic amyotrophic lateral sclerosis patients. *Brain Res.* **917,** 97–104 (2001).
18. Fitzmaurice, P. S. *et al.* Evidence for DNA damage in Amylotrophic Lateral Sclerosis. *Muscle Nerve* **19,** 797–798 (1996).
19. Lenaz, G. *et al.* Role of mitochondria in oxidative stress and ageing. *Biochim. Biophys. Acta* **959,** 199–213 (1998).
20. Genova, M. L. *et al.* The Mitochondrial Production of Reactive Oxygen Species in Relation to Aging and Pathology. *Ann. N. Y. Acad. Sci.* **1011,** 86–100 (2004).
21. Beal, M. F. Oxidatively modified proteins in aging and disease. *Free Radic. Biol. Med.* **32,** 797–803 (2002).
22. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated

with familial amyotrophic lateral sclerosis. *Nature* **364,** 59–62 (1993).

23. Barber, S. C. & Shaw, P. J. Oxidative stress in ALS: key role in motor neuron injury and therapeutic target. *Free Radic. Biol. Med.* **48,** 629–41 (2010).

24. Strange, R. W., Yong, C. W., Smith, W. & Hasnain, S. S. Molecular dynamics using atomic-resolution structure reveal structural fluctuations that may lead to polymerization of human Cu-Zn superoxide dismutase. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 10040–10044 (2007).

25. Andrus, P. K., Fleck, T. J., Gurney, M. E. & Hall, E. D. Protein oxidative damage in a transgenic mouse model of familial amyotrophic lateral sclerosis. *J. Neurochem.* **71,** 2041–8 (1998).

26. Battistini, S., Benigni, M., Ricci, C. & Rossi, A. SOD1 mutations in amyotrophic lateral sclerosis. *Eur. Neurol. J.* 1–11 (2010). at <http://www.slm-psychiatry.com/uploads/media/SOD1_Mutations_in_Amyotrophic_Lateral_Sclerosis.pdf>

27. Beckman, J. S., Carson, M., Smith, C. D. & Koppenol, W. H. ALS, SOD and peroxynitrite. *Nature* **364,** 584 (1993).

28. Borchelt, D. R. *et al.* Superoxide dismutase 1 with mutations linked to familial amyotrophic lateral sclerosis possesses significant activity. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 8292–8296 (1994).

29. Yim, M., Kang, J. & Yim, H. A gain-of-function of an amyotrophic lateral sclerosis-associated Cu, Zn-superoxide dismutase mutant: an enhancement of free radical formation due to a decrease in Km for hydrogen peroxide. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 5709–5714 (1996).

30. Borchelt, D. & Lee, M. Superoxide dismutase 1 with mutations linked to familial amyotrophic lateral sclerosis possesses significant activity. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 8292–8296 (1994).

31. Reaume, A., Elliott, J. & Hoffman, E. Motor neurons in Cu/Zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury. *Nat. Genet.* **13,** 43–47 (1996).

32. Bruijn, L. I. *et al.* Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1. *Science (80-.).* **281,** 1851–1854 (1998).

33. Barber, S. C., Mead, R. J. & Shaw, P. J. Oxidative stress in ALS: a mechanism of neurodegeneration and a therapeutic target. *Biochim. Biophys. Acta* **1762,** 1051–67 (2006).

34. Yim, M. B., Chock, P. B. & Stadtman, E. R. Enzyme function of copper, zinc superoxide dismutase as a free radical generator. *J. Biol. Chem.* **268,** 4099–105 (1993).

35. Yim, M. B., Chock, P. B. & Stadtman, E. R. Copper, zinc superoxide dismutase catalyzes hydroxyl radical production from hydrogen peroxide. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 5006–10 (1990).

36. Crow, J. P., Sampson, J. B., Zhuang, Y., Thompson, J. A. & Beckman, J. S. Decreased zinc affinity of amyotrophic lateral sclerosis-associated superoxide dismutase mutants leads to enhanced catalysis of tyrosine nitration by peroxynitrite. *J. Neurochem.* **69,** 1936–1944 (1997).

37. Estévez, A. G. *et al.* Induction of nitric oxide-dependent apoptosis in motor neurons by zinc-deficient superoxide dismutase. *Science (80-.).* **286,** 2498–2500 (1999).

38. Roe, J., Wiedau-Pazos, M. & Moy, V. In vivo peroxidative activity of FALS-mutant human CuZnSODs expressed in yeast. *Free Radic. Biol. Med.* **32,** 169–174 (2002).

39. Wiedau-Pazos, M. *et al.* Altered reactivity of superoxide dismutase in familial amyotrophic lateral sclerosis. *Science (80-.).* **271,** 515–518 (1996).

40. Singh, R. J. *et al.* Reexamination of the mechanism of hydroxyl radical adducts formed from the reaction between familial amyotrophic lateral sclerosis-associated Cu,Zn superoxide dismutase mutants and H2O2. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6675–6680 (1998).

41. Smith, A. P. & Lee, N. M. Role of zinc in ALS. *Amyotroph. lateral Scler. Off. Publ. World Fed. Neurol. Res. Gr. Mot. Neuron Dis.* **8,** 131–143 (2007).

42. Lyons, T. J. *et al.* Mutations in copper-zinc superoxide dismutase that cause amyotrophic lateral sclerosis alter the zinc binding site and the redox behavior of the protein. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12240–12244 (1996).

43. Ermilova, I. P. *et al.* Protection by dietary zinc in ALS mutant G93A SOD transgenic mice. *Neurosci. Lett.* **379,** 42–46 (2005).

44. Wang, J. *et al.* Copper-binding-site-null SOD1 causes ALS in transgenic mice: aggregates of non-native SOD1 delineate a common feature. *Hum. Mol. Genet.* **12,** 2753–64 (2003).

45. Subramaniam, J. R. *et al.* Mutant SOD1 causes motor neuron disease independent of copper chaperone-mediated copper loading. *Nat. Neurosci.* **5,** 301–7 (2002).

46. Kirby, J. *et al.* Mutant SOD1 alters the motor neuronal transcriptome: implications for familial ALS. *Brain* **128,** 1686–706 (2005).

47. Bedard, K. & Krause, K. The NOX Family of ROS-Generating NADPH Oxidases : Physiology and Pathophysiology. *Physiol. Rev.* **87,** 245–313 (2007).

48. Harraz, M. M. *et al.* SOD1 mutations disrupt redox-sensitive Rac regulation of NADPH oxidase in a familial ALS model. *J. Clin. Invest.* **118,** 659–670 (2008).

49. Graham, M., Kennedy, K., Giampapola, M. & Bilyk, M. Revisiting oxidative damage in ALS : microglia , Nox , and mutant SOD1. *Production* **118,** 474–478 (2008).

50. Wu, D.-C., Ré, D. B., Nagai, M., Ischiropoulos, H. & Przedborski, S. The inflammatory NADPH oxidase enzyme modulates motor neuron degeneration in amyotrophic lateral sclerosis mice. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 12132–12137 (2006).

51. Marden, J. J. *et al.* Redox modifier genes in amyotrophic lateral sclerosis in mice. *J. Clin. Invest.* **117,** 2913–2919 (2007).

52. Van Damme, P., Dewil, M., Robberecht, W. & Van Den Bosch, L. Excitotoxicity and amyotrophic lateral sclerosis. *Therapie* **3,** 348–359 (2005).

53. Olney, J., McGeer, E. & McGeer, P. in *Kainic Acid as a Tool Neurobiol.* 95–121 (Raven Press, 1978).

54. Shaw, P. J., Ince, P. G., Johnson, M., Perry, E. K. & Candy, J. The quantitative autoradiographic distribution of [3H]MK-801 binding sites in the normal human spinal cord. *Brain Res.* **539,** 164–168 (1991).

55. Williams, T. L., Ince, P. G., Tyne, U. N. & Tyne, N. An immunocytochemical study of the distribution of AMPA selective glutamate receptor subunits in the normal human motor system. *Neuroscience* **74,** 185–198 (1996).

56. Niciu, M. J., Kelmendi, B. & Sanacora, G. Overview of glutamatergic neurotransmission in the nervous system. *Pharmacol. Biochem. Behav.* **100,** 656–64 (2012).

57. Estevez, A. G., Stutzmann, J. M. & Barbeito, L. Protective effect of riluzole on excitatory amino acid-mediated neurotoxicity in motoneuron-enriched cultures. *Eur. J. Pharmacol.* **280,** 47–53 (1995).

58. Talbot, K., Turner, M. R., Marsden, R. & Botell, R. *Motor Neuron Disease a practical manual*. (Oxford University Press, 2010).

59. Choi, D. W. Ionic dependence of glutamate neurotoxicity. *J. Neurosci.* **7,** 369–79 (1987).

60. Manev, H., Favaron, M., Guidotti, A. & Costa, E. Delayed increase of Ca2+ influx elicited by glutamate: role in neuronal death. *Mol. Pharmacol.* **36,** 106–112 (1989).

61. Meldrum, B. & Garthwaite, J. Excitatory amino acid neurotoxicity and neurodegenerative disease. *Trends Pharmacol. Sci.* **11,** 379–387 (1990).

62. Lees, G. J. Contributory mechanisms in the causation of neurodegenerative disorders. *Neuroscience* **54,** 287–322 (1993).

63. Bondy, S. C. & Lee, D. K. Oxidative stress induced by glutamate receptor agonists. *Brain Res.* **610,** 229–33 (1993).

64. Murphy, T. H., Miyamoto, M., Sastre, A., Schnaar, R. L. & Coyle, J. T. Glutamate toxicity in a neuronal cell line involves inhibition of cystine transport leading to oxidative stress. *Neuron* **2,** 1547–1558 (1989).

65. Shaw, P. J., Forrest, V., Ince, P. G., Richardson, J. P. & Wastell, H. J. CSF and plasma amino acid levels in motor neuron disease: elevation of CSF glutamate in a subset of patients. *Neurodegeneration* **4,** 209–216 (1995).

66. Rothstein, J. D., Martin, L. J. & Kuncl, R. W. Decreased glutamate transport by the brain and spinal cord in amyotrophic lateral sclerosis. *N. Engl. J. Med.* **326,** 1464–1468 (1992).

67. Foran, E. & Trotti, D. Glutamate Transporters and the Excitotoxic Path to Motor Neuron Degeneration in Amyotrophic Lateral Sclerosis. *Antioxidants redox Signal.* **11,** 1587–1602 (2009).

68. Henze, K. & Martin, W. Essence of mitochondria. *Nature* **426,** 13–14 (2003).

69. McBride, H. M., Neuspiel, M. & Wasiak, S. Mitochondria: more than just a powerhouse. *Curr. Biol.* **16,** R551–R560 (2006).

70. Mattiazzi, M. *et al.* Mutated human SOD1 causes dysfunction of oxidative phosphorylation in mitochondria of transgenic mice. *J. Biol. Chem.* **277,** 29626–33 (2002).

71. Wiedemann, F. R., Manfredi, G., Mawrin, C., Beal, M. F. & Schon, E. a. Mitochondrial DNA and respiratory chain function in spinal cords of ALS patients. *J. Neurochem.* **80,** 616–25 (2002).

72. Wong, P. C. *et al.* An adverse property of a familial ALS-linked SOD1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria. *Neuron* **14,** 1105–1116 (1995).

73. Liu, J. *et al.* Toxicity of familial ALS-linked SOD1 mutants from selective recruitment to spinal mitochondria. *Neuron* **43,** 5–17 (2004).

74. Vande Velde, C., Miller, T. M., Cashman, N. R. & Cleveland, D. W. Selective association of misfolded ALS-linked mutant SOD1 with the cytoplasmic face of mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 4022–7 (2008).

75. Pasinelli, P. *et al.* Amyotrophic lateral sclerosis-associated SOD1 mutant proteins bind and aggregate with Bcl-2 in spinal cord mitochondria. *Neuron* **43,** 19–30 (2004).

76. Damiano, M. *et al.* Neural mitochondrial Ca2+ capacity impairment precedes the onset of motor symptoms in G93A Cu/Zn-superoxide dismutase mutant mice. *J. Neurochem.* **96,** 1349–61 (2006).

77. Grosskreutz, J., Van Den Bosch, L. & Keller, B. U. Calcium dysregulation in amyotrophic lateral sclerosis. *Cell Calcium* **47,** 165–74 (2010).

78. Sathasivam, S., Grierson, a J. & Shaw, P. J. Characterization of the caspase cascade in a cell culture model of SOD1-related familial amyotrophic lateral sclerosis: expression, activation and therapeutic effects of inhibition. *Neuropathol. Appl. Neurobiol.* **31,** 467–85 (2005).

79. Menzies, F. M. *et al.* Mitochondrial dysfunction in a cell culture model of familial amyotrophic lateral sclerosis. *Brain* **125,** 1522–1533 (2002).

80. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80,** 155–165 (1995).

81. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (80-.).* **314,** 130–3 (2006).

82. Sreedharan, J. *et al.* TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science (80-.).* **319,** 1668–72 (2008).

83. Mackenzie, I. R., Rademakers, R. & Neumann, M. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol.* **9,** 995–1007 (2010).

84. Ferraiuolo, L., Kirby, J., Grierson, A. J., Sendtner, M. & Shaw, P. J. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* **7,** 616–630 (2011).

85. Liu-Yesucevitz, L. *et al.* Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. *PLoS One* **5,** e13250 (2010).

86. Dormann, D. *et al.* ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *EMBO J.* **29,** 2841–57 (2010).

87. Ito, D., Seki, M., Tsunoda, Y., Uchiyama, H. & Suzuki, N. Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS. *Ann. Neurol.* **69,** 152–62 (2011).

88. Chen, Y.-Z. *et al.* DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.* **74,** 1128–1135 (2004).

89. Greenway, M. J. *et al.* ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nat. Genet.* **38,** 411–3 (2006).

90. Kieran, D. *et al.* Control of motoneuron survival by angiogenin. *J. Neurosci.* **28,** 14056–61 (2008).

91. Chang, Y. *et al.* Messenger RNA oxidation occurs early in disease pathogenesis and promotes motor neuron degeneration in ALS. *PLoS One* **3,** e2849 (2008).

92. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–56 (2011).

93. Renton, A. E. *et al.* A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron* **72,** 257–68 (2011).

94. Friedland, R. P. *et al.* Behavioral variant frontotemporal lobar degeneration with amyotrophic lateral sclerosis with a chromosome 9p21 hexanucleotide repeat. *Front. Neurol.* **3,** 136 (2012).

95. Sha, S. & Boxer, A. Treatment implications of C9ORF72. *Alzheimers Res. Ther.* **4,** 46 (2012).

96. Kwiatkowski, T. J. *et al.* Mutations in the FUS/TLS gene on chromosome 16 cause

familial amyotrophic lateral sclerosis. *Science (80-.).* **323,** 1205–1208 (2009).

97. Vance, C. *et al.* Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. *Science (80-.).* **323,** 1208–1211 (2009).

98. Yang, Y. *et al.* The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nat. Genet.* **29,** 160–165 (2001).

99. Chow, C. Y. *et al.* Deleterious Variants of FIG4, a Phosphoinositide Phosphatase, in Patients with ALS. *Am. J. Hum. Genet.* **84,** 85–88 (2009).

100. Nishimura, A. L. *et al.* A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am. J. Hum. Genet.* **75,** 822–831 (2004).

101. Maruyama, H. *et al.* Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **465,** 223–6 (2010).

102. Mitchell, J. *et al.* Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 7556–7561 (2010).

103. Johnson, J. O. *et al.* Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* **68,** 857–864 (2010).

104. Deng, H.-X. *et al.* Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature* **477,** 211–215 (2011).

105. Hutton, M. Missense and splice site mutations in tau associated with FTDP-17: multiple pathogenic mechanisms. *Neurology* **56,** S21–S25 (2001).

106. Orlacchio, A. *et al.* SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. *Brain* **133,** 591–598 (2010).

107. Luty, A. A. *et al.* Sigma nonopioid intracellular receptor 1 mutations cause frontotemporal lobar degeneration-motor neuron disease. *Ann. Neurol.* **68,** 639–649 (2010).

108. Al-Saif, A., Al-Mohanna, F. & Bohlega, S. A mutation in sigma-1 receptor causes juvenile amyotrophic lateral sclerosis. *Ann. Neurol.* **70,** 913–9 (2011).

109. Hosler, B. A. *et al.* Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22. *J. Am. Med. Assoc.* **20,** 68–75 (2000).

110. Hand, C. K. *et al.* A novel locus for familial amyotrophic lateral sclerosis, on chromosome 18q. *Am. J. Hum. Genet.* **70,** 251–256 (2002).

111. Sapp, P. C. *et al.* Identification of Two Novel Loci for Dominantly Inherited Familial Amyotrophic Lateral Sclerosis. *Am. J. Hum. Genet.* **73,** 397–403 (2003).

112. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* 10–17 (2004). doi:10.1038/nm1066

113. Piao, Y. & Wakabayashi, K. Neuropathology with Clinical Correlations of Sporadic Amyotrophic Lateral Sclerosis : 102 Autopsy Cases Examined Between 1962 and 2000. *Brain Pathol.* **13,** 10–22 (2003).

114. Shibata, N. *et al.* Cu/Zn superoxide dismutase-like immunoreactivity in Lewy body-like inclusions of sporadic amyotrophic lateral sclerosis. *Neurosci. Lett.* **179,** 149–152 (1994).

115. Rakhit, R. *et al.* An immunological epitope selective for pathological monomer-misfolded SOD1 in ALS. *Nat. Med.* **13,** 754–759 (2007).

116. Giordana, M. T. *et al.* TDP-43 Redistribution is an Early Event in Sporadic

Amyotrophic Lateral Sclerosis. *Brain Pathol.* **20,** 351–360 (2010).

117. Hall, E. D., Oostveen, J. O. A. & Gurney, M. E. Relationship of Microglial and Astrocytic Activation to Disease Onset and Progression in a Transgenic Model of Familial ALS. *Glia* **23,** 249–256 (1998).

118. Henkel, J. S., Beers, D. R., Zhao, W. & Appel, S. H. Microglia in ALS : The Good , The Bad , and The Resting. *J. Neuroimmune Pharmacol.* **4,** 389–398 (2009).

119. Stanley, H. A., G, R., Jozsef, I. E. & Stefani, E. Evidence for autoimmunity in amyotrophic lateral sclerosis. *J. Neurol. Sci.* **118,** 169–174 (1993).

120. Hensley, K. *et al.* Message and protein-level elevation of tumor necrosis factor alpha (TNFalpha) and TNFalpha-modulating cytokines in spinal cords of the G93A-SOD1 mouse model for amyotrophic lateral sclerosis. *Neurobiol. Dis.* **14,** 74–80 (2003).

121. Poloni, M. *et al.* Circulating levels of tumour necrosis factor-alpha and its soluble receptors are increased in the blood of patients with amyotrophic lateral sclerosis. *Neurosci. Lett.* **287,** 211–214 (2000).

122. Cereda, C. *et al.* TNF and sTNFR 1/2 plasma levels in ALS patients. *J. Neuroimmunol.* **194,** 123–131 (2008).

123. Mitchell, R. M. *et al.* A CSF biomarker panel for identification of patients with amyotrophic lateral sclerosis. *Neurol.* **72 ,** 14–19 (2009).

124. Mantovani, S. *et al.* Immune system alterations in sporadic amyotrophic lateral sclerosis patients suggest an ongoing neuroinflammatory process. *J. Neuroimmunol.* **210,** 73–79 (2009).

125. Kuhle, J. *et al.* Increased levels of inflammatory chemokines in amyotrophic lateral sclerosis. *Eur. J. Neurol.* **16,** 771–774 (2009).

126. Hensley, K. *et al.* Temporal patterns of cytokine and apoptosis-related gene expression in spinal cords of the G93A-SOD1 mouse model of amyotrophic lateral sclerosis. *J. Neurochem.* **82,** 365–374 (2002).

127. Rentzos, M. *et al.* Interleukin-15 and Interleukin-12 Are Elevated in Serum and Cerebrospinal Fluid of Patients with Amyotrophic Lateral Sclerosis. *Eur. Neurol.* **63,** 285–90 (2010).

128. Yoshihara, T., Ishigaki, S., Yamamoto, M., Liang, Y. & Niwa, J. Differential expression of inflammation- and apoptosis-related genes in spinal cords of a mutant SOD1 transgenic mouse model of familial amyotrophic lateral sclerosis. *J. Neurochem.* **80,** 158–167 (2002).

129. Weydt, P., Yuen, E. C., Ransom, B. R. & Mo, T. Increased Cytotoxic Potential of Microglia From ALS-Transgenic Mice. *Glia* **48,** 179–182 (2004).

130. Gowing, G., Dequen, F., Soucy, G. & Julien, J.-P. Absence of tumor necrosis factor-alpha does not affect motor neuron disease caused by superoxide dismutase 1 mutations. *J. Neurosci.* **26,** 11397–402 (2006).

131. Hensley, K. *et al.* Primary glia expressing the G93A-SOD1 mutation present a neuroinflammatory phenotype and provide a cellular system for studies of glial inflammation. *J. Neuroinflammation* **3,** 1–9 (2006).

132. Clement, A. M. *et al.* Wild-type nonneuronal cells extend survival of SOD1 mutant motor neurons in ALS mice. *Science (80-.).* **302,** 113–7 (2003).

133. Rg, M., Jd, M., Lyon, M. & Dh, M. Riluzole for amyotrophic lateral sclerosis (ALS)/ motor neuron disease (MND). *Cochrane Database Syst. Rev.* (2007).

134. Rothstein, J., Martin, L. & Kuncl, R. Decreased glutamate transport by the brain and spinal cord in amyotrophic lateral sclerosis. *N. Engl. J. Med.* **326,** 1464–1468 (1992).

135. Gurney, M. E., Fleck, T. J., Himes, C. S. & Hall, E. D. Riluzole preserves motor function in a transgenic model of familial amyotrophic lateral sclerosis. *Neurology* **50,** 62–6 (1998).

136. Gurney, M. E. *et al.* Benefit of vitamin E, riluzole, and gabapentin in a transgenic model of familial amyotrophic lateral sclerosis. *Ann. Neurol.* **39,** 147–157 (1996).

137. Bruijn, L. I. *et al.* ALS-linked SOD1 mutant G85R mediates damage to astrocytes and promotes rapidly progressive disease with SOD1-containing inclusions. *Neuron* **18,** 327–38 (1997).

138. Petajan, J. H., Ravits, J. M., Ringel, S. P., Ross, M. A. & Group, S. Placebo-controlled trial of gabapentin in patients with amyotrophic lateral sclerosis. *Neurology* **47,** 1383–1388 (1996).

139. Desnuelle, C., Dib, M., Garrel, C. & Favier, A. *A double-blind, placebo-controlled randomized clinical trial of alpha-tocopherol (vitamin E) in the treatment of amyotrophic lateral sclerosis. ALS riluzole-tocopherol Study Group. Amyotroph. lateral Scler. other Mot. neuron Disord. Off. Publ. World Fed. Neurol. Res. Gr. Mot. Neuron Dis.* **2,** (2001).

140. Cudkowicz, M. E. *et al.* A randomized, placebo-controlled trial of memantine for functional disability in amyotrophic lateral sclerosis. *Neurology* **61,** 456–460 (2003).

141. Bensimon, G., Lacomblez, L. & Meininger, V. A controlled trial of riluzole in amyotrophic lateral sclerosis. *N. Engl. J. Med.* **330,** 585–591 (1994).

142. Cheah, B. C., Vucic, S., Krishnan, a V & Kiernan, M. C. Riluzole, neuroprotection and amyotrophic lateral sclerosis. *Curr. Med. Chem.* **17,** 1942–199 (2010).

143. Holloway, R. G. *et al. Pramipexole vs levodopa as initial treatment for Parkinson disease: a 4-year randomized controlled trial. Arch. Neurol.* **61,** (2004).

144. Gribkoff, V. K. & Bozik, M. E. KNS-760704 [(6R)-4,5,6,7-tetrahydro-N6-propyl-2, 6-benzothiazole-diamine dihydrochloride monohydrate] for the treatment of amyotrophic lateral sclerosis. *CNS Neurosci. Ther.* **14,** 215–26 (2008).

145. Bozik, M. E., Mather, J. L., Kramer, W. G., Gribkoff, V. K. & Ingersoll, E. W. Safety, tolerability, and pharmacokinetics of KNS-760704 (dexpramipexole) in healthy adult subjects. *J. Clin. Pharmacol.* **51,** 1177–1185 (2011).

146. Samson, B. Y. K. New ALS drug shows dose-dependant efficacy in phase 2 trial. *Neurol. Today* **10,** 1, 34 (2010).

147. Biogen-Idec. EMPOWER trial fails to demonstrate efficacy in primary and key secondary endpoints. (2013). at <http://www.biogenidec.com/PRESS_RELEASE_DETAILS.aspx?ID=5981&ReqId=1770780>

148. Orrell, R., Lane, R. & Ross, M. Antioxidant treatment for amyotrophic lateral sclerosis or motor neuron disease. *Cochrane Database Syst. Rev.* (2007).

149. De Boer, A. G. & Gaillard, P. J. Strategies to improve drug delivery across the blood-brain barrier. *ClinPharmacokinet* **56,** 553 –576 (2007).

150. Bova, M. P. & Kinney, G. G. Emerging drug targets in amyotrophic lateral sclerosis. *Expert Opin. Orphan Drugs* **1,** 5–20 (2013).

151. MND Treatment trials. *MNDA* (2015). at <http://www.mndassohttp//www.mndassociation.org/research/mnd-research-and-you/treatment-trials/>

152. Morren, J. A. & Galvez-Jimenez, N. Current and prospective disease-modifying

therapies for amyotrophic lateral sclerosis. *Expert Opin. Investig. Drugs* **21,** 297–320 (2012).

153. Hsueh, K.-W., Hsieh, A.-C., Harn, H.-J. & Lin, S.-Z. Stem cell therapy in amyotrophic lateral sclerosis. *BioMedicine* **2,** 58–63 (2012).

154. Genervon Press Releases. (2015). at <http://www.genervon.com/genervon/about_pressreleases.php>

155. Smith, R. & Miller, T. Antisense oligonucleotide therapy for neurodegenerative disease. *J. Clin. Invest.* **116,** 4–10 (2006).

156. Nizzardo, M. *et al.* Research advances in gene therapy approaches for the treatment of amyotrophic lateral sclerosis. *Cell. Mol. Life Sci.* **69,** 1641–50 (2012).

157. Magesh, S., Chen, Y. & Hu, L. Small Molecule Modulators of Keap1-Nrf2-ARE Pathway as Potential Preventive and Therapeutic Agents *. *Med. Res. Rev.* **32,** 687–726 (2012).

158. Moi, P., Chan, K., Asunis, I., Cao, a & Kan, Y. W. Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 9926–30 (1994).

159. Itoh, K. *et al.* An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. *Biochem. Biophys. Res. Commun.* **236,** 313–322 (1997).

160. Itoh, K. *et al.* Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev.* **13,** 76–86 (1999).

161. Nguyen, T., Yang, C. S. & Pickett, C. B. The pathways and molecular mechanisms regulating Nrf2 activation in response to chemical stress. *Free Radic. Biol. Med.* **37,** 433–41 (2004).

162. Baird, L. & Dinkova-Kostova, A. T. The cytoprotective role of the Keap1-Nrf2 pathway. *Arch. Toxicol.* **85,** 241–72 (2011).

163. Hochmuth, C. E., Biteau, B., Bohmann, D. & Jasper, H. Redox regulation by Keap1 and Nrf2 controls intestinal stem cell proliferation in Drosophila. *Cell Stem Cell* **8,** 188–99 (2011).

164. Surh, Y.-J., Kundu, J. K. & Na, H.-K. Nrf2 as a master redox switch in turning on the cellular signaling involved in the induction of cytoprotective genes by some chemopreventive phytochemicals. *Planta Med.* **74,** 1526–39 (2008).

165. Brigelius-Flohé, R. & Flohé, L. Basic principles and emerging concepts in the redox control of transcription factors. *Antioxid. Redox Signal.* **15,** 2335–81 (2011).

166. Suh, J. H. *et al.* Decline in transcriptional activity of Nrf2 causes age-related loss of glutathione synthesis, which is reversible with lipoic acid. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3381–6 (2004).

167. Mimoto, T. *et al.* Impaired antioxydative Keap1/Nrf2 system and the downstream stress protein responses in the motor neuron of ALS model mice. *Brain Res.* **1446,** 109–18 (2012).

168. Vargas, M. R., Johnson, D. A., Sirkis, D. W., Messing, A. & Johnson, J. A. Nrf2 activation in astrocytes protects against neurodegeneration in mouse models of familial amyotrophic lateral sclerosis. *J. Neurosci.* **28,** 13574–13581 (2008).

169. Guo, Y. *et al.* The modest impact of transcription factor Nrf2 on the course of disease in an ALS animal model. *Lab. Invest.* **93,** 825–33 (2013).

170. Sarlette, A. *et al.* Nuclear erythroid 2-related factor 2-antioxidative response element signaling pathway in motor cortex and spinal cord in amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.* **67,** 1055–1062 (2008).

171. Tanji, K. *et al.* Keap1 is localized in neuronal and glial cytoplasmic inclusions in various neurodegenerative diseases. *J. Neuropathol. Exp. Neurol.* **72,** 18–28 (2013).

172. Duan, W. *et al.* Mutant TAR DNA-binding protein-43 induces oxidative injury in motor neuron-like cell. *Neuroscience* **169,** 1621–9 (2010).

173. Guo, Y. *et al.* HO-1 induction in motor cortex and intestinal dysfunction in TDP-43 A315T transgenic mice. *Brain Res.* **1460,** 88–95 (2012).

174. Collins, M. *et al.* The RNA-binding motif 45 (RBM45) protein accumulates in inclusion bodies in amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP) patients. *Acta Neuropathol.* **124,** 717–732 (2012).

175. Bakkar, N., Kousari, A., Kovalik, T., Li, Y. & Bowser, R. RBM45 modulates the antioxidant response in ALS through interactions with KEAP1. *Mol. Cell. Biol.* MCB.00087–15 (2015). doi:10.1128/MCB.00087-15

176. Kobayashi, A., Kang, M. & Watai, Y. Oxidative and electrophilic stresses activate Nrf2 through inhibition of ubiquitination activity of Keap1. *Mol. Cell. Biol.* **26,** 221–229 (2006).

177. Villeneuve, N., Lau, A. & Zhang, D. D. Regulation of the Nrf2–Keap1 antioxidant response by the ubiquitin proteasome system: An insight into Cullin-ring ubiquitin ligases. *Antioxid. Redox Signal.* **13,** 1699–1712 (2010).

178. Eggler, A. L., Gay, K. a & Mesecar, A. D. Molecular mechanisms of natural products in chemoprevention: induction of cytoprotective enzymes by Nrf2. *Mol. Nutr. Food Res.* **52 Suppl 1,** S84–94 (2008).

179. Nguyen, T., Nioi, P. & Pickett, C. B. The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *J. Biol. Chem.* **284,** 13291–5 (2009).

180. Kwak, M.-K., Itoh, K., Yamamoto, M. & Kensler, T. W. Enhanced Expression of the Transcription Factor Nrf2 by Cancer Chemopreventive Agents: Role of Antioxidant Response Element-Like Sequences in the nrf2 Promoter. *Mol. Cell. Biol.* **22,** 2883–2892 (2002).

181. Ogura, T. *et al.* Keap1 is a forked-stem dimer structure with two large spheres enclosing the intervening, double glycine repeat, and C-terminal domains. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2842–7 (2010).

182. Kobayashi, A. *et al.* Oxidative Stress Sensor Keap1 Functions as an Adaptor for Cul3-Based E3 Ligase To Regulate Proteasomal Degradation of Nrf2 Oxidative Stress Sensor Keap1 Functions as an Adaptor for Cul3-Based E3 Ligase To Regulate Proteasomal Degradation of Nrf2. *Mol. Cell. Biol.* **24,** 7130–7139 (2004).

183. Zipper, L. M. & Mulcahy, R. T. The Keap1 BTB/POZ dimerization function is required to sequester Nrf2 in cytoplasm. *J. Biol. Chem.* **277,** 36544–52 (2002).

184. Velichkova, M. & Hasson, T. Keap1 Regulates the Oxidation-Sensitive Shuttling of Nrf2 into and out of the Nucleus via a Crm1-Dependent Nuclear Export Mechanism. *Mol. Cell. Biol.* **25,** 4501–4513 (2005).

185. Tong, K., Katoh, Y. & Kusunoki, H. Keap1 Recruits Neh2 through Binding to ETGE and DLG Motifs: Characterization of the Two-Site Molecular Recognition Model. *Mol. Cell. Biol.* **26,** 2887–2900 (2006).

186. Wakabayashi, N. *et al.* Protection against electrophile and oxidant stress by

induction of the phase 2 response: fate of cysteines of the Keap1 sensor modified by inducers. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2040–5 (2004).

187. Dinkova-Kostova, A. T. *et al.* Direct evidence that sulfhydryl groups of Keap1 are the sensors regulating induction of phase 2 enzymes that protect against carcinogens and oxidants. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11908–13 (2002).

188. Katsuoka, F. *et al.* Genetic evidence that small maf proteins are essential for the activation of antioxidant response element-dependent genes. *Mol. Cell. Biol.* **25,** 8044–51 (2005).

189. Nioi, P., Nguyen, T., Sherratt, P. J. & Pickett, C. B. The Carboxy-Terminal Neh3 Domain of Nrf2 Is Required for Transcriptional Activation. *Mol. Cell. Biol.* **25,** 10895–10906 (2005).

190. Katoh, Y. *et al.* Two domains of Nrf2 cooperatively bind CBP, a CREB binding protein, and synergistically activate transcription. *Genes to Cells* **6,** 857–868 (2001).

191. McMahon, M. *et al.* The Cap'n'Collar basic leucine zipper transcription factor Nrf2 (NF-E2 p45-related factor 2) controls both constitutive and inducible expression of intestinal detoxification and glutathione biosynthetic enzymes. *Cancer Res.* **61,** 3299–3307 (2001).

192. Rada, P. *et al.* SCF/{beta}-TrCP promotes glycogen synthase kinase 3-dependent degradation of the Nrf2 transcription factor in a Keap1-independent manner. *Mol. Cell. Biol.* **31,** 1121–1133 (2011).

193. Cuadrado, A. Structural and functional characterization of NRF2 degradation By glycogen synthase kinase 3/β-TrCP. *Free Radic. Biol. Med.* (2015). doi:10.1016/j.freeradbiomed.2015.04.029

194. Nioi, P., McMahon, M., Itoh, K., Yamamoto, M. & Hayes, J. D. Identification of a novel Nrf2-regulated antioxidant response element (ARE) in the mouse NAD(P)H:quinone oxidoreductase 1 gene: reassessment of the ARE consensus sequence. *Biochem. J.* **374,** 337–348 (2003).

195. Dhakshinamoorthy, S., Jain, A. K., Bloom, D. A. & Jaiswal, A. K. Bach1 competes with Nrf2 leading to negative regulation of the antioxidant response element (ARE)-mediated NAD(P)H:quinone oxidoreductase 1 gene expression and induction in response to antioxidants. *J. Biol. Chem.* **280,** 16891–16900 (2005).

196. Zhang, D. D. *et al.* Ubiquitination of Keap1, a BTB-Kelch substrate adaptor protein for Cul3, targets Keap1 for degradation by a proteasome-independent pathway. *J. Biol. Chem.* **280,** 30091–30099 (2005).

197. Tong, K. I., Kobayashi, A., Katsuoka, F. & Yamamoto, M. Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. *Biol. Chem.* **387,** 1311–1320 (2006).

198. Hong, F., Sekhar, K. R., Freeman, M. L. & Liebler, D. C. Specific patterns of electrophile adduction trigger Keap1 ubiquitination and Nrf2 activation. *J. Biol. Chem.* **280,** 31768–75 (2005).

199. Nguyen, T., Sherratt, P. J., Huang, H.-C., Yang, C. S. & Pickett, C. B. Increased protein stability as a mechanism that enhances Nrf2-mediated transcriptional activation of the antioxidant response element. Degradation of Nrf2 by the 26 S proteasome. *J. Biol. Chem.* **278,** 4536–4541 (2003).

200. Canning, P. *et al.* Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases. *J. Biol. Chem.* **288,** 7803–7814 (2013).

201. Choo, Y. Y. & Hagen, T. Mechanism of cullin3 E3 ubiquitin ligase dimerization. *PLoS One* **7,** (2012).

202. Zhang, D. & Hannink, M. Distinct Cysteine Residues in Keap1 Are Required for Keap1-Dependent Ubiquitination of Nrf2 and for Stabilization of Nrf2 by Chemopreventive Agents and Oxidative Stress. *Mol. Cell. Biol.* **23,** 8137–8151 (2003).

203. Rachakonda, G. *et al.* Covalent modification at Cys151 dissociates the electrophile sensor Keap1 from the ubiquitin ligase CUL3. *Chem. Res. Toxicol.* **21,** 705–710 (2008).

204. Baird, L. & Dinkova-Kostova, A. T. Diffusion dynamics of the Keap1-Cullin3 interaction in single live cells. *Biochem. Biophys. Res. Commun.* **433,** 58–65 (2013).

205. Sun, Z., Zhang, S., Chan, J. Y. & Zhang, D. D. Keap1 Controls Postinduction Repression of the Nrf2-Mediated Antioxidant Response by Escorting Nuclear Export of Nrf2. *Mol. Cell. Biol.* **27,** 6334–6349 (2007).

206. Kobayashi, M. *et al.* The antioxidant defense system Keap1-Nrf2 comprises a multiple sensing mechanism for responding to a wide range of chemical compounds. *Mol. Cell. Biol.* **29,** 493–502 (2009).

207. Zhuang, C., Narayanapillai, S., Zhang, W., Sham, Y. Y. & Xing, C. Rapid Identification of Keap1-Nrf2 Small-Molecule Inhibitors through Structure-Based Virtual Screening and Hit-Based Substructure Search. *J. Med. Chem.* **57,** 1–7 (2014).

208. Sun, H.-P. *et al.* Novel protein–protein interaction inhibitor of Nrf2–Keap1 discovered by structure-based virtual screening. *Medchemcomm* **5,** 93 (2014).

209. Jiang, Z.-Y. *et al.* Discovery of potent Keap1-Nrf2 protein-protein interaction inhibitor based on molecular binding determinants analysis. *J. Med. Chem.* **57,** 2736–45 (2014).

210. Marcotte, D. *et al.* Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorg. Med. Chem.* **21,** 4011–4019 (2013).

211. Hu, L. *et al.* Discovery of a small-molecule inhibitor and cellular probe of Keap1-Nrf2 protein-protein interaction. *Bioorg. Med. Chem. Lett.* **23,** 3039–3043 (2013).

212. Jnoff, E. *et al.* Binding Mode and Structure-Activity Relationships around Direct Inhibitors of the Nrf2-Keap1 Complex. *ChemMedChem* **9,** 1–7 (2014).

213. Dinkova-Kostova, a T., Massiah, M. a, Bozak, R. E., Hicks, R. J. & Talalay, P. Potency of Michael reaction acceptors as inducers of enzymes that protect against carcinogenesis depends on their reactivity with sulfhydryl groups. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3404–9 (2001).

214. Tayyem, R. F., Heath, D. D., Al-Delaimy, W. K. & Rock, C. L. Curcumin content of turmeric and curry powders. *Nutr. Cancer* **55,** 126–131 (2006).

215. Dinkova-Kostova, A. T. & Talalay, P. Relation of structure of curcumin analogs to their potencies as inducers of Phase 2 detoxification enzymes. *Carcinogenesis* **20,** 911–914 (1999).

216. Scapagnini, G. *et al.* Curcumin activates defensive genes and protects neurons against oxidative stress. *Antioxidants redox Signal.* **8,** 395–403 (2006).

217. Copple, I. M. *et al.* Chemical Tuning Enhances Both Potency Toward Nrf2 and In Vitro Therapeutic Index of Triterpenoids. *Toxicol. Sci.* **140,** 462–469 (2014).

218. Mead, R. J. *et al.* S[+] Apomorphine is a CNS penetrating activator of the Nrf2-ARE pathway with activity in mouse and patient fibroblast models of amyotrophic lateral sclerosis. *Free Radic. Biol. Med.* **61,** 438–452 (2013).

219. Calabrese, E. J. Neuroscience and Hormesis: Overview and General Findings. *Crit.*

Rev. Toxicol. **38,** 249–252 (2008).

220. Venci, J. V. & Gandhi, M. A. Dimethyl Fumarate (Tecfidera): A New Oral Agent for Multiple Sclerosis. *Ann. Pharmacother.* **47,** 1697–1702 (2013).

221. Yore, M. M., Kettenbach, A. N., Sporn, M. B., Gerber, S. a. & Liby, K. T. Proteomic analysis shows synthetic oleanane triterpenoid binds to mTOR. *PLoS One* **6,** 1–10 (2011).

222. Baell, J. B. & Holloway, G. a. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53,** 2719–2740 (2010).

223. Baell, J. & Walters, M. Chemical con artists foil drug discovery. *Nature* **513,** 8–10 (2014).

224. Kern, J., Hannink, M. & Hess, J. Disruption of the Keap1-containing ubiquitination complex as an antioxidant therapy. *Curr. Top. Med. Chem.* **7,** 972–978 (2007).

225. Go, P. & Bellanger, J. M. BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3. *Nature* **425,** 316–321 (2003).

226. Padmanabhan, B. *et al.* Structural Basis for Defects of Keap1 Activity Provoked by Its Point Mutations in Lung Cancer. *Mol. Cell* **21,** 689–700 (2006).

227. Beamer, L. J., Li, X., Bottoms, C. A. & Hannink, M. Conserved solvent and side-chain interactions in the 1.35 Angstrom structure of the Kelch domain of Keap1. *Acta Crystallogr. Sect. D* **61,** 1335–1342 (2005).

228. Li, X., Zhang, D., Hannink, M. & Beamer, L. J. Crystal Structure of the Kelch Domain of Human Keap1. *J. Biol. Chem.* **279,** 54750–54758 (2004).

229. Lo, S., Li, X., Henzl, M. T., Beamer, L. J. & Hannink, M. Structure of the Keap1 : Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J.* **25,** 3605–3617 (2006).

230. Cheung, Y. L., Lewis, D. F. V, Ridd, T. I., Gray, T. J. B. & Ioannides, C. Diaminonaphthalenes and related aminocompounds: Mutagenicity, CYP1A induction and interaction with the ah receptor. *Toxicology* **118,** 115–127 (1997).

231. Tarumoto, T. *et al.* Ascorbic acid restores sensitivity to imatinib via suppression of Nrf2-dependent gene expression in the imatinib-resistant cell line. *Exp. Hematol.* **32,** 375–381 (2004).

232. Wang, X. J., Hayes, J. D., Henderson, C. J. & Wolf, C. R. Identification of retinoic acid as an inhibitor of transcription factor Nrf2 through activation of retinoic acid receptor alpha. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 19589–19594 (2007).

233. Tang, X. *et al.* Free Radical Biology & Medicine Luteolin inhibits Nrf2 leading to negative regulation of the Nrf2 / ARE pathway and sensitization of human lung carcinoma A549 cells to therapeutic drugs. *Free Radic. Biol. Med.* **50,** 1599–1609 (2011).

234. Dinkova-Kostova, A. T., Holtzclaw, W. D. & Wakabayashi, N. Keap1, the sensor for electrophiles and oxidants that regulates the phase 2 response, is a zinc metalloprotein. *Biochemistry* **44,** 6889–6899 (2005).

235. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* (2012). doi:10.1021/ci3001277

236. Palm, K., Stenberg, P., Luthman, K. & Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **14,** 568–571 (1997).

237. Baurin, N. *et al.* 2D QSAR consensus prediction for high-throughput virtual screening. An application to COX-2 inhibition modeling and screening of the NCI database. *J. Chem. Inf. Comput. Sci.* **44,** 276–285 (2004).

238. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **44,** 235–249 (2001).

239. Walters, W. P. & Murcko, M. A. Prediction of 'drug-likeness'. *Adv. Drug Deliv. Rev.* **54,** 255–271 (2002).

240. Nendza, M. & Müller, M. Screening for low aquatic bioaccumulation (1): Lipinski's 'Rule of 5' and molecular size. *SAR QSAR Environ. Res.* **21,** 495–512 (2010).

241. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The Penultimate Rotamer Library. *Proteins Struct. Funct. Genet.* **40,** 389–408 (2000).

242. Pope, A. J., Haupts, U. M. & Moore, K. J. Homogeneous fluorescence readouts for miniaturized theory and practice. *Drug Discov. Today* **4,** 350–362 (1999).

243. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **437,** 55–75 (1948).

244. Schaap, M., Hancock, R., Wilderspin, A. & Wells, G. Development of a steady-state FRET-based assay to identify inhibitors of the Keap1-Nrf2 protein-protein interaction. *Protein Sci.* **22,** 1812–9 (2013).

245. Baird, L., Llères, D., Swift, S. & Dinkova-Kostova, A. T. Regulatory flexibility in the Nrf2-mediated stress response is conferred by conformational cycling of the Keap1-Nrf2 protein complex. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 15259–64 (2013).

246. Tong, K. I. *et al.* Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response. *Mol. Cell. Biol.* **27,** 7511–21 (2007).

247. Lea, W. A. & Simeonov, A. Fluorescence Polarization Assays in Small Molecule Screening. *Expert Opin. Drug Discov.* **6,** 17–32 (2012).

248. Inoyama, D. *et al.* Optimization of fluorescently labeled Nrf2 peptide probes and the development of a fluorescence polarization assay for the discovery of inhibitors of Keap1-Nrf2 interaction. *J. Biomol. Screen.* **17,** 435–47 (2012).

249. Hancock, R. *et al.* Peptide inhibitors of the Keap1-Nrf2 protein-protein interaction. *Free Radic. Biol. Med.* **52,** 444–51 (2012).

250. Wu, P., Brasseur, M. & Schindler, U. A high-throughput STAT binding assay using fluorescence polarization. *Anal. Biochem.* **249,** 29–36 (1997).

251. Nasir, M. S. & Jolley, M. E. Fluorescence polarization: an analytical tool for immunoassay and drug discovery. *Comb. Chem. High Throughput Screen.* **2,** 177–190 (1999).

252. Lynch, B. A., Loiacono, K. A., Tiong, C. L., Adams, S. E. & MacNeil, I. A. A fluorescence polarization based Src-SH2 binding assay. *Anal. Biochem.* **247,** 77–82 (1997).

253. Jadhav, A. *et al.* Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease. *J. Med. Chem.* **53,** 37–51 (2010).

254. Lokesh, G. L., Rachamallu, A., Kumar, G. D. K. & Natarajan, A. High-throughput fluorescence polarization assay to identify small molecule inhibitors of BRCT domains of breast cancer gene 1. *Anal. Biochem.* **352,** 135–141 (2006).

255. Perez-Torrado, R., Yamada, D. & Defossez, P. A. Born to bind: The BTB protein-protein interaction domain. *BioEssays* **28,** 1194–1202 (2006).

256. Cleasby, A. *et al.* Structure of the BTB domain of Keap1 and its interaction with the

triterpenoid antagonist CDDO. *PLoS One* **9,** e98896 (2014).

257. Lodish, H.; Berk, A.; Zipursky, S. L. et al. DNA Cloning with Plasmid Vectors. *Mol. Cell Biol.* **4,** (2000).

258. Yokoyama, S. Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7,** 39–43 (2003).

259. Lackner, A. *et al.* A bicistronic baculovirus vector for transient and stable protein expression in mammalian cells. *Anal. Biochem.* **380,** 146–148 (2008).

260. Rosser, M. P. *et al.* Transient transfection of CHO-K1-S using serum-free medium in suspension: A rapid mammalian protein expression system. *Protein Expr. Purif.* **40,** 237–243 (2005).

261. Kost, T. a, Condreay, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23,** 567–75 (2005).

262. Malys, N., Wishart, J. a., Oliver, S. G. & McCarthy, J. E. G. Protein production in Saccharomyces cerevisiae for systems biology studies. *Methods Enzymol.* **500,** 197–212 (2011).

263. Cregg, J. M., Cereghino, J. L., Shi, J. & Higgins, D. R. Recombinant Protein Expression in Pichia pastoris. *Mol. Biotechnol.* **16,** 23–52 (2000).

264. Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **10,** 411–421 (1999).

265. Healthcare, G. E. Recombinant Protein Purification Handbook. *Methods* **41,** 1–306 (2009).

266. Graham, F. L. & van der Eb, A. J. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52,** 456–467 (1973).

267. Promega. Protocols & Applications Guide - Transfection. (2015).

268. Cui, S. *et al.* The Mechanism of Lipofectamine 2000 Mediated Transmembrane Gene Delivery. *Engineering* **04,** 172–175 (2012).

269. Karra, D. & Dahm, R. Transfection techniques for neuronal cells. *J. Neurosci.* **30,** 6171–6177 (2010).

270. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6,** 365–387 (2011).

271. Sittampalam GS, Gal-Edd N, Arkin M, et al. *Assay Guidance Manual*. *Eli Lilly Co. Natl. Cent. Adv. Transl. Sci.* (2004).

272. Fukutomi, T., Takagi, K., Mizushima, T., Ohuchi, N. & Yamamoto, M. Kinetic, Thermodynamic and Structural Characterizations of Association between Nrf2-DLGex Degron and Keap1. *Mol. Cell. Biol.* (2013). doi:10.1128/MCB.01191-13

273. Jameson, D. M. & Ross, J. A. Fluorescence polarization/anisotropy in diagnostics and imaging. *Chem. Rev.* **110,** 2685–2708 (2010).

274. Dawes, T. D. *et al.* Compound Transfer by Acoustic Droplet Ejection Promotes Quality and Efficiency in Ultra-High-Throughput Screening Campaigns. *J. Lab. Autom.* (2015). doi:10.1177/2211068215590588

275. Naylor, J., Rossi, A. & Hornigold, D. C. Acoustic Dispensing Preserves the Potency of Therapeutic Peptides throughout the Entire Drug Discovery Workflow. *J. Lab. Autom.* (2015). doi:10.1177/2211068215587915

276. Bickerton, G. R., Paolini, G. V, Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **4,** 90–8 (2012).

277. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077,** 101–113 (2014).

278. Chen, Y., Inoyama, D., Kong, A. N., Beamer, L. J. & Hu, L. Kinetic analyses of Keap1-Nrf2 interaction and determination of the minimal Nrf2 peptide sequence required for Keap1 binding using surface plasmon resonance. *Chem Biol Drug Des* **78,** 1014–1021 (2011).

279. Núñez, S., Venhorst, J. & Kruse, C. G. Target-drug interactions: first principles and their application to drug discovery. *Drug Discov. Today* **17,** 10–22 (2012).

280. Silverman, R. B. & Holladay, M. W. *The Organic Chemistry of Drug Design and Drug Action*. (Elsevier, 2014). doi:10.1016/B978-0-12-382030-3.00002-7

281. Krogsgaard-Larsen, P., Stromgaad, K. & Madsen, U. *Textbook of Drug Design and Discovery*. (CRC Press, Taylor & Francis Group, 2010).

282. Hunter, C. A. & Sanders, J. K. M. The Nature of π-π Interactions. *J. Am. Chem. Soc.* **112,** 5525–5534 (1990).

283. Xu, J. & Hagler, A. Chemoinformatics and Drug Discovery. *Molecules* **7,** 566–600 (2002).

284. Anderson, A. The Process of Structure-Based Drug Design. *Chem. Biol.* **10,** 787–797 (2003).

285. Loew, G. H., Villar, H. O. & Alkorta, I. Strategies for indirect computer-aided drug design. *Pharm. Res.* **10,** 475–486 (1993).

286. Veselovsky, A. V & Ivanov, A. S. Strategy of computer-aided drug design. *Curr. Drug Targets. Infect. Disord.* **3,** 33–40 (2003).

287. Anderson, A. C. The Process of Structure-Based Drug Design. *Chem Biol* **10,** 787–797 (2003).

288. Wermuth, C., Ganellin, C., Lindberg, P. & Mitscher, L. Glossary of terms used in medicinal chemistry (IUPAC recommendations 1997). *Pure Appl. Chem.* **70,** 1129–1143 (1998).

289. Patrick, G. L. *An Introduction to Medicinal Chemistry*. (Oxford University Press, 2001).

290. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **32,** 335–373 (2003).

291. Cross, J. B. *et al.* Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **49,** 1455–74 (2009).

292. Warren, G. L. *et al.* A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **49,** 5912–31 (2006).

293. Marsden, P. M., Puvanendrampillai, D., Mitchell, J. B. O. & Glen, R. C. Predicting protein-ligand binding affinities: a low scoring game? *Org. Biomol. Chem.* **2,** 3267–73 (2004).

294. Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P. & DesJarlais, R. L. Docking: successes and challenges. *Curr. Pharm. Des.* **11,** 323–333 (2005).

295. Dickson, M. & Gagnon, J. P. The cost of new drug discovery and development. *Discov. Med.* **4,** 172–179 (2004).

296. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45,** 2615–2623 (2002).

297. van de Waterbeemd, H. & Gifford, E. ADMET in silico modelling: towards prediction paradise? *Nat. Rev. Drug Discov.* **2,** 192–204 (2003).

298. Egan, W. J., Merz, K. M. & Baldwin, J. J. Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* **43,** 3867–3877 (2000).

299. Egan, W. J. & Lauri, G. Prediction of intestinal permeability. *Adv. Drug Deliv. Rev.* **54,** 273–289 (2002).

300. Cheng, A. & Merz, K. M. Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships. *J. Med. Chem.* **46,** 3572–3580 (2003).

301. Susnow, R. G. & Dixon, S. L. Use of robust classification techniques for the prediction of human cytochrome P450 2D6 inhibition. *J. Chem. Inf. Comput. Sci.* **43,** 1308–1315 (2003).

302. Xia, X., Maliski, E. G., Gallant, P. & Rogers, D. Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem.* **47,** 4463–4470 (2004).

303. Dixon, S. L. & Merz, K. M. One-dimensional molecular representations and similarity calculations: methodology and validation. *J. Med. Chem.* **44,** 3795–809 (2001).

304. Votano, J. R. *et al.* QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation. *J. Med. Chem.* **49,** 7169–7181 (2006).

305. Congreve, M. & Carr, R. A 'rule of three'for fragment-based lead discovery? *Drug Discov. Today* **8,** 876–877 (2003).

306. Suzuki, T., Motohashi, H. & Yamamoto, M. Toward clinical application of the Keap1-Nrf2 pathway. *Trends Pharmacol. Sci.* **34,** 340–346 (2013).

307. Hur, W. & Gray, N. S. Small molecule modulators of antioxidant response pathway. *Curr. Opin. Chem. Biol.* **15,** 162–73 (2011).

308. Sekhar, K. R., Rachakonda, G. & Freeman, M. L. Cysteine-based regulation of the CUL3 adaptor protein Keap1. *Toxicol. Appl. Pharmacol.* **244,** 21–26 (2010).

309. Levonen, A.-L. *et al.* Cellular mechanisms of redox cell signalling: role of cysteine modification in controlling antioxidant defences in response to electrophilic lipid oxidation products. *Biochem. J.* **378,** 373–382 (2004).

310. Eggler, A. L., Liu, G., Pezzuto, J. M., van Breemen, R. B. & Mesecar, A. D. Modifying specific cysteines of the electrophile-sensing human Keap1 protein is insufficient to disrupt binding to the Nrf2 domain Neh2. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10070–5 (2005).

311. Stogios, P. J., Downs, G. S., Jauhal, J. J. S., Nandra, S. K. & Privé, G. G. Sequence and structural analysis of BTB domain proteins. *Genome Biol.* **6,** R82 (2005).

312. Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G. & Leprince, D. The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth Differ.* **6,** 1193–1198 (1995).

313. Bardwell, V. J. & Treisman, R. The POZ domain: A conserved protein-protein interaction motif. *Genes Dev.* **8,** 1664–1677 (1994).

314. Zollman, S., Godt, D., Privé, G. G., Couderc, J. L. & Laski, F. A. The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10717–10721 (1994).

315. Kreusch, A., Pfaffinger, P. J., Stevens, C. F. & Choe, S. Crystal structure of the tetramerization domain of the Shaker potassium channel. *Nature* **392,** 945–948 (1998).

316. Espinás, M. L. *et al.* The N-terminal POZ domain of GAGA mediates the formation of

oligomers that bind DNA with high affinity and specificity. *J. Biol. Chem.* **274,** 16461–16469 (1999).

317. Errington, W. J. *et al.* Adaptor Protein Self-Assembly Drives the Control of a Cullin-RING Ubiquitin Ligase. *Structure* **20,** 1141–1153 (2012).

318. Van Geersdaele, L. K. *et al.* Structural basis of high-order oligomerization of the cullin-3 adaptor SPOP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69,** 1677–1684 (2013).

319. Yachdav, G. *et al.* PredictProtein-an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* **42,** W337–43 (2014).

320. Hu, C., Eggler, A. L., Mesecar, A. D. & Van Breemen, R. B. Modification of Keap1 cysteine residues by sulforaphane. *Chem. Res. Toxicol.* **24,** 515–521 (2011).

321. Van Geersdaele, L. K. *et al.* Structural basis of high-order oligomerization of the cullin-3 adaptor SPOP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69,** 1677–1684 (2013).

322. Williamson, T. *et al.* Discovery of Potent, Novel Nrf2 Inducers Via Quantum Modeling, Virtual Screening, and In Vitro Experimental Validation. *Chem. Biol. Drug Des.* **80,** 810–820 (2012).