THE UNIVERSITY of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

Approaches to target WD40 proteins and synthesis and evaluation of chemical tools for on-bead screening

Thesis Submitted in Accordance with the Requirement of The University of Edinburgh for

the Degree of Master of Philosophy

By

Nicholas Fethers

School of Biological Sciences

2017

Declaration

This is to certify that the work contained within this thesis has been composed by me and is entirely my own work. No part of this thesis has been submitted for any other degree or professional qualification.

Signed

(Nicholas Fethers)

Date:

Acknowledgements

First and foremost I would like to thank my principle supervisor, Prof. Manfred Auer for his guidance and advice throughout my studies.

To all members of the Auer lab, past and present, I extend my most sincere gratitude. Much of the knowledge that I gathered over the course of my studies originates from advice offered from the friendly Auer lab members. Special mentions go to Peter Dodd who first took the time to introduce me to solid phase methods, Gemma Mudd who developed a novel synthetic route to rhodamine dyes, and Steve Shave who was and continues to be a key contributor to a great number of projects in the Auer lab.

I would also like to thank Bennets bar for supplying Auer lab members with a locale to attend in those rare moments of free time.

Finally I would like to thank my girlfriend Gwen. Who lead by example, showing great positivity in the face of her own adversity and supporting me throughout my own difficulties.

List of Abbreviations

2D-FIDA	2D-Fluorescence Intensity Distribution Analysis
Alloc	Allyloxycarbonyl
APAF1	Apopototic protease-activating factor
Ar	Aryl
Boc	Butoxycarbonyl
BREAD	Bead Ring Evaluation and Analysis of Data
BSA	Bovine Serum Albumin
CarbPhe	Carbamoylphenylalanine
CD	Circular Dichroism
CDC20	Cell Division Cycle protein 20
CNBr	Cyanogen Bromide
CONA	Confocal Nanoscanning
COPAS	Complex Object Parametric Analyzer and Sorter
CPC	Chromosomal Passenger Complex
Cy5	Cyanine5 dye
d	Doublet
Da	Daltons
DBCO	Dinbezylcyclooctyne
DIPEA	*N,N*-Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DOA	Fmoc-8-amino-3,6-dioxaoctanoic acid
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

ESI MS	Electron spray ionisation – mass spectrometry
EtOAC	Ethyl acetate
EtOH	Ethanol
FCS	Fluorescence Correlation Spectroscopy
FIDA	Fluorescence Intensity Distribution Analysis
Fmoc	9-Fluorenylmethyloxycarbonyl
FRED	Fast Rigid Exhaustive Docking
h	Hour(s)
HATU	(1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate)
HCl	hydrochloride
HDAC	Histone Deacetylase
HPLC	High performance liquid chromatography
K_D	Disociation constant
LC-MS	Coupled liquid chromatography - mass spectrometry
LRRK2	Leucine Rich Repeat Kinase 2
m	Multiplet
m/z	Mass to charge ratio
MALDI	Matrix-Assisted Laser Desorption/Ionization
MeCN	Acetonitrile
MeOH	Methanol
mHz	Megahertz
min	Minute(s)
MLL1	Mixed Lineage Leukaemia 1
MS	Mass spectrometry

NHS	N-Hydroxysuccinimide
Ni-NTA	Nickel-Nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
OBOC	One bead one compound
PCR	Polymerase chain reaction
PD	Parkinson's Disease
PEGA	Acrylamide-PEG co-polymer V
PPI	Protein-Protein Interaction
ppm	Parts per million
Pra	Propargylglycine
r.t.	Room temperature RAM Rink Amide
Rt	Retention time
s	Singlet
SCAL	Safety-Catch Acid-Labile
SDS	Sodium dodecyl sulphate Sodium dodecyl sulphate
SDS-PAGE	SDS-Polyacrylamide Gel electrophoresis
SEC	Size-Exclusion Chromatography
SOBOC	Scanning One-Bead One-Compound
t	Triplet
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TG	TentaGel
TIS	Triisopropylsilane

TMR	Tetramethylrhodamine
TMR-N$_3$	Tetramethylrhodamine-azide
TNBS	Trinitrobenzenesulfonic acid
TSA	Thermal Shift Analysis
UFSCAT	UltraFast Shape recognition with Credo Atom Types
UFSRAT	UltraFast Shape Recognition with Atom Types
USR	Ultrafast Shape Recognition
WDR5	WD40 Repeat containing protein 5

1. Abstract

A database consisting of information on human WD40 domains was compiled from literature sources. Data collected included information on function, structure, links to disease and information on molecules known to bind to WD40 domain containing proteins. Curation of the data collected suggested that 21% of WD40 domain containing proteins are linked to cancer, and that only 6% had known small molecule binders. From the database a shortlist of WD40 domain containing proteins that were considered of interest as research targets was produced. It was determined that WD40 Domain Containing Protein 5 (WDR5) was a potential cancer target open to several targeting methods.

WDR5 normally plays a structural roll in the formation of a complex containing WDR5, RbBP5, ASH2L, and DPY-30. This complex is required for methylation of H3K4, when MLL1 joins the complex it is able to methylate H3K4me2. It was also recently determined that WDR5 complexes with MYC, another protein with roles in transcriptional control. Both MLL1 and MYC are known to be prominent cancer targets.

His-tagged WDR5 was successfully expressed in BL21 (DE3) cell line and purified by a 2 step method. First the protein was purified via His tag - Ni-NTA agarose affinity chromatography, the eluted protein was then further purified via Size Exclusion Chromatography.

The first approach targeting WDR5 consisted of a combination of an *in-silico* approach and a small molecule binder screen. Two *in-silico* methods, Q-mol and USRCAT, were used to determine small molecules that would potentially bind to WDR5. From this suggested set, 81 compounds were screened against WDR5. Thermal denaturation fluorescence (TDF) was chosen as assay technique. A single compound increased the thermal stability of WDR5 in repeated experiments. This compound, NCI292249, was further characterised in microdialysis experiments where it was determined to have a low affinity of 564 µM to WDR5.

In a second approach, in order to target the MYC-WDR5 interaction, a series of truncated peptides derived from the WDR5 binding motif from MYC were produced testing a variation of the One-Bead One-Compound (OBOC) synthesis method derived in this work. These peptide fragments were synthesised using Fmoc solid phase peptide synthesis on TentaGel micobeads experimenting with the SOBOC technique (Scanning OBOC) adapted to produce all possible fragments of a peptide in parallel. The peptide fragments were fluorescently

labelled with tetramethylrhodamine and screened against 6XHis-WDR5 isolated on Ni-NTA functionalised agarose beads. The peptide fragments were ranked based on their affinity to the WDR5-coated micro-beads assayed by Confocal Scanning (CONA). The highest affinity peptide found in the CONA screen was further tested for WDR5 binding in solution by fluorescence anisotropy which resulted in an affinity of 96 µM to WDR5. This 7-mer truncate of the MYC peptide was used as input for an *in-silico* method of peptide optimisation named MorPH. MorPH is a technique developed in the Auer Lab in which amino acids in a peptide are systematically replaced by all commercially available non-natural amino acids in a sequential manner. Each of ~ 1000 modified peptidomimetics are docked *in-silico* to the target structure. The suggestions from MorPH for the MYC peptide truncate were analysed and the potential replacements discussed in order to plan a possible future synthesis.

The MorPH technique was tested experimentally in this thesis in a second example, targeting of Survivin. Survivin is followed as cancer target in the Auer lab and it is found in significantly high concentrations in cancer cell lines and in stem cells. Increased Survivin expression has also been linked to a poor prognosis and reduced patient survivability in the clinic. Several suggestions resulting from the MorPH *in-silico* screen were synthesised and screened against Survivin. The best-in-series peptide was shown to have a Kd of 2.5 µM, with significantly increased plasma stability.

Several chemical tools were developed and characterised for use with on-bead synthesis methods. A contribution was made to a novel synthetic method for isomerically pure rhodamine dyes and their functionalisation[1] (Tetramethylrhodamine was azide-functionalised for use in peptide labelling, based on the azide-alkyne Huisgen cycloaddition reaction). The Auer lab synthesises many of its compounds and libraries using solid phase synthesis techniques. Several compounds exist in literature for the linking of chemicals to a solid-support, all of which are stable to different chemistries and require different conditions to cleave the reaction product from the solid phase. The use of methionine as a linker is described in literature as being highly specific in its cleavage conditions. A series of literature cleavage conditions were tested, the method that offered the highest purity was selected to be improved through further testing. The improved cleavage method was then characterised by cleaving the 20 natural amino acids from the methionine linker. In this experiment it was

determined that all amino acids tolerated the new conditions with the exception of methionine, cysteine and tryptophan, which were expected to react poorly to the harsh conditions. This verified that methionine was a suitable alternative to current lab standards for bead linkage.

Contents

Chapter 9. Materials and Methods

2. Introduction

2.1. Protein-Protein Interactions

The cellular pathways of the human body are a complicated interconnected series of interactions, balancing one another in an effort to maintain function. Enzymes are proteins that bind and interact with ligands, processing them into a new product or products. However, many biological processes are carried out by more complicated structures, formed of several proteins. These large protein structures are known as protein complexes. These protein complexes are held together by protein-protein interactions (PPIs), their formation dependant on the affinity of the interactions and local cellular concentrations. As these molecular machines are made of several proteins, as opposed to enzymes that are often single unit, there is the possibility of inhibition of these complexes through the targeting of the interactions that allow for the formation of the complex. While this is a reasonable assumption to be made the practice has proven difficult. PPIs have been shown to be difficult to target, interaction interfaces tend to be large, flat surfaces with few charged residues or pockets that could hold a small molecule inhibitor.[1] This is in contrast to enzymes that have represented the collection of targets focussed upon by research for decades, which typically have a well-defined active site/cofactor site. In more recent years it has been determined that PPIs tend to be modulated by interaction hotspots.[2] These interaction hotspots are small patches of residues that contribute significantly more to the interaction than the other residues in the interaction interface.

The field of PPI targeting has seen significant growth in the last decade or so, arguably starting in 2004 with the discovery that the small molecule Nutlin and its derivatives modulate the activity of the P53 pathway via binding to MDM2, an inhibitor of P53.[3] To date around 50 PPIs having been successfully targeted, both for inhibition and for stabilisation.[4] Information on these successfully targeted interactions is maintained in online databases such as TIMBAL[4] and 2P2I.[5] According to TIMBAL 49 PPIs have been targeted with small molecules for modulation. 9 of these examples have been found to stabilise PPIs interactions. Lipinski famously outlined guidelines for the characteristics of a drug-like chemical structure.[6] However, the structures determined to bind to PPI interfaces tend to be larger than 400 Daltons, more hydrophobic and with a larger number of hydrogen bonds.[4, 5, 7] As classical

methods of targeting proteins are biased towards Lipinskis guidelines and the current state of PPI targeting small molecules break many of these guidelines it could be interpreted that one of the challenges of targeting PPIs is due to the requirement of shifting the rules of small molecule discovery.

Despite impressive steps in the improvement of targeting PPIs human PPI networks are vast, therefore there are many therapeutically interesting interactions and protein complexes that, as of yet, have not been successfully targeted. Many of these protein complexes are formed around PPI scaffold proteins. Scaffold proteins serve as a core around which large complexes can assemble, an example of a protein domain that behaves as a scaffold for complexation is the WD40 domain.

2.2.   WD40 Domains

WD40 domains are large, made up of several repeated units named WD40 repeat domains. The repeat domains are around 40 amino acid residues in length and often terminate in a tryptophan-aspartic acid sequence, hence the name WD40. A single WD40 domain will generally have 5-7 WD40 repeat domains.[8] Structurally, WD40 domains resemble a propeller, with the repeat domains forming the blades of the propeller. Within these propeller structures is a powerful hydrogen bonding network that holds the distinctive shape of the WD40 domain.[9]
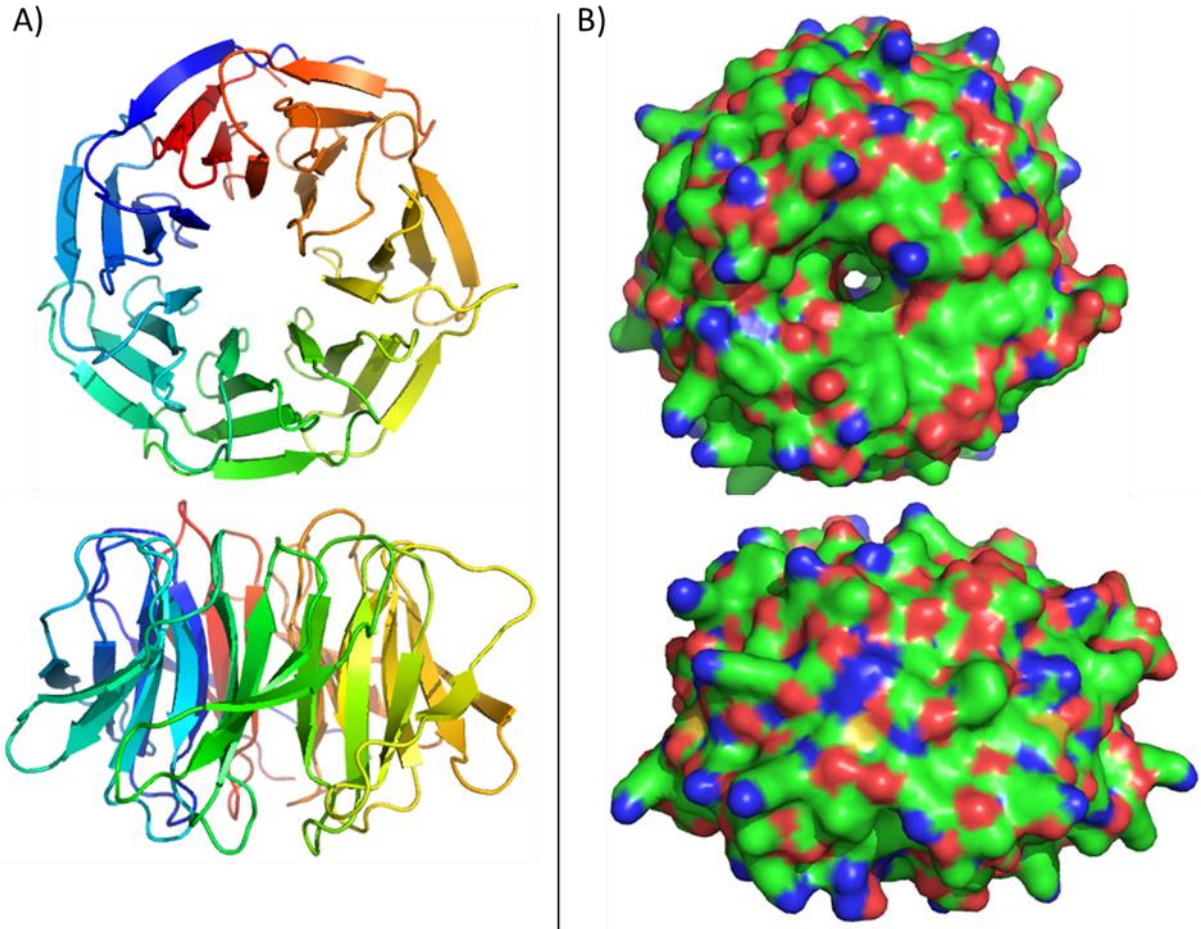
Figure 1 - Example WD40 domain structure. A) The protein chain is coloured from blue to red. The folding of the domain is highly compact with distinct "blades". B) The same structure with the surface shown. Oxygen is shown in red, nitrogen is shown in blue.

Despite having common structures WD40 domains are each highly specific towards their own series of binding partners. These interactions generally occur on the smaller top interface of the WD40 domain with the bottom interface having fewer associated interactions and the sides fewer still.[8] Despite the stereotypes of PPI interfaces previously described in this chapter (primarily that the interfaces are flat), the top face of WD40 domains tend to resemble a binding pocket. WD40 domains are remarkably prevalent within the human proteome, preliminary searches into WD40 domains and their interactions revealed that WD40 domains appear in around 1 % of all proteins in the human proteome.[8] Being as prominent as they are the domains appear in a broad variety of different proteins and pathways, covering an equally broad variety of functions. WD40 domain containing proteins have roles in cell signalling, cell cycle control and gene expression to give a few examples.

2.3. Databases

Databases have become a valuable source of information for researchers in a great many fields. In the field of biological sciences the range of databases covers information on structures, expression levels, and general summaries of protein function. These databases were utilised in a broad literature search at the beginning of this research project in an effort to better understand the nature of WD40 domains and their potential as research targets. In order to better investigate the interactions and cellular roles of hundreds of proteins the information found was initially catalogued as a spreadsheet in Microsoft Excel. Overtime this spreadsheet grew to incorporate information from other databases such as NCI databases[10] and ChEMBL databases[11] and itself was compiled into a database for ease of use. The information gathered was then used to determine human WD40 domain containing proteins of possible interest for further research. In order to better determine possible targets for further research a set of criteria were defined that would guide the decision making process. It was determined that the biology of the target proteins should be well understood, that inhibition of the interactions of the target would lead to a predictable effect. The target should have 3D structural data available for the application of *in-silico* methods, preferably data where a structure obtained through co-crystallisation with a known binding partner reveals the binding site. The structural information would allow for the application of structure based ligand screening methods. Finally for the application of ligand-based screening methods the structures of known binders to the target WD40 domain are required. These criteria were applied and a small number of potentially interesting WD40 domain containing targets were determined.

2.4. WD40 Repeat Domain Containing Protein 5

Of the potential targets identified WD40 repeat domain containing protein 5 (WDR5) met all criteria. WDR5 is known to form complexes with cancer related proteins, has 3D structural information available, and has been targeted for chemical intervention before, providing a selection of known binders. WDR5 also had the added advantage of being well documented as being expressible in bacterial cell lines, as opposed to other potential targets of interest that required more complex cell lines. WDR5 is made up of 334 amino acids, forming 7 WD40

repeat domains. WDR5 is a protein comprised of only a WD40 domain, its only function is to behave as a scaffold for complexes to form around. This makes WDR5 an excellent example of a prototypical PPI scaffold protein. WDR5 is found at the core of several different complexes, the most notable complexes are the methyltransferase complexes that are formed around this core subunit. The methyltransefase group of complexes all contain ASH2L, RBBP5, WDR5, DPY30 plus an additional specific methyltransferase subunit, for example for the methylation of H3K4. One of the binding partners to WDR5 in this complex, mixed lineage leukaemia 1 protein (MLL1), has been associated with a variety of leukaemias through its mutations. Another WDR5 interacting protein is MYC, a prominent cancer target, it was recently shown that this interaction is required to drive tumorigenesis.[12] Structural information available on WDR5 includes a variety of apo structures and structures with small molecules and peptides based on biological partners. These structures were used as input for structure-based *in-silico* docking experiments to determine possible novel binders to WDR5. Structures of WDR5 and its interaction partners showed the interaction between MLL1 and the top interface of WDR5, while this fulfilled an important criterion for future experiments over the course of this research project further structural information regarding the bottom interface of WDR5 and the WDR5-MYC interaction was published. This bottom interface interaction would go on to become the subject of a peptide fragment based interaction characterisation, as opposed to the better characterised top interface. The top interface of WDR5 has been a subject of interest for 2 research groups that began publishing their findings soon after the start of this research project. The two groups had each determined their own series of high affinity binders to the top interface of WDR5. These ligands would go on to be used for our own ligand-based *in-silico* screening in an effort to determine novel binders to WDR5.

2.5. In-silico Methods for the Determination of Novel Small Molecule Ligands to WDR5, Screened Via Label Free Affinity Methods

Having decided on WDR5 as a target, *in-silico* methods were applied towards the development of a novel chemical probe for the modulation of WDR5. Ligand-based methods were applied using the structures of known binders to WDR5 as input. Ligand-based methods are widely used for the discovery of novel ligands to protein targets, their advantage being that there is no requirement for a target structure for docking studies. The lack of structural docking studies also makes the methods less computationally intensive. Many ligand-based methods are available, however molecular similarity was selected as it has been reported to have a relatively high success rate compared to structure-based methods and due to the experience that the Auer lab has with molecular similarity methods, specifically USRCAT. The structural information gathered on WDR5 was used as input for structure-based *in-silico* screening, with collaboration partner Anton Cheltsov using the Q-mol software package of his own design.[13] The small molecule suggestions from both USRCAT and Q-mol were tested using label-free affinity detection methods. For the targeting of PPIs enzymatic assays are not an option, due to the lack of activity in the absence of a complete complex. Fluorescence techniques are popular for the detection and characterisation of affinity, such as fluorescence anisotropy, however for the screening of the suggestions label-free methods were better suited. Label-free methods allow researchers to move directly from suggestions to the screening phase without the need for costly or potentially time consuming labelling step, it also removes the chance of the label moiety interfering with an assay. As there are a variety of *in-silico* techniques for the identification of novel ligands for proteins, so too is there a variety of label-free techniques available for the screening of these potential binders. Methods such as isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) are powerful techniques for the characterisation of binding interactions. Thermal shift analysis is an exceptionally low cost screening method, compared to many other label-free screening methods, which utilises a quantitative real-time PCR machine and allows for screening in a 96 well format. The method determines the temperature at which a protein denatures, upon incubation with a ligand the stability of the ligand will increase, reflected as an increase in

thermal stability and an increase in the temperature of denaturation. Any compounds found to increase the thermal stability of the target protein require a second assay to verify that the potential binder is truly active. Microdialysis was selected as the second label-free method. Microdialysis has a distinct advantage as an assay technique, as the results can be used to determine if the test compound is precipitating under the assay conditions or binding to the assay vessel walls.

2.6.   The MYC-WDR5 Interaction

The second method of targeting WDR5 was to use structural information that illustrated the MYC-WDR5 interaction[12]. As briefly mentioned above, MYC binds to the bottom interface of WDR5. MYC proteins have been linked to a variety of different cancers, however the interaction between a MYC protein and WDR5 is required for MYC to bind to chromatin[12]. The bottom interface of WDR5 is also involved in binding interactions with KANSL2 and RBBP5, interactions required for the formation of methyltransferase protein complexes[14]. As these interactions occupy the same space on the bottom of WDR5 they cannot occur in parallel. This suggests that MYC does not bind with WDR5 in the context of a methyltransferase or NSL complex. MYC proteins are over expressed in a large number of different malignancies. MYC forms a complex with MAX to form a DNA binding domain that is able to regulate the transcription of target genes. It was recently determined that WDR5 is also required for the targeting of these genes. Wang *et al*[12] determined that MYC binds directly to WDR5 and that this complex binds to chromatin. 80 % of genomic sites occupied by MYC were also occupied by WDR5. Mutations in MYC that inhibited the ability of MYC to bind to WDR5 were found to have no effect on binding to DNA. However, binding to chromatin was negatively affected and its tumorigenic potential was attenuated in mice. There are a few different possible conclusions on the nature of the MYC-WDR5 interaction that could be drawn from the results of this investigation. Current research suggests that modulation of the behaviour of MYC could be possible through the targeting of the MYC-WDR5 interaction. Although MYC has existed as a cancer target for years[15] targeting the protein has proven to be difficult as both MYC and its known binding partner MAX are unstructured in solution.[16] The current state of the art in MYC inhibition are BET bromodomain inhibitors that instead target the expression of MYC, however the effectiveness is limited by MYC expression being driven by a number

of different proteins.[15] The binding of MYC to WDR5 does not appear to alter the structure of WDR5 on binding.[16] This could indicate that the targeting of the two interfaces of WDR5 is mutually exclusive, requiring that targeting of the two interfaces be carried out independently.

2.7. Targeting the MYC-WDR5 interaction

An alanine scan from literature suggested that the central 4 residues "IDVV" contributed significantly more to the interaction than other residues, making this area a hotspot.[12] To better evaluate the nature of the interaction it was decided that a series of truncates of this known binder should be synthesised and screened against WDR5. However rather than simply producing single peptides built around the central IDVV sequence a method was trialled whereby all possible truncates of a certain size range are synthesised in parallel. The basic method is inspired by one-bead one-compound library synthesis and deconvolution methods. One Bead, One Compound (OBOC) library synthesis produces its library products with only a single compound occurring an each bead. For the purposes of deconvolution throughout the synthesis of an OBOC library a sample of resin can be removed from each precursor sub-library after each coupling reaction. In the final library sub-libraries are separated based on the last coupling reaction carried out on the resin. These sub-libraries can then be screened in an on-bead format, using confocal nanoscanning (CONA) with fluorescently labelled protein. In peptide synthesis these precursor sub-libraries are N-terminal truncates, meaning that isolation of N-terminal truncates over the course of a peptide synthesis is a simple matter of removing resin after each coupling step. By synthesising a series of C-terminal truncates in parallel and removing N-terminal truncates as the synthesis progresses all truncates can be isolated easily. Having established the synthesis method of the peptide fragments, a screening method needed to be established. A two colour CONA screen with His-tagged protein isolated on Ni-NTA agarose resin was planned, whereby hit peptides could be ranked based on the relative fluorescence intensities of the labelled peptide and protein on bead. Further characterisation of peptides of interest was to be carried out using fluorescence anisotropy. To compliment this investigation *in-silico* studies were also carried out using structural based methods, Fast Rigid Exhaustive Docking (FRED, OpenEye Scientific Software) and hotpoint.[17] FRED is a part of the OEDocking suite and was used to dock fragments of the MYC peptide

to its interaction site on WDR5 to evaluate the contributions of individual residues to the MYC-WDR5 interaction. Hotpoint is an online server that analyses the change in potential and solvent accessibility of binding pocket to determine potential hotspots on either side of the PPI. The information provided from this research provides valuable guidance in the development of a novel binders based on the WDR5 interaction sequence of the MYC peptide.

The Auer lab has a unique method for the development of peptidomimetics, utilising in-house proprietary software, MorPH.

2.8. MorPH – An In-silico Tool for Optimising Known Peptide Binders and Protein-Protein Interactions

The MorPH software uses structural information of peptide-protein or protein-protein interactions as input, systematically replacing the amino acid residues of each position with a selection of building blocks that are suitable for solid phase synthesis methods. These replacements are incorporated into the structure of the peptide and the modified peptide docked to the target protein and scored. This produces a list of suggestions for replacements for each position, ranked by the docking scoring function. In order to test the method the protein Survivin was used as an example target. Survivin is not a WD40 domain containing protein, however it has been a target of interest for the Auer lab for several years. Survivin is a part of the chromosomal passenger complex (CPC), along with Aurora B kinase, Borealin and inner centromere protein. The CPC is responsible for the proper localisation of microtubules to kinetochores during mitosis. Survivin also inhibits apoptosis in healthy cells by binding to caspases 3 and 7 with nanomolar affinity.[18] The cellular concentration of Survivin is highly regulated. Usually Survivin is found to be in a very low abundance, except in the G_2M phase of the cell cycle there is a significant increase in cellular concentrations. Survivin is found to be in significantly higher concentrations in many cancer cell lines, particularly in lung and breast cancer cell lines.[19-21] Higher levels of Survivin expression are also linked to reduced patient survivability in the clinic.[19, 20] The compilation of these factors has made Survivin an important cancer target. Survivin and its interactions have been the subject of extensive investigation. These investigations have led to the development of binders to Survivin which were determined to increase cell death and delay the cell cycle in cellular

assays. Shugoshin 1 (hSgo1) is known to bind to histone H2A when the histone is phosphorylated by the kinase Bub1, in turn hSgo1 is known to bind to Borealin. Interestingly a fragment of hSgo1, a peptide with the sequence AKER, was found to also bind to Survivin in a co-crystallisation experiment.[22] This sequence, AKER, was used as a starting point for a MorPH investigation targeting Survivin. A series of recommendations for the targeting of Survivin had already been developed by the Auer lab by the time this thesis had begun. In this prior work the best-in-class peptides representing the best replacements for each position were determined. In the following work described in this thesis these best-in-class peptides were synthesised for further characterisation, including plasma stability analysis and fluorescence anisotropy. For the purpose of targeting WDR5 MorPH software was applied to the MYC-WDR5 structure. The output from this was reviewed position by position, creating a list of possible replacements for targeting the bottom interface of WDR5 that could be incorporated into a future synthesis to develop novel binders to targeting the MYC-WDR5 interaction.

2.9.   An Introduction to Solid Phase Peptide Synthesis

The Auer lab uses on-bead synthesis methods for many of its projects. In this thesis solid phase peptide synthesis was used to produce labelled peptides. Both using more classical methods for the synthesis of labelled Survivin MorPH peptides and via the scanning one-bead one-compound library production method, for the probing of the MYC-WDR5 interface. Solid phase peptide synthesis was first developed in 1963 by Robert Merrifield,[23] in the decades that followed the method evolved and improved into a technique that sees wide application in the synthesis of peptides and peptidomimetics. Solid phase synthesis uses polymer beads made of a semi-porous resin with many repeated functional groups throughout the polymer, to these functional groups a solid phase linker is attached. The linker acts as a chemical attachment point between solid phase and the reaction products. To this linker products are synthesised through the iterative process of coupling N-Fmoc protected amino acid building block, then deprotecting the newly attached building block ready for the attachment of the next. The coupling reaction can be broken down into two steps. The first step is the activation of the amino acid building block that is going to be coupled onto the peptide chain. 1-Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxid

hexafluorophosphate (HATU) is the most frequently used coupling reagent used in this thesis. The carboxylic acid is converted into an OAt-active ester through incubation with HATU. After the formation of this ester the activated amino acid is added to resin holding the N-deprotected peptide, where the free N-terminus attacks the carbonyl carbon in a nucleophilic substitution reaction Figure 2.
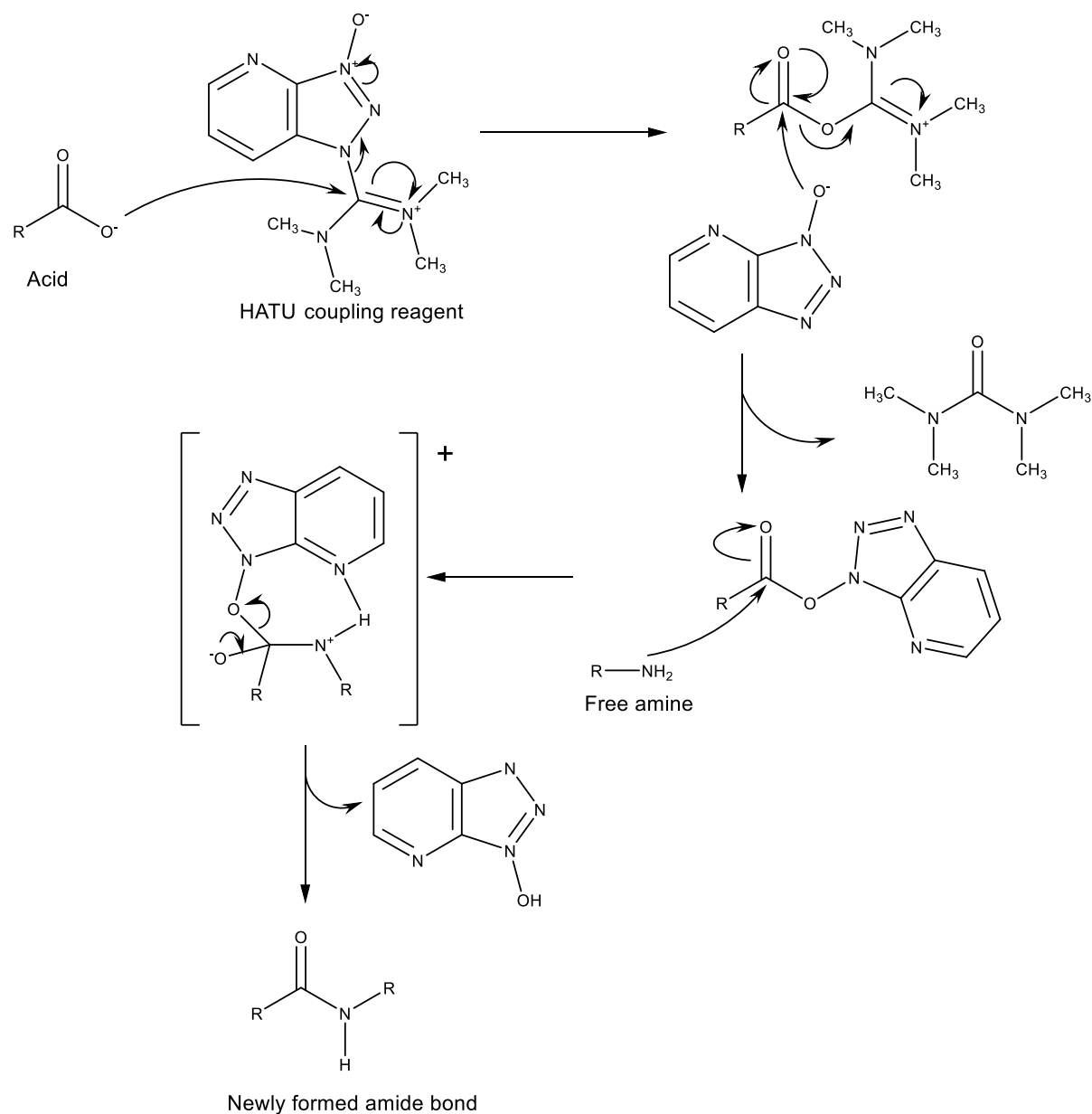

Figure 2 - The amide coupling mechanism using HATU as the coupling reagent.

Following the coupling of an amino acid building block the Fmoc protecting group is removed. The Fmoc protecting group has a short half-life in the presence of a secondary

amine, in the case of piperidine the halflife has been determined to be 6 seconds in a 20 %
solution.[24] In this work 20 % piperidine in dimethylformaldehyde (DMF) was used as the
Fmoc deprotection reagent of choice for peptide synthesis Figure 3.
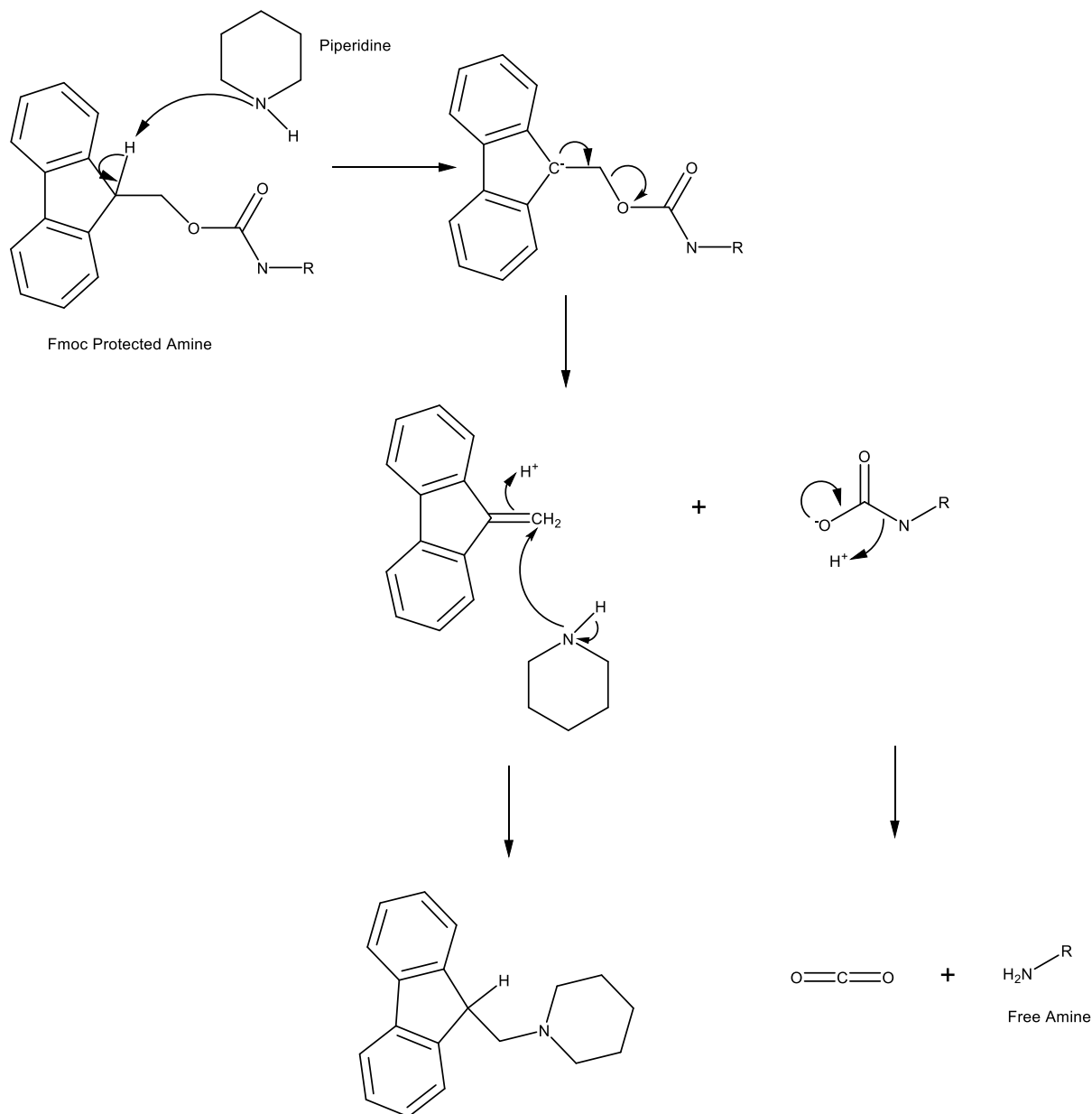
Figure 3 - Deprotection of an Fmoc protected amine using piperidine.

Both the coupling and deprotection reaction are carried out on the immobilised reagent or
product, leaving the unreacted reagents and side products in the solution. This is the great
advantage of solid phase synthesis; purification is a simple matter of thoroughly washing the
resin with a suitable solvent. These two reactions make for a versatile core from which a great
variety of chemicals can be produced.

2.10.   Tetramethylrhodamine-azide

Tetramethylrhodamine (TMR) is a commonly used fluorescent label by the Auer lab, often used in conjunction with solid phase techniques to produce TMR labelled peptides. Due to the frequency of its use it is necessary to synthesise the label in-house, often in the scale of grams. Previously published synthetic routes to TMR and other rhodamine cored labels would frequently produce mixed isomers that would prove difficult to separate.[25] Gemma Mudd et al. determined that an alternative synthetic route was possible that produced single isomer rhodamine cored labels, forgoing the difficult and often low-yielding purification process.[26] As a part of this work other researchers in the Auer lab with chemistry training verified that the method was reproducible on a variety of scales. One of the more frequent applications of TMR in the Auer lab is in the labelling of peptides via the Huigen azide-alkyne click reaction.[27] In this generic method an azide-functionalised fluorescent label is attached to a peptide immobilised on solid phase that has been functionalised with an alkyne group. In order to carry out this reaction the TMR synthesised using the novel synthetic route must be azide functionalised. Fortunately the 5 or 6 position carboxylic acid on the pendent ring can be coupled with no significant impact of the fluorescence of the dye. Coupling reagents HATU, DIC, PyBop, EDC and TSTU were trialled in a series of time course experiments in order to determine optimal conditions for the formation of an amide coupling reaction with 3-azidopropylamine. The best method would be employed as the standard synthesis method to produce all azide-functionalised TMR used in this thesis and in the experiments of other Auer lab members.

2.11.   Methionine Linker

For the development of chemical libraries for on-bead screening the linker used must be highly specific, it must be inert to all reactions required to produce and deprotect the library of compounds but must also be able to release the synthesis products from the solid phase under conditions that are inert to the products of the synthesis. In other words the method of

library synthesis and the method of cleavage must be sufficiently different from one another to ensure no premature cleavage or side reactions occur. A number of solid phase linkers that meet this criteria are known as safety catch linkers. These linkers equire a 2 step cleavage or the inclusion of a reagent will not be used under any normal synthesis conditions. At the time this research was carried out the lab standard safety catch linker required an 11 step synthesis and gives a final yield of 22 %. The time and reagent cost of such a synthesis makes the linker less appealing for frequent use. A search of the literature revealed that methionine had previously been used as a linker for the solid phase synthesis of libraries.[28] Cyanogen bromide (CNBr) is known to react with methionine residues in proteins, causing the cleavage of the peptide bond at the C-terminus of the methionine residue.[29] It is reasonable to expect that labs that frequently carry out peptide synthesis will often have a complete selection of all 20 natural amino acids as N-Fmoc protected building blocks. The application of the Fmoc-methionine building block as a linker is a significantly faster and cheaper option compared to the in house synthesis of a safety catch linker. In order to validate the use of methionine as a solid phase linker a series of experiments were carried out to verify the best cleavage methods for the linker. These methods would evaluate the reaction time, yield and purity. The purity of the reaction would be further characterised through the testing of all 20 natural amino acids under the finalised cleavage conditions for the methionine linker.
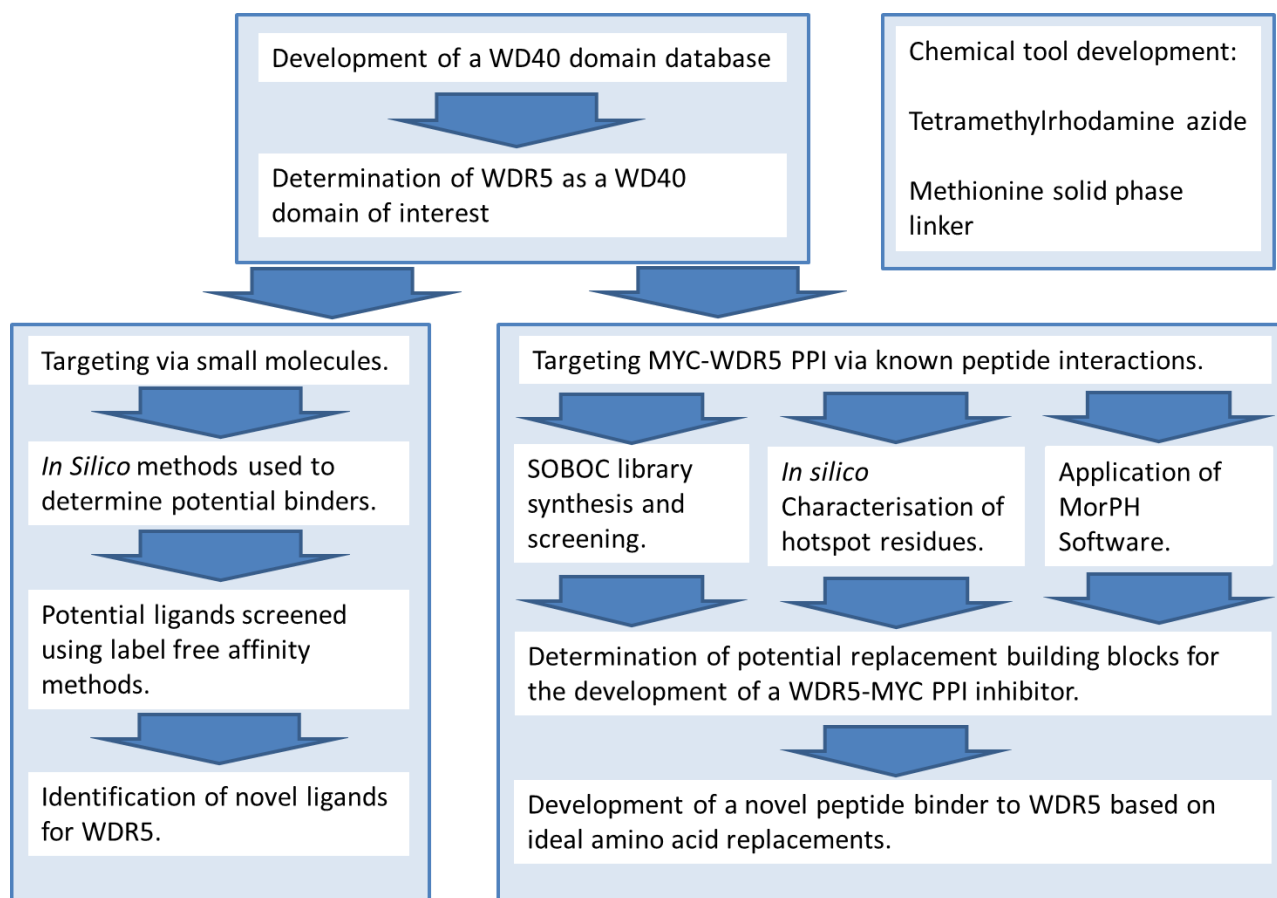
2.12. Summary of Thesis Work

Figure 4 - Chart showing illustrating the outline of work carried out during this thesis.

The work carried out in this thesis begins with the development of a WD40 domain containing protein database, from this database a list of potential protein targets was defined. Of these targets WDR5 was judged to be most auspicious and selected as the target protein to focus on in this work. In order to find small molecule inhibitors for WDR5 two routes were taken. The first used *in-silico* methods to identify small molecules with the potential to bind to WDR5. These small molecules were screened using the label-free affinity methods thermal shift analysis and microdialysis to identify WDR5 binders. The second route focused on the WDR5-MYC interaction interface, evaluating the contributions of residues to the interaction and creating a possible starting point for the development of a peptide inhibitor of the WDR5-MYC interaction. A library of peptide truncates was synthesised and a sample of these truncates were screened against WDR5. In parallel with this *in-silico* techniques were utilised to evaluate the contributions of individual residues to the PPI to determine possible hotspot residues. Finally MorPH software was utilised using the WDR5-MYC structural information

from literature as input. The suggestions from MorPH were evaluated and a short list of potential reasonable building block replacements determined.

3. The Gathering of Information on WD40 Domains and a Review of Literature

3.1.   Introduction

WD40 domains are amongst the most prominent within the human proteome, appearing in 1% of all human proteins[8] and are involved in a variety of pathways and cellular processes including apoptosis, the cell cycle and cell signalling. In turn this makes it possible to create a connection between a WD40 domain containing protein, sometimes several, and a variety of disease states. The current literature on WD40 domains reflects this hypothesis, indicating WD40 domain containing proteins linked to cancer developmental disorders, neurological disorders and others.[30] To better understand WD40 domains and their roles in both health and disease, a data gathering effort was undertaken and a database generated for the evaluation of WD40 domains as drug targets. As a part of this work a review of the current literature binders to WD40 domains and methods used to find these binders is described in this chapter.

3.2.   Aims

Aims of this section of work were as follows; i) To gather and curate data on WD40 domain containing proteins. ii) To assemble the data into a digestible format for future reading and ease of use. Data gathered should be as complete as possible, including information of protein structure, known binders and relevant biological information, referring to literature references where applicable. iii) From the gathered and curated information a list of WD40 domain containing proteins of interest was to be determined. These proteins would fulfil criteria such as having a perceived possible therapeutic value, having 3D structural data available, known protein-protein interaction partners and known binders.

3.3.   Results

The Universal Protein Resource (UniProt) database (uniprot.org)[10] was searched for all proteins tagged as WD40 domain containing. The UniProt Knowledgebase contains information on all known proteins from a variety of organisms. The database has unique identifiers for each protein and includes information such as alternative names, sequences and some biological information[10]. Records for these proteins were exported and used as a list of

WD40 domain containing proteins. Proteins from this list were then searched for in literature to determine their roles in pathways and disease states. Through searching however, it was determined that a number of proteins flagged as containing WD40 domains may have been falsely flagged. In total 311 proteins were listed as WD40 domain containing, 283 were confirmed to have WD40 domains as labelled domains within their sequences[10]. The list of confirmed WD40 domain containing proteins was used as a starting point for the search for further information. Searching other databases using the list of names and identifiers that was taken from the Uniprot knowledgebase[10] significantly increased the rate at which information could be gathered from other sources. Biogrid (Biogrid.org)[31] is a repository for biological interaction datasets. The database contains information on thousands of interactions between proteins in a variety of organisms. The information provided includes interaction partners and the method used to determine their interaction. The Biogrid database was used to determine interaction partners of all WD40 domains that were suggested by the UniProt database. The European Bioinformatics Institute (EBI) ChEMBL database (ebi.ac.uk/chembl/)[11] was searched for binders to the list of WD40 domain containing proteins. The ChEMBL database contains information on the results of the screening of compounds against proteins and protein complexes. The information includes structural information on the binding compound, the affinity and the method in which the interaction was measured. Searching this database provided a list 1747 binders for 17 WD40 domain containing proteins and protein complexes. These binders are further discussed in the database summary below. Reactome (reactome.org)[32], another EBI database, uses visual tools to illustrate the links between proteins within pathways and the links between pathways. Using the Uniprot identifiers of the confirmed WD40 domain containing proteins led to 138 WD40 domain containing proteins being described across 475 pathways. However, many pathways entered into the Reactome database overlap, possibly increasing the number significantly.

3.4.   Discussion

The original intention of the WD40 domain containing protein database was to catalogue all human WD40 domain containing proteins and all relevant information about them. This catalogue was then to be used to determine which WD40 domain proteins could be considered high priority for finding hits due to valuable information being available; for example, known

binders or structural data, and information on disease states that are linked to these proteins. It was determined during the course of information gathering that many WD40 domain containing proteins, such as CDC20, WDR5 and LRRK2 have been reported as valuable targets in literature with strong links to disease. As these proteins are more closely associated with disease, there was also more data available with more research carried out. It could be argued that proteins such as WDR5 and LRRK2 were highly cited and had more reported binders than others due to their links with disease. This investigation into the literature showed which WD40 domain containing proteins were of the centre of larger research efforts. Similarly it shows which WD40 domain containing proteins have been the subject of less research. Many WD40 domain proteins in the database are missing information due to a lack of published literature on their function.

3.4.1. Database Summary

The CTB-WD40 database contains entries for 283 proteins as listed in the UniProt database as being flagged as WD40 domain containing. Of these 278 are confirmed to contain a WD40 domain. Of these confirmed WD40 domain containing proteins, 59 (21%) are, as of writing, linked to cancer according to literature with the most common cancer type being colorectal cancer. 16 (6%) are linked to neurological disorders such as Parkinson's disease and Alzheimer's disease. There is a total number of 2664 unique binders with an affinity of < 100 µM for 17 entities as determined from the ChEMBL database and by literature searches. The total number of binders to WD40 domains is significantly inflated by 3 entries. Of these WD40 domains it was determined there was a subset that were of particular interest for studies. These WD40 domain containing proteins generally had significantly more information available. 3D structural information and affinity information on binders were key datapoints for WD40 target selection for this work.

3.4.2. WD40 domain containing pathways

Data from Reactome gathered by searching for the Uniprot identifiers of WD40 domains. The search returned 475 different named pathways containing varying numbers of WD40 domain containing proteins. Tables 1 and 2 describe in which pathways WD40 domain containing proteins are most highly concentrated and which pathways the most WD40 domains reside in.

Table 1 - Pathways containing WD40 domains sorted by ratio of WD40 domains in pathway, extracted from Reactome[32].

Pathway name	WD40 domain containing proteins	Total number of proteins	Ratio of WD40 domains to all proteins
Signal Transduction	38	2569	0.245
Immune System	33	2302	0.220
Metabolism	16	2113	0.202
Gene Expression	50	1790	0.171
Innate Immune System	9	1429	0.137
Metabolism of proteins	33	1399	0.134
Signaling by GPCR	6	1326	0.127
Disease	12	1110	0.106
Adaptive Immune System	21	1072	0.102
Developmental Biology	11	1061	0.101
GPCR downstream signaling	5	1003	0.096
Generic Transcription Pathway	10	894	0.085
Post-translational protein modification	17	866	0.083
Vesicle-mediated transport	14	795	0.076
Metabolism of lipids and lipoproteins	5	793	0.076
Cytokine Signaling in Immune system	9	740	0.071
Transmembrane transport of small molecules	10	680	0.065
Cell Cycle	21	603	0.058
Membrane Trafficking	14	602	0.058
Hemostasis	9	601	0.057

Table 2 below is also the results of a Reactome database search for WD40 domain containing proteins, but sorted by total number of WD40 domains rather than ratio of WD40 domain containing proteins to all proteins.

Table 2 - Pathways containing WD40 domains sorted by number of WD40 domains in pathway, extracted from Reactome database[32].

Pathway name	WD40 domain containing proteins	Total number of proteins	Ratio of WD40 domains to all proteins
Gene Expression	50	1790	0.171
Signal Transduction	38	2569	0.245
Metabolism of proteins	33	1399	0.134
Immune System	33	2302	0.220
Cell Cycle	21	603	0.058
Adaptive Immune System	21	1072	0.102
Cellular responses to stress	19	420	0.040
Class I MHC mediated antigen processing & presentation	18	437	0.042
Cell Cycle, Mitotic	18	498	0.048
Post-translational protein modification	17	866	0.083
Antigen processing: Ubiquitination & Proteasome degradation	16	306	0.029
Metabolism	16	2113	0.202
Major pathway of rRNA processing in the nucleolus and cytosol	15	172	0.016
rRNA processing in the nucleus and cytosol	15	183	0.017
rRNA processing	15	193	0.018
Processing of Capped Intron-Containing Pre-mRNA	14	236	0.023
Organelle biogenesis and maintenance	14	350	0.033
Membrane Trafficking	14	602	0.058
Vesicle-mediated transport	14	795	0.076
Chaperonin-mediated protein folding	13	95	0.009

The pathways listed must be interpreted with the caveat that many pathways overlap or even in some cases contain other pathways that may be listed. For example, the immune system appears four times in table 1 and twice in table 2. Searching within the immune system pathway and checking the roles of WD40 domain containing proteins appearing within shows that many are connected to cell cycle regulation and gene expression. While it is of course true that these proteins play an important role in the immune system, their role would normally be defined as what their specific pathway is, e.g. cell cycle regulation and gene expression. Table 1 and Table 2 show that WD40 domains appear in proteins in a wide variety of human

cell pathways. It is valuable to have this information available to better understand the role of a WD40 domain containing protein and to better assess its suitability as the subject of further investigation.

As this compiled data can be used to determine WD40 domain targets so too can it be used to determine which WD40 domain containing proteins should be avoided. Taking an example at random, the protein WDR83 acts as a scaffold for several different complexes, including complexes in the MAPK and ERK pathway,[33] a highly validated cancer related pathway[34]. However WDR83 is also known to be linked to insulin signalling.[32] This example demonstrates that the database allows researchers interested in a specific WD40 domain containing protein can check for effects outside of the desired pathway. As WD40 domains are often a scaffold to several complexes it could be thought that some WD40 domains could be found in many different biological pathways. Using the results from the Reactome database and checking the occurrences of the different WD40 domain containing proteins in each different pathway reveals that the proteins are almost exclusively involved in interactions in specific pathways. Several pathways related to cancer were investigated using information gathered from Reactome and literature to determine pathways of interest that contain WD40 domains and to assess the role of WD40 domain containing proteins within these pathways.

3.4.3. WD40 domains in cancer pathways

Several different pathways that are linked to cancer were investigated possible pathways that could be investigated to find WD40 domain containing proteins of interest. In this section several examples of cancer pathways that contain WD40 domain containing proteins are described. The NOTCH pathway is associated with numerous functions including maintaining stem cell populations, cell fate, promoting cell survival and inhibiting apoptosis. The NOTCH pathway has been targeted due to its connections to cancer, most famously by the targeting of gamma secretase as NOTCH activation depends largely on gamma secretase activity.[35] Several WD40 domain containing proteins appear within the NOTCH pathway[11]. Mutations in the WD40 domain containing protein FBXW7-binding pocket of Fbw27/Sel-10 E3 ubiquitin ligase can lead to Notch pathway activation. The effects of modulation of FBXW7 are likely to be complicated however, as FBXW7 is the substrate recognition subunit of an E3

ubiquitin ligase that interacts with a variety of different phosphorylated proteins[36] there is likely to be other pathways affected.

Another notable cancer related pathway is the PI3K/AKT1/mTOR pathway, specifically the constitutive signalling of AKT1[E17K] in cancer, featuring the WD40 domain protein LST8 (target of rapamycin complex subunit LST8). LST8 is a subunit of both MTORC1 and MTORC2. MTORC2 phosphorylates AKT1. AKT1 is a kinase and is activated in the event of this phosphorylation. In turn AKT1 phosphorylates several cellular proteins involved in cell cycle entry and cell survival. Several small molecules have been developed to inhibit the activities of different proteins in the pathway, however the inhibition of other proteins is seen to be less effective when the E17K mutation of AKT1 is present. While LST8 has value as a potential target for future research, other WD40 domain containing proteins were determined have more favourable qualities as a starting point as LST8 lacks known binders.

17 WD40 domain containing proteins appear in pathways linked to control of the cell cycle. Control of the mitotic cell cycle is a complicated, highly regulated series of interactions involving hundreds of proteins. Research into the roles of different proteins as a part of the cell cycle is at the heart of many cancer research efforts. The role of WD40 domains in cell cycle regulation is varied, with many different WD40 domain containing proteins appearing in complexes promoting or inhibiting the advance of the cell cycle. CDH1 and CDC20 act as substrate recognition subunits of the anaphase promoting complex (APC). The APC ubiquitinates key targets such as securing and the S and M cyclins to tag them for degradation. The degradation of these targets promotes the transition of the cell cycle from metaphase to anaphase. For the purposes of cell cycle regulation, the WD40 domain containing protein BUB3 exists as part of a complex with BUB1. The BUB1/BUB3 complex binds to and phosphorylates free CDC20, inhibiting its ability to form the APC.

3.4.4. Known Binders to WD40 domain containing proteins

The ChEMBL database[11] (version 19) was searched for WD40 domain containing proteins and protein complexes. The data set was supplemented with data found in the literature, returning a total of 23 entries with 4,335 total binders. A large portion of these reported binders however, have very low affinities. By applying a cut off of a minimum K_D of 100 µM the dataset is reduced to 2,664 binders for 17 WD40 domain containing targets (Table 3, following page).

The dataset is clearly skewed towards a few targets with a very high number of reported binders, Leucine rich repeat kinase (LRRK2) and homeodomain-interacting kinase 2 (HIPK2) making up half of all reported binders (45% of the unfiltered dataset, 53% with the 100 μM filter). With the 100 μM filter in place HIPK2 and LRRK2 have 828 and 596 unique compounds respectively. Both of these high-hit rate WD40 domain containing proteins are kinases. Kinases represent a prominent cancer target with 25 drugs in use against cancers and many more in different stages of clinical trials.[37] HIPK2 is a kinase that regulates the tumour suppressor p53 via phosphorylation of Ser46. The role that HIPK2 plays in cancer with regards to p53 has been reviewed extensively.[38] LRRK2 is also a kinase, however despite the link of kinases to cancers LRRK2 is widely known for its connections to Parkinsons disease. After the kinases the histone deacetylase (HDAC) family of proteins, HDAC1, HDAC3 etc, has the highest number of binders. The HDAC family deacetylate histones, mediating gene expression. The HDAC family as a whole has been targeted for intervention due to its role in cancer and psychological disorders such as depression. At the time of writing 5 HDAC inhibitors have been approved for use in the clinic (vorinostat, romidepsin, chidamide, panobinostat and belinostat) with several more in clinical trials. Apopototic protease-activating factor (APAF1) also holds a large number of known binders to its name with 640 unique entries in ChEMBL (198 with an affinity >100 μM). APAF1 Mediates apoptosis via the cytochrome c and ATP-dependant activation of caspase 3 via pro-caspase-9. APAF1 is associated with a variety of diseases including multiple myeloma,[39] kidney neoplasms,[40] cardiovascular diseases[41] and depressive disorder.[42] The number of binders for each protein can be interpreted as a reflection of how successful the research efforts to target the protein have been to date. A large number of binders can be interpreted as the target is of great interest and therefore has been targeting in many screens, while a small number of binders could reflect that the protein is of interest however there has been difficulty targeting the protein or its value as a target has only recently been established. This interpretation provides another criterion for target determination of a WD40 domain-containing protein that has a small number of known binders.

Table 3 - List of binders to WD40 domain containing targets extracted from ChEMBL database

| Chembl ID	Name	unique binders	binders < 100 µM
1829	Histone deacetylase 3	607	319
4284	Serine/threonine protein phosphatase 2A, 55kDa regulatory subunit B, alpha isoform	2	2
4576	Homeodomain-interacting protein kinase 2	987	828
6114	Phospholipase A2 activating protein	37	15
1075104	Leucine-rich repeat serine/threonine-protein kinase 2	983	596
1075317	WD repeat-containing protein 5	68	40
1163119	U4/U6 small nuclear ribonucleoprotein Prp4	29	7
1250378	Splicing factor 3B subunit 3	15	13
1795093	Apoptotic protease-activating factor 1	640	198
2093865	Histone deacetylase	534	356
2096976	Peroxisome proliferator-activated receptor gamma/Nuclear receptor corepressor 2	239	177
2111363	Histone deacetylase 3/Nuclear receptor corepressor 2	73	66
2157851	Ubiquitin carboxyl-terminal hydrolase 47	14	12
2176773	PH-interacting protein	37	0
2189144	Phosphoinositide 3-kinase regulatory subunit 4	1	1
3038483	Histone deacetylase 1/3/5/8	25	2
3038484	Histone deacetylase 3/NCoR1	38	31
3120040	Regulatory-associated protein of mTOR	1	0
3137261	PRMT5-MEP50	1	0
3137282	MLL1-ASH2L-RBBP5-WDR5 complex	1	0
3137286	EZH2-SUZ12-EED complex	1	0
3137287	EZH1-SUZ12-EED-AEBP2-RBBP4 complex	1	0
-	Fbox and WD40 domain containing protein 7	1	1

3.4.5. WD40 Domains of Interest

The following domains are notable WD40 domain containing proteins for several reasons. These WD40 domain containing proteins were short-listed for further investigation as they have more data available, including structural information, known binders and known PPI binding partners.

3.4.6. FBW7

The first known inhibitor of a WD40 domain with a co-crystal structure was discovered via the high-throughput screening of a 50,000-compound Maybridge Screening Collection library.[43] Initially the compound was discovered to bind to Cell division control protein 4 in yeast. However, the compound, known as SCF-I2 (Figure 5), also interacts with to the human

homologue, F-box/WD repeat-containing protein 7 (FBW7), at higher concentrations (an IC_{50} of 274 ± 3 μM to FBW7 compared to an IC_{50} of 1.9 ± 1 μM with Cdc4).[43]
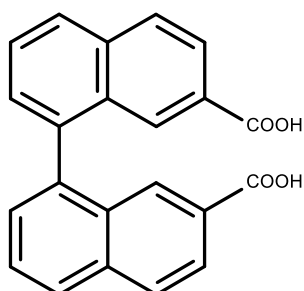
Figure 5 - SCF-I2, the first known WD40 domain inhibitor with a determined co-crystal structure

Despite being a weaker binder to the human homologue the discovery was still ground breaking, as it is the first binder to a WD40 domain. First FBXW7 is a target protein of interest in leukaemia research,[44] but also because this is the first instance of a small molecule binding to a human WD40 domain. After obtaining a crystal structure of the interaction **1** was shown to bind in an inter-blade space on Cdc4 (Figure 6).
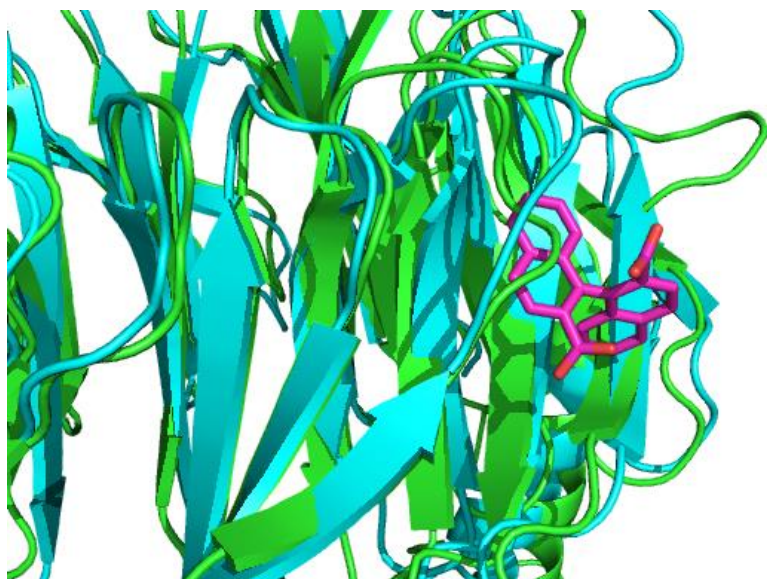
Figure 6 - Overlay of CDC4 (cyan) and FBW7 (green), showing the binding location of SCF-I2 (magenta) on the side interface of CDC4. This predicts the binding site on FBW7.

This is interesting as most commonly WD40 domains are known to be involved in PPIs via their top face (reviewed and researched in reference[45, 46]). This means the SCF-I2 is the first example of a small molecule binding to a WD40 domain and the first example of a crystal structure showing a ligand binding to the "side" interface of a WD40 domain. In some WD40

domains an inter-blade site is also the binding site of natural ligands, as was determined to be the case with CDH1[47] and CDC20.[48] However the structure of SCF-I2 is inherently unfavourable. The structure is largely hydrophobic, formed of two napthyl ring systems with only two carboxylic acids as functional groups. This provides a weak starting point for further development, it was therefore determined that this would not be used in this research project.

3.4.7. CDC20

Cell Division Cycle homologue 20 (CDC20) is the substrate recognition subunit of the E3 ubiquitin ligase named the Anaphase Promoting Complex (APC). The APC is responsible for driving the cell cycle from metaphase to anaphase. To achieve this the APC ubiquitinates key targets such as securing and the S and M cyclins to tag them for degradation.

CDC20 has strong links to a variety of cancers and is reviewed in depth by Kidokoro et al.[49] The review summarises research on 19 different cancers with regards to the level of overexpression of CDC20 recorded. Out of 445 tissue samples 194 (44%) were found to have greater than threefold baseline expression of CDC20. Bladder, choleangiocellular and pancreatic cancers were found to have overexpression in all samples tested, although n was small (20, 7 and 6 respectively). 59% of colon and rectum carcinoma (CRC) samples (16 of 27) were found to have elevated expression of CDC20. CDC20 was also studied in another statistical analysis that found that CRC patients with high levels of CDC20 show lower survival rates at all stages across 5 years .[50]

Table 4 - Survival rates of patients with CRC across 2, 3 and 5 years as a function of CDC20 expression levels

Years survival	2	3	5
High CDC20	61.4%	45.8%	27.2%
Low CDC20	81.5%	72.3%	53.8%

A prevailing theory is that overexpression of CDC20 drives the cell cycle forward more quickly. This leads to the improper splitting of sister chromatids. This improper splitting leads to aneuploidy. Drugs like taxanes and vinca alkaloids affect the cell cycle by interacting with microtubules. These drugs put cells into mitotic arrest which should lead to cell death. However, some cells survive, these cells undergo "mitotic slippage" and fall into interphase before continuing with their cell cycle. It is thought that the degradation of cyclin B could promote mitotic slippage. As cyclin B is a substrate of the APC it may be possible that

inhibition of the interactions of CDC20 as a part of the APC could lower the chances of mitotic slippage occurring.

In 2014 a small molecule was determined as an inhibitor of cyclin proteolysis in mitotic Xenopus egg extract. It was later determined by the King group that the molecule, named Apcin (APC INhibitor), was binding to the CDC20 subunit of the APC [51].
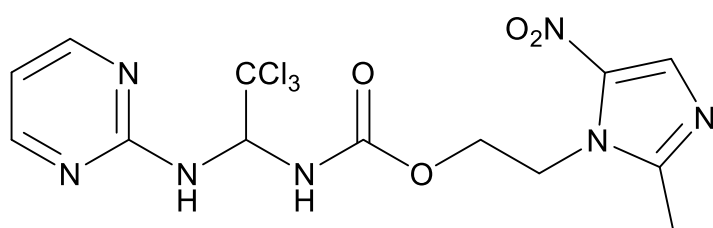
Figure 7 - Structure of Apcin

This discovery comes after the group determined that Tosyl-argenine methyl ester (TAME) was an inhibitor of the APC via displacement and blocking of CDC20 from the APC complex, also by the King group.[52]
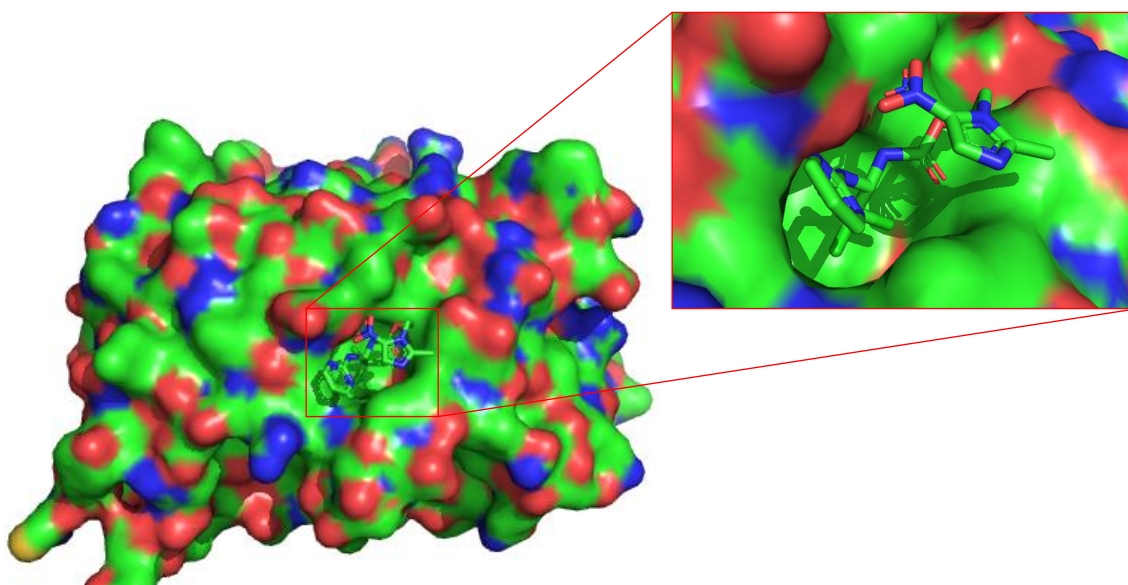
Figure 8 - Apcin (APC inhibitor) in D-box binding site of CDC20 (structure from PDB entry 4N14[10]

The binding site of apcin was determined via crystallography to be in the predicted Destruction box (D-box) binding site (Figure 8). The D-box is a motif that follows the amino acid sequence RXXLXXXXN where X is any amino acid. The motif binds with the D-box binding site on the side of CDH1 or CDC20. This binding interaction attaches the substrate protein to the APC complex for ubiquitination. The D-box binding site of CDC20 was

determined by overlaying the available crystal structures of the CDH1-ACM1 interaction and the CDC20-BUB1B structures. Apcin directly inhibits the ability of the APC to recognise substrates via the D-box motif, this is valuable as the King group also showed that Apcin works synergistically with TAME towards the inhibition of the APC [51]. The determination that Apcin binds to CDC20

3.4.8. LRRK2

It has been determined that the leucine-rich repeat serine/threonine-protein kinase 2 protein has 596 unique man-made binders of affinities lower than 100 µM affinity. LRRK2 is associated with Parkinson's disease, its mutants are thought to be responsible for the oxidation and phosphorylation of key anti-oxidant proteins. This leads to oxidative stress, one of the hallmarks of Parkinson's disease. As mentioned above kinases represent a prominent cancer target. As the protein is a kinase it has appeared as a protein in many large-scale kinase inhibitor screens likely screened along with many other kinases in parallel. These larger scale screens are where a large portion of the list of LRRK2 binders comes from. Despite this there is still a respectable list of compounds that have been found to bind to LRRK2 as a result of focused investigations.[53] The most potent binder to LRRK2 found by this research effort was determined by such an investigation. However no structural information for LRRK2 was available at the start of this project (information has since become available[54]). A focus on LRRK2 would not allow for the utilisation of structure-based in-silico methods used by the Auer lab, while ligand-based methods and library screening methods have already been used extensively to discover new ligands to LRRK2.

3.4.9. WDR5

WDR5 provides an excellent example of a WD40 domain containing protein and has also been successfully targeted by 2 approaches that using two different starting points. WDR5 has been targeted for investigation with the goal of developing a chemical probe due to the roll the protein plays in cancer, specifically leukaemia. WDR5 normally plays a structural roll in the formation of a complex containing WDR5, RbBP5, ASH2L, DPY-30[55] where WDR5 acts as the substrate recognition subunit. This complex is required for methylation of H3K4, when MLL1 joins the complex it is able to methylate H3K4me2. Chromosome translocations on the MLL1 gene lead to MLL1 fusion proteins that are linked to acute lymphoid leukaemia (ALL) and

acute myeloid leukaemia (AML).[56] MLL rearrangements produce around 80 different possible fusion pairings[57] and appear in >70% of infant leukaemia, in adults this reduces to approximately 10% of leukemias.[58] Topoisomerase inhibitors are often used in the treatment of acute leukaemia, however MLL rearrangements are commonly seen after therapy.[59] This is clearly counterproductive to the treatment of leukaemia patients and drives research for a suitable replacement. WDR5 also interacts with MYC, a biomarker for several cancers. This interaction has been shown to be required to drive tumorigenesis.[12] The interaction between MYC proteins and WDR5 occurs on the opposite face of WDR5 to the MLL1 interaction. At the time of writing the majority of research has been carried out on the MLL1 interaction interface, as summarized in the following sections.

3.4.9.1. Small molecule screening and optimisation

A high-throughput screen of a sub set of the 16000 diverse compounds from the Ontario Institute for Cancer Research general screening library led to the discovery of three compounds that bind to WDR5[60]. The best of these was WDR5-0101 (Figure 3), with a K_D of 5.5 ± 0.6 μM by ITC and a K_D of 23 ± 1 μM when displacing a fluorescein-labelled WIN peptide (sequence GSARAEVHLRKS), WDR5-0101 was the only compound that was discovered to displace this peptide. The group then carried out a commercial expansion to determine similar structures to WDR5-0101. This resulted in a screen of 119 compounds that were assayed. The most promising result of this screen was WDR5-0103, with a K_D of 0.45 ± 0.02 μM and a K_D of 3.0 ± 0.1 μM when displacing the WIN peptide.
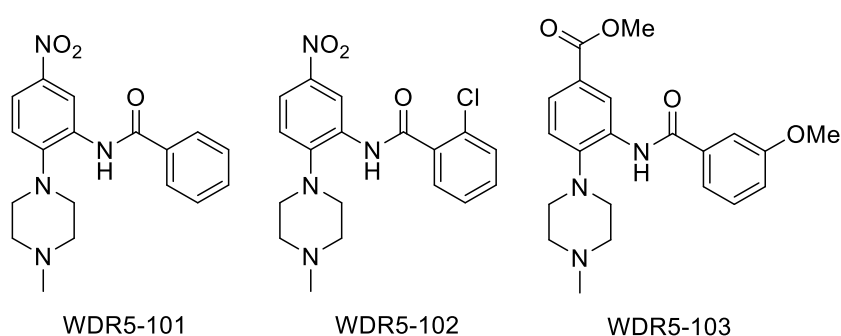
Figure 9. Small molecule binders to WDR5, 101 discovered by a high throughput screening effort. 102 and 103 from the screening of commercially available similars.

The group then worked to improve on the structure of WDR5-103, 69 compounds were produced to investigate the SAR at 3 positions[61].
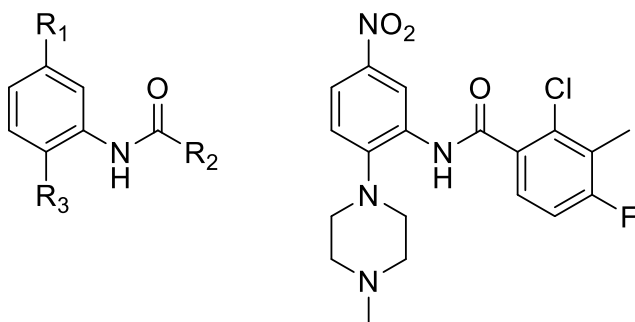
Core used for SAR   47

Figure 10. (Left) core of WDR5 binders used for SAR experiments. (Right) Structure of compound 47, optimised binder to WDR5

Compounds were assayed to determine K_D, compound 47 scored the best with a K_D of 0.27 ± 0.1 μM. Further optimization was carried out by the SGC-OICR collaboration to further develop a binder to WDR5. This improved binder, OICR-9429, has a K_D of 52 nM by ITC and an $IC_{50} < 1$ μM. Li *et al*. explores the SAR of compound 47 further[62].
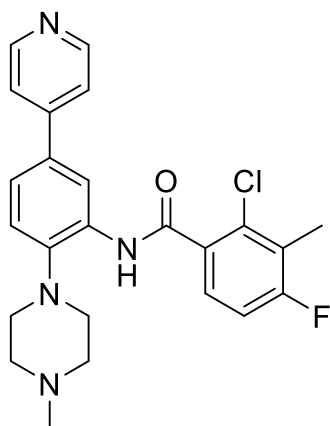

Figure 11. Compound 23 from Li *et al*

In the publication the authors specify that the affinity could likely be further improved by the introduction of a polar group at the 4-fluoro position to allow for a hydrogen bond interaction for ASP107. This theory was later explored by the SGC-OICR collaboration. In the final compound the 4-fluoro is substituted with a carbonyl (Figure 12).
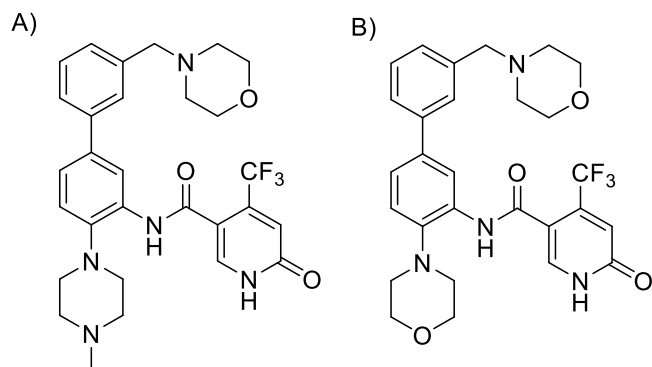
Figure 12. (A) WDR5 inhibitor OICR-9429, (B) negative control compound OICR-0547. Both compounds developed by the SGC and OICR collaboration.

OICR-9429 was determined to bind to WDR5 with high affinity (K_D = 93 ± 28 nM by fluorescence polarisation in a competition experiment against a labelled peptide).[63] OICR-9429 was also tested against primary human AML cells. At 5 μM the compound caused a decrease to 53% mean cell viability (n = 8). In AML patient cells without the C/EBPα mutation viability remained higher at 86% (n=8). Interestingly the almost identical control compound OICR-0547 shows no affinity to WDR5 and showed no activity on OICR-9429 sensitive cells at tested concentrations.[64] Figure 13 shows OICR-9429 in the top interface of WDR5, the methylpiperazine points down into the pocket. With OICR-0547 it is possible that the lack of affinity comes from the oxygen of the morpholine being too negatively charged, repelling against the negative charge of the pocket wall.
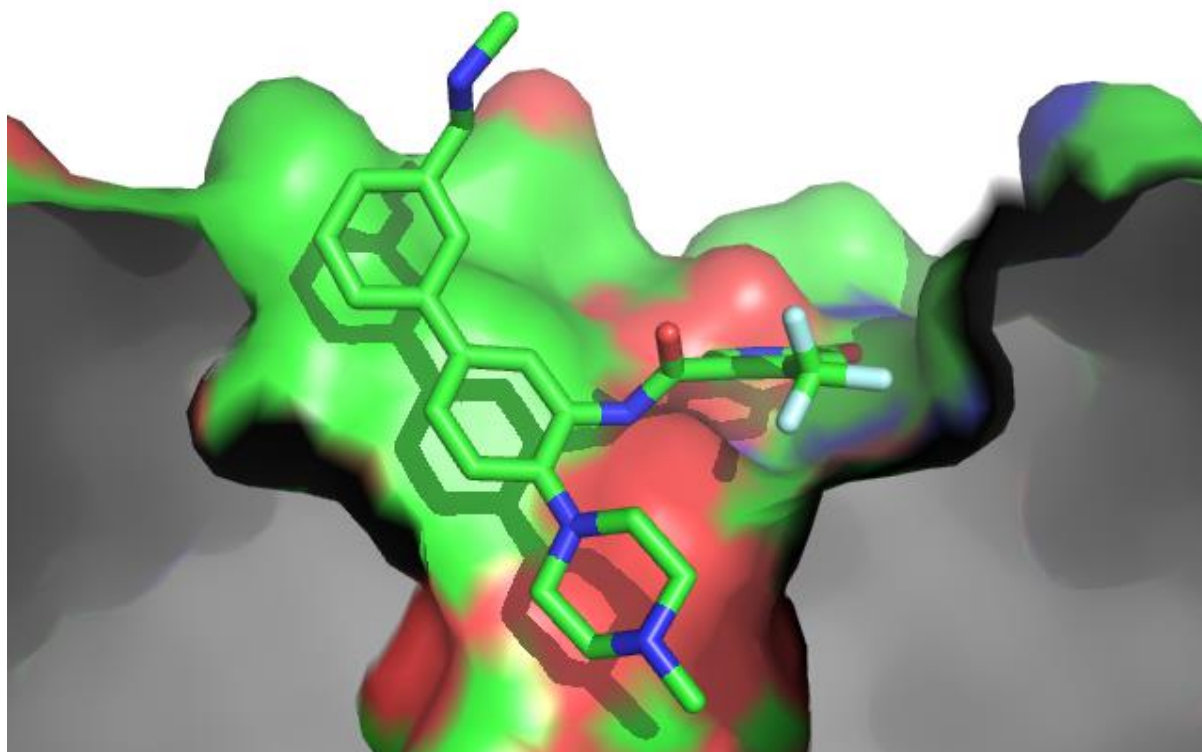
Figure 13 - Crystal structure of OICR-9429 in the top binding interface of WDR5

The development of OICR-9429 shows that a classical library screening approach followed by optimisation via medicinal chemistry is capable of producing a strong binding compound to a PPI target. In the following section a different method for the development of a binder is described, for comparison of these as methods to develop binders to WD40 domains.

3.4.9.2. Peptidomimetics Derived From Protein-Protein Interactions

The second series of binders to WDR5 discussed in this chapter comes from the Wang group of the University of Michigan. Initial research carried out by the group narrowed down the key residues of the WDR5 interfacing (WIN) motif (sequence: GSARAEVHLRKS) of MLL1 and the Histone 3 N-terminus (sequence: ARTKQTARKS) by synthesizing peptide fragments of sequence and screening via competition experiments against labelled full length peptides. It was determined that both of these motifs could be shortened down to the trimeric peptides ARA and ART respectively (Table 5 and Table 6)[65]. The research indicated that only a small loss in inhibition could be achieved, even when shortening to a trimeric peptide, as long as the key arginine residue and flanking residues were maintained. The WIN motif of MLL1 is needed for the attachment of MLL1 to a preexisting complex formed of WDR5, RbBP5,

ASH2L, DPY-30 (WRAD complex). This complex is capable of H3K4 methylation, with MLL1 it is capable of methylation of H3K4me2.

Table 5. Affinity of truncates of the WIN motif peptide

| Sequence	IC$_{50}$ ± SD (μM)	K$_I$ ± SD (μM)
Ac-GSARAEVHLRKS-NH2	0.75 ± 0.10	0.16 ± 0.02
Ac-SARAEVHLRKS-NH2	1.04 ± 0.14	0.20 ± 0.03
Ac-ARAEVHLRKS-NH2	0.02 ± 00004	0.003 ± 0.001
H2N-SARAEVHLRKS-NH2	0.08 ± 0.01	0.02 ± 0.002
Ac-RAEVHLRKS-NH2	29 ± 4	6.30 ± 0.80
Ac-ARAEVHL-NH2	0.16 ± 0.03	0.03 ± 0.01
Ac-ARAEVH-NH2	0.40 ± 0.10	0.09 ± 0.02
Ac-ARAEV-NH2	0.75 ± 0.10	0.16 ± 0.03
Ac-ARAE-NH2	0.40 ± 0.05	0.08 ± 0.01
Ac-ARA-NH2	0.54 ± 0.03	0.12 ± 0.01
Ac-AR-NH2	125 ± 6	27 ± 1.4

The N-terminus of H3 is also able to directly bind WDR5. This interaction stimulates methylation of H3K4. The H3 N-terminus was also explored via truncation around the key arginine residue (Table 6).

Table 6. Affinity of truncates of the Histone 3 peptide

Sequence	IC$_{50}$ ± SD (μM)	K$_I$ ± SD (μM)
H2N-ARTKQTARKS-NH2	70 ± 6	15.10 ± 1.30
Ac-ARTKQTARKS-NH2	0.006 ± 0.002	<0.001
H2N-ART-NH2	127 ± 12	27.30 ± 2.50
Ac-ART-NH2	0.08 ± 0.003	0.02 ± 0.001
H2N-ARTK(Me)QTARKS-NH2	69 ± 7	15.00 ± 1.60
H2N-ARTK(Me2)QTARKS-NH2	42 ± 6	9.00 ± 1.30
H2N-ARTK(Me3)QTARKS-NH2	73 ± 6	15.60 ± 1.20

The N-acetylated trimer peptides were used as a starting point for SAR-driven optimization. First by replacing amino acid residues around the key arginine residue, which determined that a peptide of Ac-XRV-NH2 (where x is an unknown residue) would be the best starting point for further SAR.[66] The optimization led to the development of compounds MM-101, MM-102 and MM-103.
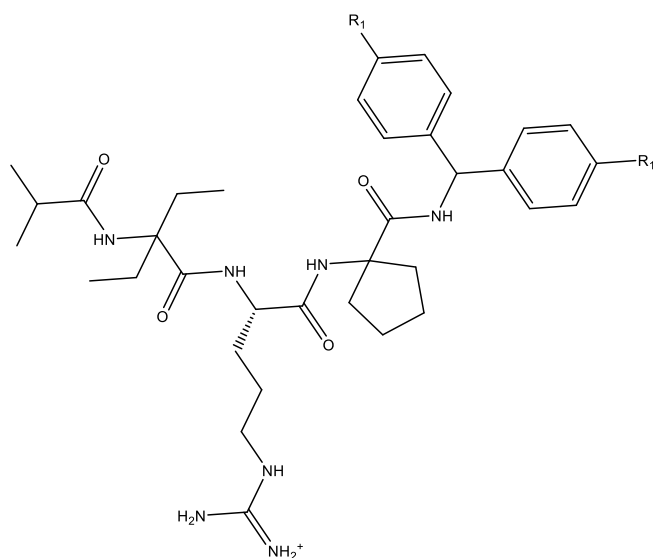
Compound	R1	IC$_{50}$ ± SD (nM)	K$_I$ (nM)
MM-101	-H	2.9 ± 1.4	<1
MM-102	-F	2.4 ± 1.7	<1
MM-103	-Cl	4.5 ± 0.6	<1

Figure 14. SAR of MM-10X Series

Compound MM-102 was found to improve the IC$_{50}$ 250 times to that of the Ac-ARA-NH2 peptide. MM-102 was also shown to inhibit cell growth in 2 leukaemia cell lines, MV4;11 and KOPN8. Further optimisation of MM-102 led to the development of MM-401, a cyclic analogue of the MM-10X series (Figure 15)[67].
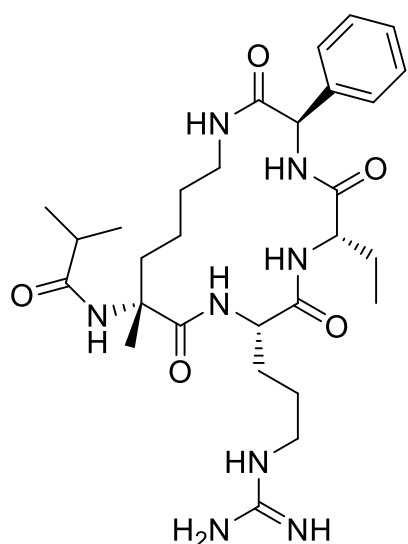

Figure 15. Structure of WDR5 inhibitor MM-401

MM-401 was determined to have an IC$_{50}$ of 0.9 ± 0.2 nM to WDR5, a significant increase in affinity compared to the MM-10X compounds and better than OICR-9429 (Figure 12).

Similarly to MM-102, MM-401 was screened against leukaemia cell lines. MM-401 was shown to inhibit the growth of MV4:11, MOLM13 and KOPN8 leukemia cells, cell lines that contain a single allele MLL1 translocation. This method, taking a natural interaction and improving upon it, was shown to produce a binding compound with a higher affinity than the small molecule described in the prior section. It is rare to be able to compare these two methods in the context of targeting WD40 domains. At the time of writing WDR5 is the only WD40 domain containing protein to have been targeted via two methods on the same interaction interface.

3.5.   Conclusions and Further work

Initially the database was started to determine WD40 domains of interest. During its creation proteins of interest were determined, more data was gathered. Based on the information collected on human WD40 domains it was determined that WDR5 would be an interesting target for further research. At the time at which this project was started WDR5 had no known small molecule binders, but the biological information available suggested that the protein could make for an attractive target for further investigation (discussed above). As mentioned previously the links between WDR5 and MLL1 and MYC strongly suggest that WDR5 and its interactions may play an important role in a number of cancers. During the course of the research project the publications discussed above were published. It was shown that MM-401 and OICR-9429 are active in cancer cell lines, inhibiting cell growth and reducing cell viability.[63] These results further verified the possibility that WDR5 could be an interesting target to investigate. The structural information that had been published encouraged further research. Crystallographic data of interactions on both the top and bottom interfaces would later prove to be invaluable in the development of a new binder to WDR5. This information would later prove invaluable when used as input for the MorPH process as described in chapter 6. MorPH is a computational tool which uses a peptide sequence for input and outputs a selection of amino acid replacements for each. This method will improve the stability of the peptide and through screening amino acid replacements improve the affinity between protein and peptide can be determined. Structural information on the interactions of WDR5 allows for the comparison of the output from MorPH with the results from the published optimisation efforts that led to the development of MM-401.[66]

The database was maintained throughout the research project. This makes the database useful for future work in several ways. The database contains alternate names and identifiers that can be used to search for the proteins in literature or other databases. Data on interaction partners, biology and disease allows researchers to easily evaluate a WD40 domain containing protein as a target. The database includes references to known binders to proteins that can be used to evaluate screening research that has been carried out on a protein. The information on binders can also be used to determine if a positive control is available for screening experiments against a target protein.

The compiled database largely consists of information from other databases, the value of this database is in the aggregation of information into one database allows users to quickly find all available information on a given WD40 domain containing protein or group of proteins.

Future work utilising the database requires that the database itself is maintained. Without updating the database still holds value, however by keeping relevant information such as disease associations updated the value of the collected and curated information will increase over time. Further research can also be carried out curating the data to search for patterns in WD40 domains and relevant diseases. For example if a project is focusing on the discovery of novel small molecules for the targeting of a specific pathway the database can be used to find WD40 domain containing proteins in the pathway and information on relevant biology.

An alternative could be to consider research into the WD40 domains with the lowest amount of information available in the database. A low level of detail in the database is a reflection of the lack of data available in literature. These overlooked WD40 domain containing proteins could be researched with the purpose of developing a molecular probe to investigate their biological functions.

4. Affinity Selection Testing of Small Molecules From Computational Techniques

4.1.   Introduction

The application of *in-silico* techniques can contribute greatly towards the discovery of novel binders. Many computational tools exist that can be accessed online, websites for the determination of hotspots in protein-protein interactions,[17] to aid in the filtering of pan assay interference compounds (PAINS) in compound datasets[68] and many tools for the 3D modelling of protein structures[11, 17, 68, 69]. The most challenging, and potentially rewarding, application of computational tools to the drug discovery process is the suggesting of potential binders and inhibitors. In this chapter two distinct approaches to *in-silico* drug discovery are discussed and applied in an effort to discover novel small molecule binder to WDR5. The first, method to be discussed does not involve any protein structures or docking. Instead USRCAT is a ligand-based method that suggests compounds based on their having similar structural descriptions to that of known binders. The second method is the structure-based docking method, the Q-mol platform. This is a structure-based method of discovering potential binding sites on a protein surface. A binding site is a specific region of a protein that is capable of entering into a stabilising reaction with another entity.[70] After potential binding sites are found small molecule libraries can be docked against the newly discovered binding sites and suggested for screening experiments.

4.1.1. Ligand-based techniques

The application of ligand based techniques requires small molecule structural information for known binders to the target. They do not require target information, instead, looking for common binding motifs, molecular shape and electrostatic similarities. The rationale behind this is the molecular similarity principle,[71] which simply states that similar molecules should make similar interactions. Substructure techniques involve a search of a compound database for 2D or 3D structures that contain a substructure of interest, while similarity techniques search for compounds that are 'similar' in their features.[72] Often substructure techniques will involve the aggregation of the structures of known ligands into a pharmacophore. A pharmacophore describes the ensemble of the spatial arrangement of steric and electronic

features in a binding interaction site.[70] Note that a pharmacophore is not itself a chemical entity. From this, a pharmacophore can be used as a template to find compounds with similar structures. Molecular similarity analysis allows for the comparison of chemical structures in terms of similar chemical moieties and their relative distance from one another. This description of the structure can be presented in 1, 2 or 3 dimensions, where 1 dimension is simply a vector of features, 2 dimensions includes graph connectivity and 3 dimensions include the distributions of charges.[73] The simplest description of a chemical structure can be presented as a bit string fingerprint, a series of 1s and 0s that simply describe whether certain moieties are present in a structure. There is an expectation that the 3-dimensional similarity methods would be more successful in their ability to determine novel binding partners than description based similarity methods, however the opposite has been proven in some studies against certain targets true.[72] There are a huge variety of methods available for the conversion of molecules into a set of descriptors.[72, 74] For example, one open source software package PaDel-descriptor is able to calculate 663 1D and 2D descriptors and 134 3D descriptors.[75]

For analysis a chemical structure can be converted into its descriptors, these descriptors are then be compared to one another through simple mathematics to determine how structurally similar they are. For example the most common approach for the comparison of molecular fingerprints is the Tanimoto index, more complex 3D comparisons and scoring methods are discussed later in this chapter.

Equation 1 - The Tanimoto index

$$Tanimoto = \frac{Fa}{Ft + Fq + Fa}$$

Where Fa refers to all of the features present in both the target and the query molecule, Ft is the number of features present in the target and Fq is the number of features present in the query molecule. The result of this equation will be a value between 0 and 1, where a score of 0 indicates that the two structures have nothing in common and a score of 1 suggests that the two are identical.

The popularity of the Tanimoto scoring method could be partly due to its adaptability to many different input criteria and gives an easy to compare score as an output. Due to its prominence the Tanimoto index is often compared to other methods of investigating similarity,[76] generally

it is accepted that it is the best method for group fusion.[77] Group fusion is a method combining data from several different sources, allowing for broader comparisons of structures.[78] The Tanimoto index however has also received some criticism, the most notable statement being that the method has a bias to score small molecules more favourably.[79, 80] Molecular similarity depends on a simple principle, that similar compounds with similar properties should have similar activities. This is known as the similarity property principle.[81] According to the similar property principle suitably similar compounds should show similar affinity. However the similar property principle does not perfectly describe the relationship between similar structures and the same target. Consider investigations into structure activity relationships (SAR). Classical SAR approaches make small, progressive changes to a larger structure and monitor the effect that the change has on the affinity. Over the course of some of these investigations however it can be found that very small changes in structure can lead to very large changes in affinity, for example in chapter 2 OICR-9429 shows an affinity of 93 nM to WDR5, however the remarkably similar structure OICR-0547 shows no affinity and is used as a negative control in later literature experiments. Despite this caveat molecular similarity can allow for the change from one core structure (sometimes referred to as a scaffold) to another.[82] This would be an ideal result from a molecular similarity project, as it would allow for SAR exploration and optimisation around a different structural core.

4.1.2. Molecular similarity analysis using URSCAT (carried out by Steven Shave)

In this study Ultrafast Shape Recognition with CREDO Atom Types (USRCAT)[83] was used to determine potential similar structures from the Vichem nested chemical library (NCL, https://vichemchemie.com/nested-chemical-library-ncl/). Vichem stipulated that, as a part of a collaboration with the company, their compound libraries will be shared with the Auer lab in an encoded format. USRCAT is a moment-based virtual screening method that uses the relative position of bonded atoms to describe molecular shape.[83] Moment-based methods use approximations to convert shape information into an invariant numerical form, allowing alignment free comparison of small molecule structures. These methods provide data in a simple numerical form it allows for the storage of information of many compounds in a database. URSCAT is the most recent iteration of a series of moment-based methods. Ultrafast Shape Recognition (USR) is the original method, devised to use 4 reference points for each

structure from which a set of descriptors can be calculated.[84] For each of these reference positions a description formed of the mean, variance and skew of the molecular distance between the reference point and all atoms in the structure. The 4 reference points are at the centre of the molecular structure, the atom closest to the centre point, the atom farthest from the centre point and the atom farthest from the point farthest from the centre point. For a structure this will output 12 values; 4 points with 3 values for each (mean, variance and skew). Comparison of the 2 structures via their descriptions is calculated by the sum of least absolute differences of their respective moments. The USR method was validated and improved upon. UltraFast Shape Recognition with Atom Types (UFSRAT) also calculates descriptor sets for structures, however it also takes into account atom types. The method considers all atoms, hydrogen bond acceptors, hydrogen bond donors and hydrophobic atoms. While USR uses 12 descriptions for each molecule UFSRAT uses 48, giving a more descriptive definition of the structure.[85]
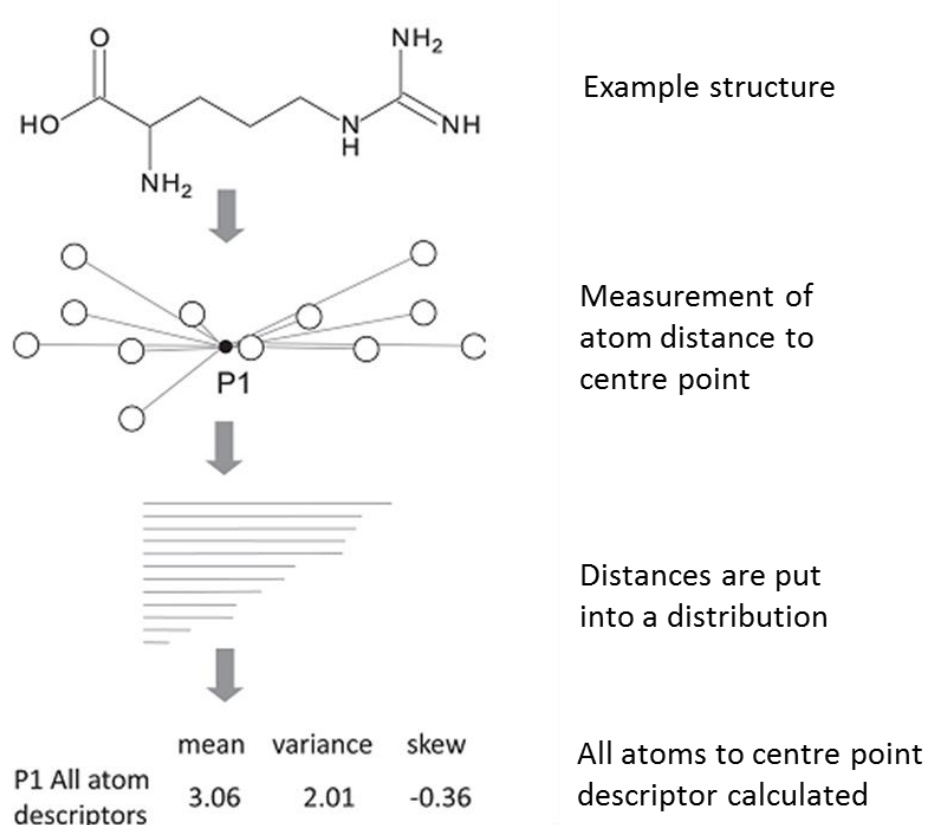

Figure 16 - Example of how the description of all atoms is determined in UFSRAT and USRCAT. Figure adapted from reference.[85]

USRCAT also accounts for atom types in calculations of descriptor sets, the method uses the same 4 atom criteria as UFSRAT but also aromatic atoms and defines point distributions in a different way to UFSRAT, defining P1 to P4 points and keeping them static between distributions. In summary USRCAT describes a structure as 60 numerical values. The calculation of similarity between two structures is performed through the sum of the squares of the differences between each descriptor. An important factor for consideration regarding some ligand based methods, such as molecular similarity, is that the work can be carried out with no knowledge of the structures involved, only the numerical descriptions. In many cases it is not possible to work backwards and discern the structure of a molecule from the description. While this means an additional step is required to ensure a structure suggested is suitably similar to the reference structure, it also allows for compound suppliers to provide virtual libraries of compounds to collaboration partners be used in virtual ligand screening without disclosing the structures included in their library. As is the case with the encoded Vichem library that was used to determine possible novel binders to WDR5 in this chapter.

4.1.3. Structure based techniques

In contrast to ligand-based methods structure based techniques use known structural information of a target protein (usually an X-ray structure) and positions, or "poses" potential ligands to defined binding sites on the protein surface. The clear advantage of structure-based virtual screening is that there is no requirement for there to be known binders to a target. However structural data is not always available for the target of a project, structure-based methods are also significantly more computationally intensive compared to ligand-based methods.

Structural based screening has several prerequisites for an experiment. First a database of compounds must be available. Often, the database must have the compounds available as a 3D structure or the ability to convert the data into a 3D structure. Next the target must be prepared using available 3D structural information. At the heart of this stage is definition of the binding site, narrowing down the possible binding location and reducing computational search space and avoiding exhaustive docking of the compound library to the surface of the entire protein. Methods to carry out structure based docking without data on the target structure have also been developed. Protein homology methods use known structures of

similar proteins to predict the structure of target binding sites for example the software FINDSITE[comb] [86]. Despite discrepancies between a homology model and a protein target, homology modelling methods have been found to be reasonably competent at producing accurate representations of target binding sites[87].

Following the preparation of the target structure and compound structure dataset docking is carried out. There are a great many docking methods that can be grouped in several different ways. In this section methods will be discussed based on their ability to account for the flexibility of the ligand or the protein. Predicting the flexibility of the ligand can be carried out several ways. For example FlexX (BioSolvIT) is a method that breaks a structure down into fragments and systematically builds the structure in the binding pocket based on a database of known possible bond angles.[88] Glide also accounts for ligand flexibility but instead exhaustively trials rotamer states for each bond in the ligand, removing unsuitable conformations as it progresses[89]. Trying to account for flexibility can lead to a great increase in computational intensity, however rigid ligand structures cannot describe the nature of a binding interaction. A compromise that some methods use is to take several conformations of each ligands structure which are docked rigidly to the protein separately, this is known as conformation ensemble docking and is used by techniques such as DOCK, FRED and SLIDE[90, 91]. A similar compromise can be made with regards to the protein flexibility by using molecular dynamics simulations, several example structures from a single molecular dynamics experiment can be used as a rigid structure for the docking of potential ligands[92]. Although the sampling rate of the molecular dynamics study will have a large effect on the results of such work. Alternatively more intense methods attempt to more accurately represent the binding interface, SLIDE for example simulates the relaxation of residues in the binding pocket upon docking of a structure[91].

When docking a potential ligand to a protein there are many possible metrics that can be used to score the predicted interaction. These scoring methods can be grouped as knowledge-based, physical-based or empirical. Knowledge-based methods use statistical-mechanics to calculate an approximation of the binding free energy of the complex, based on observed binding interactions from databases. Physical-based scoring functions are sometimes referred to as force-field scoring functions. These functions use a force-field function for the interaction

partners, this gives an energy value that is used as the score for the ligand. DOCK is an example method that uses physical scoring methods[93]. Empirical-based functions determine the binding free energy as a function of the contributions of different interactions detected, such as hydrogen-bonding or hydrophobic interactions, and scores them based on known-experimental binding energies for these interactions. Examples of methods that use empirical-based scoring functions include Flexx,[88] Chemscore[94] and SCORE[95].

4.1.4. Q-mol (carried out by Anton Cheltsov of Q-mol LLC)

Q-mol is an *in-silico* method that uses two steps to determine novel possible ligands to protein targets. The first step determines possible binding interfaces on the surfaces of proteins.

This method moves amino acids across the surface of the target protein and predicts sites where there is an increase in affinity. In the output possible binding sites are visualised as these areas of the protein surface will be surrounded with spheres with each sphere representing an amino acid binding site and the size and colour being proportional to the affinity of an amino acid to that region. Q-mol uses all-atom optimized potential for liquid simulations (OPLS), a combination of molecular dynamics and Monte Carlo approximations, to simulate the environment of a protein in solution[96]. Q-mol then simulates a binding interaction and possible conformational changes in the protein that occur on the binding of a small molecule[13, 97, 98]. As a whole the Q-mol package is able to define new possible binding sites to proteins, then use a docking simulation to determine possible ligands to these sites.

The Q-mol virtual ligand screening method has been used to determine a possible binding site on importin alpha 5 with the goal of evaluating the proteins role in influenza. After this possible binding site was determined the National Cancer Institute compound library was used for a preliminary virtual ligand based screening effort. This virtual screen produced many suggestions including several known anti-viral agents[98]. The Q-mol approach was also used to target protease region of the multifunctional NS3 protein. The NS3 protease domain (NS3pro) is responsible for viral polyprotein production. Virtual ligand screening determined 76 compounds that included ligands with a previously identified phenyl(sulfonyl-pyrazol scaffold. The ligands were fully characterised and their mechanism of action investigated.[13]

Given the relative success of the Q-mol software package it was determined that WDR5 should be targeted in a similar manner to the targets described above.

4.1.5. Label Free Affinity Methods

The Auer lab has developed a fully integrated platform of label free affinity selection methods for the screening of small molecules against protein targets in-vitro. There is no requirement to express the target as a fluorescent protein conjugate or carry out chemical labelling of the protein or ligand, these methods also avoid the requirement of assay development. Both of these barriers can impose a cost or time demand. In this section several methods that were considered for label free screening of compounds against WDR5 are discussed.

4.1.5.1. Surface Plasmon Resonance

Surface plasmon resonance (SPR) uses an experimental set up is known as the Kretschmann configuration[99]. A prism is attached to one side of a metal plate, usually gold. On the other side of the metal plate is an attachment point for protein, usually antibodies for the protein. A polarised light source fires a beam through the prism, the beam is reflected off the metal plate at a specular angle and into a detector. Given the experimental set up there will be an angle at which a significant decrease in reflected light, this is known as the resonance angle. At this angle the light resonantly couples with the electrons on and near the surface of the metal. This creates oscillating electrons near the surface of the metal called plasmons.
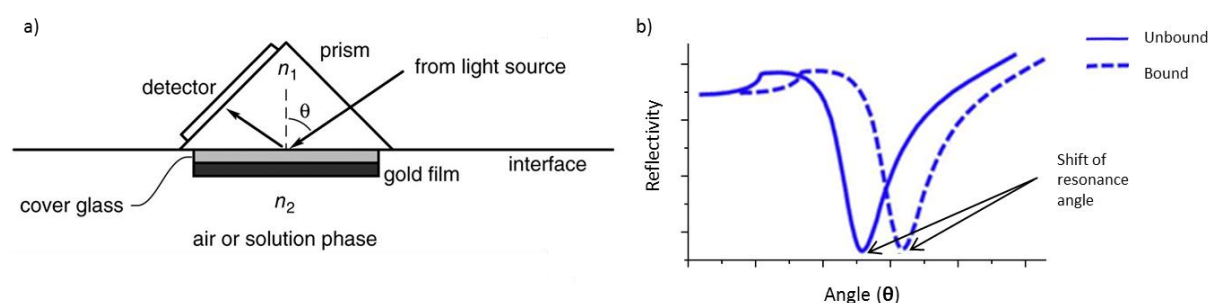

Figure 17 - a) diagram showing the Kretschmann configuration commonly used in SPR, adapted from reference.[100] b) plot of reflectivity vs angle of light source relative to the incident surface, a large dip in reflectivity indicates the resonance angle. The angle shifts on the binding of small molecules to the surface

A flow of buffer containing the ligand is passed over this face of the plate. As ligand is flowed over the plate binding interactions with the immobilized protein will leads to a change in the mass of the surface of the plate. This change in mass, indicating a binding event is recorded as a change the resonance angle due to change in the refractive index of the metal surface Figure 17b.

SPR is able to characterise the on and off rates of a ligand to an immobilised target. As buffer containing ligand is flowed over the immobilised protein there is an abundance of binding sites, the on-rate is high. After some time an equilibrium will be reached where the on and off rates are equal, seen as a plateau in the SPR response curve. Following this the buffer will start to wash away the ligand as it becomes unbound from the binding sites, at this point the off-rate is significantly higher than the on-rate.
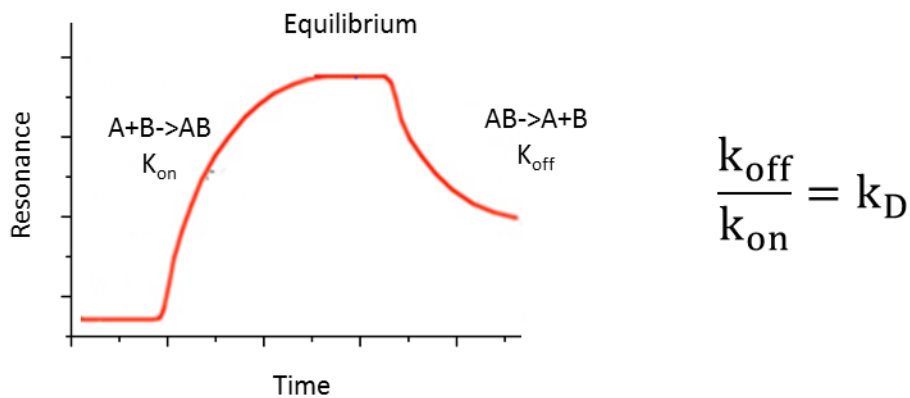
Figure 18 - Example SPR resonance curve showing the different stages of binding, equilibration and dissociation.

From the on and off rates the dissociation coefficient K_D can be calculated. While this method provides a large amount of information, the method is limited as it has a lower throughput than other methods discussed in this section. This makes the method less favourable for application as a primary screening technique.

4.1.5.2.   Isothermal titration calorimetry

In Isothermal Titration Calorimetry (ITC) 2 chambers are set up, a standard chamber and a chamber containing the target protein. Both chambers are heated with extremely high precision to the same temperature. A sample of test compound is added to the chamber and agitated.
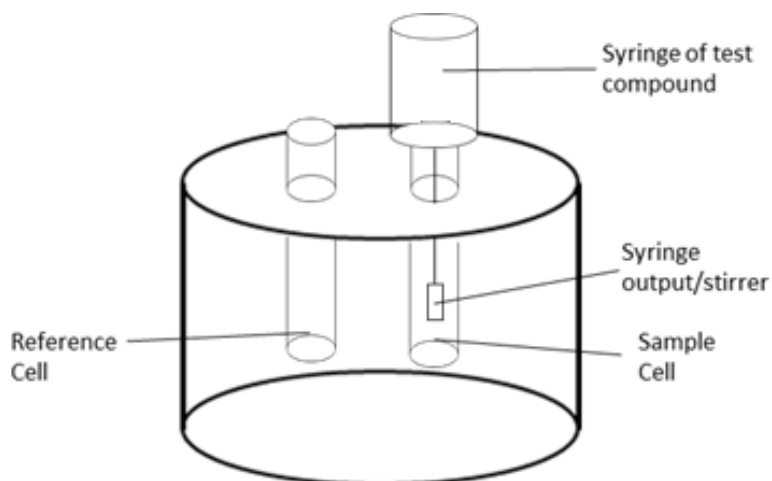
Figure 19 - Simplified schematic of and ITC calorimeter.

Upon addition and mixing there will be a temperature change dependant on the nature of the interaction between protein and ligand. The sample chamber will control the heating or cooling of the sample in order to maintain the temperature as close as possible to the standard chamber. The energy required to maintain this temperature is plotted over time.
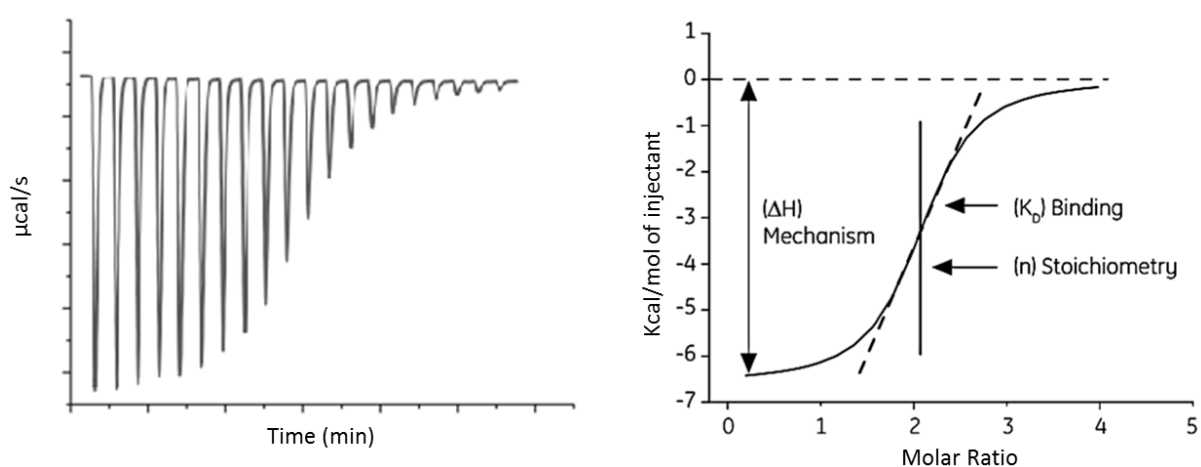

Figure 20 - Left) example output of an ITC experiment, each dip in the plot represents the addition of ligand prior to allowing the temperature to return to normal. Right) Wiseman plot example.

Once the sample temperature no longer requires intervention from the temperature controller another sample of ligand is added. This pattern of adding ligand and monitoring the energy required to maintain the temperature is repeated several times, each time the energy required to normalise the temperature of the chamber reduces. The output of an ITC experiment can reveal much about the nature of a binding interaction. Plotting the data in a Wiseman plot (Figure 20) will reveal the K_D (from the slope of the curve), the stoichiometry of the interaction

(the ratio of ligands binding to proteins is equal to the molar ratio at the middle of the slope) and the vertical size of the curve can suggest whether the mechanism of the interaction is entropy or enthalpy driven. Similarly to SPR, ITC provides a large amount of information at the sacrifice of throughput.

4.1.5.3.   Protein Thermal Shift Analysis

Thermal shift analysis (TSA), sometimes referred to as thermal denaturation analysis, thermal denaturation fluorescence or differential scanning fluorimetry, is another thermal based screening method that takes advantage of the stabilising effects imparted upon proteins by the binding of small molecules or other proteins to the target protein surface. In a TSA experiment the protein is incubated with SYPRO orange dye, this dye binds non-specifically to hydrophobic patches of the protein. When in a hydrophobic environment SYPRO orange has a much higher fluorescence than in aqueous conditions (ex/em 485/575 nm). The sample of protein and SYPRO orange dye is then heated, the protein denatures and in doing so reveals its hydrophobic interior. The SYPRO orange dye fluoresces in the hydrophobic environment presented, allowing for the detection of the denaturation of the protein as in increase in fluorescence. In the event of a binding interaction between the protein and a ligand the thermal stability of the protein will increase, seen as a delay in the fluorescence increase compared to a control experiment. The difference in melting temperature (T_m) in the presence of a ligand is compared to the (T_m) of the control experiment is referred to as the ΔT_m. The stabilising effect of a binding interaction increases as a function of affinity and concentration, theoretically allowing for the ranking of hits from a TSA screen.[101] This allows TSA to be used as a preliminary label-free screening method. The experiments in this chapter were carried out using a Biorad IQ5 ICycler, a real-time quantitative PCR machine. Many quantitative PCR instruments can be configured to be used for a TSA assay allowing for screening in a 96 well format with relatively low concentration of protein per assay (5-15 µM in an assay volume of 50 µl). The output from the experiment is a plot showing the change in fluorescence as temperature increases, Figure 21 shows an example result of a TSA experiment.
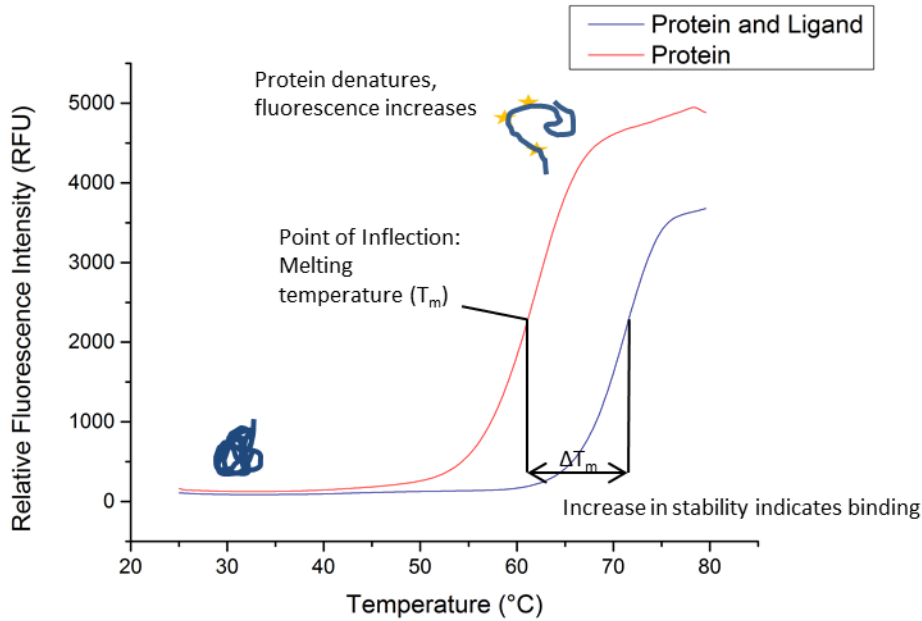
Figure 21 - Example of the results of a TSA experiment. As temperature increases the protein denatures revealing its hydrophobic interior. SYPRO orange dye binds and its fluorescence intensity increases significantly in the hydrophobic environment, allowing for determination of the protein melting temperature (Tm).

From the resulting melting curve it can be difficult to reproducibly determine the melting point of the protein, as it is defined as the point of inflection in the protein melting curve. The negative differential of the data (-d(RFU)/dT) however transforms the melting curve into a peak (note that some programs alternatively use d(RFU)/dT). The melting point (T_m) is then determined as the x value for the lowest y value in the data set. Figure 22 below shows the negative differential of Figure 21above.
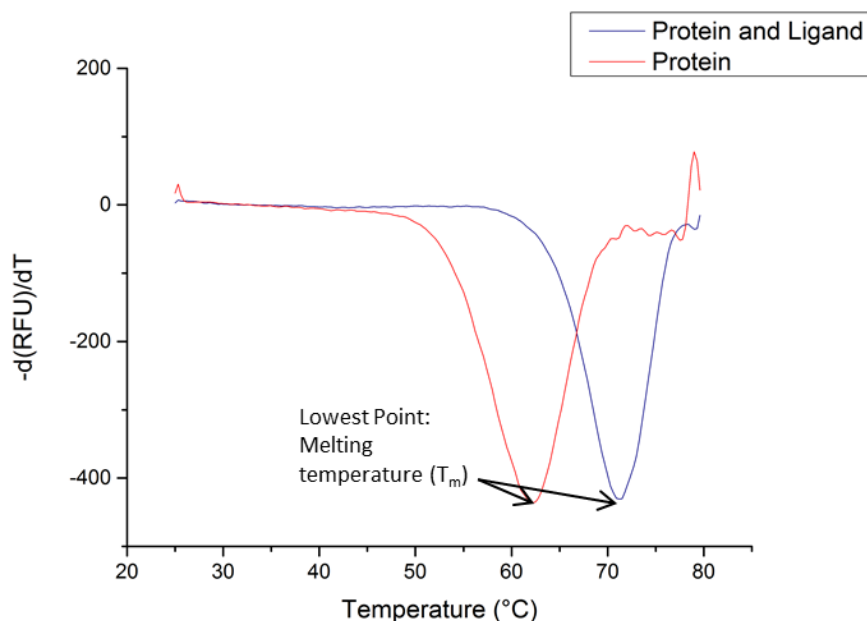
Figure 22 - Example plot of -d(RFU)/dT data. This plot is used to determine the protein melting temperature of a sample.

However TSA has several downsides that must be considered in experimental planning. Mainly that the use of a dye that fluoresces in a hydrophobic environment leads to a poor tolerance of detergents in the assay buffer. It was previously stated that the increase in protein stability is a function of the affinity between the protein and ligand, however the relative stabilising effect of a ligand to a protein will vary from protein to protein. Generally proteins with a higher thermal stability show a small increase of stability on ligand binding, making the determination of lower affinity hits more difficult. Compared to ITC and SPR TSA has much higher throughput, screening in a 96 well-plate format with acquisition times that can be varied depending on the desired resolution. As the degree of stabilisation is reflected as a shift in melting temperature results of TSA screens can also be ranked. These reasons make TSA a favourable initial screening technique.

4.1.5.4. Microdialysis

Microdialysis allows for affinity determination while also evaluating the ligands behaviour in the assay environment. In a microdialysis experiment a rapid equilibrium dialysis (RED) cartridge (Thermo Scientific) is held in a specialised RED plate, the cartridge is made up of 2 chambers separated by a nitrocellulose membrane with an 8000 dalton molecular weight cut off. Compound and protein are mixed in a single chamber at the start of incubation (referred

to as the incubation chamber), to the connected chamber (the diffusion chamber) buffer is added. The principle of microdialysis is that the free ligand will pass freely through the membrane into the second chamber (referred to as the diffusion chamber).
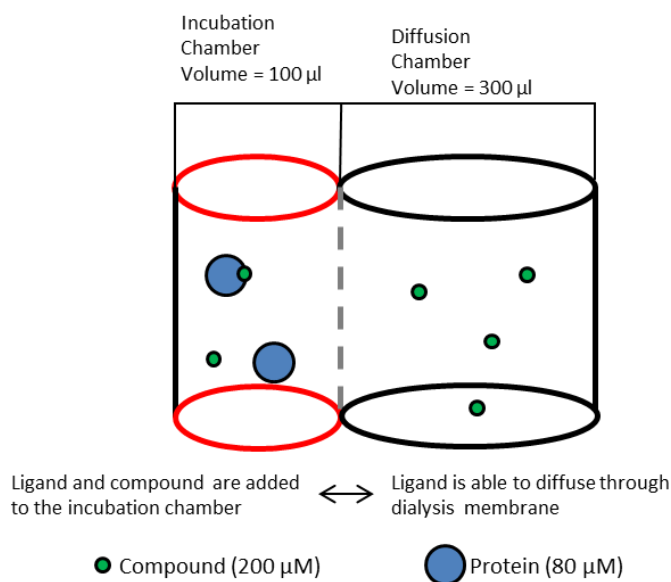
Figure 23 - Schematic of a microdialysis experiment. Protein and ligand are added to the red chamber, after being agitated for a time the ligand will pass through the membrane and an equilibrium concentration established.

Microdialysis experiments can reach equilibrium in 2-4 hours of agitation, in experiments discussed in this chapter the duration was increased to 16 hours to allow in order to avoid precipitation of protein. An equilibrium will be reached where the concentrations of free compound are constant on both sides of the membrane. Should the ligand have an affinity for the protein in the incubation chamber then the different concentrations of ligand in each chamber at equilibrium will reflect this. This is because the ligand bound to the target protein is trapped in the incubation chamber with the protein, while the free compound can move freely. As a result the incubation chamber will have a higher concentration of ligand, as a function of the affinity between the ligand and the protein, the volume assay volume and the concentrations of ligand and protein.[102]

Microdialysis Simulation: The effect of Kd on ligand distribution
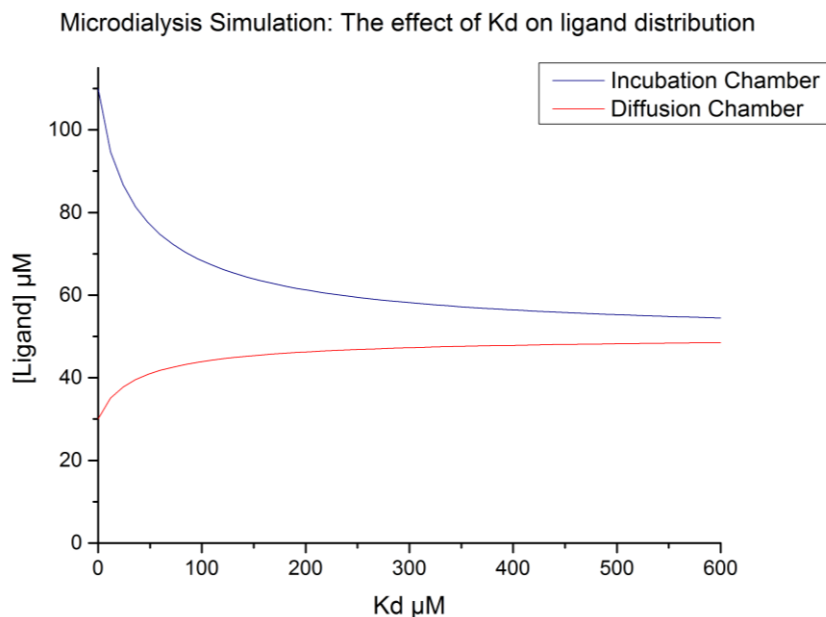
Figure 24 - The effect of increasing affinity on the distribution of ligand in the incubation and diffusion chambers. Concentration of protein and compound are 80 µM and 200 µM respectively.

Figure 24 shows that as affinity increases (K_D decreases) the difference in ligand concentration becomes smaller, this plot illustrates that at higher affinities small changes in ligand concentration reflect larger changes in K_D. When discussing the equilibrium of a microdialysis experiment it is easiest to define the equilibrium numerically, for this the partition coefficient (P) is used. The partition coefficient is the concentration of ligand in the incubation chamber (the starting chamber) divided by the diffusion chamber. In this chapter the partition coefficient of the control experiment is referred to as Pc, the partition coefficient is referred to a Pt. Other label free assay methods discussed above provide limited information on the behaviour of the ligand in the assay environment. Some compounds may, for example bind to the walls of the assay vessel or precipitate out of solution. This would lead to a lower free concentration of the compound in solution and skew the perceived affinity of the compound to the target. In a microdialysis investigation a control experiment is carried out in parallel that evaluates the behaviour of the ligand is carried out. In this control experiment a known volume and concentration of compound alone is added to the incubation chamber and a known volume of buffer is added to the diffusion chamber. The plate is sealed and agitated for 16 hours and the concentration of compound in the two chambers is determined using HPLC and a calibration curve. From the results of the control experiment the behaviour of the ligand in the experiment is determined. For example if 100 µl of a 200 µM solution is added

to the incubation chamber and 300 μl of buffer is added to the diffusion chamber the total volume of the system is 400 μl. Therefore it is expected, with ideal behaviour, that the concentration of ligand in both chambers will be 50 μM.

The concentration in each well is determined by HPLC by using a calibration curve, area under HPLC trace peak vs known concentration of sample. If the value is lower than 50 μM in each well this would suggest that the compound is aggregating on the vessel walls. The compound is expected to be in equal concentrations either side of the dialysis membrane, the Pc should be ≈ 1, this indicates that the ligand is able to pass freely through the membrane. .The results of the control experiment are also used in the interpretation of the results from the main experiment. Next we extend the example above to include an experiment carried out in parallel that uses the same conditions but includes the target protein in the incubation chamber (200 μM ligand and 80 μM protein in 100 μl added to the incubation chamber, 300 μl of buffer added to the diffusion chamber). As the incubation chamber contains protein it is more difficult to evaluate the concentration of ligand in the chamber. HPLC analysis using some systems would require the removal of protein from the sample as protein can lead to blockages in columns. In summary 3 experiments are run in parallel: The test experiment where test compound and test protein are incubated together in order to determine the affinity between the two. A control experiment, where just the ligand in buffer is left to equilibrate in the microdialysis cartridge, is used to determine what the expected total concentration of ligand in the microdialysis system is after equilibration. From this total the amount of ligand in the diffusion well of the test cartridge is subtracted to calculate the amount of ligand in the incubation chamber, from this Pt can be calculated. This control also provides information on the behaviour of the ligand in the assay environment. A second control experiment is carried out where Bovine Serum Albumin (BSA) is used in place of the test protein. This control is used to test the small molecule for non-specific binding.

4.2.   Aims

This project aimed to utilise 2 different *in-silico* methods, the structure-based docking of Q-mol and ligand-based molecular similarity method USRCAT, to determine potential binders to the WD40 domain containing protein WDR5. The compounds suggested by these *in-silico* methods were to undergo screening via label free affinity methods. Initial screening was carried out using thermal shift analysis as it allowed for the screening of all 81 compounds suggested by the *in-silico* tools. Further characterisation of potential hit compounds was carried out using microdialysis.

4.3.   Results

4.3.1. Suggestions from Q-mol software

In-silico work carried out by Anton Cheltsov of Q-mol LLC (http://q-mol.com/).
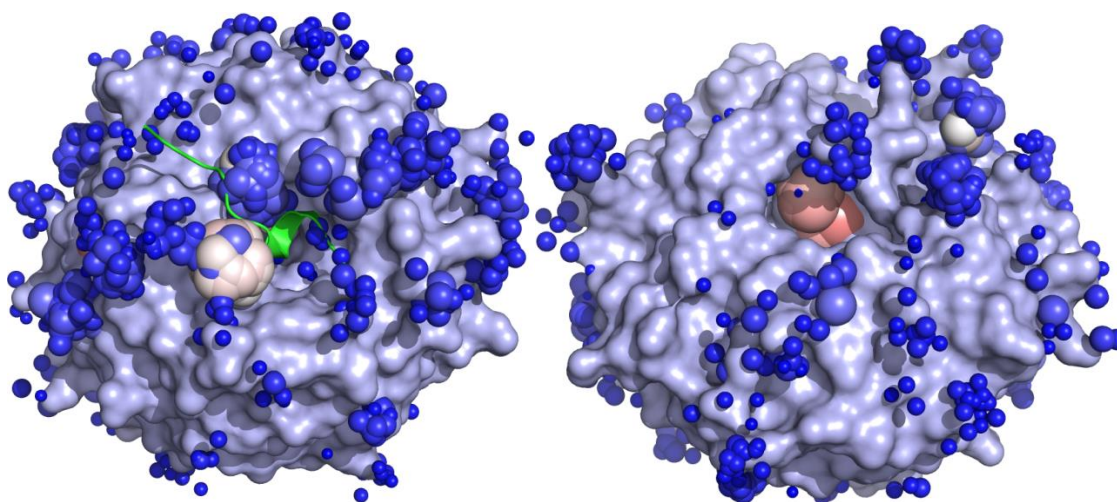

Figure 25 - Top (left) and bottom (right) interfaces of WDR5 (blue). MLL1 peptide (green) shows top binding interface. Spheres (blue to white, to red) indicate possible binding sites, note the predicted binding site on the top interface: The binding site of peptide ARA.

Figure 25 shows the results of the Q-mol software amino acid binding site search. The spheres across the surface indicate where amino acids showed the highest affinity in virtual screening. The colour and size of the spheres indicates the affinity of amino acids in this region. Generally low affinity regions are spread across the surface of the protein. However in the top and bottom interface sites larger spheres of white to red are highly concentrated, indicating higher affinity of amino acids in these known binding sites.

The compound library of the National Cancer Institute (NCI) were used for virtual ligand screening. Potential hit compounds from this virtual screen were forwarded to the Auer lab by collaboration partner Anton Cheltsov of Q-mol LLC. Compounds were filtered for possible pan-assay interference compounds (PAINS) and the structures checked again by eye to ensure no unreasonable compounds were screened. 66 unique compounds were ordered from the NCI. The NCI specifies that compounds sent from their library are not guaranteed to match the structures ordered or to be of high purity. With this in mind all batches of compounds were tested by HPLC and LCMS on arrival for quality control to ensure they matched with the structures suggested by the Q-mol package and were of reasonable purity.
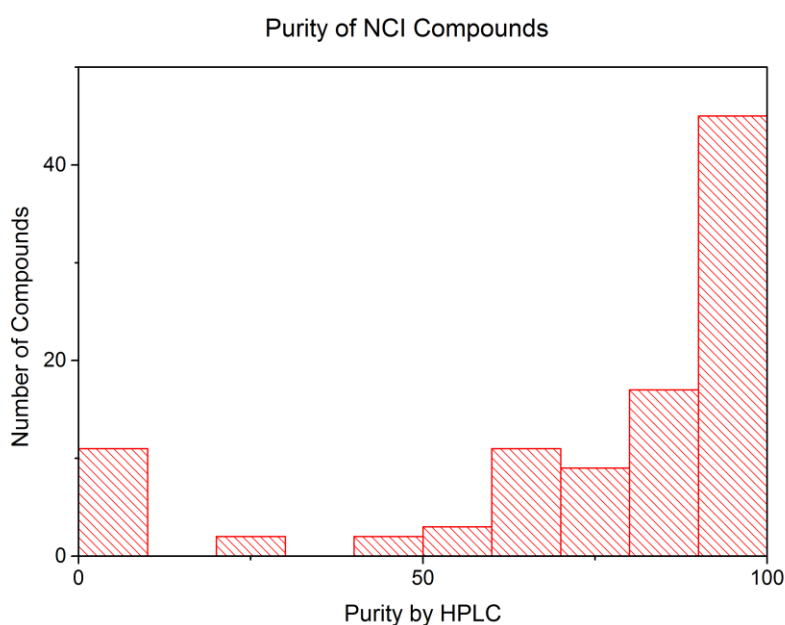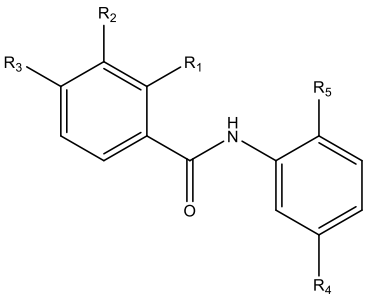

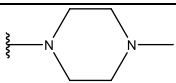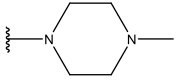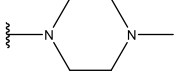**Figure 26 - Histogram showing the purity of compounds received from NCI compound library**
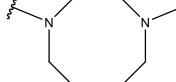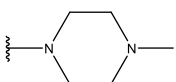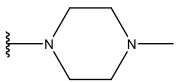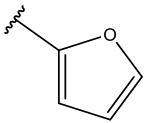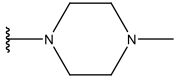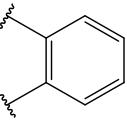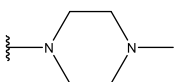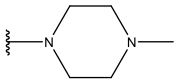
62% of compounds received were of above 80% purity. All compounds were incorporated into the screening process and their purity accounted for.

4.3.2. Suggestions from USRCAT software (carried out by Steve Shave)

For the purposes of targeting WDR5 the structures of 9 known binders were used as starting points. Binders were each analogues from the same SAR series that was carried out by the Structural Genomics Consortium (SGC), targeting the top interface of WDR5[103]. These input structures are summarised in Table 7.

Table 7 - Structures of molecules used as input for USRCAT

CHEMBL ID	R1	R2	R3	R4	R5
CHEMBL1407785	H	H	H	NO_2	piperazine
CHEMBL2337031	H	OMe	F	NO_2	piperazine
CHEMBL2337029	Cl	H	F	NO_2	piperazine
CHEMBL2337035	Cl	H	H	NO_2	diazocane
CHEMBL2337032	Cl	Me	F	NO_2	piperazine
CHEMBL2337021	Cl	H	H	CO_2Me	piperazine
CHEMBL2337026	Cl	H	H	furan	piperazine
CHEMBL2337017	fused benzene		H	NO_2	piperazine
CHEMBL1457864	H	OMe	H	NO_2	piperazine

The descriptors of the compounds in Table 7 were extracted and compared to descriptors of compounds in the Vichem NCL. The structures of the Vichem NCL were not released. 60 compounds were suggested to be structurally similar by USRCAT. Table 8 lists the Vichem compounds determined to be similar their corresponding known WDR5 binders.

Table 8 - 60 Vichem suggestions and their known active analogue ID.

Known active ID	Vichem ID	USRCAT score	Known active ID	Vichem ID	USRCAT score
CHEMBL2337031	753739	19.886	CHEMBL2337031	831487	58.776
	569490	21.742		227480	59.279
	26098	21.912		989582	64.435
	540656	22.886		901864	64.882
	919704	28.443		633975	67.363
	200748	30.643	CHEMBL2337017	231498	37.080
	651631	31.694		767674	38.404
	426460	34.551	CHEMBL1457864	205790	27.766
	16281	35.290		48729	31.486
	368132	35.568		198490	33.264
	45851	35.913		465960	33.357
	106710	38.751		237847	36.354
	656584	39.532		859152	37.600
	716300	39.900		256638	41.835
	842952	41.901		931779	44.085
	632479	41.925		629613	49.154
	825861	42.282		135975	49.773
	103983	43.488	CHEMBL2337032	867989	28.347
	924807	43.903		966518	28.843
	645779	44.726		8936	30.879
	741125	44.858		442333	64.321
	393581	46.679	CHEMBL2337035	133829	47.172
	763039	50.336	CHEMBL2337021	176995	40.810
	88484	50.412	CHEMBL2337026	284066	62.748
	878668	53.636	CHEMBL2337029	966518	25.208
	506057	53.677		867989	31.069
	279910	56.513		8936	34.328
	282276	56.659	CHEMBL1407785	541316	30.303
	448902	57.274		775754	43.872
	356408	57.588			

USRCAT scoring is a function of the difference between the descriptors of the known binder from CHEMBL and the Vichem compound. A lower score reflects that the two compounds have more similar descriptors.

Of the 60 suggestions the 15 compounds with the lowest scores (highest similarity to known binders) that were available from Vichem were ordered and assayed against WDR5 by TSA in the same manner as the Q-mol suggested compounds.

Table 9 - Final list of 15 compounds suggested by USRCAT screened against WDR5

Vichem ID	CHEMBL ID of known active	USRCAT score
753739	CHEMBL2337031	19.886
966518	CHEMBL2337029	25.208
569490	CHEMBL2337031	21.742
205790	CHEMBL1457864	27.766
26098	CHEMBL2337031	21.912
26098	CHEMBL2337031	21.912
867989	CHEMBL2337029	31.069
867989	CHEMBL2337032	28.347
867989	CHEMBL2337029	31.069
540656	CHEMBL2337031	22.886
540656	CHEMBL2337031	22.886
919704	CHEMBL2337031	28.443
541316	CHEMBL1407785	30.303
426460	CHEMBL2337031	34.551
8936	CHEMBL2337029	34.328
8936	CHEMBL2337032	30.879
16281	CHEMBL2337031	35.290
651631	CHEMBL2337031	31.694
45851	CHEMBL2337031	35.913
465960	CHEMBL1457864	33.357

The USRCAT score is a dissimilarity score, i.e a score of 0 would be given for two identical structures. Compounds underwent quality control as with the Q-mol suggested compounds, all compounds were determined to be of >95% purity.
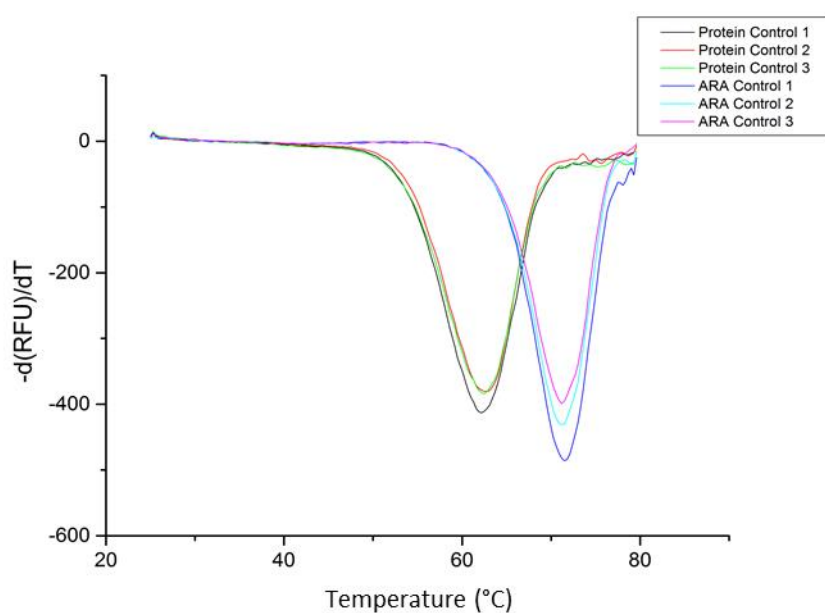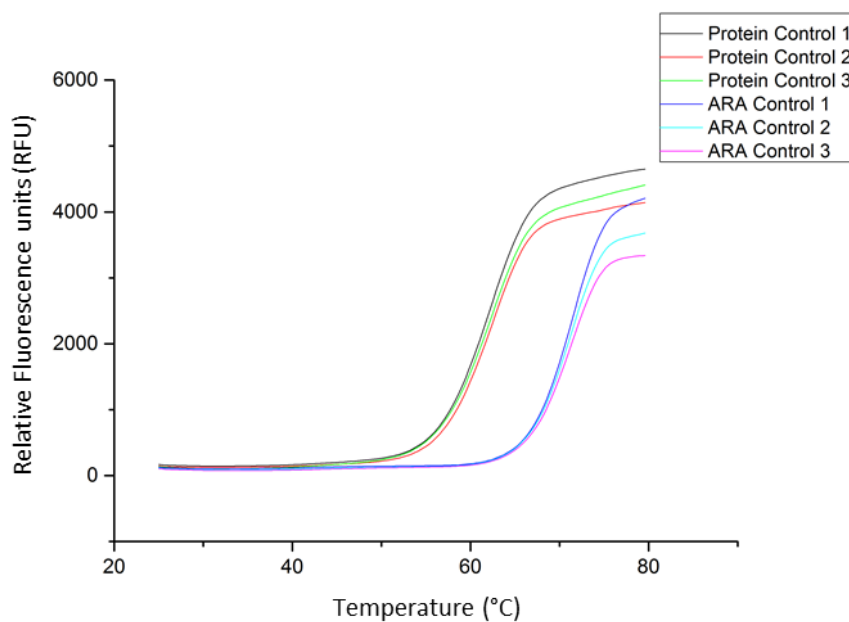
4.3.3. Screening of Q-mol Suggested Compounds

4.3.4. Thermal Denaturation Fluorescence

4.3.4.1. Control Results

Figure 27 shows that the melting temperature of WDR5 alone is 62.5 °C. There is a 9 °C stabilisation of the melting temperature of the protein on inclusion of the control peptide Ac-ARA-CONH$_2$ (literature affinity: $K_D = 0.12 \pm 0.01$ μM by competition).[104]

	TM1 (°C)	TM3 (°C)	TM3 (°C)
Control	62.2	62.8	62.5
ARA	71.5	71.2	71.2

Figure 27 - Control experiment results. Melting experiments were repeated in triplicate. On inclusion of control peptide ARA melting curves shifted significantly to the right, indicating an increase in thermal stability.

4.3.4.2. Screening Results

Screening of the NCI compounds gave a single hit, NCI-292249 Figure 29. Compounds with a similar structure, NCI292250 and NCI292251 (Figure 32) showed no increase in stability. 10 compounds also saw a significant decrease in stability.
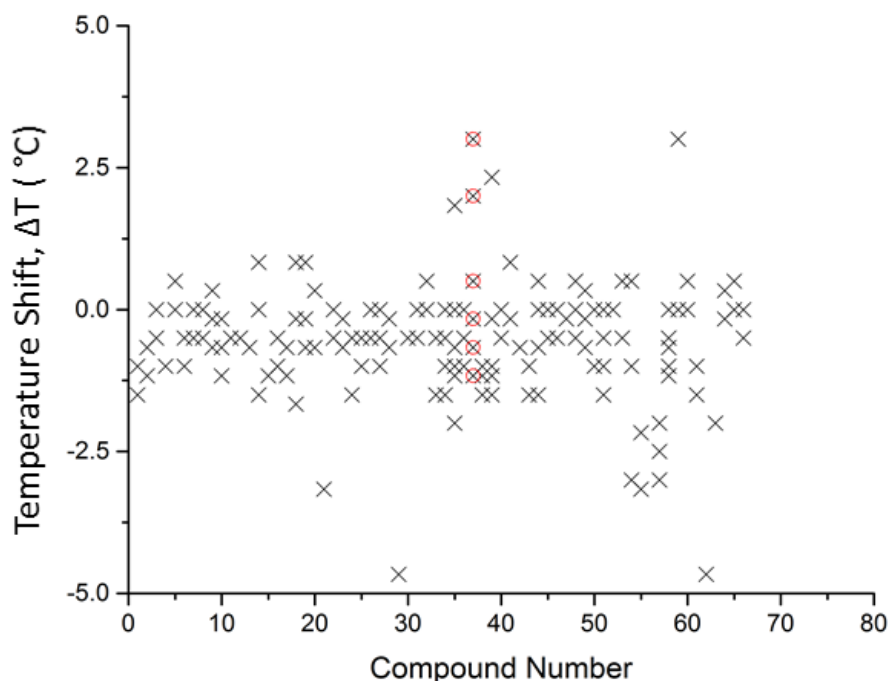
Figure 28 - Summary of temperature shifts caused by incubation of WDR5 with NCI compound selection. NCI292249 is highlighted in red.

Compounds NCI292251, NCI710574 and NCI247023 were also seen to have similar stabilising effects but only in singular experiments, while 292249 was seen to increase the melting temperature of WDR5 in multiple experiments. Repeated screens of 292249 against WDR5 showed no notable increase in stabilisation. Table 10 below cross references the compound number of Figure 28 with the corresponding NCI number.

Table 10 - Complete list of NCI compounds screened against WDR5, NCI292249 highlighted in grey.

#	NCI#	ΔT	#	NCI#	ΔT	#	NCI#	ΔT
1	4945	-1.5	11	71016	-0.5	19	119728	-0.2
1	4945	-1.5	11	71016	-0.5	19	119728	0.8
1	4945	-1	11	71016	-0.5	19	119728	-0.7
2	16881	-0.7	12	78039	-0.5	20	134126	-0.7
2	16881	-1.2	12	78039	-0.5	20	134126	0.3
2	16881	-1.2	12	78039	-0.5	20	134126	-0.7
3	26808	-0.5	13	79629	-0.7	21	134142	-12.2
3	26808	0.0	13	79629	-8.7	21	134142	-6.7
3	26808	-0.5	13	79629	-0.7	21	134142	-3.2
4	29598	-1	14	101590	-12.2	22	152524	-0.5
4	29598	-1	14	101590	0.8	22	152524	0.0
4	29598	-1	14	101590	17.3	22	152524	-0.5

Table 10 cont.

#	NCI#	ΔT	#	NCI#	ΔT	#	NCI#	ΔT
5	41312	0.5	14	101590	-35.5	23	163571	-0.2
5	41312	0.0	14	101590	0.0	23	163571	-0.2
5	41312	0.0	14	101590	-1.5	23	163571	-0.7
6	45031	-1	15	106502	-12.2	24	165704	-1.5
6	45031	-0.5	15	106502	-11.2	24	165704	-0.5
6	45031	-1	15	106502	-1.2	24	165704	-1.5
9	45939	-0.2	17	114360	-1.2	26	167707	0.0
9	45939	-0.7	17	114360	-0.7	26	167707	-0.5
9	45939	0.3	17	114360	-12.2	26	167707	-35.5
10	46210	-0.7	18	119689	0.8	26	167707	-35.5
10	46210	-0.2	18	119689	-1.7	26	167707	0.0
10	46210	-1.2	18	119689	-0.2	26	167707	-35.5
27	174930	0.0	35	247023	1.8	40	294166	0.0
27	174930	-1.0	35	247023	0.0	40	294166	-0.5
27	174930	-0.5	35	247023	-1.0	40	294166	-0.5
28	174931	-0.7	35	247023	-1.2	41	295516	-0.2
28	174931	-0.7	35	247023	-0.7	41	295516	0.8
28	174931	-0.2	35	247023	-2.0	41	295516	-0.2
29	185348	-4.7	36	290753	-0.5	42	321518	-0.7
29	185348	-5.2	36	290753	-1.0	42	321518	-0.7
29	185348	-11.7	36	290753	-0.5	42	321518	-0.7
30	201914	-0.5	37	292249	-0.2	43	329276	-1.5
30	201914	-0.5	37	292249	-0.7	43	329276	-1.5
30	201914	-0.5	37	292249	0.5	43	329276	-1.0
31	202116	0.0	37	292249	2.0	44	338076	0.0
31	202116	-0.5	37	292249	3.0	44	338076	0.0
31	202116	-0.5	37	292249	-1.2	44	338076	0.0
32	202428	0.5	38	292250	-1.5	44	338076	-0.7
32	202428	0.0	38	292250	-1.0	44	338076	-0.7
32	202428	0.0	38	292250	-1.2	44	338076	-0.7
33	202477	-0.5	38	292250	-1.2	44	338076	-1.5
33	202477	-1.5	38	292250	-1.0	44	338076	0.5
33	202477	-0.5	38	292250	-1.2	44	338076	0.0
34	215624	-1.0	39	292251	-1.5	45	339182	-0.5
34	215624	-0.5	39	292251	-1.0	45	339182	0.0
34	215624	-0.5	39	292251	-1.0	45	339182	0.0
34	215624	-0.5	39	292251	-1.2	46	340965	-0.5
34	215624	-0.5	39	292251	-0.2	46	340965	0.0
34	215624	-1.5	39	292251	2.3	46	340965	-0.5
34	215624	-0.5	40	294166	0.0	47	348965	-0.2
34	215624	0.0	40	294166	-0.5	47	348965	-0.2
34	215624	-0.5	40	294166	-0.5	47	348965	-0.2

Table 10 continued

#	NCI#	ΔT	#	NCI#	ΔT	#	NCI#	ΔT
48	367458	0.0	54	613679	-3.0	63	710766	-35.5
48	367458	0.0	54	613679	-1.0	63	710766	-2.0
48	367458	-0.5	54	613679	0.5	63	710766	-35.5
48	367458	0.5	55	622608	-2.2	64	715809	-0.2
48	367458	0.5	55	622608	-3.2	64	715809	0.3
48	367458	0.0	55	622608	-2.2	64	715809	0.3
48	367458	-0.5	56	622622	-12.2	65	724956	0.5
48	367458	-0.5	56	622622	-8.2	65	724956	0.5
48	367458	-0.5	56	622622	-12.2	65	724956	0.0
49	373089	-0.2	57	635302	-3.0	66	1000129	-0.5
49	373089	0.3	57	635302	-2.0	66	1000129	-0.5
49	373089	-0.7	57	635302	-2.5	66	1000129	0.0
49	373089	-0.2	58	635443	-0.5			
49	373089	-0.2	58	635443	-1.0			
49	373089	-0.2	58	635443	-1.2			
50	379379	0.0	58	635443	-1.2			
50	379379	-1.0	58	635443	-0.7			
50	379379	-1.0	58	635443	0.0			
51	401293	-0.5	59	656633	0.0			
51	401293	-0.5	59	656633	0.0			
51	401293	-1.0	59	656633	3.0			
51	401293	-1.5	60	665606	0.5			
51	401293	-0.5	60	665606	0.0			
51	401293	0.0	60	665606	0.0			
52	401294	-35.5	61	667453	-1.0			
52	401294	-35.5	61	667453	-1.0			
52	401294	0.0	61	667453	-1.5			
53	401302	0.5	62	710574	-8.2			
53	401302	0.5	62	710574	7.3			
53	401302	-0.5	62	710574	-4.7			

NCI292249 (Figure 29) was the only screened compound to show a borderline stabilising effect in more than a single screen ($\Delta T > 3$ °C) and so was selected for further investigation.

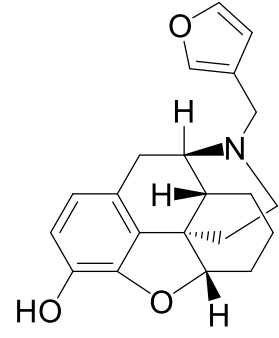|	NCI-292249
	Purity by HPLC: >95%
	Mass: 337.41 m/z by LCMS: 338.34 [M+H]$^+$

Figure 29 - Structure and quality control data of NCI292249

NCI-292249 was further characterised by microdialysis below.

4.3.4.3.   Microdialysis of NCI292249

To determine the affinity of NCI292249 microdialysis was used. First a series of concentrations of NCI292249 solutions were analysed by HPLC to determine a concentration gradient.

Table 11 - Results of HPLC analysis of concentration gradient of NCI292249

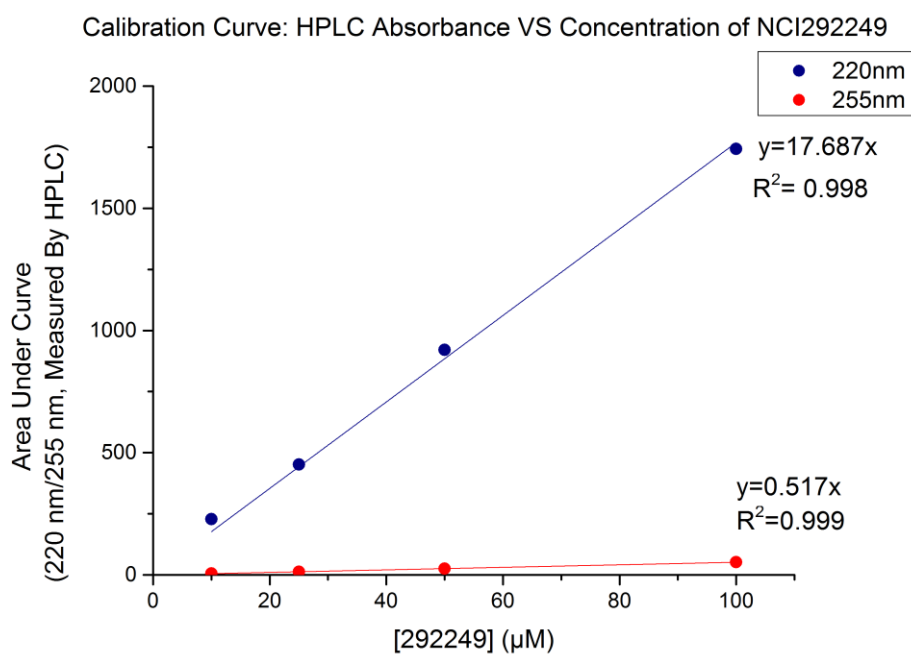|		Peak Area	
[NCI292249] µM	RT (min)	220 nm	255 nm
10	14.47	228.6	6.2
25	14.46	451.4	12.8
50	14.47	920.4	25.5
100	14.45	1743.2	51.8

Figure 30 - Calibration curve of NCI292249, analysed by HPLC

A blank experiment where NCI292249 was incubated alone in a microdialysis cartridge was carried out. This was to verify that the compound dialysed across the nitrocellulose membrane with no loss of material. Table 12 shows the results of HPLC analysis after incubation for 16 hours at 25 °C.

Table 12 - Results of NCI292249 microdialysis equilibration evaluation. 100 µl of 200 µM NCI292249 was injected into the incubation well, 300 µl of buffer was injected into the diffusion well. Plate was sealed, shaken 20°C for 16 hours. Concentration in each well was determined by HPLC with a concentration curve.

Sample	RT (min)	Peak Area 220nm	Peak Area 255nm	Total Peak Area 220 nm	Total Peak Area 255 nm	Concentration (µM) 220nm	Concentration (µM) 255nm	Partition Coefficient (Pc) 220nm	Partition Coefficient (Pc) 255nm
Incubation Chamber 1	14.47	963.5	27.5	1919.0	54.7	54.48	53.19	1.008	1.011
Diffusion Chamber 1	14.47	955.5	27.2			54.02	52.61		
Incubation Chamber 2	14.47	994.5	28.2	1973.4	55.7	56.23	54.55	1.016	1.025
Diffusion Chamber 2	14.47	978.9	27.5			55.35	53.19		
			Average total peak area	1946.2	55.2				

Using the calibration curve it was determined that the concentration of compound on either side of the membrane was slightly higher than expected. 100 µl of a 200 µM solution was diluted into a 400 µl solution, giving an expected concentration of 50 µM either side of the membrane. The value is close to the expected value, error from pipetting, evaporation etc can lead to errors. This indicates that the compound is not binding to the membrane or walls of the vessel. The partition coefficient, Pc, (incubation chamber concentration divided by diffusion chamber concentration) after 16 hours incubation is close to 1.0. This indicates that the compound is able to pass freely through the membrane.

Once it had been verified that NCI292249 behaves well under the applied assay conditions the compound was assayed against WDR5 and a control protein, BSA. Table 13 shows the results of this assay. Values for protein chamber (incubation chamber) concentrations of NCI292249 were calculated by subtracting the peak area of the diffusion chamber from the average total peak area calculated in table 6.

Table 13 - Results of the microdialysis of NCI292249 against WDR5 and BSA. 100 µl of 200 µM NCI292249 and 80 µM protein was injected into the incubation well, 300 µl of buffer was injected into the diffusion well. Plate was sealed, shaken 20°C for 16 hours. Concentration of material was determined by HPLC concentration curve.

Sample	RT (min)	Area 220 nm	255 nm	Concentration (µM) 220 nm	255 nm	Partition Coefficient (Pt) 220 nm	255 nm
WDR5 – Incubation Chamber		1035.2*	29.8*	58.53*	57.64*	1.136	1.173
WDR5 – Diffusion Chamber	14.45	911.0	25.4	51.51	49.13		
BSA – Incubation Chamber		1048.5*	30.1*	59.28*	58.22*	1.168	1.199
BSA – Diffusion Chamber	14.46	897.7	25.1	50.75	48.55		

*calculated through subtraction of diffusion chamber values from the average total concentration calculated from the control experiment (Table 12).

The partition coefficient of NCI292249 when incubated with either BSA or WDR5 is > 1, showing that NCI292249 has an affinity to both WDR5 and BSA. The partition coefficient is converted into a K_D using an equation derived by the Mathmatica software package by Steve Shave. A partition coefficient of 1.136 gives a K_D of 564 µM. The microdialysis results suggest that NCI-292249 is a very weak binder to WDR5, showing a similar affinity to BSA (K_D calculated to be 452 µM). As discussed in the introduction small changes in ligand concentration reflect larger changes in K_D, making it possible that the affinity of NCI292249 and BSA, and NCI292249 and WDR5 is similar.
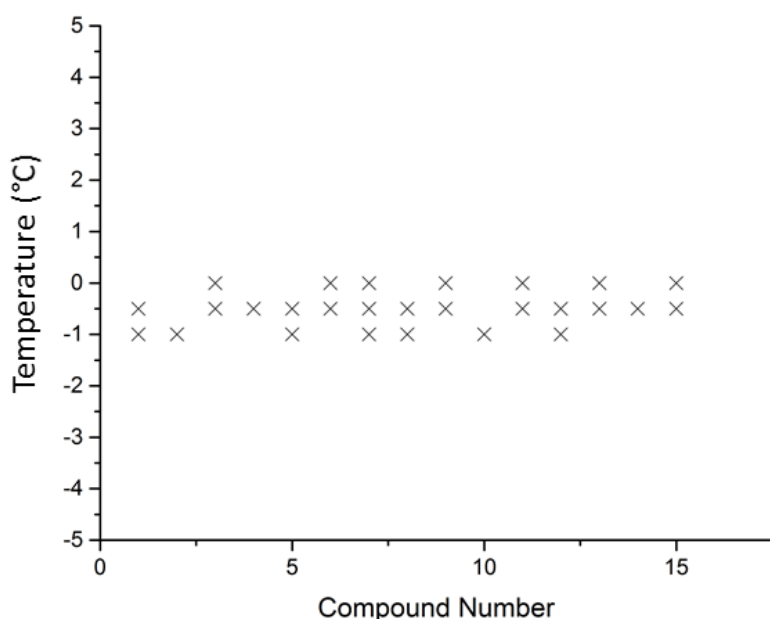
4.3.5. Screening of USRCAT suggested compounds.

Figure 31 - Summary plot of the change in melting temperature of WDR5 when incubated with USRCAT suggested Vichem compounds

Data shown in Figure 31 reveals that the compounds assayed against WDR5 from Vichem showed a much smaller range of temperature shifts compared to the Q-mol suggested compounds from the NCI library. Table 8 below cross references the compound number of Figure 31 with the corresponding Vichem number. No compounds showed any significant change in the melting temperature of WDR5.

Table 14- Results of TSA assay of Vichem compounds against WDR5

#	Vichem Number	ΔT	#	Vichem Number	ΔT	#	Vichem Number	ΔT
1	867989	-0.5	6	16281	0.0	11	753739	0.0
1	867989	-1.0	6	16281	-0.5	11	753739	-0.5
1	867989	-0.5	6	16281	0.0	11	753739	-0.5
2	45851	-1.0	7	569490	0.0	12	541316	-0.5
2	45851	-1.0	7	569490	-0.5	12	541316	-1.0
2	45851	-1.0	7	569490	-1.0	12	541316	-0.5
3	26098	0.0	8	8936	-1.0	13	919704	-0.5
3	26098	-0.5	8	8936	-0.5	13	919704	-0.5
3	26098	-0.5	8	8936	-0.5	13	919704	0.0
4	651631	-0.5	9	966518	0.0	14	540656	-0.5
4	651631	-0.5	9	966518	-0.5	14	540656	-0.5
4	651631	-0.5	9	966518	0.0	14	540656	-0.5
5	205790	-0.5	10	426460	-1.0	15	465960	0.0
5	205790	-0.5	10	426460	-1.0	15	465960	-0.5
5	205790	-1.0	10	426460	-1.0	15	465960	0.0

4.4. Discussion and Future Work

The Structural Genomics Consortium carried out a TSA screening project between WDR5 and 1600 compounds. In this screen compounds were considered hits when they stabilised the protein melting temperature by ≥3°C.[105] Comparatively the only compound seen to have a stabilising effect of this significance in repeated screens was NCI292249. Further testing showed that this binding was likely low affinity and possibly non-specific as the compound a similar affinity to BSA in microdialysis. It was determined that the interaction between NCI292249 was too weak to pursue any optimisation efforts from the structure. The structure of NCI292249 shows little similarity to any known WDR5 binding compounds. NCI292250 and NCI292251 (Figure 32) were also assayed against WDR5 in the TSA experiment. Due to their similar structure and consecutive numbering it is possible the compounds were a part of a series screened against a different target prior to their addition to the NCI library. Despite their similarity NCI292250 and NCI292251 showed no affinity to WDR5.
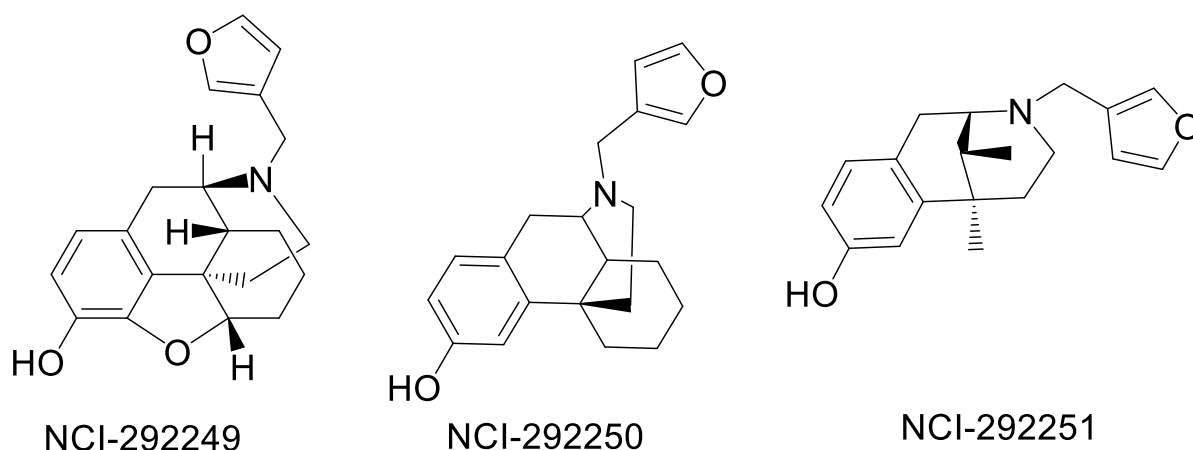
Figure 32 - Structures of NCI292249, NCI292250 and NCI292251

Searching for the structure of NCI292249 finds that compounds that are the most similar in structure to opioid antagonists and agonists such as naltrexone and morphine (Figure 33).
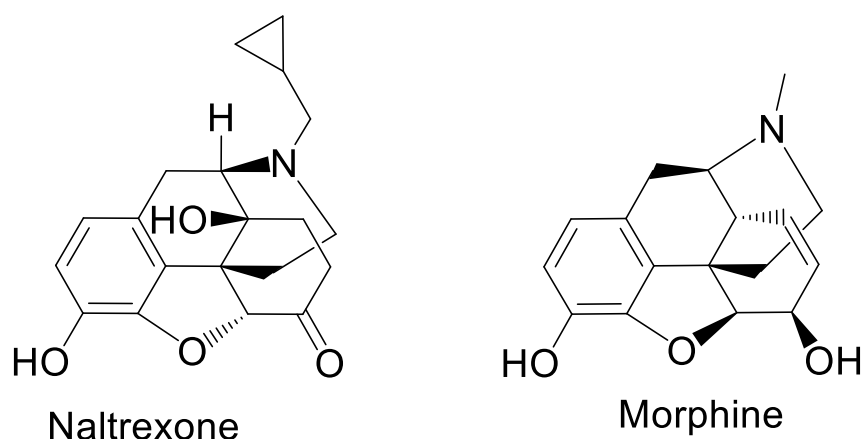
Figure 33 - Structures of naltrexone and morphine. Compounds with similar molecular structures to NCI292249

Compounds supplied by the NCI compound library were variable in quality, in any screening experiment this is an unfavourable situation as it can draw into question any results of the experiments. Several compounds screened from the NCI library were recorded as causing a negative thermal shift. These compounds seem to cause a destabilisation of the protein, literature suggests that this destabilisation is likely concentration dependent and could be a symptom of the compounds displaying low solubility.[106]

4.4.1. Ligand Based **In-Silico** Methods Q-Mol and USRCAT

Despite the unfavourable results of the compound screening experiments the Q-mol software package did determine the top and bottom interface sites of WDR5 (Figure 25), showing that the software is capable of determining sites for its docking experiments. The compounds provided by Vichem based on suggestions from similarity analysis are much harder to compare as their structures are unavailable. Having the structures of compounds suggested by a molecular similarity method allows for researchers to compare structures to their input. In this instance it is unknown whether the compounds screened were reasonable analogues of the known binders to screen against WDR5. However the results of the similarity analysis can be discussed. The similarity scores of the Vichem compounds screened against WDR5 had a range of 20 – 36. Given that the USRCAT compares the distances of similar atoms to calculate a score these values are high, suggesting that the screened compounds are not as similar to the known binders as would be desired. It is also important to consider the results of SAR work that was carried out to develop the known binders for WDR5 that was used as input for USRCAT. The Structural Genomics Consortium determined that affinity of the compounds

developed in their SAR series was highly sensitive to very small changes in structure.[103] Without being able to compare the output of USRCAT to known binders to WDR5 or carry out any *in-silico* docking investigations it is impossible to assess the output any further than the results of screening experiments.

Ligand-based virtual screening methods provide an additional tool for the discovery of novel ligands to protein targets. In this instance a new potential binder to WDR5 was determined, however due to the poor affinity and complex structure it was determined that NCI292249 was limited in terms of opportunities for further development.

4.4.2. Microdialysis

Microdialysis provided a simple method of characterising the affinity of NCI292249 to WDR5 and BSA. Part of the characterisation process was the determination of the behaviour of the compound in the assay vessel. Ligands can be promiscuous and bind with the negatively charged nitrocellulose membrane or the walls of the incubation chamber, or may precipitate out of solution. Without correcting for these factors the concentration in of ligand in the protein incubation chamber would be calculated to be higher, resulting in inaccurate K_D determination. In the control experiment it was determined that the concentration of ligand was higher than expected, this could have been the result of evaporation reducing the sample volume or an error in pipetting. As the experiments were carried out in parallel the average total peak area from the control experiment was used as the total peak area for the affinity determination experiment. From this the peak area for the incubation chamber was calculated. **Error! Reference source not found.** shows the expected equilibrium concentration of compound in microdialysis chambers as a function of the affinity between the protein and the ligand. The plot uses the concentrations of protein and ligand used in the experiment described in this chapter, 80 μM and 200 μM respectively. The plot shows that as K_D increases (affinity decreases) the respective concentrations in either chamber plateau. This makes affinity determination at lower affinities less accurate, as small changes in concentration correspond to large changes in K_D.

NCI292249 was determined to have a K_D of 564 μM to WDR5 and 452 μM to BSA. Shows that a simulation of the experiment puts these results at the upper limit of the simulation, where

small changes in the concentration of ligand equate to large variations in the calculated affinity. This means that the calculated K_D values have a large error associated with them.

4.4.3. Future Work

4.4.3.1. **In-silico** Methods

Future screening experiments could be made more successful by combining structure-based and ligand-based techniques using the same database of ligands. For example using molecular similarity techniques to create a short-list of potential compounds, followed by docking studies using this priority set of compounds, preferably using different docking programs. Multiple techniques could also be applied to the same dataset to develop a consensus scoring approach. A key advantage of molecular similarity methods is that they allow for the reduction of a structure to a string of numbers. This string of numbers is significantly simpler than a structure and allows for a great many structures to be collected in the same format in a database. A database of molecular descriptions from CHEMBL entries, the NCI or other databases or libraries could be used for a broader search for compounds with similar structures to known binders. Searching many thousands of compounds could potentially return many possible similar compounds to use as input for a structure based methods.

4.4.3.2. Screening Methods

TSA was shown to be only be a useful label free affinity technique if the affinity of the compounds to the target protein is high enough, the thermal shift of weaker binders is difficult to detect. Though the method is cost effective with reasonable throughput thanks to the 96 well plate format. In the future TSA could be further utilised as a label free screening method for WDR5 or other WD40 domain containing proteins to determine potential hits for further characterisation. The method could also be used to characterise protein-protein interactions of WD40 domains.[107] Methods exist for the determination of K_D from the results of TSA assays,[108] however these methods require further characterisation of thermal denaturation of the protein. While determination of affinity from a primary screen is advantageous, without the K_D determination it is still possible to rank hits by their thermal shift. As discussed in the introduction, the use of microdialysis as a K_D determination technique has the distinct advantage of characterising the behaviour of ligands in the assay environment. This makes it a staple of the Auer lab screening process, ensuring that it will be used for future screening

experiments. In order to better characterise a binder an alternative method of K_D determination should be used as well as microdialysis. ITC or SPR (as discussed in the introduction) are reasonable choices for K_D determination and can also offer more information on the mechanism of binding. An alternative to these, given a compound shown to bind to a target protein through label free techniques, would be to carry out a competition experiment where a labelled, known binder to WDR5 or another WD40 domain containing protein is displaced by the unlabelled ligand. In the case of WDR5 both the top and bottom interfaces have known peptide binders that can easily be synthesised in a labelled form using methods well characterised by the Auer lab.

5. Synthesis and Screening of a Scanning One-Bead, One-Compound (SOBOC) Peptide Library

5.1.   Introduction

5.1.1. WDR5 – Structural core to therapeutically relevant complexes

In recent years protein-protein interactions (PPIs) have become increasingly targeted as a result of better understanding of their nature both, from the view of biological pathways and structural data.[109] Classically PPIs have been defined as difficult targets due to their characteristic flat, hydrophobic surfaces. Despite this generalisation structural data of interactions between proteins indicates that relatively short stretches of protein are responsible for interactions.[46] These areas that have a more significant contribution to binding are referred to as hotspots.[2] Through alanine scanning it is possible to determine the relative contributions of amino acid residues to the PPIs. Using this information it is possible to use peptides, derived from known interacting sequences in the development of a new or optimized peptide-derived binder for target PPIs. In this chapter a method is described whereby a known binding peptide is fragmented into all possible peptide truncates of the original full length sequence. By reducing a long peptide to a smaller fragment it becomes much more manageable to improve on the interaction through SAR studies, screening all possible fragments is the most thorough method of achieving this. WDR5 is an example of WD40 domain involved in several different protein-protein interactions including two validated cancer targets, MLL1 (mixed linked leukemia 1) protein and MYC proteins. The WDR5 crystal structure revealed that its 334 amino acid sequence exhibits two major interaction surfaces, the top and bottom faces. The smaller top interface is the interaction site for mixed lineage leukemia 1 (MLL1), as discussed in chapter 3 translocation in the MLL1 gene can lead to the formation of MLL fusion proteins that are linked to leukemia. A notable example is the MLL1-AF9 fusion protein that has been shown to cause acute leukemia in mice when introduced as a fusion gene.[110] Inhibition of these mutant fusion complexes from binding with WDR5 has been the subject of extensive research regarding the top interface of WDR5. MLL1 dependant cancers, including AML (Acute Myeloid Leukemia) and ALL (Acute Lymphoblastic Leukemia).[111] MYC is a highly validated target known to bind to WDR5. MYC is a prominent cancer biomarker that is seen in a variety of cancers.[15, 16] Recently it has been

determined that the sequence H$_2$N-DEEEIDVVSVE-CO$_2$H from MYC binds to the larger bottom interface of WDR5 (Figure 34).[12]
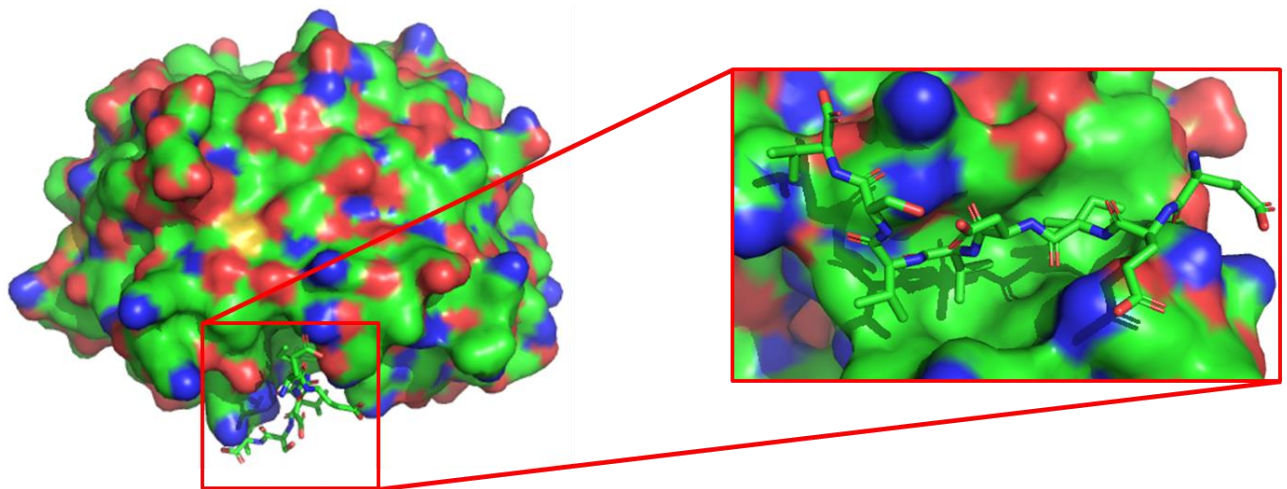
Figure 34 - Crystal structure of MYC peptide binding to the bottom interface of WDR5[12]

MYC is not the first protein found to bind to the bottom interface of WDR5, retinoblastoma-binding protein-5 (RbBP5) also binds to the flat interface as a part of the MLL1 complex where the absence of WDR5 and RbBP5 significantly inhibits methylation of H3K4.[112] KAT8 Regulatory NSL Complex Subunit 2 (KANSL2) also binds to the bottom interface as a part of the non-specific lethal (NSL) complex, a complex that is responsible for the regulation of gene expression.[113]

5.1.2. MYC

The MYC gene encodes for 3 different oncoprotein transcription factors, c-MYC, N-MYC and L-MYC. MYC proteins bind with the protein Max to form a transcription factor complex. The 3 MYC proteins are linked to a huge variety of cancers and are expressed at different levels in different cancers.[15] The following are a few examples of the diverse range of both blood-borne and solid tumours that different MYC proteins appear deregulated in. C-MYC is closely related to Burkitt's lymphoma, appearing in deregulated in the majority of cases.[114] In breast cancer C-MYC amplification was found in 28 % of tumours in a study of 457 primary breast cancers,[115] amplification of C-MYC in breast cancer was also found in an earlier study in 1986, where a 2-15 fold increase in amplification was found in 32 % of 121 tumours.[116] In 1985 a 5-40 fold increase in expression was found in a study of primary human colon adenocarcinoma.[117] N-MYC has been linked to neuroblastoma since 1984, reported as being

overexpressed in a third of cases.[118] More recent studies confirmed these findings. L-MYC, C-MYC and N-MYC have links to small cell lung cancer,[119] with L-MYC being linked as early as 1985.[120] N-MYC and C-MYC expression appears to be highly variable, the frequency of overexpression is reported to range from 2-30 %[121] depending on the nature of the sample tested. The level of overexpression also varies from 5 fold to over 100 fold (reviewed extensively in reference).[121] It has been reported that half all cancers are linked to an amplified expression of MYC (The Myc oncoprotein as a therapeutic target for human cancer). Many of the examples of misregulation of MYC in this section specify dates around the mid-80s as when MYC proteins were determined to be linked to different cancers. MYC proteins are linked to such a large number of cancers and have been recognised as a target for around 30 years, despite this there is a notable lack of small molecule binders for this prolific set of targets. There exists a small number of compounds that target MYC activity by inhibiting its interaction with MAX, either through the direct inhibition of binding or through stabilisation of Max dimers.[122] BET bromodomain inhibitors have also been determined to be inhibit MYC activity by repressing MYC transcription. These methods all avoid direct binding to MYC, an indication of the challenges of directly targeting of MYC proteins. This is partly due to the three MYC derivatives having different structures.
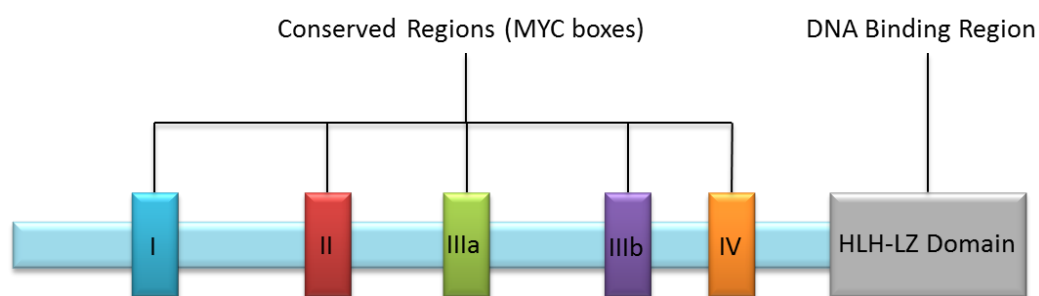
Figure 35 - Visual representation of MYC proteins and the relative locations of their conserved sequences. Note that MYC proteins vary in size and that region IIIa is absent in L-MYC.

Although MYC proteins are different sizes with different sequences several regions are conserved across multiple MYC proteins, these sequences are sometimes referred to as MYC boxes (Figure 35). The DNA binding Helix-loop-helix leucine zipper (HLH-LZ domain) and these five MYC boxes appear in all 3 MYC proteins with the exception of MbIIIa which is absent in L-MYC. As expected these homologous regions are of significant importance to the interactions of MYC proteins. MbI and MbII have been the most widely studied of the 5 MYC

boxes. MbI is known for its accumulation of mutations in Burkitt's lymphoma. It is also the site of phosphorylation which allows for the binding of the ubiquitin ligase SCF[FBXW7] (another WD40 domain containing complex). MbII has been found to be essential for MYC proteins to drive tumorigenesis *in vivo*. MbII is required for the transcription of MYC genes through binding to the Transformation/transcription domain-associated protein (TRRAP). This recruits a histone acetyltransferase complex that opens the chromatin structure of H4 to promote the transcription of genes. MbII is also linked to the degradation of MYC proteins by acting as a binding site for the ubiquitin ligase SCF[SKP2]. Like MbI and MbII MbIIIa is linked to tumorigenesis and is a binding site for ubiquitination complexes. With MbIIIb and MbIV much less is known, though it is known that MbIV regulates DNA binding, apoptosis, transformation, and G2 arrest.[123] Little is currently known about the function of the MbIIIb region is conserved in all MYC proteins, making it particularly attractive for the subject of research in the inhibition of MYC protein activity. As mentioned in the section above it was recently determined that MYC proteins bind with WDR5 via the MbIIIb sequence, with a peptide consisting of the sequence being found to bind with an affinity of 9.3 μM.[12]

5.1.3. Targeting the WDR5-MYC Interaction Interface

Alanine scanning of the sequence reported by Lance *et al.* has revealed several residues that significantly contribute to the affinity of the 11 residue sequence.[12]

Table 15 - Results of an alanine scan of the WDR5 binding MYC peptide. K$_{DS}$ determined from competition vs labelled full length WT peptide in fluorescence polarisation anisotropy.

Name	Sequence	Kd μM
WT	DEEEIDVVSVE	9.3 ± 1.70
D258A	**A**EEEIDVVSVE	19.0 ± 5.80
E259A	D**A**EEIDVVSVE	18.1 ± 3.00
E260A	DE**A**EIDVVSVE	11.2 ± 0.35
E261A	DEE**A**IDVVSVE	19.1 ± 3.85
I262A	DEEE**A**DVVSVE	103.5 ± 11.90
D263A	DEEEI**A**VVSVE	92.3 ± 16.00
V264A	DEEEID**A**VSVE	64.2 15.10
V265A	DEEEIDV**A**SVE	64.4 ± 23.90
S266A	DEEEIDVV**A**VE	19.0 ± 2.10
V267A	DEEEIDVVS**A**E	17.2 ± 4.90
I262G	DEEE**G**DVVSVE	>500
V264G	DEEEID**G**VSVE	>500
WBM	DEEE**ED**EESVE	>500

While the alanine scan suggests the central IDVV residues are the greatest contributors to the interaction, the affinity of the full length peptide is low. By truncating the peptide to just IDVV we would expect to see a very weak affinity that would likely be an unfavourable starting point for the development of a modified molecular probe. The goals of this project are comparable to the work carried out to develop a binder for the top interface of WDR5. First an alanine scan was carried out to determine key residues, then a series of truncates around these key residues were tested. SAR was then carried out on the smallest truncate to develop a molecular probe with a significantly increased affinity.[66, 104, 111]

The articles describing this process were published over the course of 4 years (from truncate screening to optimised structure). The Scanning One Bead One Compound (SOBOC) method described in this chapter was designed to determine a reasonable starting point for the MorPH process, an *in-silico* SAR tool developed by the Auer lab to provide suggestions for possible building block replacements based on crystal structures of peptides. Using MorPH can speed up the SAR investigation process by offering possible suggestions for replacements for amino acids in a peptide starting point based on a crystal structure. However without a reasonable peptide starting point this potentially powerful tool cannot be exploited. This brings a pressure to develop a method of rapidly evaluating the shortened possible peptide inputs, based on known binding sequences, for the MorPH process. The SOBOC synthesis process discussed in this chapter was developed with this goal in mind.

5.1.4. Scanning One-Bead One-Compound (SOBOC) library

Inspired by One-Bead One-Compound (OBOC) combinatorial chemical libraries, a method was designed towards the synthesis of a library of peptide truncates based on the known WDR5 interacting sequence from MYC. OBOC methods produce a library of compounds on the solid phase with each resin bead having only one kind of chemical structure attached. One method of screening OBOC libraries on the solid phase is by confocal nanoscanning (CONA),[124] where fluorescently labelled protein is incubated with a monolayer of library beads. A confocal imaging set up scans the monolayer, exciting and detecting the fluorescently labelled protein. Beads containing hit compounds are visible via fluorescent rings around their periphery due to the higher local concentration of the bound, labelled target protein. OBOC libraries are classically produced through split-mix synthesis, allowing for the rapid

synthesis of thousands of compounds attached to the solid phase in a small number of reactions. In the last round of coupling reactions, it is usual to screen the sub-libraries as opposed to combining them. This aids deconvolution of hits from screening as in the event of a hit bead being found in a sub-library as all compounds in the sub-library experienced the same coupling reaction i.e. the moiety last added to the hit compound is already known. Deconvolution is further aided through the removal of resin after each coupling step, creating a collection of pre-cursor sub-libraries. After a sub-library is determined to contain a hit compound the precursor sub-libraries can be reacted with the known building block of the sub-library that contains the known hit. The resulting series of compounds is screened, the sub-library that contains a hit bead reveals the identity of the penultimate building block, this method can be repeated, working backwards to determine the identity of the hit.[125] This classical approach to deconvolution sees less use in an age where mass spectrometry methods are powerful enough to accurately determine the mass of different products within fractions of mass units, however this method of deconvolution did inspire the development of a SOBOC library synthesis. To synthesise an OBOC combinatorial library solid phase peptide synthesis is used. In the case of Fmoc solid phase peptide synthesis each coupling reaction adds a new amino acid residue to the N-terminus of the previously attached building block. Removal of resin after each amino acid coupling removes an N-terminal truncate of the complete peptide. By synthesising a separate C-terminal truncate of a complete peptide and isolating all N-terminal truncates of each of these C-terminal truncates all possible truncates of a peptide sequence can be isolated with the products isolated as one-bead one-compound.

5.1.5. Screening

5.1.5.1. Confocal Nanoscanning

As discussed above CONA is a method for screening OBOC libraries, invented by the Auer lab. CONA has been applied in the successful hit identification of several PPI targets[126, 127]. However in OBOC library screening each sub-library contains many different compounds on each of the different beads as a result of the split-mix library synthesis methods. As a result, screening a sub-library screens hundreds or thousands of compounds, depending on the library design. In a SOBOC library each aliquot of isolated resin contains only a single peptide. This is by design, as the objective of a SOBOC library is to characterise each of the truncated

peptides individually before evaluating the contributions of different residues towards a binding event. The relative intensity of the fluorescent rings indicating a binding event can be used to relatively rank on-bead peptides by affinity, although on bead affinities are non-linearly different to affinities of the same peptides in solution due to on-bead avidity and enthalpic effects[130]. As the loading on each bead might vary dependant on the sequence, both the protein and the ligand would be labelled using two different fluorescent dyes. This would allow for a ratiometric read out for each measurement, allowing for the accurate comparison of the ratio of ligand to protein as a function of their local concentrations. Synthesising a peptide on Tentagel resin and moving to screening on the same resin makes control of the concentration of peptide difficult, as some reaction series may have accrued more impurities than others. It must also be considered that peptides labelled via the copper-catalysed Huisgen alkyne-azide cycloaddition[27] often require a purification step to remove any unlabelled peptides. There is also the issue of the fluorophore being trapped within the resin bead, giving an erroneous signal in screening. Tentagel resin is also known to have autofluorescent qualities in some screening experiments[131], however for the purposes of CONA this has shown to be less of an issue[132]. A common alternative to Tentagel resin (a polystyrene core coated in polyethylene glycol) is PEGA resin (a polyacrylamide/polyethylene copolymer). While PEGA resin has benefit of being reported to lack autofluorescence seen in some screens using Tentagel[130] it does not solve the problem of unlabelled peptide being present on the resin during screening. It suffers from a lack of mechanical durability, making it unfavourable for synthesis. Therefore, it was decided that rather than screening the peptides isolated on beads the protein would be immobilised instead. Ni-nitriletriacetic acid (Ni-NTA) non-covalently binds His-tagged proteins with a high affinity and is usually utilised in the purification of proteins after expression.

5.1.5.2. Fluorescence anisotropy

Molecules in solution diffuse through the solution at a speed proportional to their size. With regards to its movement through solution a small molecule that is bound to protein will behave as the protein does. Fluorescence anisotropy uses this principle to characterise binding interactions. A fluorescent label attached to a small molecule of interest is excited using plane polarised light of a suitable wavelength. A distribution of these fluorophores will be suitably aligned with the polarised light and be excited. The molecule will tumble through solution

and soon after excitation will emit light at the emission wavelength of the fluorophore. Compared to a compound in solution the rate a protein diffuses through solution is significantly slower, due to the vast differences in size.

5.1.6. Assessment of the binding interactions between WDR5 and MYC peptide

While the top protein-protein interaction interface of WDR5 resembles a binding pocket, complete with a deep pocket that strongly binds arginine, the bottom interface holds no such pocket. A crystal structure showing the interaction between WDR5 and the MYC MbIIIb peptide shows that the peptide sits in a trench along the bottom interface (Figure 34). This interface is largely hydrophobic with a small number of charged residues along the edges on the surface. The screening of peptide fragments of the MYC MbIIIb peptide will help to describe the contributions of residues towards the interaction between MYC and WDR5. There are also *in-silico* techniques available that can be used to predict which peptide truncates will bind with high or low affinity (relative to one another) and which residues are potentially vital to the interaction.

5.1.6.1. Hotspot Determination with Hotpoint

Hotpoint is a hot spot prediction server for protein interfaces.[17] A hotspot is a residue or group of residues in a protein-protein interaction interface that if mutated into an alanine would see a significant drop in affinity. Experimentally this is tested using alanine scanning. As mentioned above an alanine scan of the MYC peptide has been carried out. However Hotpoint analyses contributions from both sides of the interaction, providing information on potential key residues on both WDR5 and the MYCIIIb peptide that are involved in the interaction. These residues could prove to be important guidelines should an SAR project be carried out on the bottom interface of WDR5. Hotpoint considers the solvent accessibility and contact potential of interface residues and uses structures of proteins and their protein or peptide binding partners as input. The server returns the relative complex accessible surface area (relCompASA), the relative monomer surface area (relMonoASA) and the potential of the residue. Should a residue have a potential > 18 and a relative complex accessible surface area of > 20% then it is flagged as a hotspot contributing residue.

5.1.6.2. Docking with FRED

As discussed in chapter 3 Fast Rigid Exhaustive Docking (FRED) is a part of the OEDocking suite (OpenEye scientific http://www.eyesopen.com/). FRED is a method that docks rigid ligands to rigid protein structures. To compensate for the reduced accuracy of docking rigid structures each ligand is converted into a series of different conformations for docking against the target site. FRED then systematically searches rotations and translations of each of the conformations provided within the binding site. Top scoring poses are optimised and a final score is assigned to each of the poses using the Chemgauss4 scoring method. Chemgauss scoring is a force-field method of scoring which recognises shape, hydrogen bonding (both between ligand and protein and with solvent) and metal-chelator interactions. Chemgauss scoring is only concerned with heavy atoms (atoms that are not hydrogen) and hydrogen bonding score contributions better than its predecessors.[133] Chemgauss4 will be used to score peptide fragments based on the MYC MbIIIb-WDR5 interaction structure from literature.[12] In the past the OEDocking suite has been used to discover novel binders through virtual screening.[134]

5.2. Aims

To develop a method for the synthesis of all possible truncates of a set length in a quick and efficient manner, based on known solid phase peptide synthesis techniques. This method would then be tested using the MYC-WDR5 interaction as a starting point. The synthesised library is then to be screened against WDR5 using an on-bead assay. The on bead assay is to use a 2 colour system to allow for a ratiometric measurement of the ligand-protein interaction. In order to carry out this screen WDR5 will be chemically labelled with a suitable fluorescent probe. The results of this screen are to be ranked and the shortest binder showing good affinity to the target is to be further characterised in solution phase K_D determination using fluorescence anisotropy. To use a combination of *in-silico* techniques and the SOBOC peptide fragment synthesis technique to probe the lower interface of WDR5 and assess its potential as a site for possible intervention.

5.3.   Experimental design
5.3.1. Synthesis method and rationale

Fmoc SPPS allows for the synthesis of peptides starting from the C-terminus. Amino acids are added on one residue at a time with the N-terminal protected with an Fmoc group and the side chain protected, this form is referred to as a building block. To add an amino acid, or building block, to elongate the peptide chain on the solid phase there are 2 steps. First the Fmoc group of the peptide is removed with a base, usually with piperidine, then the new building block is coupled to the peptide through the use of a coupling reagent. The nature of this method allows for manipulation in several ways that allows for the synthesis of several peptides sequentially. During the synthesis of a peptide using SPPS all N-terminal truncates (peptides that are shortened versions of the final full length peptide) are produced in the process of making the full length peptide. As the synthesis is produced on a solid phase resin isolation of these truncate intermediates is simply a matter of removing a sample of the resin after each coupling step. C-terminal truncate series cannot be synthesised in a single reaction series on a single sample of resin. The C-terminal is the attachment point to the resin, all chain elongation occurs at the N-terminus. This means that from a single solid phase synthesis only N-terminal truncates can be produced. Each sample of resin is a single C-terminal starting point, from which all N-terminus derivatives with this same C-terminus can be synthesised.

Full length sequence: H_2N-DEEEIDVVSVE-CONH$_2$
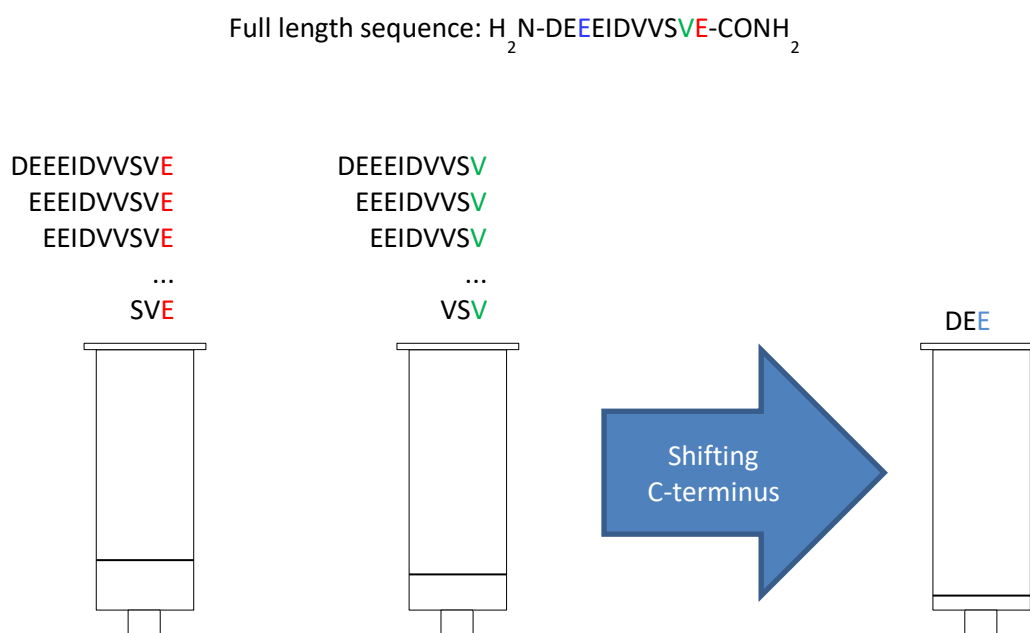
Figure 36 - Visualisation of truncates produced from each C-terminal starting point, each in a different reaction vessel.

Based on these facts it was determined that the synthesis of this complete truncate library would be produced using each of the amino acid residues in the full length peptide as a C-terminal starting point, from which a series of N-terminal truncates would be synthesised from each starting point.

5.3.1.1. Minimum truncate size
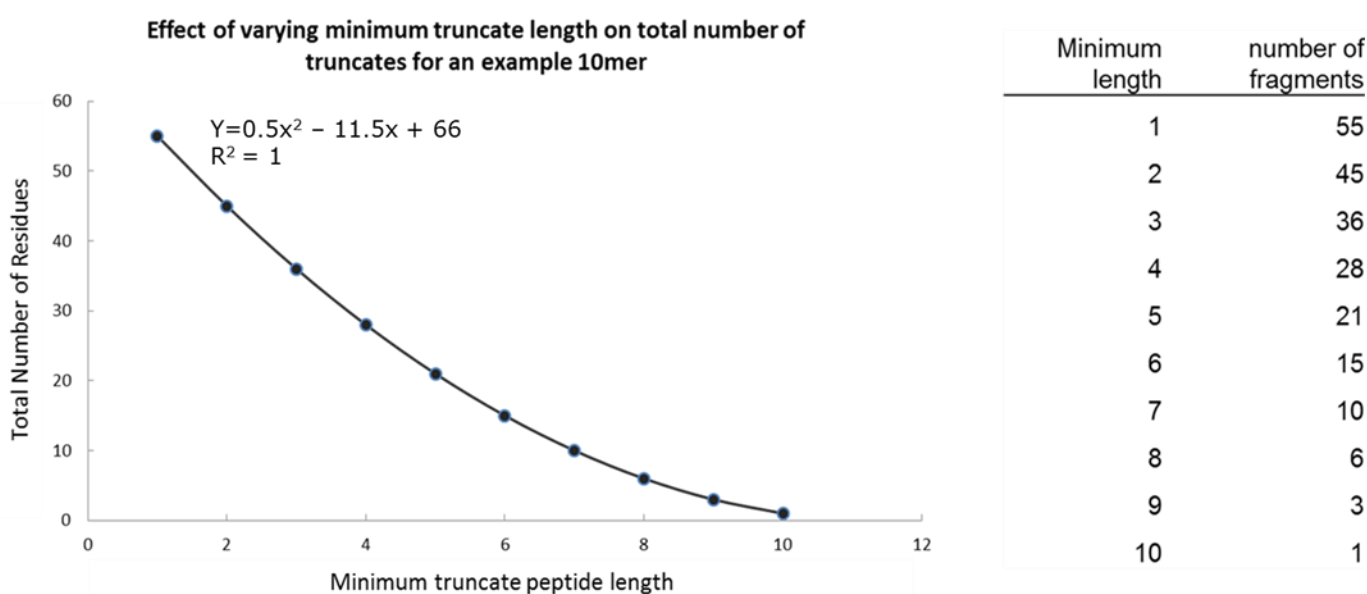
Figure 37 - Total number of peptides requiring synthesis as a function of truncate size. Example values calculated using a full length peptide 10 amino acid residues in length.

Figure 37 shows that the minimum truncate length has a large effect on the total number of truncates in a library. Smaller minimum sizes for truncates dramatically increases the number of truncates that need to be synthesised. From a practical standpoint, every decrease in truncate size by one amino acid residue another reaction vessel has to be used and all other reaction vessels must carry out one additional reaction. Increasing the number of peptides naturally will increase the amount of material required for the complete synthesis and the time required to carry out the synthesis. This includes the cleavage and quality control analysis of the crude peptides. Following this the labelling of the peptides and the purification of the labelled peptides also require an investment of time and materials. These factors would suggest that it would be best to favour a larger minimum size for truncates, however the purpose of the technique is to produce the smallest possible binding sequence. This philosophy requires that we favour the shortest possible truncates. In order to find balance between the ideal scenario and the practical synthesis of so many truncates we must consider

what size of sequence is the most likely to bind. Figure 38 shows a graph comparing pIC50 against heavy atom count, each data point representing a compound from a data set of known protein binding compounds.[135, 136]
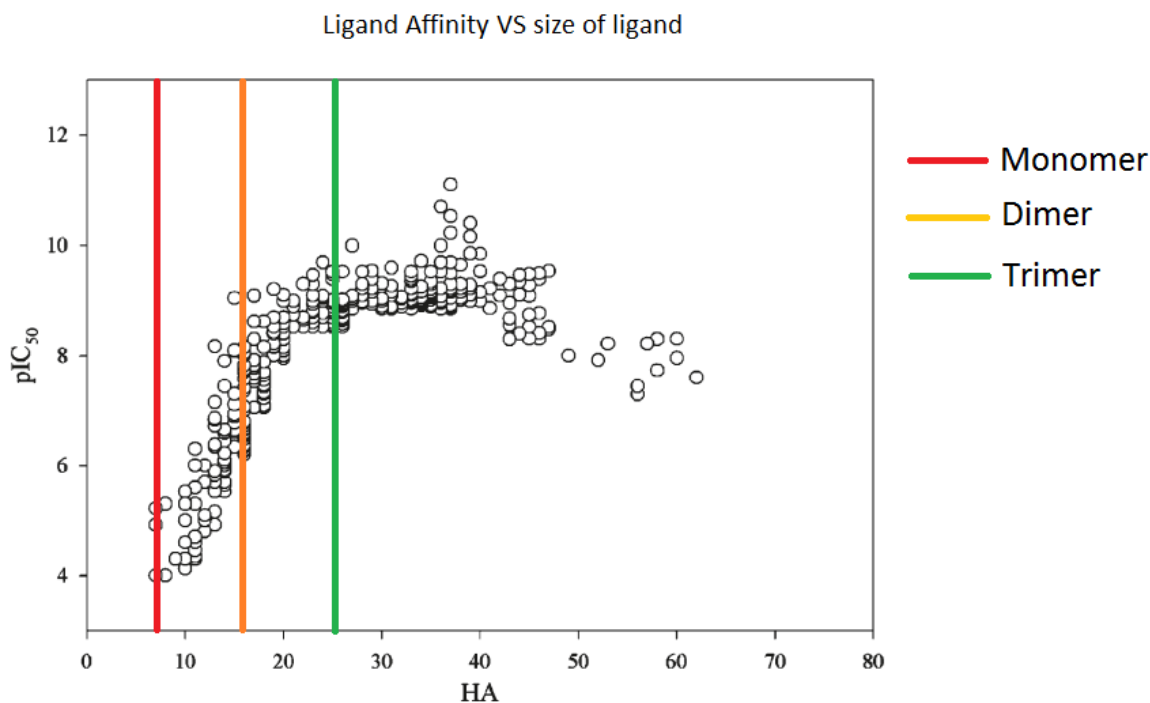
Figure 38 - Graph illustrating the relationship between ligand affinity (pIC50) vs size of ligand (heavy atom). Edited from reference[135]

Figure 38 shows that with increasing heavy (non-hydrogen) atom count ligand efficiency increases up to around 25 HA per compound. The average HA count for an amino acid is 8.3 HA (red line, Figure 38), therefore a trimer contains 24.9 HA (green line, Figure 38) – extremely close to the maximum effectiveness while keeping the total size of the compound small. Figure 39 shows the distributions of length of a peptide motif (red), number of specified positions (positions where the residue is always one of several amino acids) (green) and where the residue is always the same amino acid (blue).
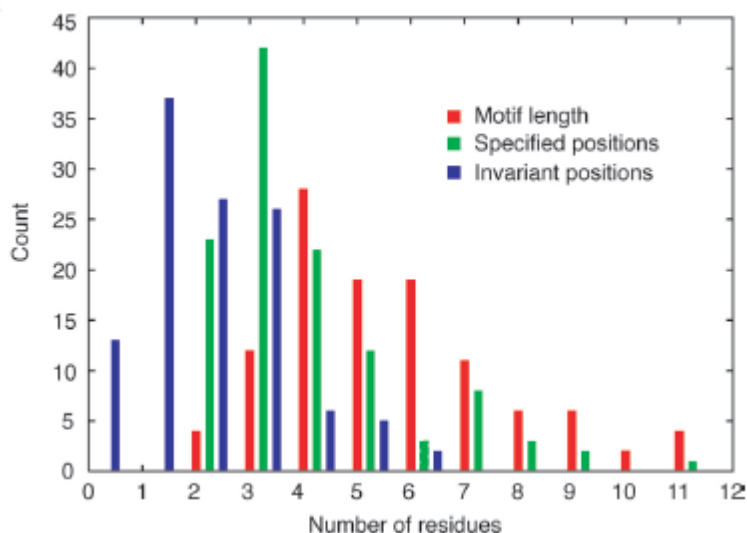
Figure 39 - Frequency of motif lengths, number of specified positions and invariant positions of peptide motifs in proteins. Calculated by DILIMOT[137]

The data in Figure 39 was determined by Discovery of Linear Motifs in Proteins (DILIMOT).[137] DILIMOT is a server that was developed to find short linear motifs in proteins that are over represented. Figure 39 shows that the majority of motifs are between 3-7 residues in length, reinforcing the suggestion of producing truncates of 3 amino acid residues or longer. In the past trimer base compound libraries have produced hits against protein targets, for example trimeric peptoids (peptides where the R group is substituted onto the N) were found to bind to GCPRs with high affinity.[138] The top interface of WDR5 was targeted started from a trimer,[66] and other interactions with WD40 domains have been considered to be dependent on motifs of 3 or 4 amino acids, for example the D-box-CDC20/CDH1 interaction.[139] Based on these points it was decided that the smallest truncate size would be 3. For the targeting of the bottom interface of WDR5 the sequence DEEEIDVVSVE is a known binding peptide from literature and was chosen as a starting point. Using a minimum truncate size of 3 with a full length peptide 11 amino acid residues in length gives 45 truncates synthesised in 9 reaction vessels. In comparison if the minimum truncate length were 2 then 55 truncates would be synthesised, labelled, purified and quality controlled. Some of these dimeric truncated peptides would also be duplicates of one another, adding an element of redundancy to their screening. However, their synthesis would still be required as they are precursors to larger truncates.

5.3.2. Design of the On-Bead Screening of SOBOC peptides against WDR5

2 possible methods for screening were considered. First the classical on-bead library screening approach used by the Auer lab. This method would allow for the rapid determination of hits from the truncate library. However, with so many peptide truncates that are likely to have a similar affinity a 2 colour method could allow for better ranking of hits via a ratiometric measurement, comparing the brightness of a labelled truncate to the brightness of a labelled protein. It was determined that altering the classical on-bead approach and simply labelling the peptides would have problems accompanying the method. Peptide synthesis and labelling methods have high conversion rates, however a purification step is still required to ensure that only the pure labelled peptide is present in screening experiments. This suggests that the synthesis and screening should be carried on different resin. As there is no requirement for a linker suitable for on-bead screening it was determined that the solid phase synthesis was to be carried out on Rink amide linker. To perform the ratiometric measurement it is required that the binding interaction must be localised to ensure the detected signal is strong enough for comparison between different truncates. It was decided that the fluorescently labelled protein would be isolated on Ni-NTA agarose beads, washed, and incubated with the peptide truncates.
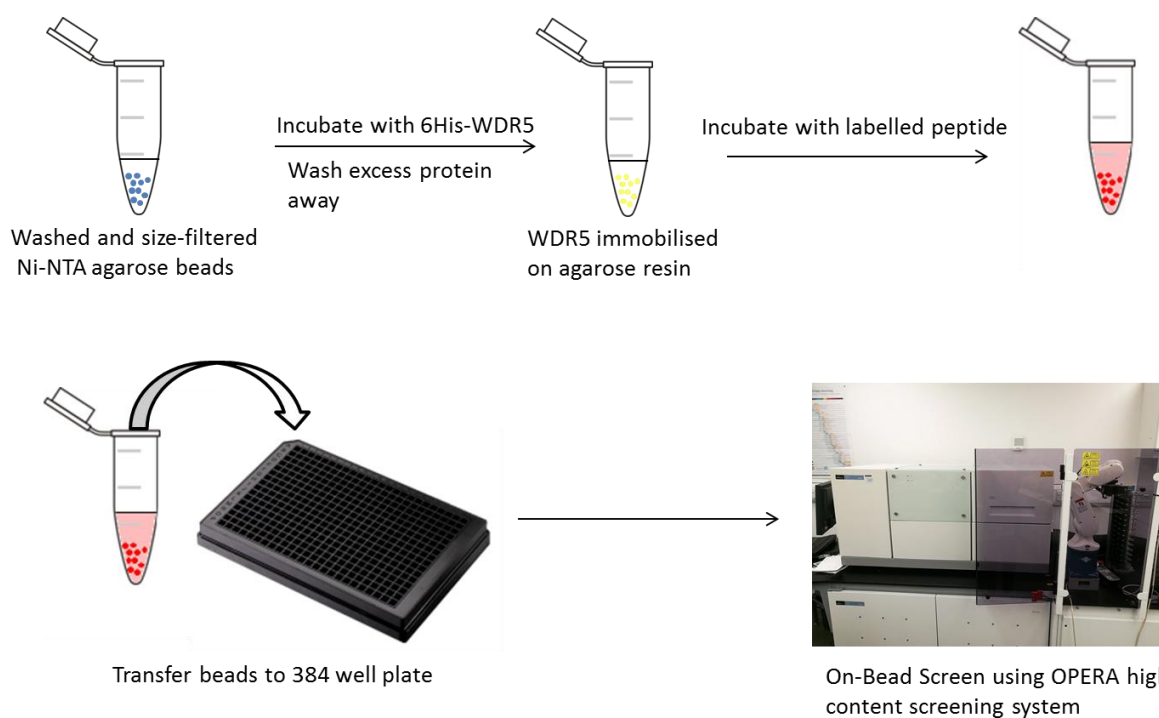
Figure 40 - Illustrated planned workflow of SOBOC library screening protocol using an OPERA screening system (Perkin Elmer).

5.4.   Results

5.4.1. Library synthesis

All truncates were synthesised in 8 working days, average purity was 73 % as determined by HPLC. Generally longer peptides showed a lower purity.
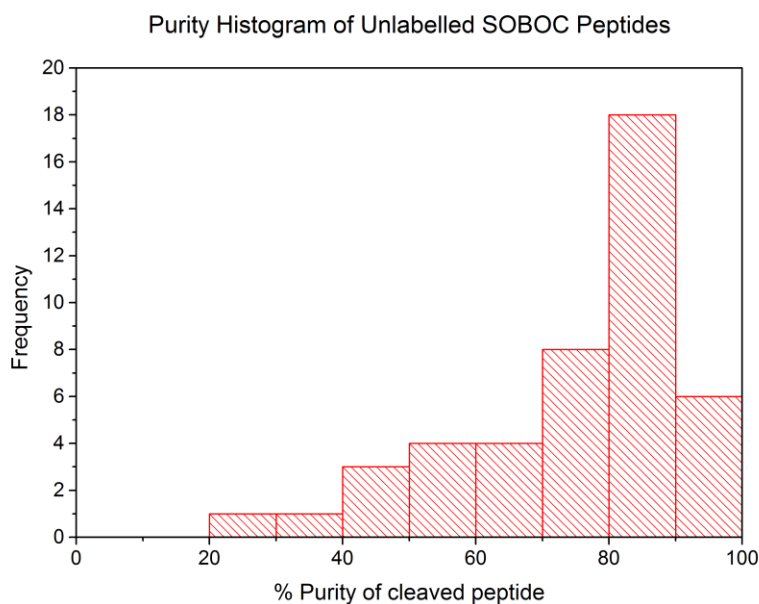

Figure 41 - Histogram of purities of crude peptides from SOBOC synthesis

The majority of peptides were of > 70% crude purity.

5.4.2. Cy5-Labelling of WDR5

To verify the labelling of WDR5 would not interfere with the bottom interface of WDR5 the crystal structure of WDR5 was investigated for possible dye amino acids, Lys and Cys which could be used for labelling with a fluorescent reagent.
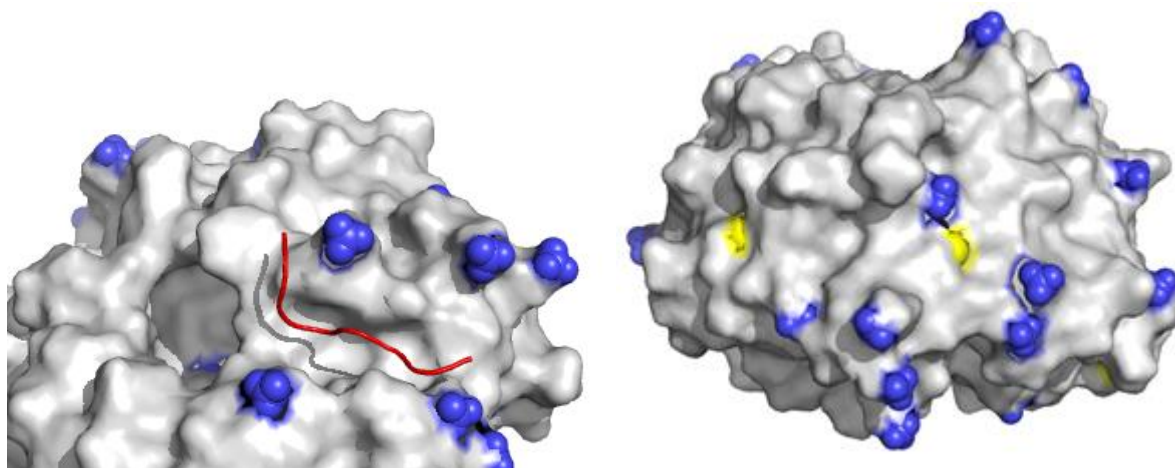
Figure 42 - The bottom interface of WDR5 (white) was determined to have several lysine residues (blue) in close proximity to the MYC peptide binding interface (left image, MYC backbone highlighted in red). Cysteine residues are shown in yellow. The bottom interface holds no surface accessible cysteines.

The bottom interface is surrounded by lysine residues, making labelling with an activated ester such as an NHS ester less favourable compared to cysteine labelling. The protein was labelled with Cy5 maleimide (details in materials and methods section 9.3).
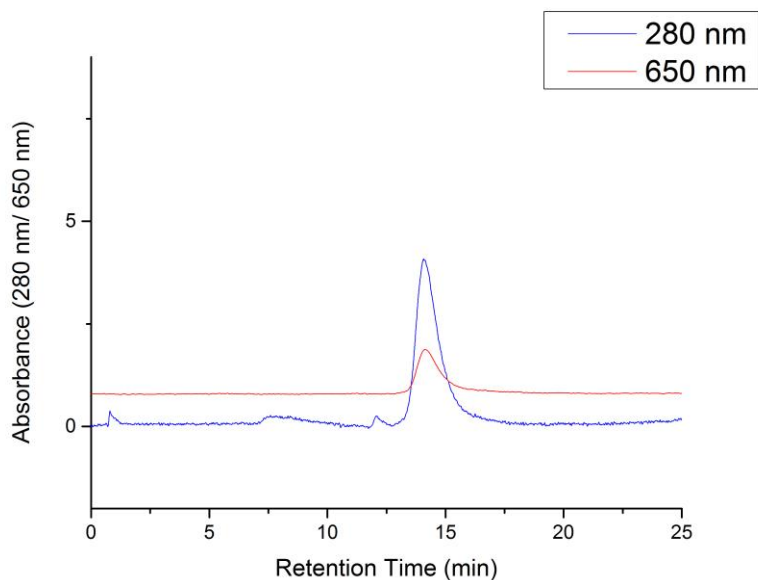
Figure 43 - Overlay of HPLC traces of Cy5-labelled WDR5. 280 nm absorbance and fluorescence intensity (Ex = 650 nm, EM = 670 nm)

The labelled protein was purified by size exclusion and characterised by HPLC and UV-VIS.

The degree of labelling was determined to be on average 1.9 Cy5 molecules per molecule of

WDR5 (calculation in section 9.3).

5.4.3. Synthesis of Labelled peptides

Peptides of lengths 11-7 were labelled with tetramethylrhodamine azide (TMR-N$_3$), thoroughly washed and cleaved. The resulting peptides were purified to > 95 % purity by preparative HPLC.

Table 16 - Results of synthesis and purification of labelled SOBOC peptides

Sequence	Expected Mass	M/Z detected [identity of ion]
DEEEIDV	1613.69	808.23 [M+2H]$^{2+}$ 1614.09 [M+H]$^+$
EEEIDVV	1583.71	792.75 [M+2H]$^{2+}$, 1584.10 [M+H]$^+$
EEIDVVS	1541.70	771.70 [M+2H]$^{2+}$, 1542.82 [M+H]$^+$
EIDVVSV	1511.73	757.18 [M+2H]$^{2+}$, 1512.00 [M+H]$^+$
IDVVSVE	1511.73	756.85 [M+2H]$^{2+}$, 1512.12 [M+H]$^+$
DEEEIDVV	1698.74	850.69 [M+2H]$^{2+}$, 1700.10 [M+H]$^+$
EEEIDVVS	1670.75	836.34 [M+2H]$^{2+}$, 1672.18 [M+H]$^+$
EEIDVVSV	1640.77	821.23 [M+2H]$^{2+}$, 1641.11 [M+H]$^+$
EIDVVSVE	1654.79	821.22 [M+2H]$^{2+}$, 798.49 [M+H]$^+$
DEEEIDVVS	1785.77	893.82 [M+2H]$^{2+}$, 1787.21 [M+H]$^+$
EEEIDVVSV	1769.81	885.74 [M+2H]$^{2+}$
EEIDVVSVE	1769.81	885.78 [M+2H]$^{2+}$
EEEIDVVSVE	1898.86	950.38 [M+2H]$^{2+}$, 1899.17 [M+H]$^+$
DEEEIDVVSV	1900.87	950.38 [M+2H]$^{2+}$
DEEEIDVVSVE	2013.88	1007.84M+2H]$^{2+}$

5.4.4. On Bead and Solution Phase Screening of SOBOC Peptides

5.4.4.1. Screening of Cy5 Labelled WDR5

Well contents	Tetramethylrhodamine (peptide)	Cy5 (labelled protein)
6xHIS-WDR5 beads incubated with TMR-labelled full length MYC peptide		
Cy5 labelled 6xHIS-WDR5 beads incubated with TMR-labelled full length MYC peptide		
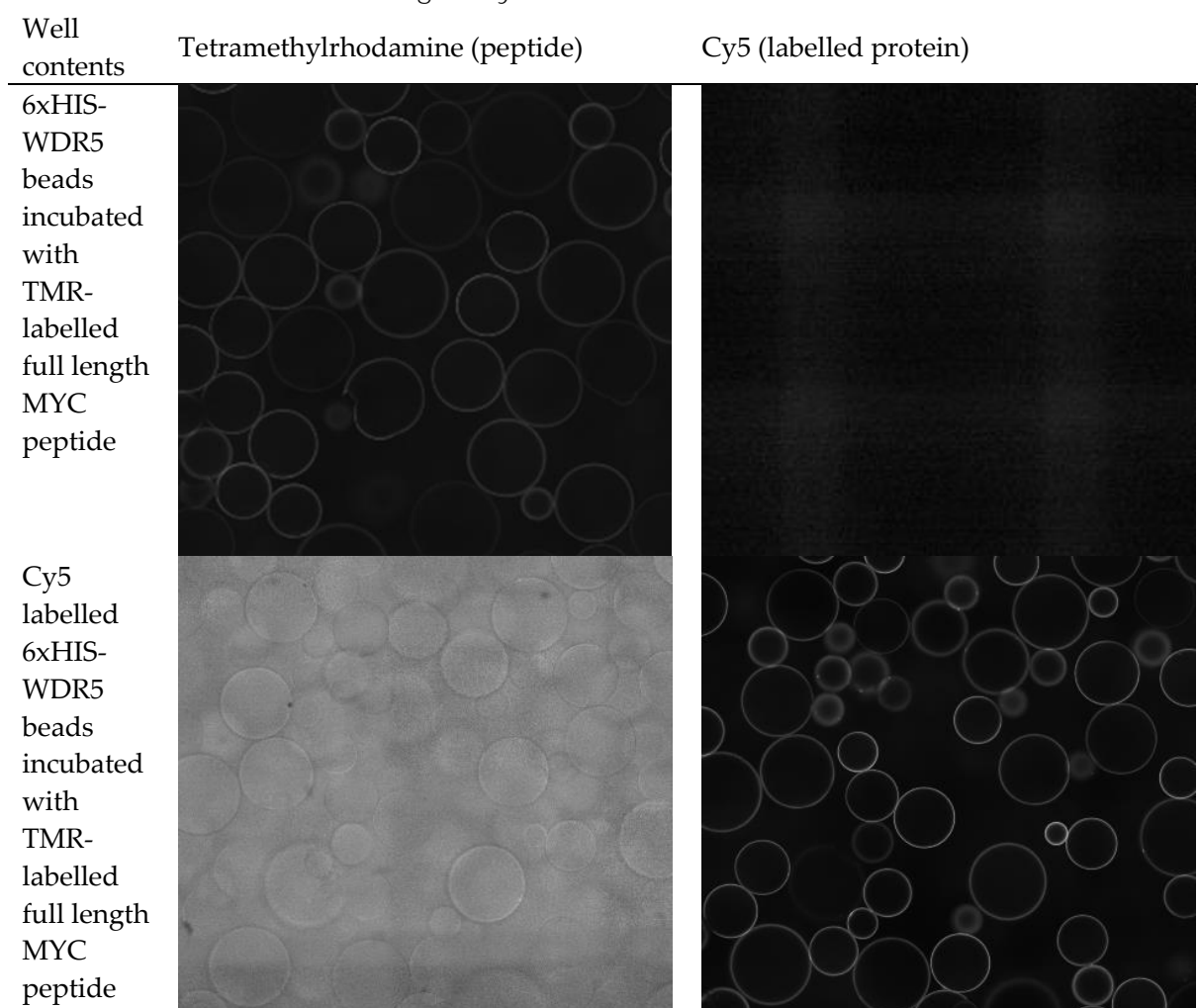
Figure 44 - Comparison of Cy5 labelled WDR5 and unlabelled WDR5 in on-bead binding assay. The Tetramethylrhodamine channel is excited at 561 nm, 1000 μW power, 200 ms exposure time, primary dichroic 445/561/640, detection dichroic 568sp, filter 585/40. The Cy5 channel was excited at 640 nm, 500 μW power, 120 ms exposure time, primary dichroic 445/561/640, detection dichroic 650sp, filter 690/70. The ligand (TMR labelled MYCIIIb peptide) concentration was 500 nM.

Cy5 labelled and unlabelled WDR5 was incubated with filtered Ni-NTA agarose resin (Quiagen) in separate Eppendorf tubes. Beads were then washed with buffer and incubated with 500 nM TMR labelled MYCIIIb peptide and screened on an OPERA unit (Perkin Elmer). The results of this experiment showed that when the Cy5 labelled protein is isolated on-bead there is a very high background signal for the MYC peptide.

Due to the lack of binding seen in the Cy5-WDR5/MYC peptide preliminary experiment further screening experiments were carried out using unlabelled WDR5 protein. While this meant that a ratiometric measurement between protein and ligand was not possible, hits could still be ranked by the intensity of the rings. Ring fluorescence intensity being a function of the concentration of protein-ligand complex.
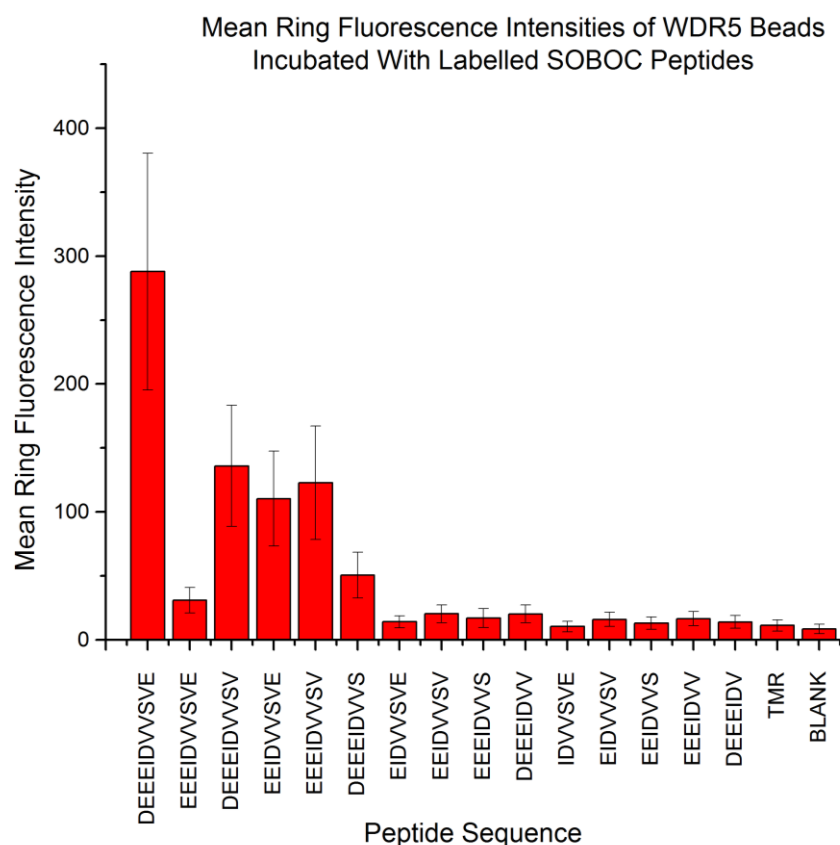

Figure 45 - Mean ring fluorescence intensities of beads incubated with WDR5 and 500 nM of TMR labelled SOBOC peptides. TMR excited at 561 nm, 1000 µW power, 200 ms exposure time, primary dichroic 445/561/640, detection dichroic 568sp, filter 585/40

Ni-NTA agarose resin beads (Quiagen) were incubated with 6xHIS WDR5, washed with buffer and the resin slurry split equally between 17 eppendorf tubes. To each of these Eppendorf tubes a labelled peptide in buffer was added (200 µl of 500 nM ligand), including control samples which contained TMR-N$_3$ or buffer alone. Further control samples were made by incubating the Ni-NTA resin beads without 6xHIS WDR5 with each of the labelled test peptides. The results of the initial screen were analysed using Auer lab proprietary on-bead screening analysis software, BREAD (Bead Ring Evaluation and Analysis of Data). The

analysis showed that some peptides had marginally higher ring intensity than in their corresponding control wells. Specifically, peptide sequence EEEIDVV-DOA-PRA(TMR) was noted as having the highest mean intensity of all 7mer peptides screened, the shortest length of peptides screened. Control experiments showed that the peptides all give similar mean ring fluorescence intensity values when labelled peptides were incubated with Ni-NTA agarose resin in the absence of 6xHis-WDR5.
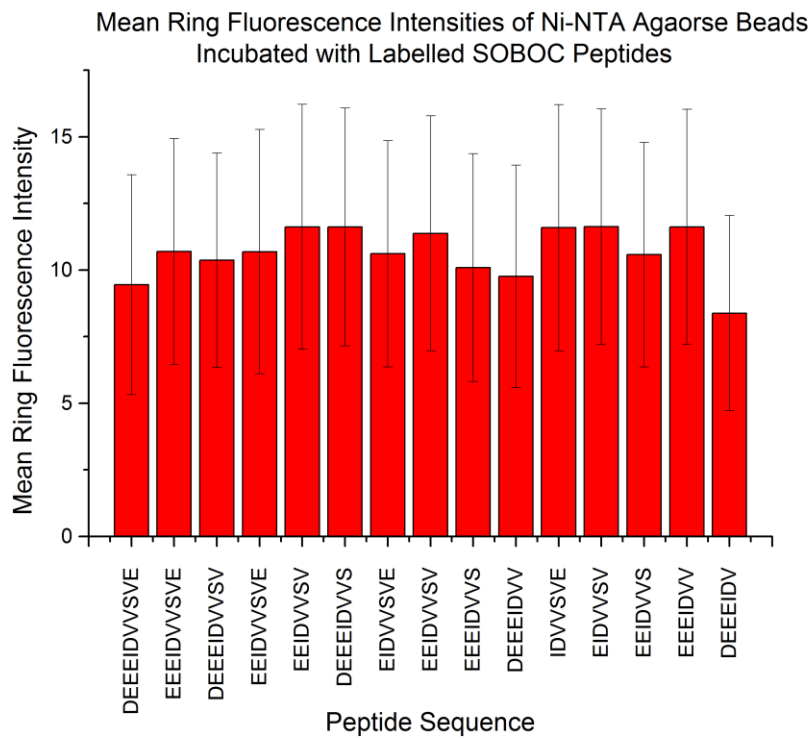

Figure 46 - Control results. Labelled SOBOC peptides incubated with Ni-NTA agarose beads that had not been incubated with 6xHIS-WDR5.

Although EEEIDVV-DOA-PRA(TMR) was seen to have a higher ring intensity compared to other peptides of the same length the intensity is remarkably low. The fluorescence intensity profiles of rings from the CONA screen were analysed using ImageJ software (https://imagej.nih.gov/ij/) and are shown in Figure 47.
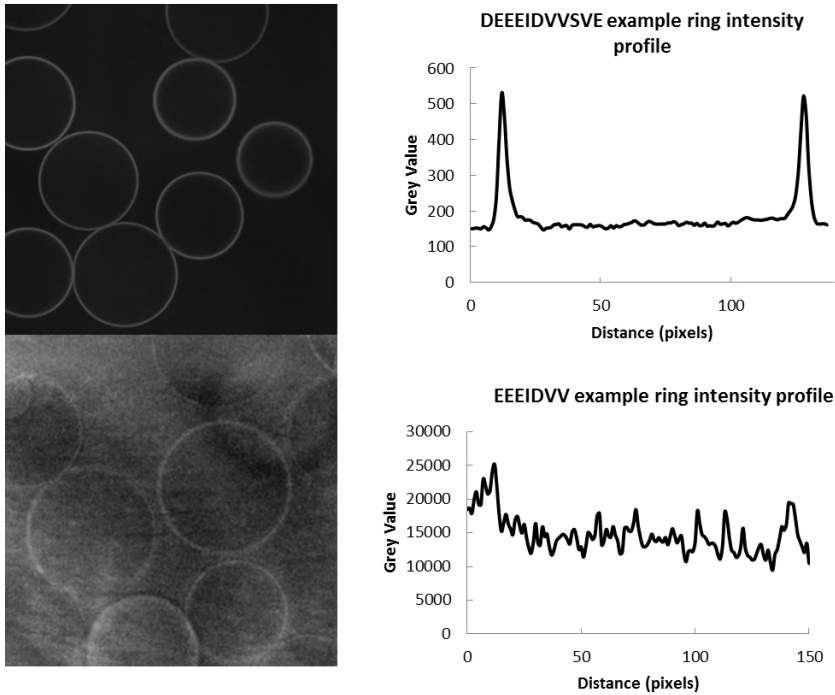
Figure 47 - Example bead profiles. Top - Full length MYC peptide shows a clear increase in fluorescence intensity at the edges of beads. Bottom - The bead profile of EEEIDVV shows no clear bead edge and a high background signal.

Evaluating the ring intensity manually the ring intensity profile is extremely noisy, although looking at the bead image (Figure 47, left) the characteristic rings indicative of a binding event are visible, however the background signal is high. To investigate further the effects of varying concentration of protein and ligand were tested.

5.4.4.3. Effects of Increasing Protein Concentration

As the initial screening experiment suggested a very low affinity for some of the truncated peptides it was first decided to investigate the effects of increasing protein concentration. Increasing ligand concentration would require the introduction of a washing step to avoid increasing the background signal. If the binding interaction between the protein and the peptide was too weak then introducing a washing step could wash the ligand from the sample.
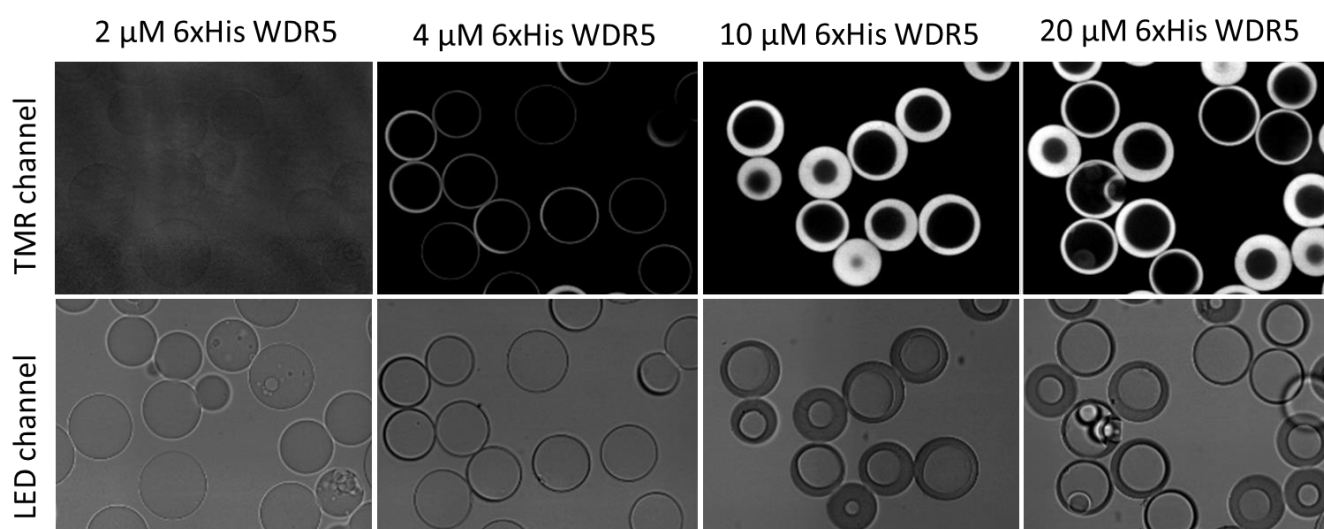
Figure 48 - Effects of increasing concentration of 6xHis-WDR5. WDR5 beads were washed and incubated with 500 nM EEEIDVV-DOA-PRA(TMR). Excess ligand was not washed away prior to imaging. TMR Imaging carried out as in section 5.4.4.2. LED channel: brightfield top illumination, 50 % LED, 200 ms exposure time, detection dichroic 650sp, filter 690/70. Details in section 9.3.

Ni-NTA agarose resin aliquots (3 µl of 50 % slurry) were incubated with 2, 4, 10 and 20 µM of 6xHIS-WDR5 (7 µl), washed with buffer. 500 nM of the TMR labelled EEEIDVV peptide (25 µl) was added to the resin which was agitated and added to a 384 well glass bottom plate. At lower concentrations an increase in fluorescent ring intensity was seen. However increasing the intensity further led to an increase in ring thickness.
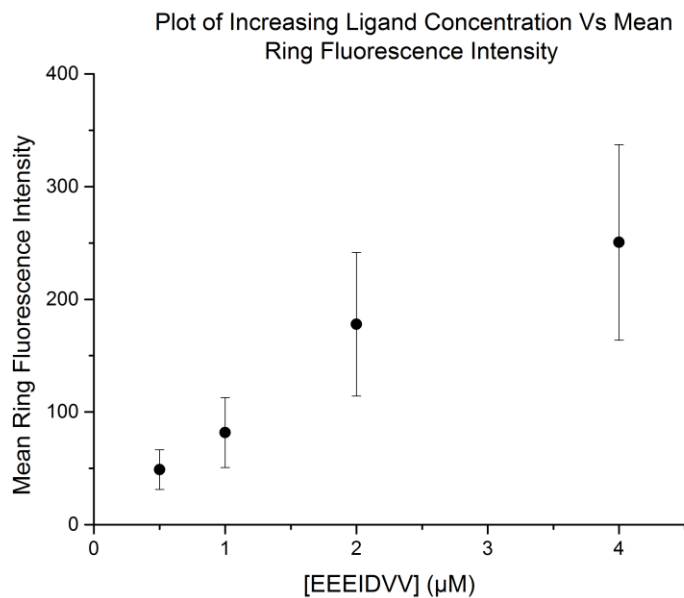
5.4.4.4. Increasing ligand concentration

Ni-NTA agarose resin aliquots (3 µl of 50 % slurry) were incubated with 2 µM of 6xHIS-WDR5 (7 µl), washed with buffer. 500, 1000, 2000 and 4000 nM of the TMR labelled EEEIDVV peptide (25 µl) for 20 minutes. The resin was quickly washed and resuspended in buffer, as the ligand concentration would otherwise be extremely high. The resin was added to a 384 well glass bottom plate and screening on an OPERA unit (Perkin Elmer, Figure 49).
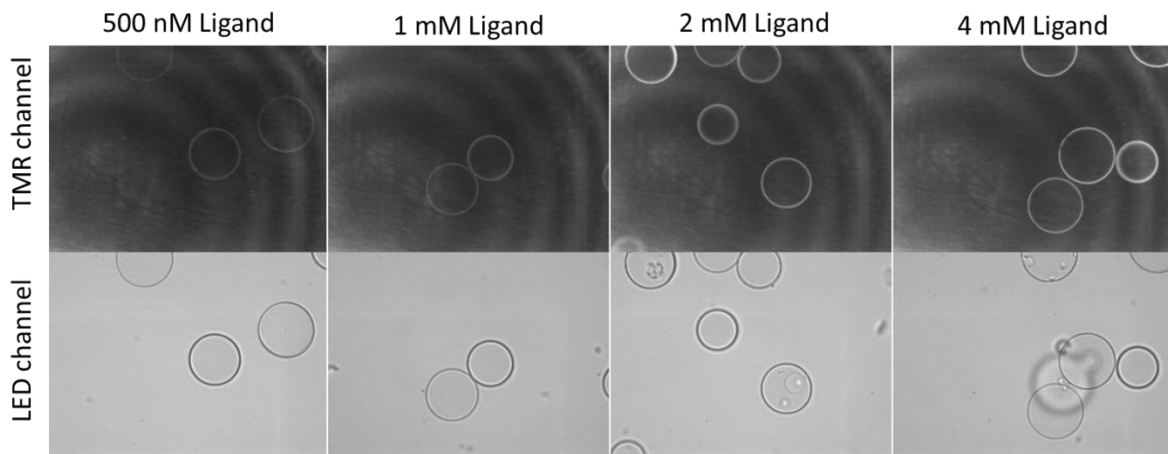
Figure 49 - Effects of increasing ligand concentration in the on-bead screen. Beads were incubated with 2 μM WDR5 and washed prior to incubation with ligand solutions. After incubation with ligand resin was washed prior to imaging. Imaging parameters as described in section 5.4.4.3.

Increasing ligand concentration was shown to lead to an increase in ring intensity. This suggests that the peptide EEEIDVV-DOA-PRA(TMR) binds to WDR5 in these on-bead assays. Figure 50 visualises this result as a plot, showing that the mean ring fluorescence intensity is concentration dependent.

Figure 50 - Plot visualising the effect of mean ring fluorescence intensity vs ligand concentration.

Due to the significantly increased concentration of ligand, controls were repeated to verify that any perceived binding interactions were not related to the TMR label portion of the ligand binding to WDR5 or either the peptide or label binding with the Ni-NTA agarose resin.
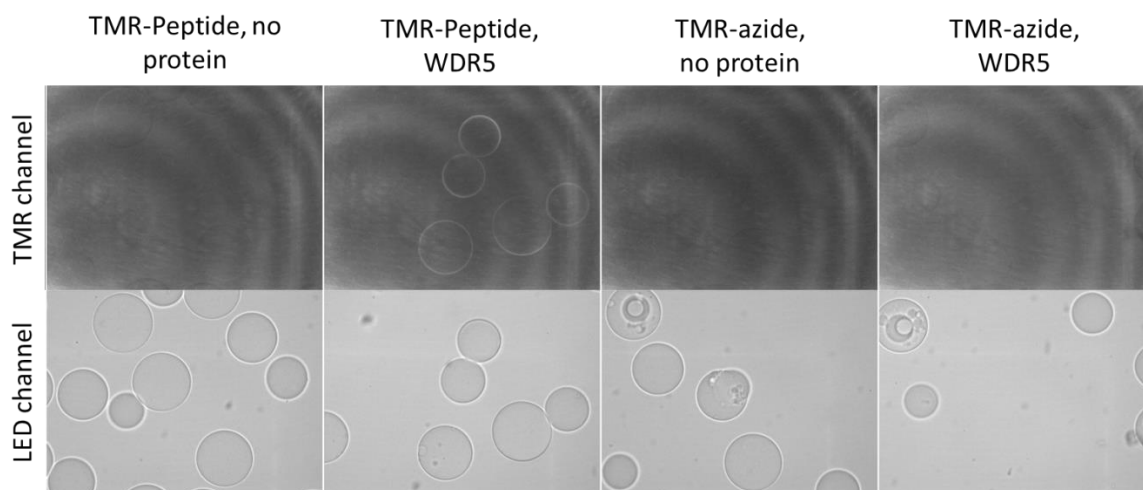
Figure 51 - Results of control experiments using 4 µM ligand concentrations and 2 µM WDR5. Only the TMR-labelled truncate is visually confirmed to show the characteristic binding rings. Screening parameters as in section 5.4.4.4.

In the TMR channel only the well holding TMR labelled peptide and WDR5 incubated beads was shown to have a ring intensity. For comparison of intensities between samples see Figure 52.
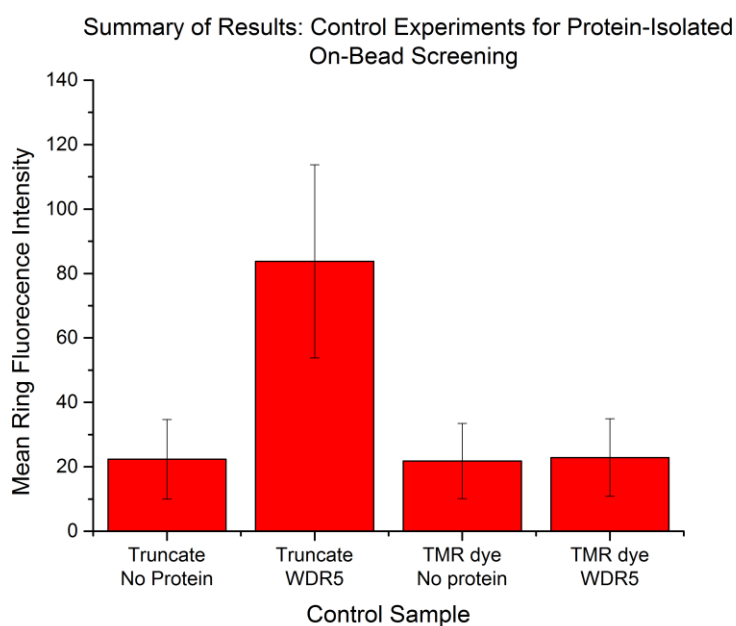
Figure 52 - Results of numerical analysis of the on-bead control experiment. TMR labelled truncate peptide (EEEIDVVSVE) screened against 6xHIS-WDR5 immobilized on NI-NTA resin.

These controls showed very low signals for negative controls compared to the well containing the test truncate, EEEIDVV-DOA-PRA(TMR). Control experiments all showed a low mean ring fluorescence intensity around 20 units, while EEEIDVV-DOA-PRA showed and intensity value 4 times higher.

5.4.4.6. Planning of Flu Anisotropy Analysis of EEEIDVV-DOA-PRA(TMR)

OriginLab version 9.0 was used to simulate a binding curve. The plot assumed a K_D of 100 μM (an order of magnitude weaker than the literature affinity of the MYC MIIIb peptide[12]) and a ligand concentration of 50 nM. This was used to determine the range of concentrations of protein which should be used in a titration experiment.
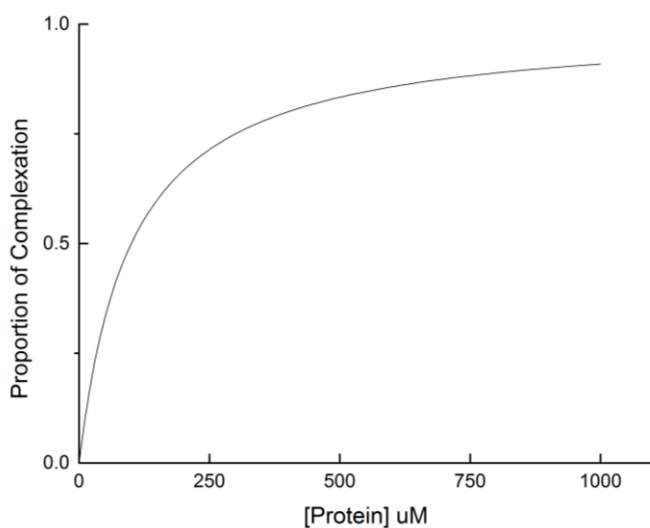

Figure 53 - Simulated curve for fluorescence anisotropy experiment. Assuming a K_D of 100 μM and a ligand concentration of 20 nM.
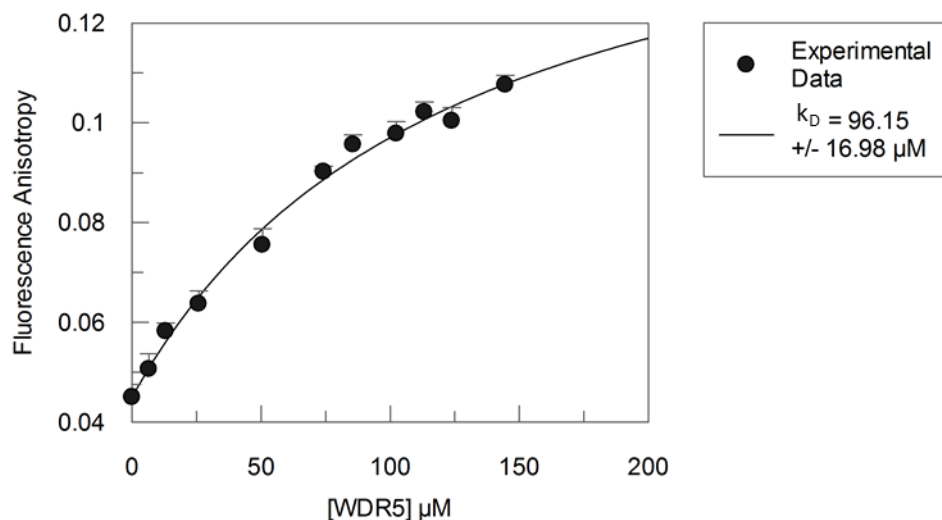
Figure 54 - Plot of fluorescence anisotropy with increasing concentration of WDR5. 50 nM ligand concentration, slits: 4 and 6 nM. Windowed 200 μl quartz cuvette.

K_D determination was carried out in a 200 μl windowed cuvette with a starting volume of 175 μl and ligand concentration of 50 nM. The end point was not reached at the concentration of protein titrated but was determined to be 0.14 by GraFit (Erithacus Software) software using Equation 15 in materials and methods.

5.4.5. **In-silico** analysis of bottom binding interface of WDR5

5.4.5.1.   Results of OEDocking suite analysis of MYC peptide fragments (carried out by Steve Shave)

Peptide fragments from literature structural data was used as input for FRED, scored by Chemgauss4 (Table 17). Lower scores equate to a higher expected affinity between fragment and protein.

Table 17 - Results of docking fragments of MYC peptide with the bottom interface of WDR5. Left: sorted by score. Right: Sorted by fragment length and score (lower values equate to higher predicted affinity).

Sequence	Chemgauss4 Score	Sequence	Chemgauss4 Score
IDVV	-4.530	**Length = 3 residues**	
IDV	-4.515	IDV	-4.514
IDVVS	-3.695	DVV	-2.690
EIDVV	-3.482	VVS	-2.490
EIDV	-3.466	EID	0.032
DVV	-2.690	EEI	1.474
EIDVVS	-2.646	VSV	2.174
VVS	-2.490	**Length = 4 residues**	
IDVVSV	-2.481	IDVV	-4.530
DDVS	-1.854	EIDV	-3.466
EIDVVSV	-1.433	DVVS	-1.854
VVSV	-1.277	VVSV	-1.277
EEIDVV	-1.276	EEID	2.237
EEIDV	-1.261	**Length = 5 residues**	
DVVSV	-0.640	IDVVS	-3.695
EEIDVVS	-0.441	EIDVV	-3.482
EID	0.032	EEIDV	-1.261
EEIDVVSV	0.773	DVVSV	-0.640
EEI	1.474	**Length = 6 residues**	
EEI	1.474	EIDVVS	-2.646
VSV	2.174	IDVVSV	-2.481
EEID	2.237	EEIDVV	-1.276
		Length = 7 residues	
		EIDVVSV	-1.433
		EEIDVVS	-0.441
		Length = 8 residues	
		EEIDVVSV	0.773

The results of this docking exercise appears to favour shorter sequences, which is in opposition to the experimental data. However the central sequence IDV appears in the majority of highly ranked peptides.

5.4.6. Results of Hotspot Determination by Hotpoint Server

Table 18 - Output from the scoring of residues in the interaction interface by the hotpoint webserver[17]. Predicted hotspot residues highlighted in bold.

Interaction partner	Residue Number	Residue ID	RelCompASA	RelMonomerASA	Potential	Prediction
WDR5	181	R	55.36	66.66	3.70	NH
	223	S	3.12	7.96	14.51	NH
	224	P	39.43	55.67	10.75	NH
	225	N	6.61	49.19	12.91	NH
	228	**Y**	**13.78**	**33.89**	**28.33**	**H**
	240	**L**	**0.00**	**5.55**	**62.96**	**H**
	249	L	44.77	53.34	26.90	NH
	250	K	7.34	28.07	9.29	NH
	266	**F**	**7.83**	**11.10**	**38.86**	**H**
	268	V	30.03	44.61	14.34	NH
	272	K	38.62	48.97	4.38	NH
	288	**L**	**0.43**	**14.56**	**38.85**	**H**
	289	Q	21.57	59.25	10.66	NH
MYC	260	E	110.12	122.29	5.66	NH
	261	E	83.11	98.75	3.45	NH
	262	**I**	**8.81**	**77.99**	**32.93**	**H**
	263	D	51.58	66.76	4.97	NH
	264	**V**	**1.21**	**86.70**	**40.24**	**H**
	265	V	40.41	79.82	12.24	NH
	266	S	48.74	66.76	4.04	NH
	267	V	87.06	134.50	8.65	NH

Table 18 shows the predicted hotspot contributing residues for WDR5 and MYC based on the structural information from PDB entry 4Y7R. RelCompASA (relative complex accessible surface area) refers to the accessible surface area of the residue in complex. RelMonomerASA (relative monomer accessible surface area) refers to the accessible surface area of the residue when the protein exists out of complex. Prediction refers to non-hotspot (NH) or hotspot (H). The output of the hotspot determination agrees with the FRED docking that MYC residues I262 and V264 are key contributors to the WDR5-MYC interaction. Unsurprisingly it also suggests that residues on WDR5 that are near to these key MYC residues are also hotspots.

5.5.   Discussion and Future Work

5.5.1. Library synthesis

This first synthesis of a scanning truncate library showed that the concept succeeds in producing all possible truncates in a timely fashion, 9 parallel reactions in place of 45, however due to time constraints the shorter peptides were not labelled or screened against WDR5. As length increased, purity decreased, an unfortunate result of the sum of multiple small impurities that were added during each coupling. However, these lower purities became of less relevance, as the truncated peptides were purified by preparative HPLC labelled. After the peptides were labelled a purification step was required prior to analysis and screening. A possible cause of the decreased purity is the removal of resin from each reaction vessel after an amino acid coupling step leading to discrepancies between the expected mass of resin in the sample. After each step the resin was dried via vacuum overnight to remove as much solvent as possible. The following day resin was weighed, an aliquot removed and the bulk of the resin weighed again in order to calculate the amounts of different reagents required for the next coupling. However, it is possible that the drying was not thorough enough and the resin was not completely dried prior to weighing. The inaccuracy between the weighed mass of the resin and actual mass due to the solvent would have led to the use of a larger equivalence of coupling reagents being used per coupling, this possibly led to the lower purities of longer peptides due to the accumulation of the effects of erroneous couplings. In a future experiment using this method of truncate synthesis it would be prudent to increase the drying time of the resin, use a more powerful vacuum or wash the resin with a more volatile solvent prior to drying.

For comparison to this method of library synthesis quotes from peptide suppliers were gathered. Quotes were generally above £700 and had constraints on minimum peptide size and minimum number of peptides. For example, one supplier limited the minimum peptide length to 6 amino acid residues and a minimum of 24 peptides per order. However using MYC as an example removing all truncates smaller than 6 amino acid residues in length draws the total number of peptides to below 24. As discussed previously there is a great deal of evidence to suggest that motifs of 3 residues in size can be key contributors to interactions between proteins. Additionally, the synthesis method described in this chapter can be carried out by

any researcher with a basic knowledge of solid phase peptide synthesis and offers complete control over the range of truncate sizes incorporated into the library.

5.5.1.1. Future Work

Future work would be to validate the synthesis method further by applying the truncation method to other peptides against other targets. In the work carried out in this chapter only the minimum truncate length was considered, however the maximum length of the peptide truncates could also be incorporated into the synthesis. If a long sequence, for example 20 or more amino acid residues, was to be broken down using this method of truncation it would lead to over a hundred different truncates to synthesise, label and quality control. With the goal of this method to be the determination of a suitable starting point for further work many of the longer truncates would be of little use but their production would increase time and cost. In the experimental design section of this chapter the occurrence of different lengths of interaction motifs in nature was discussed (Figure 39), with motifs of a size 4 appearing the most frequently and the frequency decreasing greatly around 7-9 residues in length. Future truncation experiments could be made more efficiently by limiting the range of truncate sizes to a more pragmatic window, for example 3-9 amino acids in length.

5.5.2. Screening experiments

The control screens of Cy5 labelled WDR5 showed no binding to the full length peptide, despite the work undertaken to show that there should be no interference with the bottom interface site by a dye. A search of the literature shows that generally efforts to screen WDR5 utilise a fluorescently labelled antibody as opposed to direct labelling of the protein. This could be an indication that other researchers have seen a similar result when trying to work with labelled WDR5. The bottom interface also has many lysine residues, making labelling with an activated ester unfavourable as labelling close to the binding interface may affect any binding interactions. While the screening methods described in this chapter allowed for a ranking of the truncates of the MYCIIIb peptide, the issues with labelling WDR5 are an important consideration for future work. For example, it would have made screening with the peptides unlabelled, bound on bead impossible. The use of fluorescently labelled antibodies is a reasonable solution and has a literature precedent for use in experiments investigating WDR5.

Increasing the concentration of ligand or protein resulted in an increase in the ring fluorescence intensity. This allowed for the confirmation of the activity of a weak binder to WDR5 in the on-bead assay. Increasing the incubation concentration for protein loading above 10 μM saw an increase in the internal fluorescence of the agarose resin during imaging. This indicates that the concentration of protein on the bead surface was increasing but that the ligand was diffusing further into the resin matrix. By comparison increasing the ligand concentration saw only an increase in ring intensity. When designing the experiment, a concern was that lower affinity binders would dissociate from WDR5 after washing, but this was proven not to be the case.

The results of the on-bead truncate screen suggested that there was a significant decrease in affinity of the labelled truncates of a length shorter than 9 residues. The mean ring fluorescence intensity decreases as peptide truncate length decreases (Table 19).

Table 19 - Labelled SOBOC peptide screening results sorted by mean fluorescence intensity (highest to lowest).

Length	Sequence	Mean fluorescence intensity	Standard Deviation
11	DEEEIDVVSVE	287.846	92.467
10	DEEEIDVVSV	135.988	47.434
9	EEEIDVVSV	122.799	44.312
9	EEIDVVSVE	110.397	37.035
9	DEEEIDVVS	50.592	17.842
10	EEEIDVVSVE	30.969	9.901
8	EEIDVVSV	20.331	7.055
8	DEEEIDVV	20.219	6.967
8	EEEIDVVS	17.169	7.389
7	EEEIDVV	16.595	5.539
7	EIDVVSV	16.027	5.491
8	EIDVVSVE	14.156	4.685
7	DEEEIDV	14.014	4.934
7	EEIDVVS	13.099	4.759
7	IDVVSVE	10.408	4.084

A notable unexpected result is the 10 amino acid peptide EEEIDVVSVE-DOA-PRA(TMR), despite only missing the N-terminal aspartic acid residue the mean bead ring fluorescence intensity is significantly lower than similar truncates. This could be an indication that some characteristic of the peptide is having a negative effect on its affinity, however the DOA spacer and TMR label does not appear to have such a large effect in any of the similar sized peptides. In these experiments it was found that truncates of a length of 7 amino acids were binding,

but weakly. However, when analysed, both numerically and by eye, the rings were not convincing, showing very weak brightness even compared to the control wells that contained beads with no WDR5 on bead. To clarify whether the result was evidence of a true binding interaction experiments varying the concentrations of the interaction partners was carried out. First the concentration of the protein incubated with the Ni-NTA resin was changed, this saw an increase of in the thickness of the rings. This suggests that rather than seeing an increase in the surface concentration of the protein on the bead the protein instead diffuses further into the resin. Next the concentration of the ligand was varied, increasing the concentration as high as 4mM and introducing a washing step. Initially there was concern that if the ligand is too weak a binder then after the washing step the protein-ligand complex would simply dissociate and there would be no visible binding of the ligand to the resin bound protein. Experimentally it was shown that this was not the case and rather that increasing ligand concentration led to increasing fluorescence intensity localised around the resin. The increased concentration of protein-ligand complex that gave the increased ring intensity is possibly a result of the high localised concentration of protein. As the ligand dissociates from the protein it is in an environment of high protein concentration and is more likely to bind to another protein. This is likely the same reason that His-tagged proteins remain bound to Ni-NTA resin after the repeated washing steps this screening protocol required. This experiment suggested that the interaction between the ligand and the protein was concentration dependent. To confirm this, the fluorescently labelled peptide was screened against WDR5 by fluorescence anisotropy. It was determined that the labelled truncate had an affinity of 96 μM to WDR5, for comparison the full length peptide was determined to have an affinity of 10 μM. Interestingly the sequence of the truncate is similar to a motif seen in several binders to the bottom interface of WDR5.[12]

Table 20 - Example binders to the bottom interface of WDR5 and their binding motif. "IDVV" box in red, conserved negative residues in blue.

WDR5 interacting protein	Interaction sequence
KANSL2	SDDLDVVGDG
RBBP5	DEEVDVTSVD
C-MYC	EEEIDVVSVE
Truncate	EEEIDVV

While alanine scanning suggested that the IDVV sequence was mandatory for binding the negatively charged residues of the sequence, across several different proteins, N-terminal to

IDVV gives weight to the suggestion that they have an important role to play in the binding interaction. The fluorescently labelled truncates screened in this series of experiments were determined to have higher affinities when C-terminal residues were still present (Table 19).

5.5.2.1.   Future Work

Future work using this bead based assay would likely involve screening a concentration matrix, varying the concentration of both ligand and protein. This would allow for a clearer characterisation of the method and what range of affinities is detectable using the method. Validating a 2 colour ratiometric measurement method would allow for the definitive ranking of peptides with respect to their affinities towards the target of the screen. In the case of WDR5 this could possibly be achieved using labelled antibodies. Further investigation into the low affinity of EEEIDVVSVE-DOA-PRA(TMR) would be a necessary follow up experiment, either repeating the on-bead screening or taking it through to fluorescence anisotropy to verify that the affinity is as low as perceived in this initial screen and not due to some sort of error. Another possible follow up project to this work would be to use the truncate EEEIDVV as input into the MorPH process. First the software would use the known crystal structure of the MYC peptide as input. Alternatively, and ideally, a new crystal structure would be obtained with the help of a collaborator using the truncate to construct a co-crystal structure. It is possible that the truncate has a slightly different binding mode compared to the full length peptide. Next the MorPH software would replace single amino acids of the binding peptide with commercially available building blocks, docking them and ranking them based on their score in the docking experiment. Several peptides based on these suggestions would be synthesised and screened against WDR5. Ideally the process would be carried out in multiple iterations, the strongest binder from the initial series of MorPH peptides would be incubated with WDR5 for a crystal structure. This new structure would be used as input for another MorPH and lead to another set of suggestions from the software. This would be to offset the changes in backbone structure and possibly positioning in the binding interface by the new replacement residue. The sequence EEEIDVV-DOA-PRA(TMR) has been determined to have an affinity of 96 μM to WDR5 and could be used as a starting point for the MorPH process. Further characterisation of the MYC-WDR5 interaction could be carried out, primarily further screening of truncates. This would allow for a more thorough understanding of the contribution of residues to the interaction between the MYC sequence and WDR5. While

useful there is an argument to be made that testing smaller peptides is of limited value, as they will be expected to be of lower affinity of the peptides screened in this study.

5.5.3. Analysis of the interaction interface

The OEDocking suite and the Hotpoint online service were used to evaluate potential binding contributions of different residues of the MYC peptide. These different methods of predicting the contributions to a binding interaction were in agreement in their assessment of the WDR5-MYC interaction.

5.5.3.1. OEDocking Suite (FRED, scored by Chemgauss4)

The FRED docking software was used to evaluate the expected affinities of the truncates produced by SOBOC. Interestingly the scoring algorithm determined that the sequence IDVV was the highest scoring out of all truncates. This is likely due to IDVV in the original sequence sitting deeper within the hydrophobic trench on the bottom interface of WDR5 than other residues in the sequence (Figure 55).
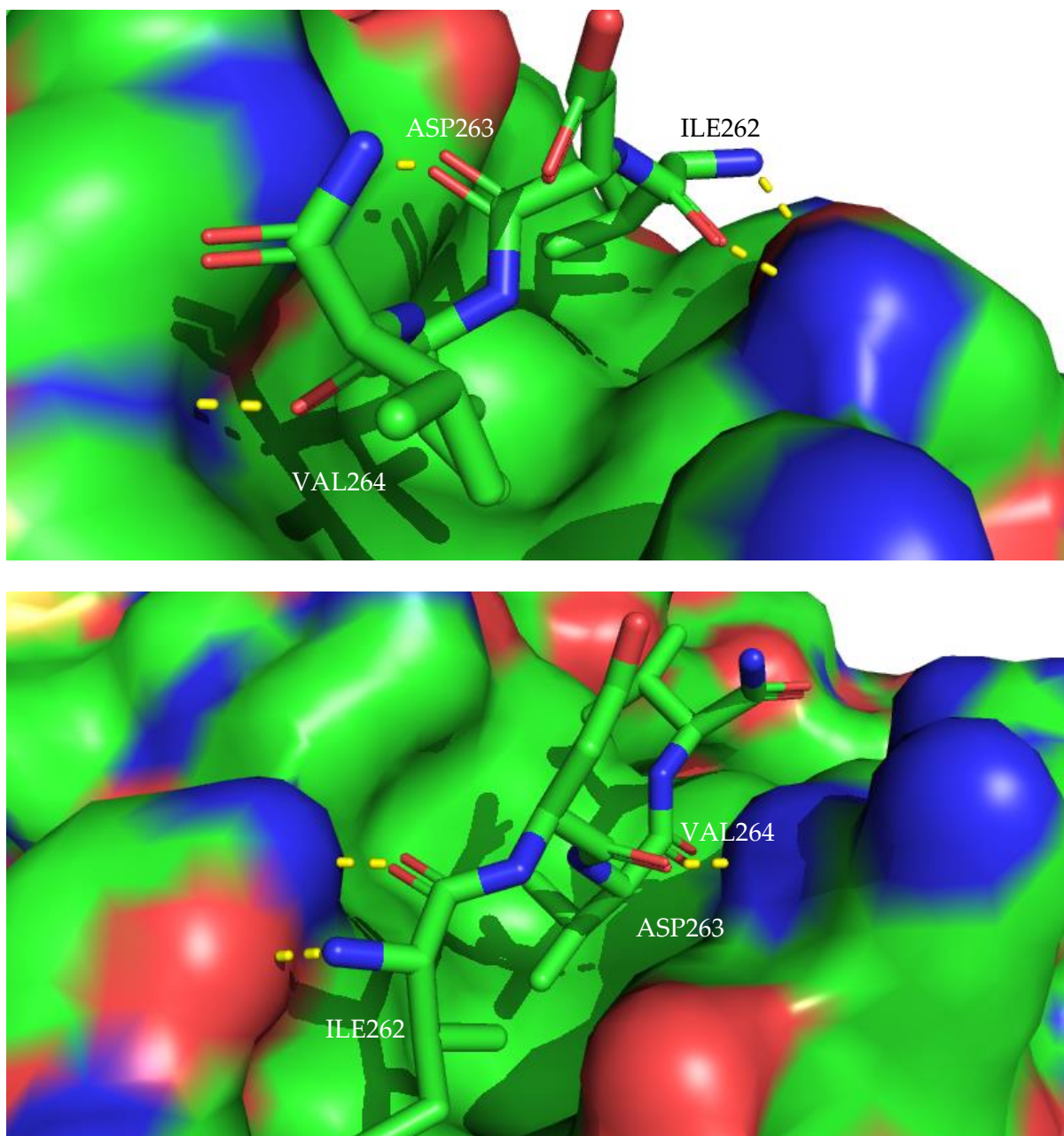
Figure 55 - H2N-IDVV-CONH2 in the WDR5 binding interface. Top: viewing from the C-terminus of the IDVV peptide. Bottom: viewing from the N-terminus of IDVV

Despite the binding interface being hydrophobic in nature the structural information available suggests that the backbone of the IDVV portion of the MYC MbIIIb peptide finds several possible polar contacts. ILE262 is predicted to interact with both its N and C terminal residues, both ASP263 and VAL264 are predicted to interact with the interface via their carbonyl groups. The residues ILE262 and VAL264 appear in all of the top 5 results scored using the Chemgauss4 algorithm. The results from Hotpoint are in agreement with the output from FRED, suggesting that ILE262 and VAL264 are key to the interaction with the bottom face of

WDR5. The results of an alanine scan also suggested that IDVV was core to the interaction between MYC and WDR5. However, it also showed that alteration of any residues of the peptide has a negative impact on affinity, even in the cases of sidechains that appear to be pointing into solvent and contributing little to the interaction. Using Pymol to highlight predicted polar and hydrogen bonding interactions for the full length peptide shows that none of the polar side chains are predicted to interact with WDR5. Measuring the distance between the polar sidechains of the peptide and protein gives a distance of 4.5 Å or 4.4 Å between heavy atoms. At this distance it would suggest that the interactions between sidechains are weak (Figure 56).
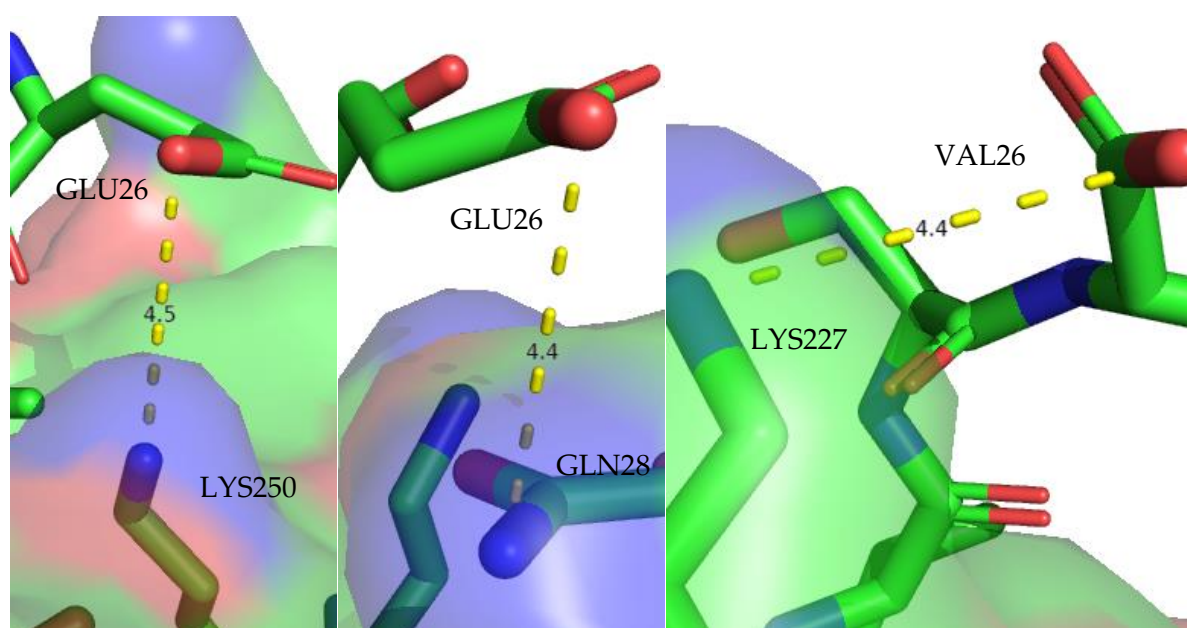
Figure 56 - Possible interactions between MYC GLU260 and WDR5 LYS250 (left), MYC GLU261 and WDR5 GLN289 (centre), MYC VAL267 and WDR5 LYS227.

While the conclusion that the core sequence of IDVV is at the heart of the interaction between these two binders is sound there are some notable anomalies in the data. The OED results also suggest that the full length peptide would be among the weaker binders to WDR5. Grouping the peptides by length and averaging their scores gives the full length peptide the worst overall average score (0.773), while on average fragments 5 amino acid residues in length have, on average, the best score (-2.269).

Comparison between the labelled SOBOC peptides and the predictions from FRED is difficult due to the spread of data available, from both the screening results described in this chapter and the structural data available for the MYC-WDR5 interaction. The screening carried out

characterises fragments of between 11 and 7 residues in length from the sequence DEEEIDVVSVE, the structural data available is for the sequence EEIDVVSV. However, the residues that were left unassigned in the crystal data were likely omitted due to poor density, suggesting that they only weakly interact with WDR5. This is contrary to the results of the labelled SOBOC peptide screening result where the removal of C or N terminal residues saw a significant decrease in ring intensity. This could be due to the screened truncates of the MYC peptide being different to the unlabelled forms tested in the docking study, due to the presence of the DOA spacer and fluorescent label.

5.5.3.2.   Hotpoint

Hotpoint is a web service that was developed specifically for the computational determination of hotspot residues in protein-protein interactions.[17] While FRED scored fragments of the MYC peptide to allow for the evaluation of the binding interaction Hotpoint analyses individual contributions of residues on both sides of the WDR5-MYC interaction. The analysis suggested that 262I and 264V of MYC are hot spots in the interactions. This is in agreement with the results from FRED and the OEDocking suite. FRED predicted that that IDV would be the highest ranking trimeric peptide, the sequence IDV also appears in each of the top 5 results as scored by FRED. The hotspot residues on the bottom interface of WDR5 were determined to be 228Y, 240L, 266F and 288L. As hotspot residues these 4 are predicted to contribute more to the binding interaction than other residues. This suggests that the bottom of the hydrophobic interface contributes more to the binding interaction than anywhere else (Figure 57). When developing tool compounds targeting this interface this result would suggest that the compound should be designed with filling the bottom of the hydrophobic trench as a priority feature.
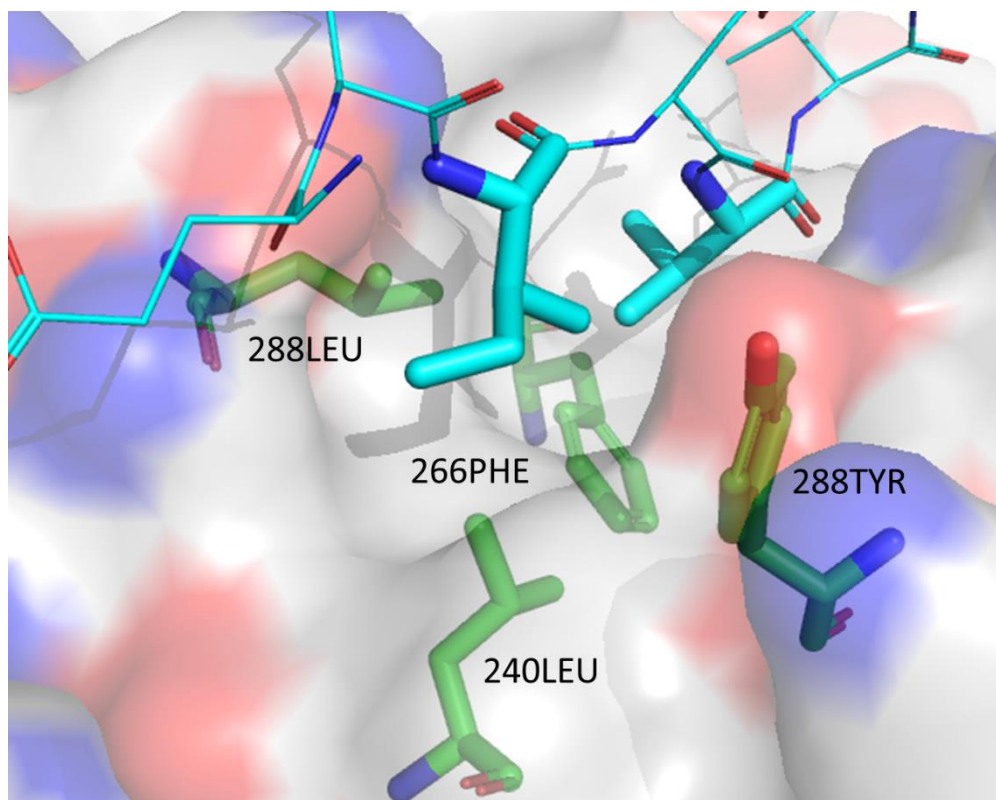
Figure 57 - Hotspot residues of the MYC-WDR5 interaction. Hotspot residues shown as sticks, MYC peptide is shown in cyan, WDR5 surface shown in white. Hotspot residues are shown in stick format. WDR5 hotspot residues are in green and labelled, MYC hotspot residues are in cyan.

5.5.3.3.   Future work

The results of this *in-silico* portion of work suggested key residues on both sides of the MYC-WDR5 interaction. Future work would be to investigate these hotspots for the targeting of the interaction. Both *in-silico* methods described in this chapter are in agreement that the bottom of the hydrophobic trench is a major contributor to the interaction, *in vitro* work suggests that the C-terminal acidic residues are also important to the interaction. These facts are of important consideration for the purposes of designing a tool compound for the probing of lower-face interactions of WDR5. If desired further information could be drawn from the investigation of other WDR5-bottom-interface binding interactions with KANSL2 and RbBP5. A similar analysis as carried out on the MYC MbIIb peptide could be carried out using known structural information on these interactions, however due to the similarities in sequence it is unlikely that there is any significantly different information to be gained.
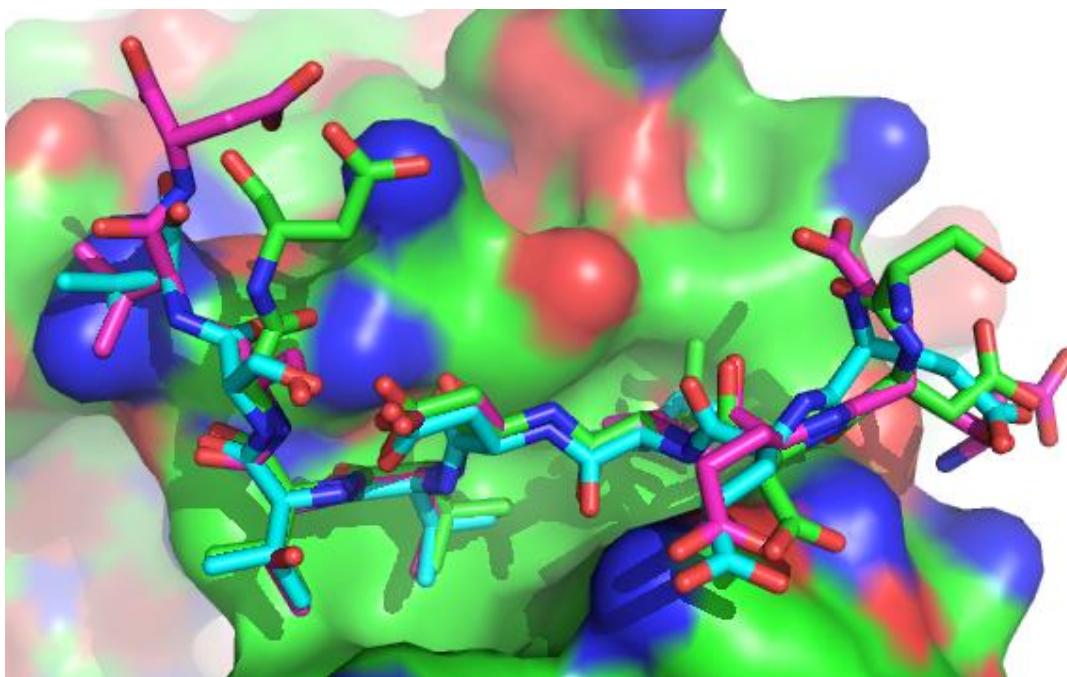
Figure 58 - WDR5 bottom-interface binding sequences of MYC (cyan), RbBP5 (magenta) and KANSL2 (green) constructed from the overlaying of PDB files 4Y7R, 2XL3 and 4CY1 respectively.

Figure 58 shows that peptides based on both RbBP5 and KANSL2 bind, as expected, in the same manner as the MYC MbIIIb peptide and have similar sequences. It is therefore reasonable to assume that the contributions of residues towards these interactions is comparable to the WDR5-MYC interaction.

6. Synthesis of Peptides Developed with MorPH Software for the Targeting of Survivin

6.1. Introduction

This chapter documents work on the non-WD40 domain containing protein Survivin. The high value target Survivin was selected as a candidate for the peptidomimetic 'MorPH' procedure. The results of this investigation will be described and its potential application in targeting the WD40 domain containing protein WDR5 is discussed.

6.1.1. Survivin

Survivin was discovered in 1997. Its initial discovery indicated that the protein was not present in healthy adult cells, only in cancer cells and embryonic cells.[140] This initial discovery directed the future of Survivin research and its validation as a target. It was later determined that Survivin was not absent in healthy cells but instead highly regulated, expressed significantly more during the G_2M phase of the cell cycle.[141] In 2000 the crystal structure of Survivin was determined (PDB code: 1F3H), the work showed the protein to exist as a dimer.[142]

Survivin is a part of the Chromosomal Passenger Complex (CPC). The CPC is made up of 4 proteins: Survivin, Borealin, Aurora B kinase, and inner centromere protein (INCENP). Survivin and Borealin are responsible for the localisation of the CPC to kinetochores during mitosis. Aurora B kinase differentially phosphorylates components of the kinetochores in response of varying tension between the kinetochore and microtubules. Connecting these is the structural unit INCENP. At the N-terminus of INCENP are the localisation proteins Survivin and Borealin, at the C-terminus is Aurora B kinase. INCENP, Survivin and Borealin form a triple helical bundle, this bundle is required for the complex to bind to spindle and mid-body localisation. Shugoshin 1 (hSgo1) is known to bind to histone H2A when the histone is phosphorylated by the kinase Bub1, in turn hSgo1 is known to bind to Borealin. Interestingly a fragment of hSgo1, a peptide with the sequence AKER, was found to also bind to Survivin via the BIR domain in a co-crystalisation.[22]
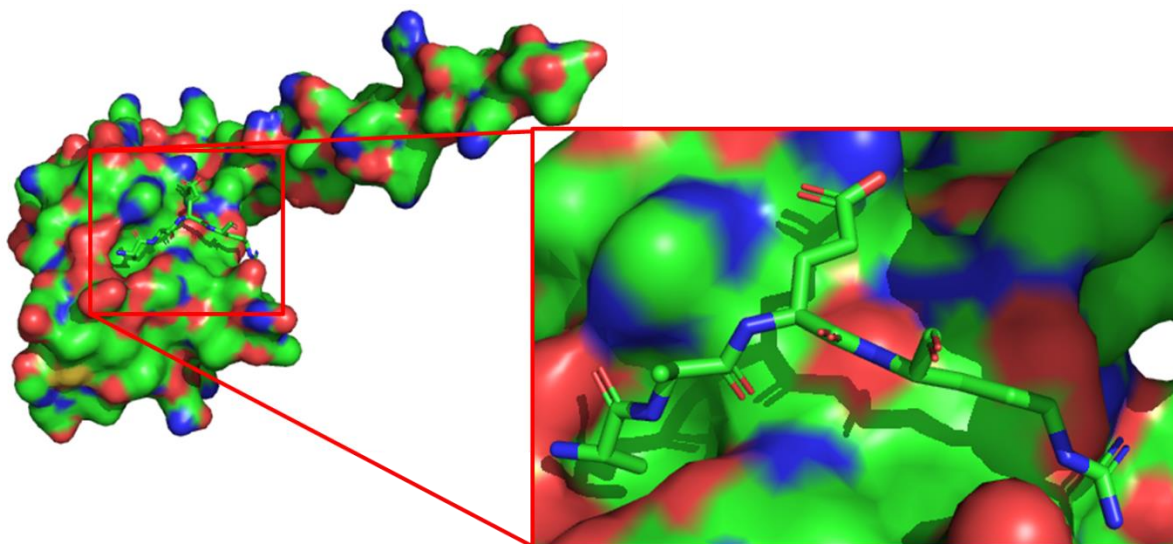
Figure 59 - 3D structure of AKER peptide binding to Survivin from PDB 4A0I. The lysine residue of the sequence does not have density assigned and appears as an alanine.

Survivin is also responsible for inhibition of apoptosis in healthy cells. During normal apoptosis caspases are responsible for the breaking down of cellular components. Caspases 3 and 7 are activated at the end of both the intrinsic and extrinsic apoptosis pathways, their activation leading to cell death. Survivin's antiapoptosis mechanism functions by inhibiting caspase 3 and 7, binding to them with nanomolar affinity.[18] However this apoptosis inhibition is also responsible for securing the survival of cancer cells overexpressing Survivin.

The protein is found in significantly higher concentrations in cancer cell lines. Out of the National Cancer Institute cancer cell lines Survivin was found in increased concentrations in all of 60 tested.[21] Particularly high levels were found in lung and breast cancer cell lines. Increased Survivin expression has been linked to a poor prognosis and reduced patient survivability in the clinic.[19, 20, 143-146] It was also determined that Survivin and another inhibitor of apoptosis; XIAP, are able to significantly reduce the rate of apoptosis in cells at elevated levels of expression.[21]

The first Survivin interference compound was an oligonucleotide discovered in 1997.[147] It was determined that the oligomer inhibited endogenous Survivin RNA expression in HeLa cells, inhibiting the expression of the protein. It was found that this interference saw an increase in apoptosis. Since this validation many more attempts at targeting Survivin, it's interactions[148] or its expression.[149]

In 2007 a publication from Abbott Laboratories described a potential new small molecule binding site for Survivin, near to the dimerization site.[150] This was an important development as previously only the binding site known on Survivin was the baculovirus inhibitor of apoptosis repeat (BIR) domain. Binders to the peptide binding site generally were shown to have a weak affinity. It was determined that this new site had a hit rate of 0.35% compared to a hit rate of 0.01% for the BIR domain by screening their libraries via NMR. Since identification of the dimer site, new small molecule binders to Survivin have been published.[151] Binders to the site located near the dimer interface site were found to increase cell death and delay the cell cycle in concentrations in the nanomolar concentration range.[150]

Referenced publications are indicative of the challenge and the value of targeting Survivin. The development of novel and tool binders to Survivin would allow for a better understanding of the roles it plays in a variety of cancers. In this chapter a novel *in-silico* technique and its application to Survivin will be discussed.

6.1.2. MorPH – Software for the iterative optimisation of peptide binders

The crystal structure of the AKER sequence of hSgo1 binding to the BIR domain of Survivin offers a valuable starting point for the development of a tool compound able to bind to Survivin to better establish the effects of chemical inhibition of the proteins interactions. Crystallographic data provides invaluable insight towards possible routes towards the development of a novel ligand, however given the number of commercially available building blocks that are available for use in Fmoc SPPS guidance towards the most promising routes of optimisation could save much time and effort. To this end, the in house proprietary peptide replacement software MorPH was developed by Steven Shave. MorPH takes an input 3D structure of a peptide and target protein and systematically replaces amino acids at each position of the peptide with commercially available building blocks, scores and ranks these replacements using a molecular force field. The current scoring process utilises the Chemgauss4 force field to evaluate multiple bound conformations for each amino acid replacement.[133]
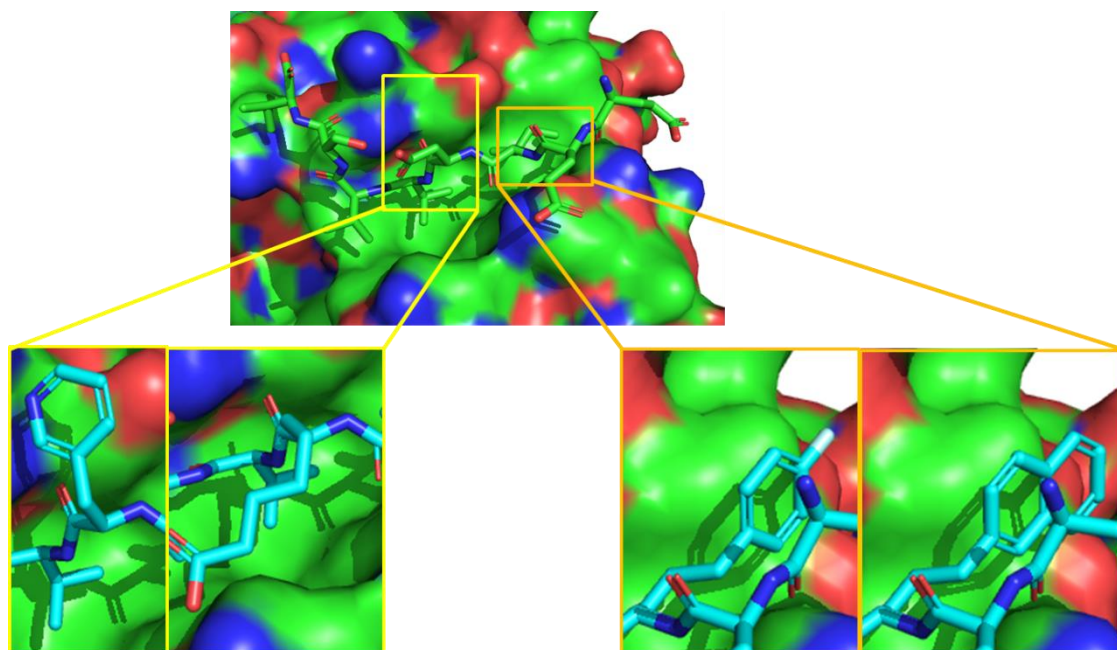
Figure 60 - Visualisation of MorPH Process. A 3D structure of a peptide-protein interaction is used as input (top). For each residue of the peptide MorPH docks and scores potential replacement residues. Example of isoleucine and glutamic acid residue replacements are shown in the yellow and orange boxes respectively.

The process outputs a ranked list of top scoring amino acid replacement suggestions for each amino acid position, along with a predicted 3D structure which may be visualised in many molecular visualisation packages. Manual inspection is then a key step in shortlisting of potential building blocks at each position which can then be ordered for synthesis and screening. When the best suggestion for a position is determined, the ideal next step would be co-crystallisation of the peptide with the protein. The structure would then be used for input for the MorPH process again. This will compensate for any changes in the peptide backbone structure that occur as a result of difficult to predict and model amino acid residue substitution or changes in the mode of binding. As well as potentially sequentially improving the affinity of a peptide one position at a time, the replacement of natural residues with non-natural amino acids is expected to improve the stability of the peptide against peptidases. This is an important consideration for the purposes of cellular assays as peptides are highly susceptible to digestion, however non-natural residues will be less likely to be recognised by enzymes. To evaluate the MorPH process a series of peptides based on MorPH suggestions for the targeting of Survivin were to be synthesised. In order to conclude the project peptides representing the best replacements for each position were synthesised and characterised. The

MorPH process was then to be applied to the MYC-WDR5 interaction discussed in the previous chapter and the results interpreted.

6.2.   Aims

To synthesise a set of example peptides based on MorPH suggestions for further testing. Testing includes affinity determination with Survivin and plasma stability characterisation. To apply the MorPH process to the MYC-WDR5 interaction and to determine potential replacement building blocks for a series of SAR syntheses.

6.3.   Results

6.3.1. Peptide synthesis

All peptides were synthesised using methods described in section 9.4.

Table 21 - Results of the synthesis of TMR labelled example Survivin Morph Peptides
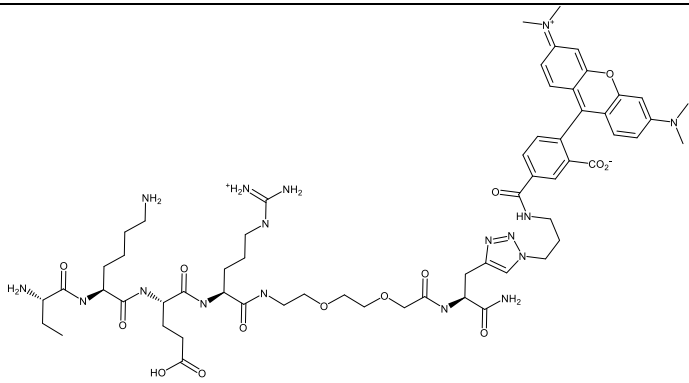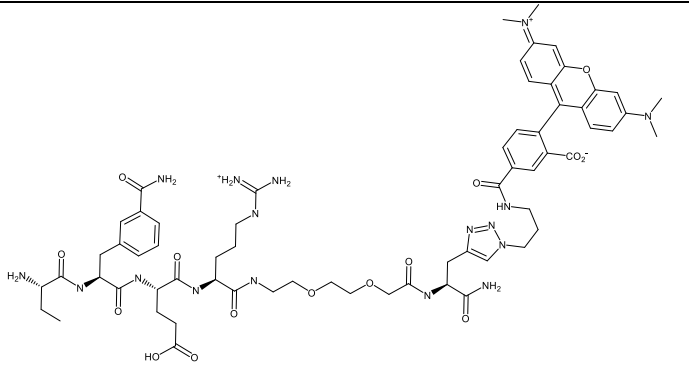
	Abu-K-E-R-DOA-PRA(TMR)
	Purity by HPLC: > 95 %
	MS results
	Mass of peptide: 1268.63,
	seen 1269.26 [M+H]$^+$,
	and 635.29 [M+2H]$^{2+}$
	Abu-CarbPhe-E-R-DOA-PRA(TMR)
	Purity by HPLC: > 95 %
	MS results
	Mass of peptide: 1331.62.
	Seen 1331.21 [M+H]$^+$
	and 666.44 [M+2H]$^{2+}$

Table 21 summaries the results of the synthesis of 2 labelled example peptides synthesised for further characterisation using on-bead methods of peptide synthesis and labelling. In the experiments described in the following section another labelled peptide, A-K-E-R-DOA-

PRA(TMR), is used in experiments. The peptide used was from a prior large scale synthesis carried out by Peter Dodd of the Auer lab.

Table 22 - Results of the synthesis of unlabelled Survivin Morph example peptides
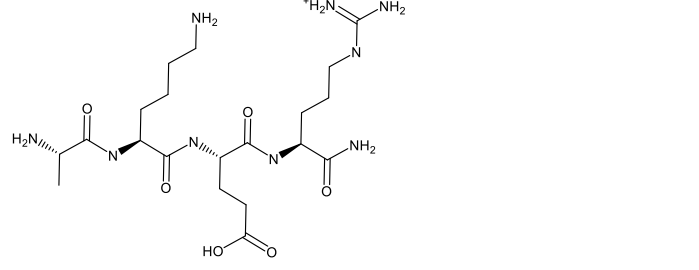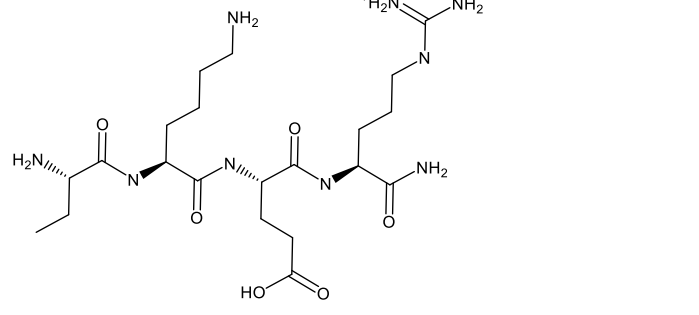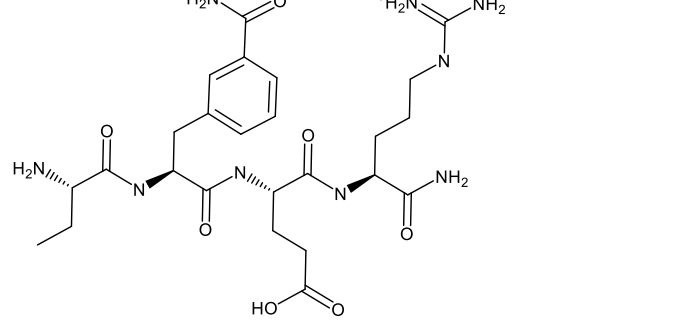
(structure of Ala-Lys-Glu-Arg peptide)	A-K-E-R
	Purity by HPLC: > 95 %
	MS results
	Peptide mass: 501.59.
	Seen: 502.57 [M+H]+
(structure of Abu-Lys-Glu-Arg peptide)	Abu-K-E-R
	Purity by HPLC: > 95 %
	MS results
	Peptide mass: 515.62.
	Seen: 516.38 [M+H]+
(structure of Abu-CarbPhe-Glu-Arg peptide)	Abu-CarbPhe-E-R
	Purity by HPLC: > 95 %
	MS results
	Peptide mass: 577.64.
	Seen: 578.55 [M+H]+

Table 22 summaries the results of the synthesis of 3 unlabelled example peptides synthesised for future planned experiments, primarily for use in competition experiments with their labelled counterparts and Survivin.

For HPLC analysis the unlabelled peptides were characterised with their Fmoc protecting group left intact. This is due to the small, polar nature causing them to elute very quickly. Leaving the N-terminal Fmoc group significantly slows the elution of the peptide and allows for more accurate determination of purity. The unlabelled peptides did not require any purification after their cleavage from the resin other than the standard Fmoc protocol of precipitating the peptides in cold ether.

6.3.2. Screening of labelled peptides against Survivin (carried out by Nhan Pham)

Note that the peptides used for K_D determination experiments were synthesised by Olivier Barbeau. Anisotropy was measured via 2D-FIDA with 2 nM labelled peptide concentrations, 9 measurements per protein concentration. Curves were plotted using Equation 15 (materials and methods).
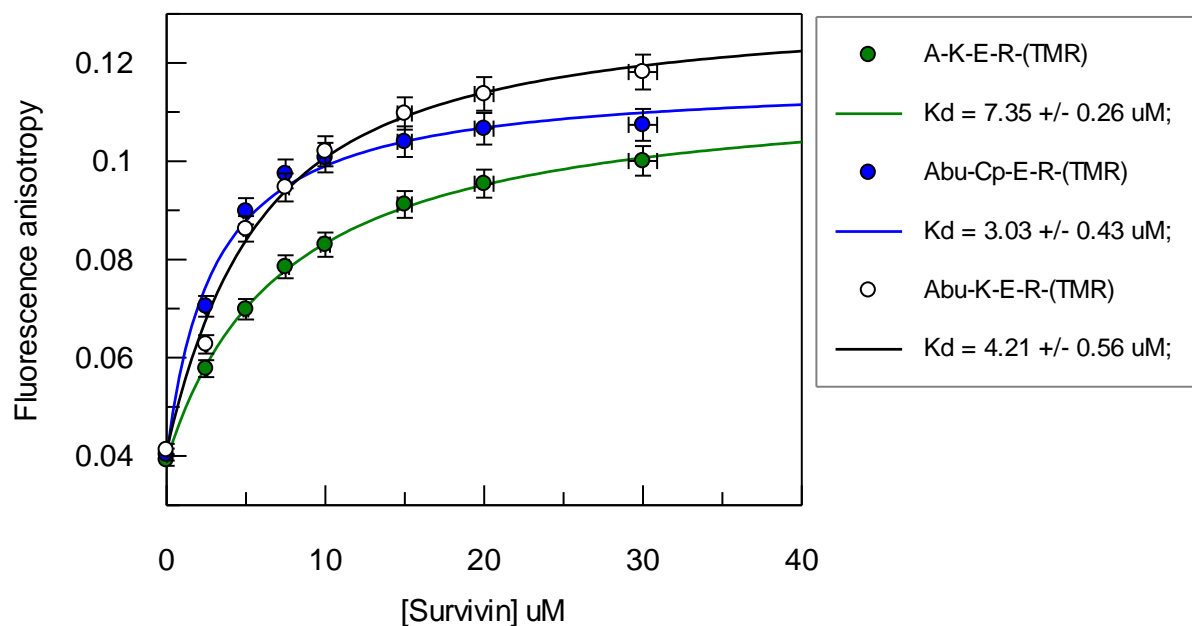

Figure 61 - Results of 2D-FIDA anisotropy experiments screening labelled peptides against Survivin.

These 2D-FIDA anisotropy experiments were carried out to compare the change in affinity as more of the natural sequence is replaced with the best-in-place suggestions from the MorPH software. The results suggest that as more of the peptide is replaced with the MorPH suggestions the K_D decreases.

Results of Plasma Stability Testing of MorPH Peptides
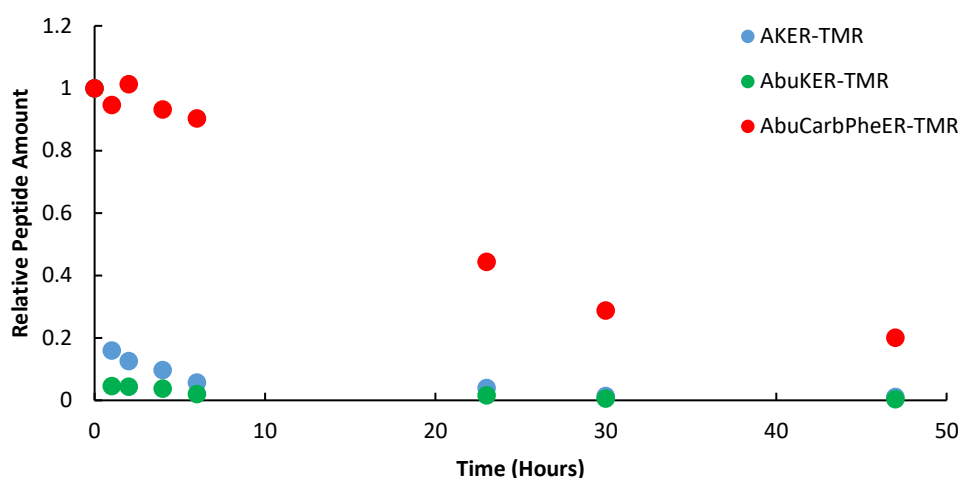
Figure 62 - Plot indicating the rate of degradation of AKER and MorPH suggested variants

Peptides are naturally broken down very quickly under cellular conditions, analysis was carried out to characterise the effects of replacing natural residues with best-in-position suggestions from MorPH on the plasma stability of the peptides. Figure 62 shows that the stability of peptides containing non-natural residues is significantly improved over the natural sequence, AKER. This shows that the incorporation of non-natural amino acids into the sequence

6.3.4. MorPH Suggestions for targeting the MYC-WDR5 interaction

3D structural data of a peptide based on the WDR5 interaction sequence from MYC was used as input for MorPH. The structure was taken from PDB entry 4Y7R. The 3D structure included information on the residues EEIDVVSV of the peptide sequence DEEEIDVVSVE binding with the bottom interface of WDR5. MorPH output the top 500 scored conformations for each of the 8 positions of the peptide that had been assigned in the structure. These results are summarised and discussed in section 6.4.2.

## 6.4.   Discussion and Future Work

6.4.1. Survivin

6.4.1.1.   Survivin MORPH Peptide Synthesis and Screening

The aim of this project was to synthesise and characterise peptides for the targeting of Survivin based on suggestions from MorPH software, in order to validate and evaluate MorPH as a tool prior to its utilisation against WDR5. Prior to the synthesis of the peptides described in this chapter many other derivatives were synthesised and characterised against Survivin, based on suggestions from MorPH. The peptides described in this chapter represent the best replacements for each position, determined by this prior work that was carried out by post-doctoral researchers Olivier Barbeau and Peter Dodd. Peptides previously synthesised based on MorPH suggestions for the Survivin-hSgo1 interaction included suggestions for the two C-terminal residues. These were tested and determined to have a detrimental effect on the affinity of the peptide. Therefore it was determined that the sequence H_2N-Abu-CarbPhe-E-R-$CONH_2$ was the highest affinity binder based on suggestions from MorPH. This final peptide synthesised based on suggestions from MorPH was found to have a similar affinity to the input sequence but a significantly increased plasma stability. The plasma stability testing experiment indicated that with increasing substitutions of amino acid residues with non-natural suggestions stability increased. This was an expected result, increasing the proportion of non-natural amino acids moves the peptide further away, in terms of structural similarity, from naturally occurring peptides that are enzymatically digested in plasma. This increased plasma lifetime confirms that peptides constructed of non-natural amino acids are stable for use in cellular assays. In the case of studying Survivin this is an important consideration as Survivin has several vital roles in the cell cycle. A tool compound that is able to withstand longer incubations in cells could provide valuable information on the effects of inhibiting the interactions of Survivin over the course of the cell cycle. Overall this initial test of MorPH suggests that there is potential for MorPH to contribute towards the tool compound discovery process.

Table 23 - Tabulated results of the affinities of MorPH suggested replacements. Kd determined by fluorescence anisotropy

Peptide	K_D (µM)
A-K-E-R-DOA-PRA(TMR)	7.35 ± 0.26
Abu-K-E-R-DOA-PRA(TMR)	4.21 ± 0.56
Abu-CarbPhe-E-R-DOA-PRA(TMR)	3.03 ± 0.43

As the proportion of suggested residues increases the affinity increases (Table 23). The effects of replacing an alanine with and Abu saw a slight decrease in stability compared to the natural residue. Replacing the lysine with a carbamoylphenylalanine however significantly increases the stability. An increased plasma stability makes a compound much more appealing for use as a tool compound in cellular assays. Peptides have notoriously short lifetimes in cells and are usually quickly degraded by enzymes, as indicated by the rapid degradation of AKER-DOA-PRA(TMR) in the stability testing experiment. Survivin is a protein that has roles at the heart of the cell cycle and cell death, should a tool compound be completely degraded after incubation for one hour the effects on the cell cycle will be very small compared to a compound with a longer lifetime. It could be assumed that by replacing the two unchanged C-terminal residues that stability could be further improved. However replacements for these positions were proven to reduce affinity. The 3D structural information for the AKER-Survivin interaction suggests that the glutamic acid and arginine residues are both involved in strong electrostatic interactions via their sidechains.
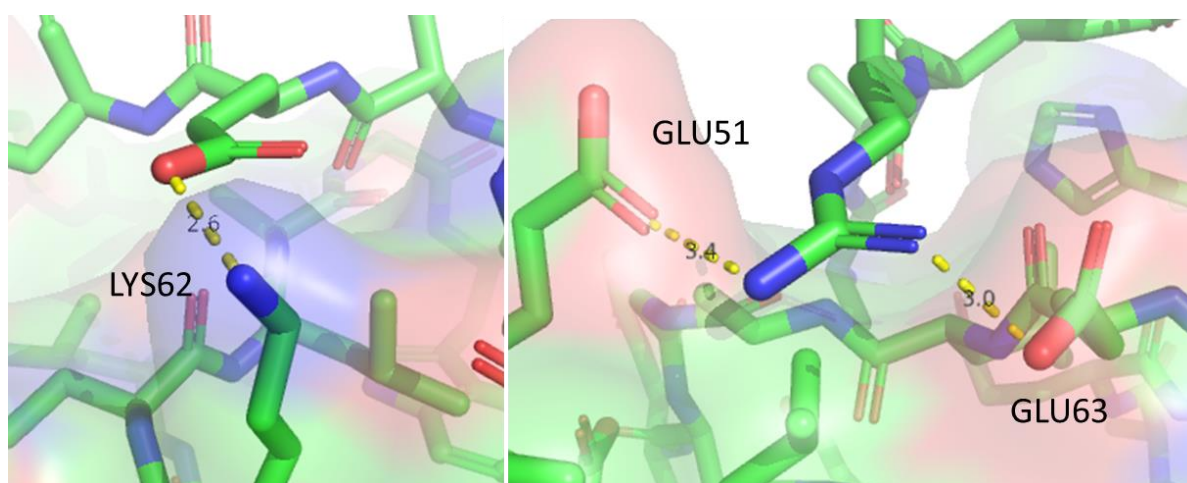
Figure 63 - Interactions of glutamic acid and arginine residues of the sequence AKER with Survivin.

In this instance these residues have been determined to be invariable. The sidechains of both the glutamic acid and the arginine side chains are specifically positioned towards oppositely charged residues on the surface of Survivin.

6.4.2. WDR5 MorPH

6.4.2.1.   Suggestions for MYC-WDR5 interaction

For each position the top 500 conformations and their Chemgauss4 score were output. In this section the commonly suggested motifs and suggestions that are thought to be of serious consideration for an SAR investigation are discussed. For reference the average Chemgauss4 score for the natural peptide (sequence: H_2N-EEIDVVSV-CO_2H) was -0.5188. The scoring for each suggestion is relative to the position, values lower than -0.5188 are considered to have a higher affinity than the input sequence according to the Chemgauss4 scoring system.

Position 1 – Glutamic Acid
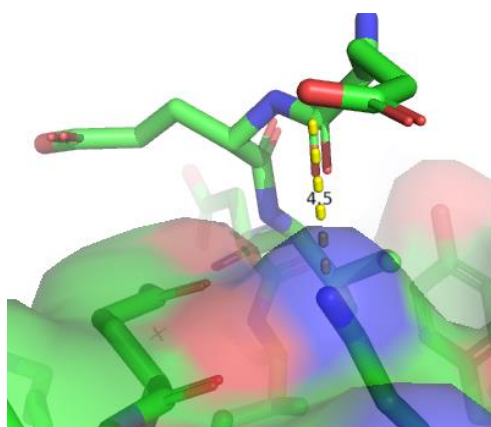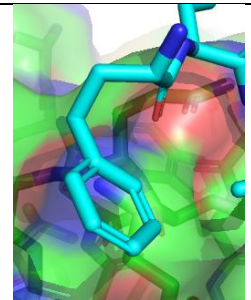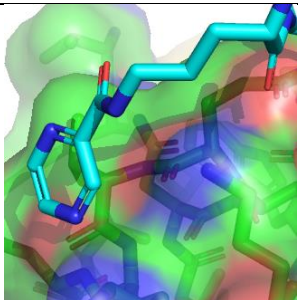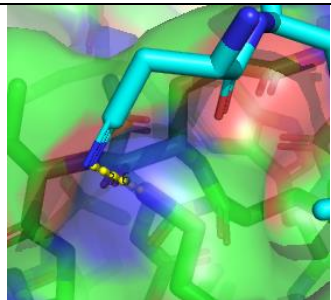
Figure 64 - Possible interaction between position 1 glutamic acid and LYS250 on WDR5

Position 1 is naturally a glutamic acid residue, it is possible that a long range interaction is occurring between the amino acid side-chain and LYS250 of WDR5.

**Table 24 - Examples of MorPH suggestions for position 1 glutamic acid**

Replacement	Phenyl	pyrazine	nitrile
score	-1.427	-1.574	-1.324
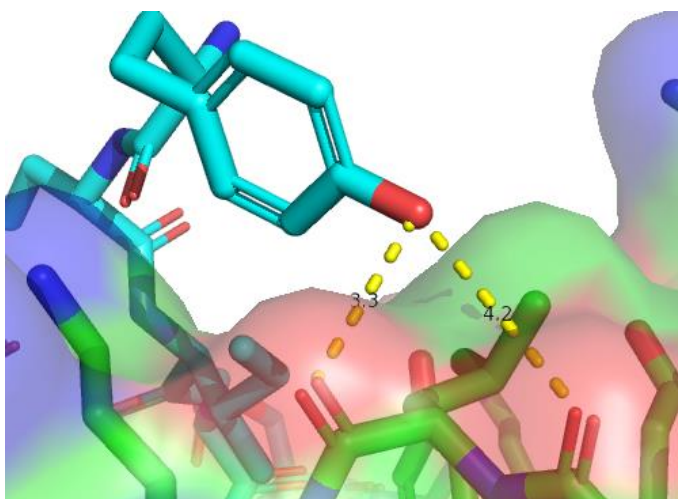
Figure 65 - Potential interactions of homotyrosine at position 1.

A large portion of suggestions from MorPH include larger side chains that have the potential to reach for alternative interactions. Figure 65 above for example shows a homotyrosine, an example of the phenyl suggestions for the position. The alcohol of which could potentially donate a hydrogen bond with the backbone carbonyl of LEU249 (3.3 Å) or, less likely, CYS248 (4.2 Å), however the phenyl ring itself does not appear to have any considerable contribution to the interaction, instead the ring is used to better position the alcohol.
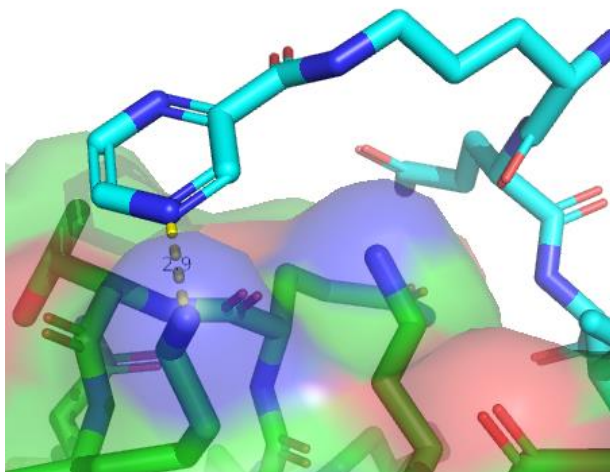
Figure 66 - (S)-2-Amino-5-(pyrazine-2-carboxamido)pentanoic acid suggested by MorPH for position 1

(S)-2-Amino-5-(pyrazine-2-carboxamido)pentanoic acid is positioned to accept a hydrogen bond from LYS291 with a distance of 2.9 Å (Figure 66). This replacement was ranked 15th with a score of -1.574.

For this position MorPH takes advantage of the charged residues surrounding the mostly hydrophobic protein-protein interaction interface.
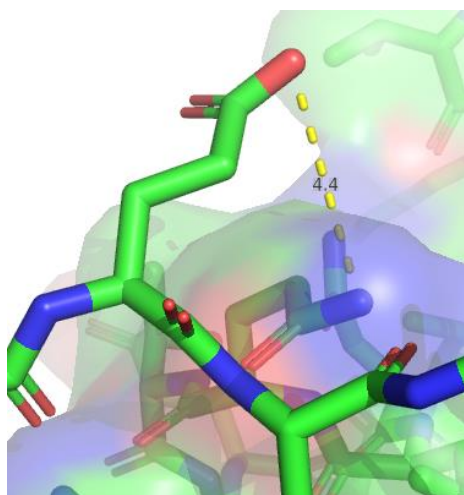
Figure 67 - Glutamic acid residue from natural MYC sequence binding to the bottom interface of WDR5.

As with the previous position the natural residue is predicted to be binding to WDR5 via an interaction with the charged residues around the hydrophobic trench. GLU2 from the MYC peptide appears as though able to bond with GLN289 in WDR5 (Figure 67). As with the previous position the distance is far (4.4 Å), making the interaction very weak.

Table 25 - Example suggestions from MorPH for position 2
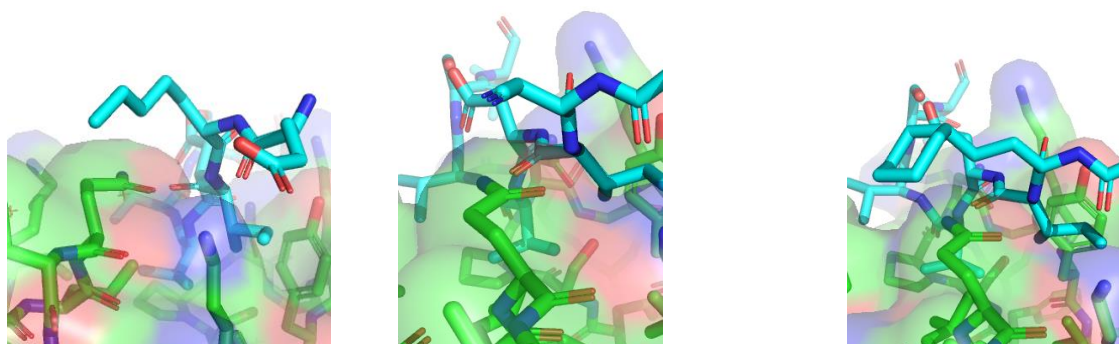
Hydrophobic chains		amines and nitriles		Cyclohexane	
Length	average score	suggestion	average score	Suggestion	average score
Hexyl	-0.655	nitrile	-0.740	ethyl spacer	-0.474
Pentyl	-0.826	amines	-0.677	no spacer	-0.463
Butyl	-0.815				
Propyl	-0.512				
Ethyl	-0.970				
methyl	-1.214				
branched	-0.696				

MorPH suggested a broad selection of suggestions for the position (Table 25). The software suggested mostly hydrophobic replacements for the position, which naturally contains an acidic residue. Of the suggestions containing hydrocarbon chains there was a tendency towards shorter chains to be scored more favourably. A small number of nitriles and amines were suggested as polar replacements for the position, which naturally holds and acid. The distance between the suggested amines and any polar groups on WDR5 were generally large, reflected in the poorer scores for the replacements.
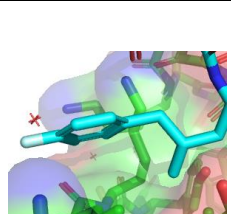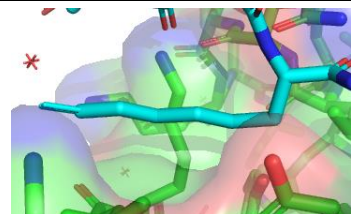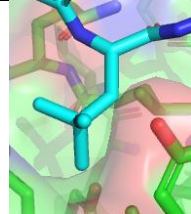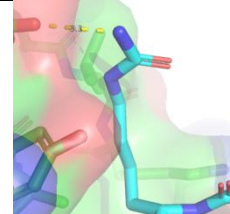
Position 3 – Isoleucine
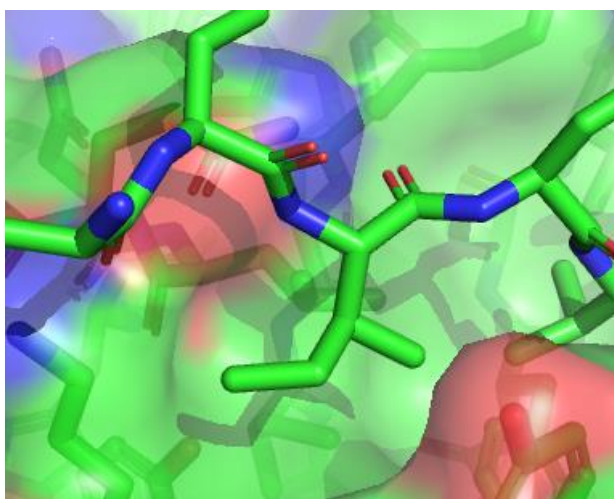
Figure 68 - Natural residue 3ILE is oriented directly towards the bottom of the hydrophobic trench of the MYC binding interface on WDR5.

Naturally an isoleucine residue points directly down into the hydrophobic base of the trench of WDR5. As expected MorPH outputs almost exclusively hydrophobic suggestions that aim to fill the space of the trench.

Table 26 - Summary of three common motifs appearing in suggestions for position 3.

Suggestion	(3S)-3-amino-5-methyl-2-oxo-6-phenyl-hexanoic acid	Linear hydrocarbon sidechains	small hydrophobic sidechains	2-Amino-6-ureidohexanoic acid
average score	-0.443	-0.570	-0.155	-0.394

The majority of the best scored suggestions fall into 2 categories; long hydrocarbon chains across the bottom of the hydrophobic interface and a series of suggestions based on (3S)-3-amino-5-methyl-2-oxo-6-phenyl-hexanoic acid (Figure 69).
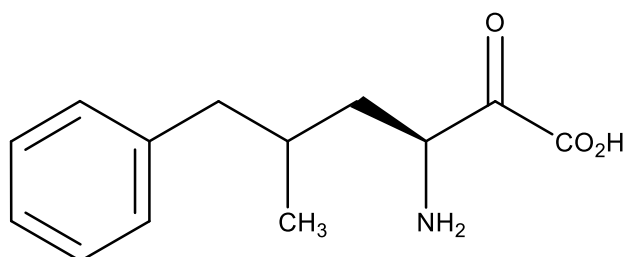

Figure 69 - Structure of the building block (3S)-3-amino-5-methyl-2-oxo-6-phenyl-hexanoic acid

All variations of the structure that appear in the suggestions appear to have modifications at the para position of the ring. The majority of suggestions based on this structure are in the top 10 % of suggestions for this position. The majority of suggestions could be classified as small hydrophobic suggestions. These suggestions generally were of a similar size to the isoleucine residue. It was established in the previous chapter that this position naturally contains a key residue in the MYC-WDR5 interaction. The predicted electron density for this core indicates that the residue is expected to fit the contours of the bottom interface. This series of suggestions would be seriously considered for incorporation into an SAR investigation.
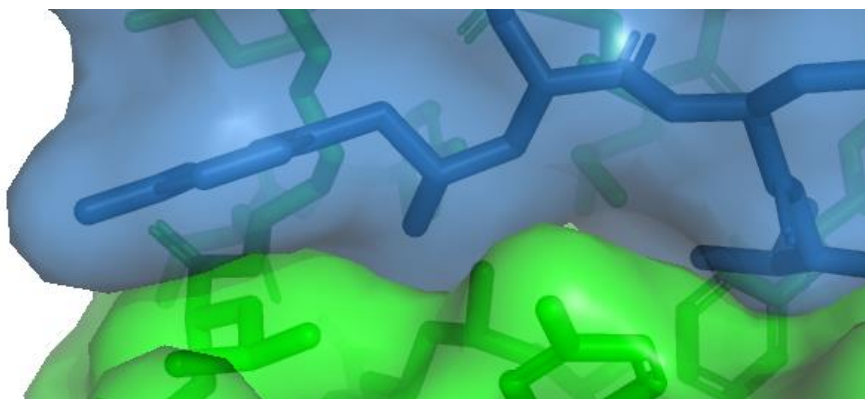
Figure 70 - Visual representation of (3S)-3-amino-5-methyl-2-oxo-6-phenyl-hexanoic acid core interacting with a large portion of the bottom of the WDR5-MYC interface. Suggested peptide shown in blue, WDR5 shown in green.

A small number of recommendations include charged groups. These generally did not score a well with the exception of 2-Amino-6-ureidohexanoic acid (ranked 10th for the position with a score of -0.394) that appears as though it may be able to interact with LYS245 of WDR5 (distance of 3.3 Å).
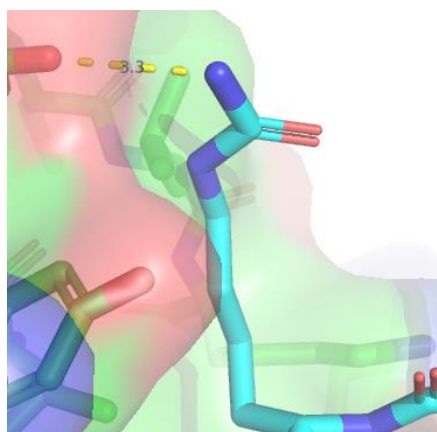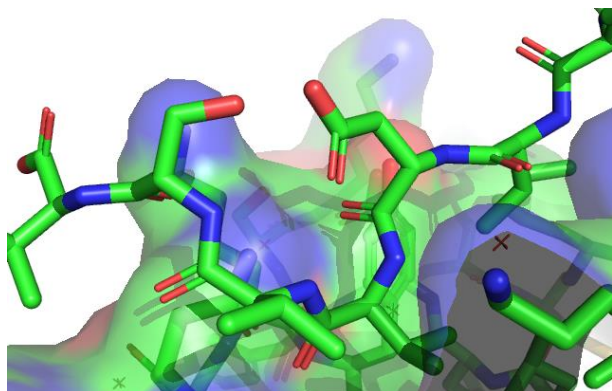

Figure 71 - 2-amino-6-ureidohexanoic acid replacement for position 3, predicted to interact with LYS245 of WDR5.

Position 4 – Aspartic acid

Figure 72 - Position 4ASP appears to point directly away from the binding interface.

The natural amino acid side chain at this position does not appear to directly interact with WDR5 according to the structural data available. MorPH failed to present many convincing suggestions for this position, however there were two suggestions of note.
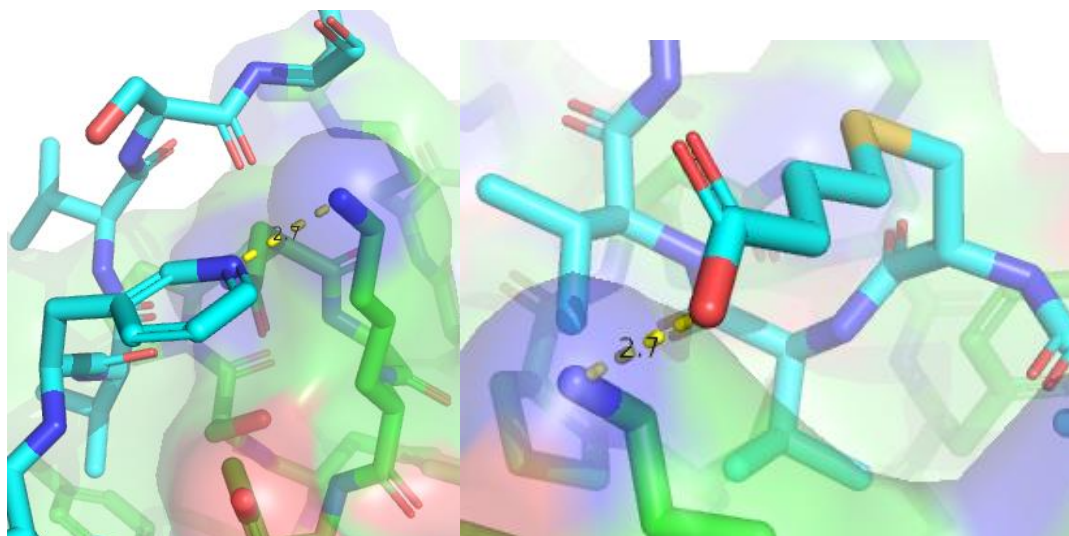
Figure 73 - Example position 4 suggestions from MorPH that are predicted to interact with LYS227 (left) or LYS272 (right).

Small number of reasonable recommendations that pick up potential electrostatic interactions with WDR5, directed towards interaction with LYS227 (Figure 73, left) or LYS272 (Figure 73, right) on WDR5.

Position 5 – Valine
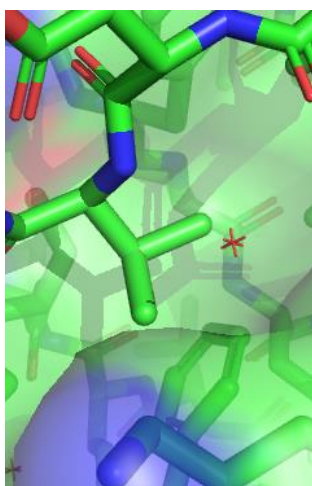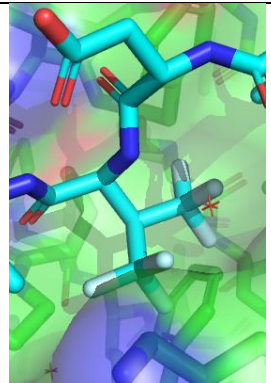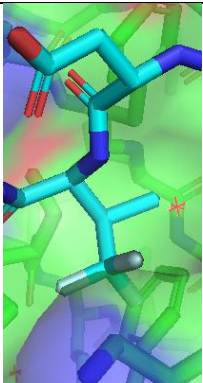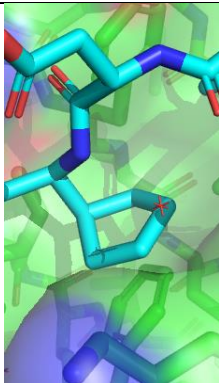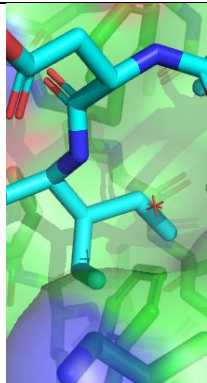
Figure 74 - Residue 5VAL is oriented directly towards the hydrophobic trench in the MYC binding interface on WDR5.

In the natural sequence this valine residue is located towards the lowest section of the hydrophobic pocket. Reflecting this, the suggestions from MorPH are suggested due to their ability to fill the pocket.

Table 27 - Suggestions for position 5VAL from MorPH.

Suggestion	Hexafluorovaline	4,4,4-trifluorovaline	cyclopentylglycine	3-ethylnorvaline
Score	-1.206	-1.222	-0.734	-0.383

Suggestions for position 3 (isoleucine) also appear to have been suggested for the purposes of filling the hydrophobic pocket, however recommendations for position 5 were generally better scored.
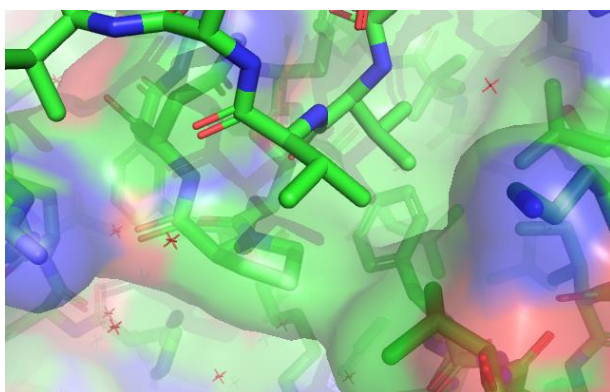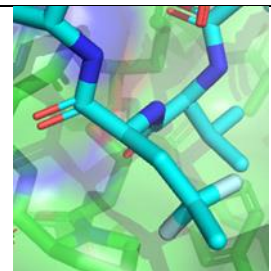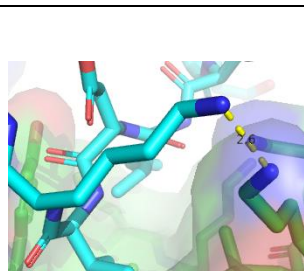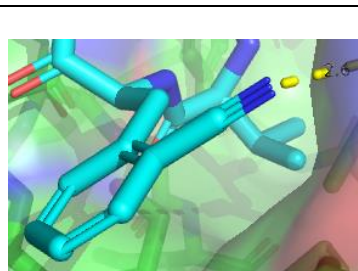
Figure 75 - Natural residue 6VAL of the MYC peptide resides in a small hydrophobic cleft at the side of the binding interface on WDR5.

While the previous valine pointed directly towards the bottom of the interaction interface this residue points towards the charged sides of the trench. MorPH takes advantage of this and suggests a number of replacements that may be able to interact with these charged residues, however hydrophobic replacements were generally better scored.

Table 28 - Summary of suggestions from MorPH for position 6.

Suggestion	Small hydrophobic side chains	Homolysine	(S)-2-Amino-3-(2-cyanophenyl)propanoic acid
average score	-1.049	-0.951	-1.088

Significantly more small hydrophobic replacements were suggested by MorPH, generally they scored better than polar replacements but lower scored hydrophobic suggestions reduced the average. The two most notable polar suggestions are (S)-2-Amino-3-(2-cyanophenyl)propanoic acid and homolysine.
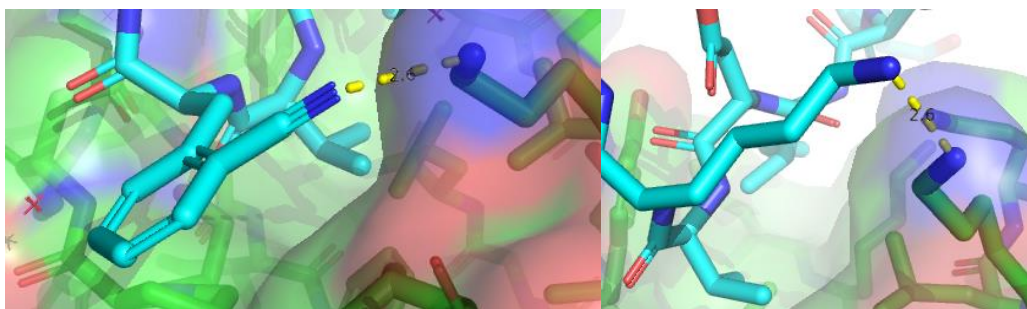
Figure 76 - The positioning of (S)-2-amino-3-(2-cyanophenyl)propanoic acid (left) and L-homolysine (right) in position 6 suggest possible interaction with the ε-amine of LYS272.
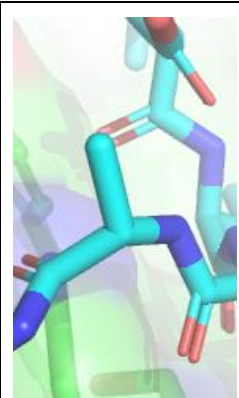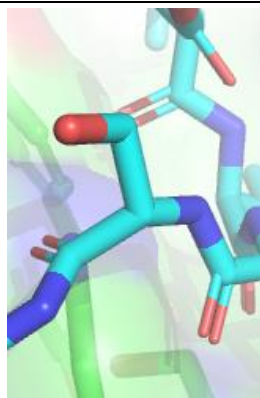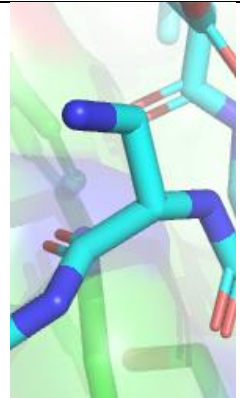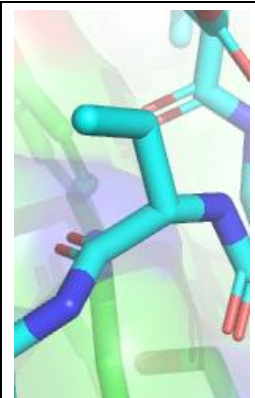
Both of these suggestions appear to be possibly interacting with LYS272 of WDR5, their distance each being 2.8 Å from the amine of the residue.

Position 7 – Serine
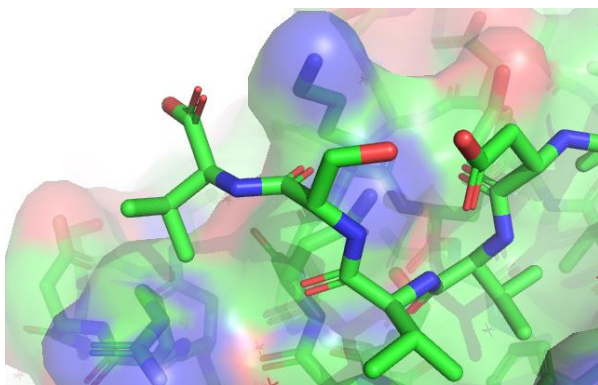
The side chain of the serine residue of EEIDVVSV appears to only be involved in an intramolecular interaction with the aspartic acid. Suggestions from MorPH for this position were very similar. The majority of replacements suggested appear to have been suggested by MorPH because there was not a clear interaction to be targeted, the top 44 scored replacements contained only 4 different suggestions, alanine, serine, homoalanine and 2, 3-diaminopropionic acid.

Table 29 - 4 best scoring suggestions for position 7S, output from MorPH

suggestion	alanine	Serine	2,3-diaminopropionic acid	homoalanine
score	-1.000	-0.626	-0.617	-0.611

From the results of the alanine scan carried out on the peptide it is known that replacing the serine with an alanine increases the K_D from 93 µM to 19 µM.[12] MorPH ranking alanine and other small sidechains so highly for this position could be evidence that MorPH was not able to suggest any better replacements for this position. Looking at lower scored suggestions reveals other apparently reasonable replacements however.
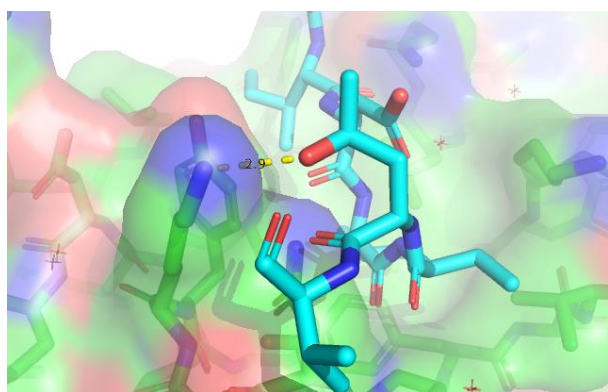
Figure 77 - Substituting serine with 2-amino-4-hydroxy-pentanoic acid at position 7 potentially allows for interaction with LYS227 on WDR5.

The most notable suggestion from MorPH is shown above. 2-amino-4-hydroxy-pentanoic acid contains a slightly longer sidechain than the natural serine. A possible interaction with LYS227 becomes available (distance of 2.9 Å) with this replacement, this suggestion was ranked 45th and scored -0.448.
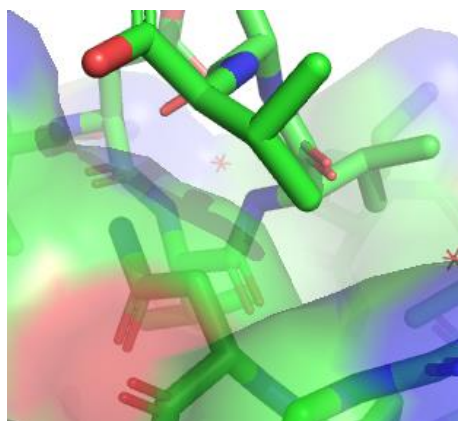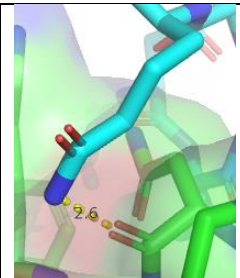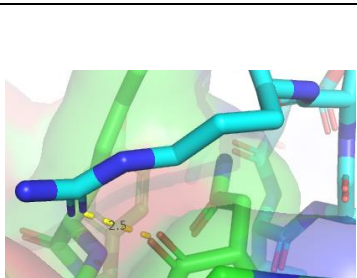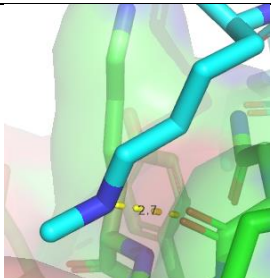
Figure 78 - Valine in the 8th position in the MYC peptide. Peptide is shown binding to the bottom interface of WDR5
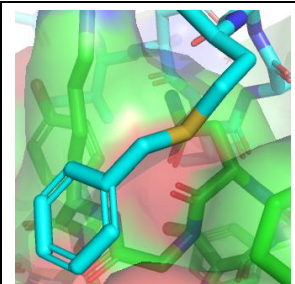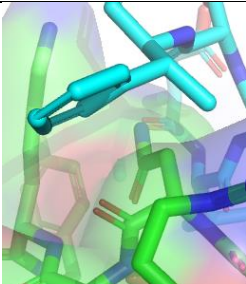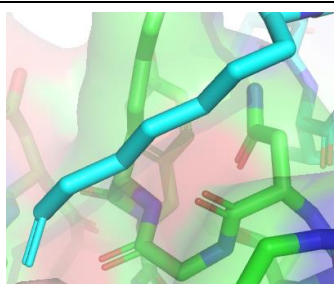
The valine at the C-terminus of the peptide for which structural data is available is at the edge of the bottom face of WDR5. The hydrophobic side chain is present slightly above the surface of WDR5.

Table 30- Example suggestions for the replacement of 8VAL and their Chemgauss4 scores

Suggestion	homoglutamine	Arginine	Methly-lysine
score	-2.785	-1.793	-1.067

Polar suggestions appeared to be focussed around interactions with the backbone carbonyl of ASN225. The amide containing homoglutamine suggestion for this position scored particularly well compared to other predicted interactors with ASN225.

Table 31 - Examples of hydrophobic suggestions for position 8 from MorPH.

suggestion	Phenyl with linker	short phenyl	Hydrocarbon Chain
score	-2.785	-1.191	-1.234

The larger hydrophobic suggestions are oriented in the same manner as the polar suggestions. These suggestions are pointed towards a separate potential binding interface on the bottom of WDR5. LYS227 points out of the bottom of WDR5, the sidechains of the best scoring building blocks wrap around this residue into a distinctly separate potential binding interface.
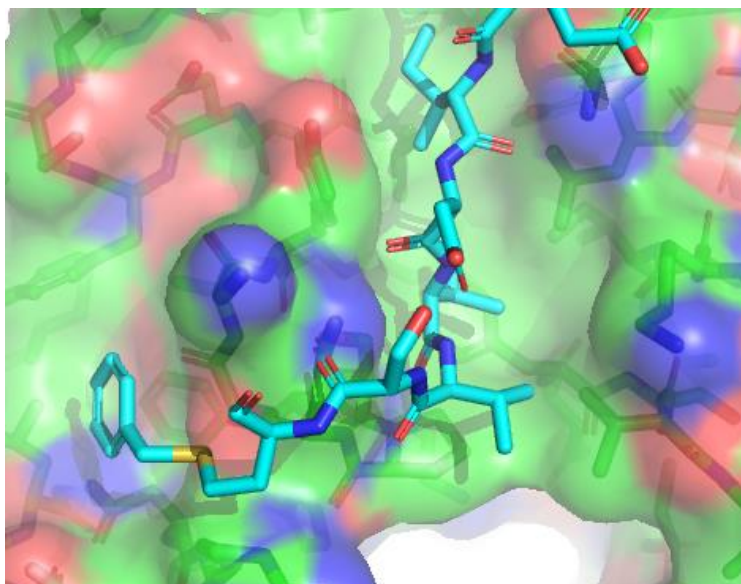
Figure 79 - C-terminal modifications suggested by MorPH exploit a separate potential binding interface on the bottom interface of WDR5.

6.4.2.2. Conclusions of MorPH Suggestions for the MYC-WDR5 interaction

MorPH was used to determine possible suggestions for an SAR investigation for the iterative modification of the WDR5 binding sequence. While positions 4D and 7S on the MYC peptide proved difficult to replace generally each position was provided with several suggestions that are considered worthy of pursuing. Most notably the (3S)-3-amino-5-methyl-2-oxo-6-phenyl-hexanoic acid based replacements for position 2 appeared ideal for filling the a large portion of the hydrophobic pocket on the bottom interface of WDR5. The scoring of MorPH appears

to favour alanine, with it appearing in the top 500 conformations for all positions and even being the highest scored for positions 1E, 2E and S7. As discussed in the previous chapter alanine scanning carried out experimentally on the peptide was demonstrated to reduce affinity,[12] alanine would be omitted from any SAR synthesis for this reason. Alanine is likely scored favourably by Chemgauss4 as it is the smallest suggestion available. This could be interpreted as the software being unable to find what it considered a reasonable replacement for the position and instead gave a better score for the replacement least likely to have any negative consequences. The scores given by Chemgauss4 are best to be taken as a guideline used to rank suggestions. Manually checking the output one position at a time, as was carried out in this section, is necessary for the proper evaluation of suggestions from MorPH. Similarly any suggestion that is scored more favourably than the natural peptide will not necessarily be a higher affinity binder *in vitro.* For the MYC-WDR5 interaction MorPH offered some suggestions worthy of serious consideration for incorporation into the development of a modified MYC peptide consisting of non-natural amino acids.

6.4.2.3. Future Work

6.4.2.3.1.   Survivin MorPH Peptides

Further attempts to improve this example MorPH peptide could be undertaken, however as previously stated this would likely prove difficult. The next step would be to put the peptide into cellular assays to evaluate its efficacy, previously compounds screened against Survivin were found to increase cell death and delay the cell cycle.[150] The Auer lab is equipped with a Laser-Enabled Analysis and Processing (LEAP) unit (Cyntellect, now Intrexon). The LEAP combines imaging (brightfield and multi-wavelength fluorescence) with laser manipulation of cells for the purpose of cell purification (destroying unwanted cells) and optoinjection. Optoinjection allows for the injection of compounds into cells by using short, intense pulses of laser light to create small pores on the surface of the cell. Given the increased plasma stability of the peptide a potential application of the MorPH peptide would be to use the LEAP to monitor the subcellular localisation of the labelled peptide in the cell. Different splice variants of Survivin have been found to localise in different parts of the cell.[152] This method could be used to identify the differences in subcellular localisation in different cell lines. A

possible challenge in such an investigation however could be that the affinity of the labelled peptide to Survivin may be too low to see localisation of the peptide to Survivin.

6.4.2.3.2. WDR5 MorPH

In the previous chapter a peptide truncate library synthesis method was described that was designed with speed and efficiency in mind. This method was developed to probe interaction interfaces and determine the relative contributions of peptide residues. After determining the relative contributions of residues one of the truncated peptides could be used as a possible peptide starting point for an SAR investigation based on known protein-protein or peptide-protein interactions. There is a clear synergy between these methods, rapid determination of a possible starting point for SAR and a tool to suggest possible replacements to guide SAR work. Future work could take advantage of both of these new methods to determine a series of potential novel binders to a protein target, potentially faster than a conventional approach. In the previous section an outline of potential amino acid replacements for the MYC peptide targeting WDR5 were described. The next step would be to select a position on the MYC peptide and to synthesise a series of peptides incorporating the suggestions from MorPH. The peptide containing the replacement with the highest affinity would be co-crystallised with WDR5 and the structure used as input for MorPH. This iterative process could be repeated until all positions have been optimised. If successful this method could develop a novel binder to the bottom interface of WDR5.

6.4.3. Discussion and Future Work Regarding MorPH Software

As MorPH is still a new tool further work must be carried out to better understand its potential contributions to the ligand discovery process. Using MorPH to design more peptides for different targets and assessing how effective the suggestions are would be the most obvious choice of progression. An alternative method of characterising MorPH would be to take a series of published SAR results that used a peptide starting point and use the Chemgauss 4 to rank the different structures based on the predicted affinity. This result could be compared to the actual rankings of the published SAR data. This could be taken further and MorPH could be used to suggest possible replacements that were not included in the original SAR series. These novel suggestions could be synthesised and compared where they were expected to fit into the ranking according to MorPH. One possible candidate for this investigation would be

the SAR carried out on the top interface of WDR5 by Karatas *et al.*[66] The publication uses a trimeric peptide as a starting point, this small peptide is ideal for rapid synthesis for assessment of peptides suggested by MorPH that were not included in the original SAR investigation.

Regardless of the outcome further characterisation is important, if MorPH is shown to be unable to predict with any certainty the general rank of potential binders then the scoring method must be scrutinised. If MorPH is shown to be able to provide a ranking that is generally close to the results of the *in* vitro analysis then it could be accepted that, with further work, the method could be used as a tool to assist in SAR projects. It is highly unlikely to expect a result where software will be able to predict with perfect accuracy what alterations would see a significant increase in affinity, small changes in structure can sometimes have huge effects on affinity. However the method should be able to suggest what variety of replacement would have a positive or negative effect on affinity. The scoring of replacements is carried out using Chemgauss4. Should it be determined that MorPH is inaccurate in its evaluation of replacement residues the replacement or supplementation of Chemgauss4 with another method could be trialled. There are a large number of different docking and scoring software options that are available for this purpose.[153]

One possible future application of MorPH would be to determine which replacements are frequently suggested to replace natural amino acid residues. This list of common suggestions could be used to develop a list of amino acid bioisosteres.

7. Chemical Tools for Practical On-Bead Methods

7.1.   Introduction

Solid phase peptide synthesis was pioneered by Merrifield in 1963 in the publication "Synthesis of a Tetrapeptide".[23] This development enabled researchers to synthesise peptides that do not express well in biological systems, including the incorporation of non-natural amino acids. A key advantage of solid phase synthesis is that the synthesis product remains attached to the resin after each reaction. To extend a peptide chain by a single amino acid residue two reactions are required, a deprotection followed by a coupling. In solution phase chemistry this would mean two potentially lengthy washing steps, however in solid phase each washing step takes only a few minutes of washing the resin with a suitable solvent. This powerful method has been adapted in a great number of ways to aid researchers in the synthesis of non-natural peptides, cyclic peptides, chemical tools and libraries of peptidic compounds.

Solid phase synthesis utilises semi-porous polymer resin beads that are functionalised depending on their application. Most commonly resin will be amine functionalised for the purposes of peptide synthesis. To the resin a linker is usually attached, this linker will be chosen based on the final requirements of the synthesis product and will dictate the conditions under which the product will be released from the resin. Generally, the linker will also be amine functionalised for the attachment of the following building blocks. An acid containing building block is incubated with a base such as DIPEA and an activation agent such as HATU. This incubation rapidly forms the activated ester that will in turn rapidly react with any suitable nucleophile, such as an amine.

HATU is used preferentially as an activation agent due to the favourable properties of the side products formed by coupling reaction, rapid reaction times and low levels of epimerisation. Following a successful coupling reaction, the resin is thoroughly washed and the attached building block is deprotected. The Fmoc protecting group has a high stability to the coupling conditions, however has a half-life of around 6 seconds in the presence of secondary amines such as piperidine.[154]

The Fmoc deprotection is reliable the resulting fulvene unit is photodetectable, allowing for the reaction to be monitored. This pattern of acid activation and coupling of an Fmoc-amino protected amino acid followed by Fmoc deprotection is repeated to produce the desired peptide. By carrying out the synthesis on solid phase the product of each reaction is easily purified by thoroughly washing away the excess reagents and side products of the completed steps. A great advantage of the use of Fmoc solid phase synthesis is the variety of reagents that can be used in the format. Suppliers offer whole catalogues Fmoc-protected amino acids.

The Auer lab has pioneered methods for the development of chemical libraries and their screening methods by using Fmoc synthesis methods.[124, 132] Because of the focus on these methods it often becomes necessary to consider commonly used reagents used and how they can be improved.

7.1.1. Tetramethylrhodamine

The chemical label of choice used by the Auer lab is tetramethylrhodamine (TMR). Rhodamine dyes have several favourable qualities for a fluorescent label, including high photo-stability, high absorption coefficients and high quantum yields. Generally fluorescent labels are expensive to purchase due to the difficulty in their synthesis and purification. Several methods of synthesis have been published and carried out by the Auer lab, but as stated previously these methods of synthesis and purification are considered difficult.
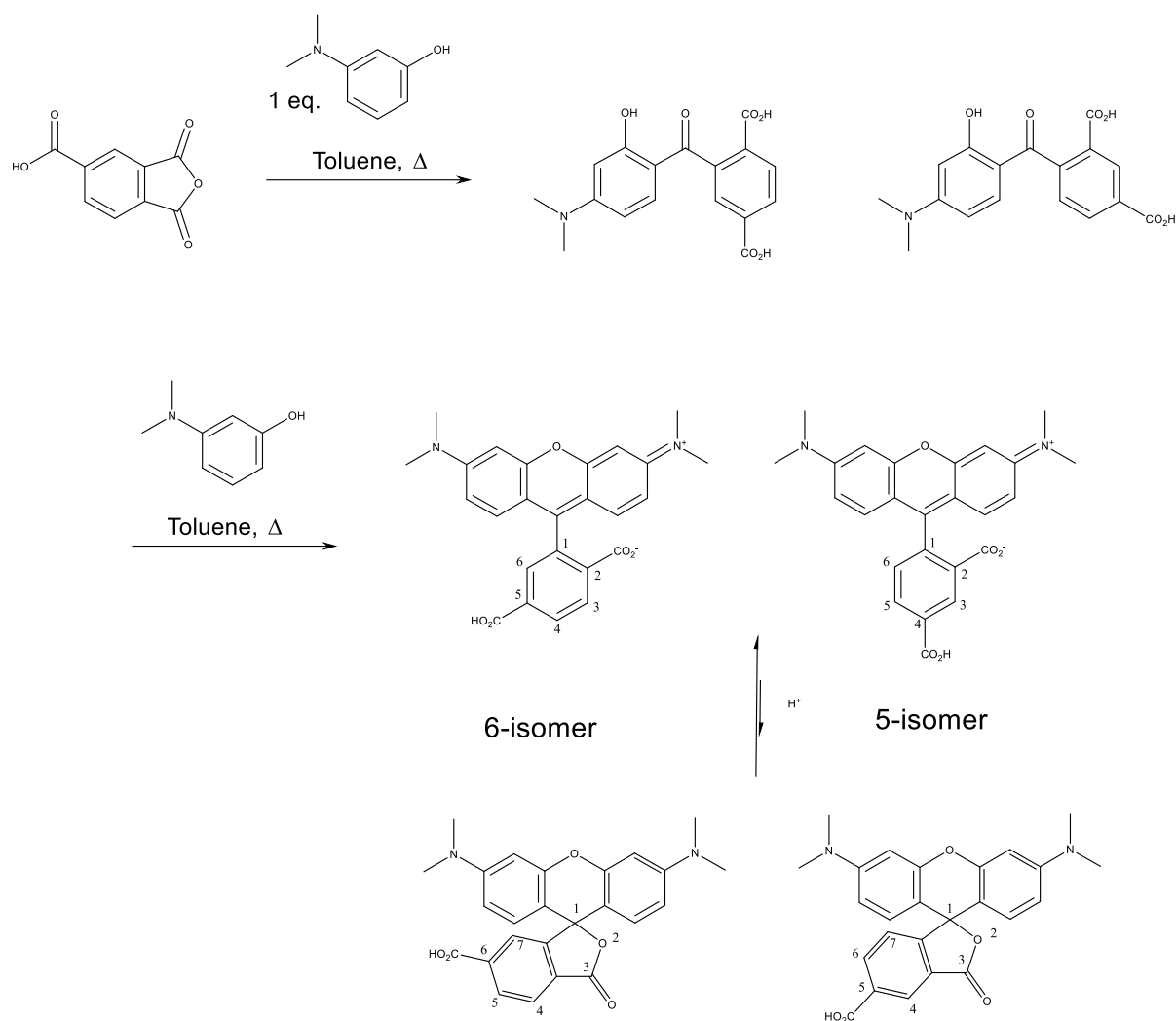
Figure 80 - General synthetic method towards rhodamines.

Rhodamines were first developed in 1888 by Maurice Ceresole.[155] Since the discovery most publications have focused on using some variation of the original process,[156] despite disadvantages in the synthetic approach. The multistep synthesis requires hash conditions and produces a mixture of isomers in relatively low yields. The general synthetic route described in Figure 80 above begins with phthalic anhydride as a starting reagent with 2 equivalents of dimethylaminophenol. Two Friedel-crafts electrophilic aromatic substitutions occur sequentially on the phthalic anhydride. The first substitution reaction results in 2 possible benzephenone intermediates, these intermediates are reacted on by the second equivalent of dimethylaminophenol to give the two regioisomers that then must be separated. While the two isomers are very similar in both their structure and characteristics it is preferable to have the isomers separated. For example the isomers have different

electrophoretic mobilities[157] and have been reported as having different activities in *in vivo* labelling experiments.[158] Note that when referring to the TMR isomers the nomenclature is based on the spirolactone form (Figure 80), a tautomer that exists under basic conditions. The separation of the 5 and 6 isomers is the topic of several publications. Prior to this project the Auer lab used a synthetic method published by Kvatch *et al.* that relied on fractional crystalisation to isolate the intermediate regioisomers. Using this method, the group reported that the 5 and 6 isomers could be isolated in 10% and 16% yields with a remaining mixture of mixed isomers that could be further purified. The purified intermediates are then reacted on to provide isomerically pure rhodamine dyes, however the low yields and potentially lengthy crystalisation method are unfavourable in a larger scale synthesis. A novel synthetic route was suggested by Gemma Mudd of the Auer lab based on the avoidance of asymetrical anhydrides as starting reagents. This route would allow for the regioselective synthesis of rhodamine dyes, eliminating the requirement of a difficult purification step. The produced TMR is functionalised in the 6 and 5 positions with a carboxylic acid. Generally, for the purposes of on-bead labelling of reagents the method of choice is the copper catalysed azide-alkyne Huisgen cycloaddition click reaction (Figure 81) due to its specificity. This requires the TMR be azide functionalised.
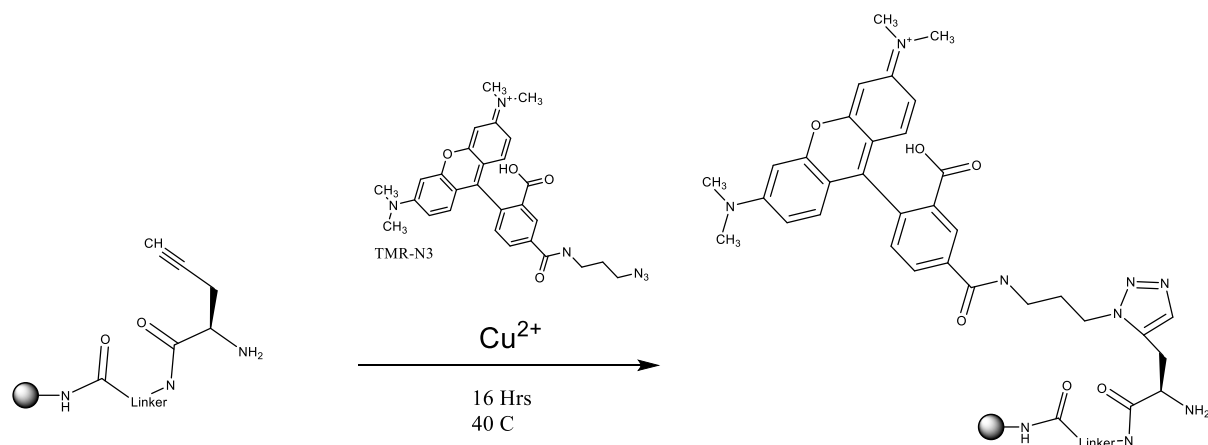
Figure 81 - Reaction scheme of the copper catalysed click reaction, used for the labelling of peptides in the Auer lab.
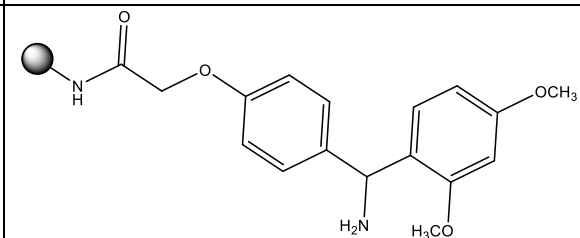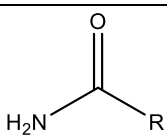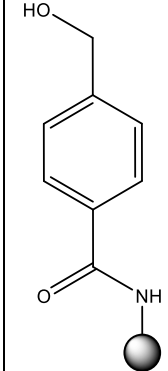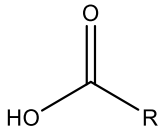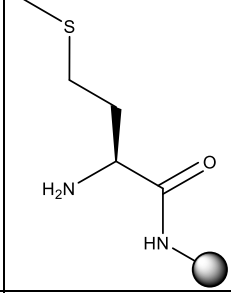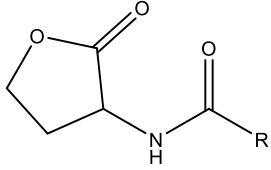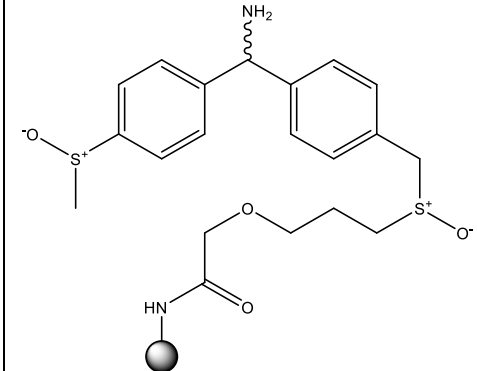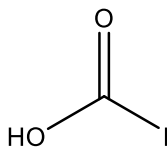
Fortunately, the acid on the 5 or 6 position can be readily activated into an ester and reacted with a reagent that has the desired functional group and a primary amine. A reaction matrix

of different acid activating agents and reactions times were tested to determine the optimal conditions for the conversion of the TMR acid group to the desired azide.

7.1.2. Methionine Linker

The first step of any on-bead synthesis is the attachment of the first building block to the resin. Before this it is important to determine what method of attachment is best suited for this synthesis. The linkage to resin must be chemically selective, stable to all conditions that the resin will be subjected to but readily free the product of the chemical synthesis when desired. There is an extremely low tolerance for side reactions that would affect the purity of the final product, if the method of cleavage must preserve the peptide sequence or the fluorescent tag.

Table 32 - Examples of different solid phase linkers, their cleavage conditions and the resulting C-terminus.

Linker Name	Structure	Cleavage method	Resulting C-terminus
Rink amide		TFA	
HMBA		1 M NaOH	
Methionine		0.1 M CNBr in acidic conditions	
SCAL		Reducing agent and acidic conditions	

If the goal of an on-bead synthesis is to simply produce a peptide via Fmoc SPPS and cleave it from the resin Rink amide is the most commonly used linker. Rink amide cleaves under the same acidic conditions used to remove conventional sidechain protecting groups from amino acid building blocks. This allows for a single step cleavage and deprotection, the peptide can then be precipitated in cold ether. Alternatively, if the aim of the synthesis is to carry out on-

bead screening, i.e. the product of the synthesis is to be left attached to the solid phase resin, the linker must be able to withstand the acidic conditions used to deprotect the amino acid sidechains. In this instance there are a number of alternative solid phase linkers that have been developed, each with associated advantages and disadvantages. The linker of choice used by the Auer lab for on-bead screening is the Safety catch - acid labile linker (SCAL linker). Safety catch linkers require 2 reaction steps to cleave synthesis products from the resin, in this instance the SCAL linker requires that the sulphoxide groups are reduced the amide linkage becomes vulnerable to cleavage in acidic conditions. The specific requirements for cleavage mean that a wide variety of chemistries are tolerated, making the SCAL linker a valuable tool for on-bead synthesis. However the SCAL linker requires a multistep synthesis, despite the efforts carried out to optimise the synthetic route at the time of writing the synthesis stands at a 22% final yield over 11 steps. This is a serious issue for a core reagent that sees use for the in-house production of all on-bead libraries. Additionally the reduction reaction has been credited as an occasional source of impurities in the cleaved products. The use of trimethylsilylbromide in cleavage has been seen to produce a secondary product with a mass 78 mass units higher than the product, corresponding with a bromination. With these issues in mind it was determined that a suitable replacement linker be found. As mentioned previously many potential linkers are described in the literature, however the use of methionine as a linker has several distinct advantages.

Most laboratories that regularly carry out peptide synthesis will likely have methionine building blocks. Methionine is highly hydrophobic and most often found on the interior of proteins, this is favourable for the investigations of protein-protein interactions as methionine has a lower propensity to appear in an interaction sequence. If methionine is used as a linker it must be omitted from the sequence of any peptide synthesised on the linker. While not often described as a safety catch linker methionine requires very specific conditions to initiate cleavage. When a peptide containing a methionine residue is treated with cyanogen bromide (CNBr) under acidic conditions the methionine residue undergoes a cyclisation to form a homoserine lactone (Figure 82).
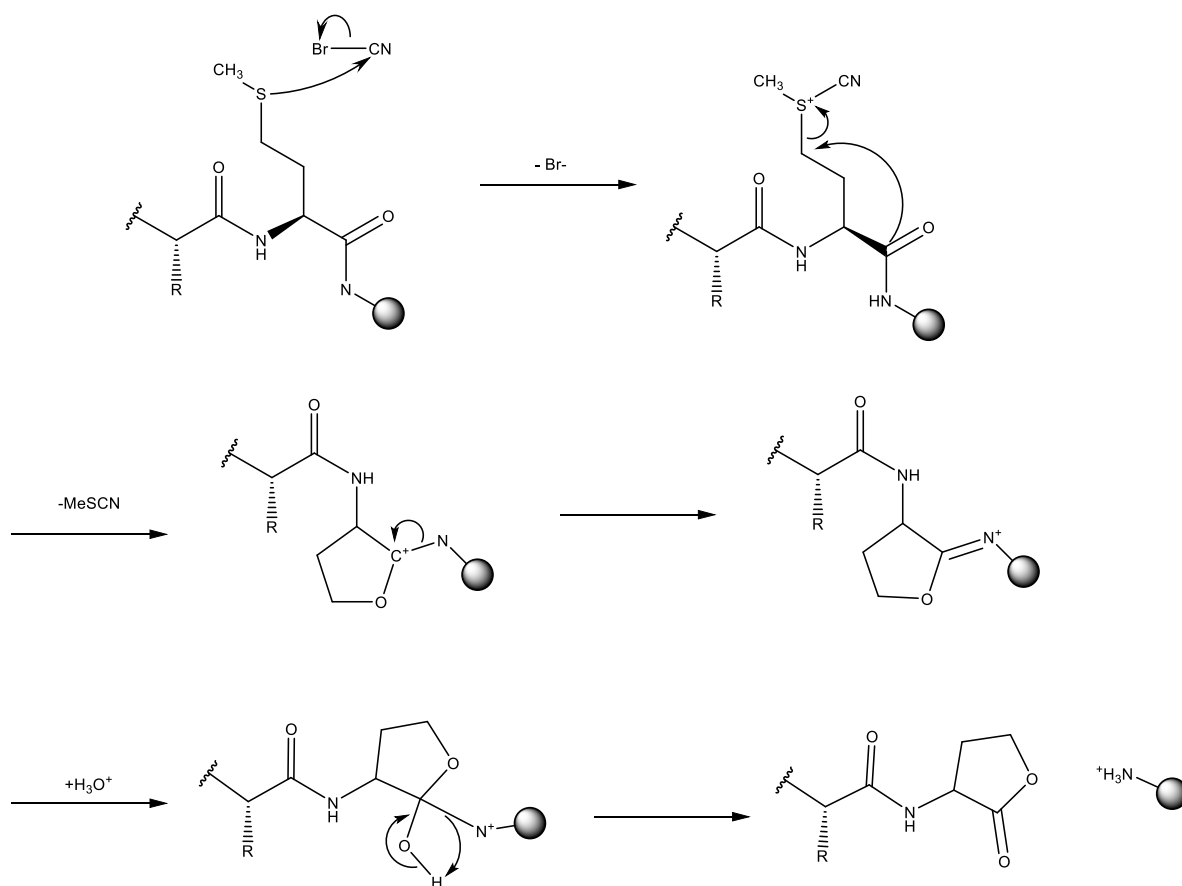
Figure 82 - Mechanism of methionine linker cleavage and formation of the C-terminal homoserine lactone.

This cyclisation separates everything N-terminal to the methionine peptide bond from the resin. The C-terminal lactone could be undesirable in the synthesis of some peptides, the structure is bulkier than the amide or amine C-terminus produced when cleaving from Rink amide or HMBA respectively. However, for the purpose of library synthesis for on-bead screening experiments or the synthesis of a C-terminally labelled peptide the structure of the C-terminus is irrelevant. In the instance of producing a library for on-bead screening cleavage is only carried out if a binding event is detected. The C-terminus is bound to the resin, the homoserine lactone is not formed and takes no part in the binding event. In the instance of the synthesis of a C-terminally labelled peptide the C-terminus and the attached fluorophore are separated from the rest of the peptide by a spacer to ensure that the label does not interfere with the binding of the peptide to the target. The Fmoc protected building block of choice used by the Auer lab as a spacer is Fmoc-8-amino-3,6-dioxaoctanoic acid (DOA). The structure is made up of polyethylene glycol (PEG) units, PEG units also make up a large portion of many solid phase resins.
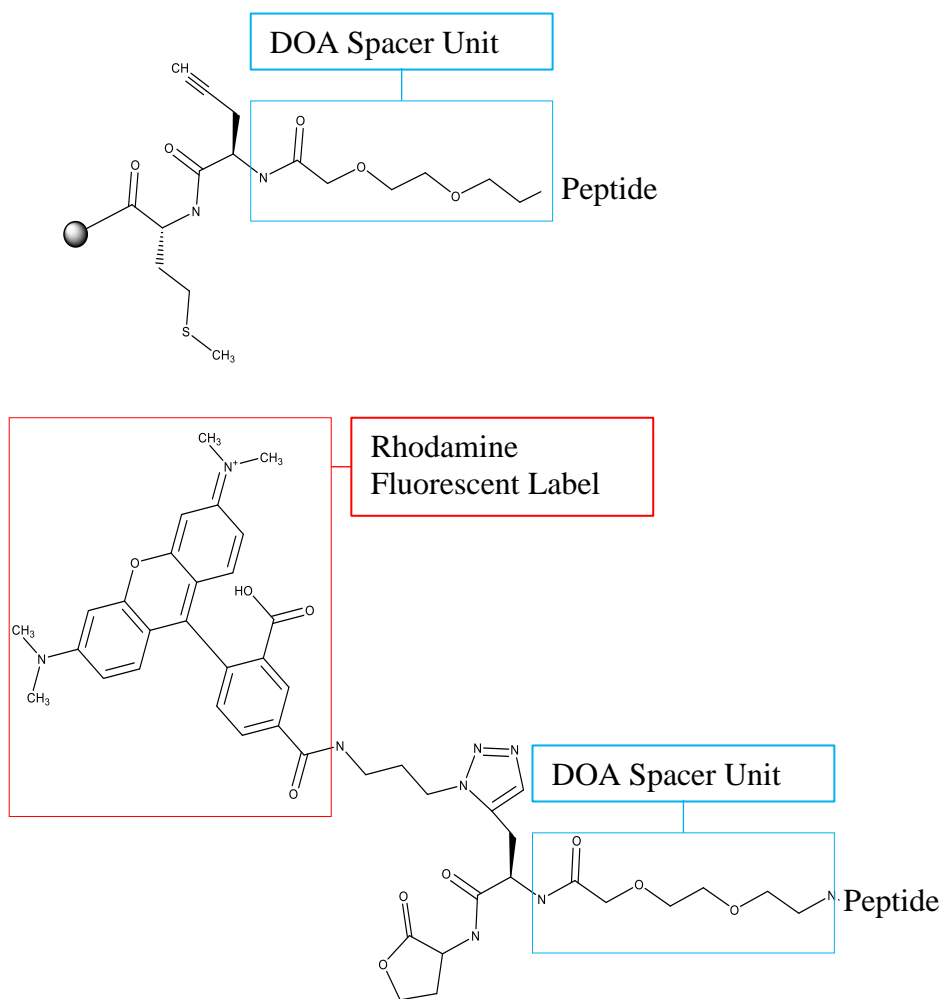
Figure 83 - The DOA spacer separates the peptide from the attachment point or label. This inhibits interference with binding to the target protein during screening.

The DOA spacer unit keeps the synthesis product away from the resin bead and the fluorescent label (TMR in this example), ensuring the resin or label do not interfere with a binding interaction between the synthesis product and target protein.

In this project the use of methionine as a solid phase synthesis linker is characterised with regards to its suitability with the common chemistries carried out in the Auer lab. To this end the methionine linker is compared directly to the SCAL linker, the current lab standard, with regards to the synthesis and cleavage of TMR labelled peptides.

7.2.   Aims

Methods commonly used by the Auer lab often utilise expensive reagents. To combat the cost of carrying out work that is core to our discovery methods it is required that the methods of producing reagents or the selection of reagents be open to criticism and improvement. Tetramethylrhodamine is an expensive yet highly valued reagent for screening experiments, however published methods of rhodamine synthesis are generally time consuming and result in low yields. A new synthetic route to tetramethylrhodamine developed by the Auer lab aims to solve this issue. The novel synthetic route was to be exemplified as a rigorous, reproducible method. Following this the isomerically pure TMR was to be functionalised with an azide group for incorporation into the copper catalysed click reaction commonly used for the labelling of peptides.

Solid phase synthesis methods are often used by the Auer lab to produce peptides, on-bead compound libraries and chemical tools. In these syntheses it is important to select an appropriate method of attachment to the solid phase. For the purposes of library production it is necessary to select a chemical linker that is resistant to the chemistries required to synthesise and deprotect the library without releasing the product in the absence of specific cleavage conditions. A variety of selectably cleavable linkers are available, however several publications have shown that the cheap amino acid building block methionine can be used as a chemical linker. The nature of methionine as a solid phase linker was to be characterised with regards to the common chemistries of solid phase synthesis and its suitability reviewed.

7.3.   Results

7.3.1. Synthesis of Tetramethylrhodamine-acid (6-isomer)
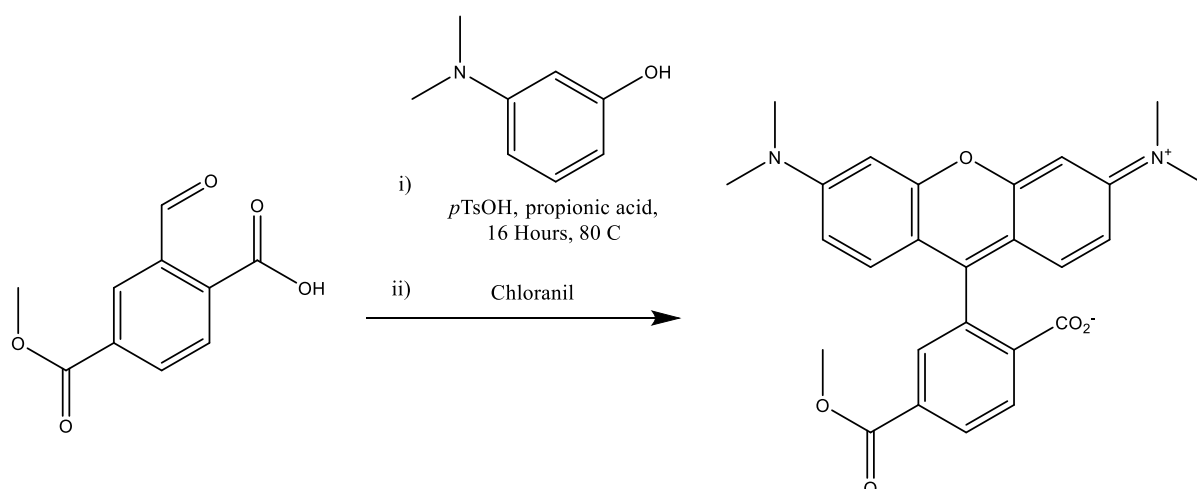
Figure 84 - Scheme for the synthesis of TMR-OMe

The use of 2-Formyl-4-methoxycarbonyl-benzoic acid as the reaction starting point in place of asymmetrical anhydride led to the isomerically pure synthesis of the TMR methyl ester in good yield HPLC analysis allows for the easy tracking of the progress of the reaction as peaks for the all intermediates resolve well and are identifiable. The methyl ester was purified by column chromatography (70 % yield).
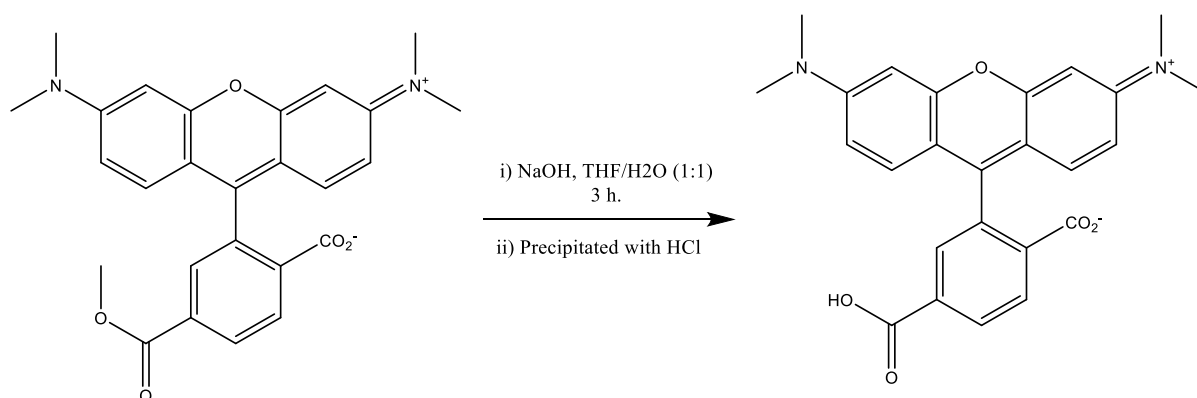
Figure 85 - Scheme for the synthesis of TMR-COOH

The methyl ester was saponified to produce the free acid form of TMR which was then precipitated with HCl and the solution filtered after 3 days. No further purification was required (89 % yield).
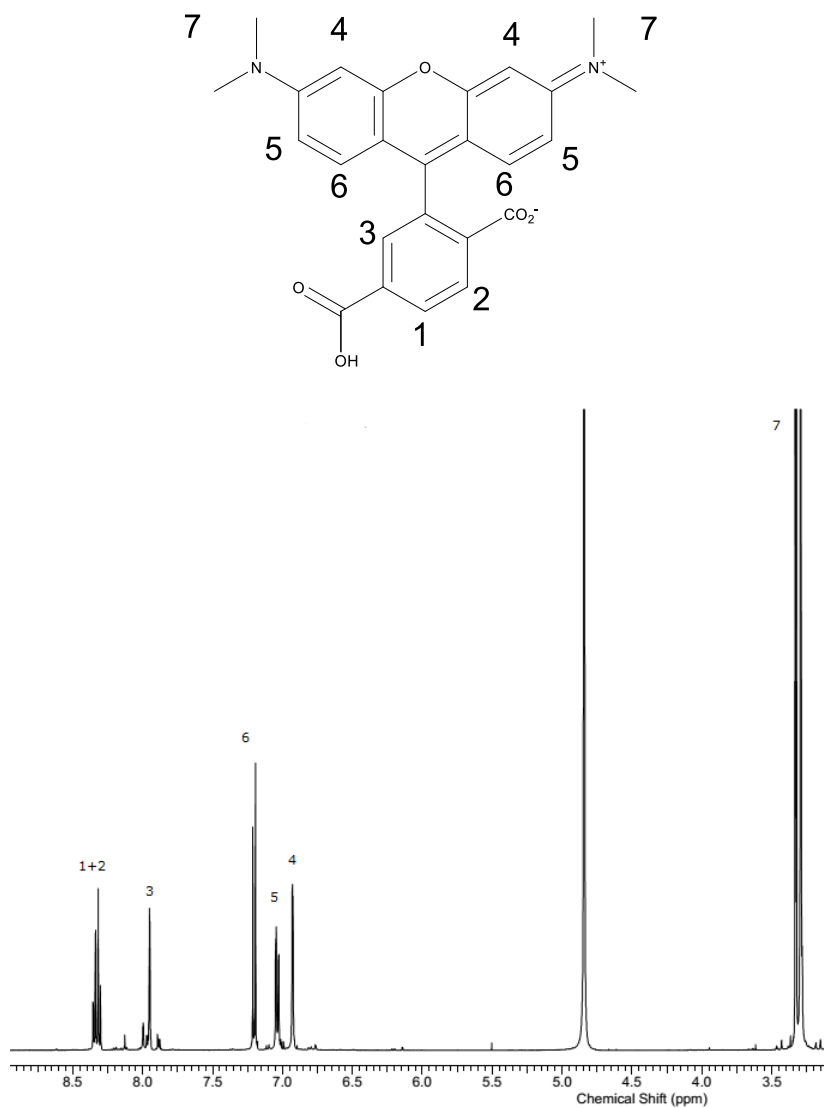
Figure 86 - H¹ NMR analysis of TMR-COOH. All peaks are assigned, indicating the presence of a single regioisomer.

NMR analysis indicates that a single isomer of TMR has been synthesised

7.3.2. Screening of reagents for amide coupling for azide functionalisation

In order to monitor the coupling reaction by HPLC TMR-COOH and TMR-N$_3$ standards were tested.

HPLC analysis of TMR (3 % solvent gradient)
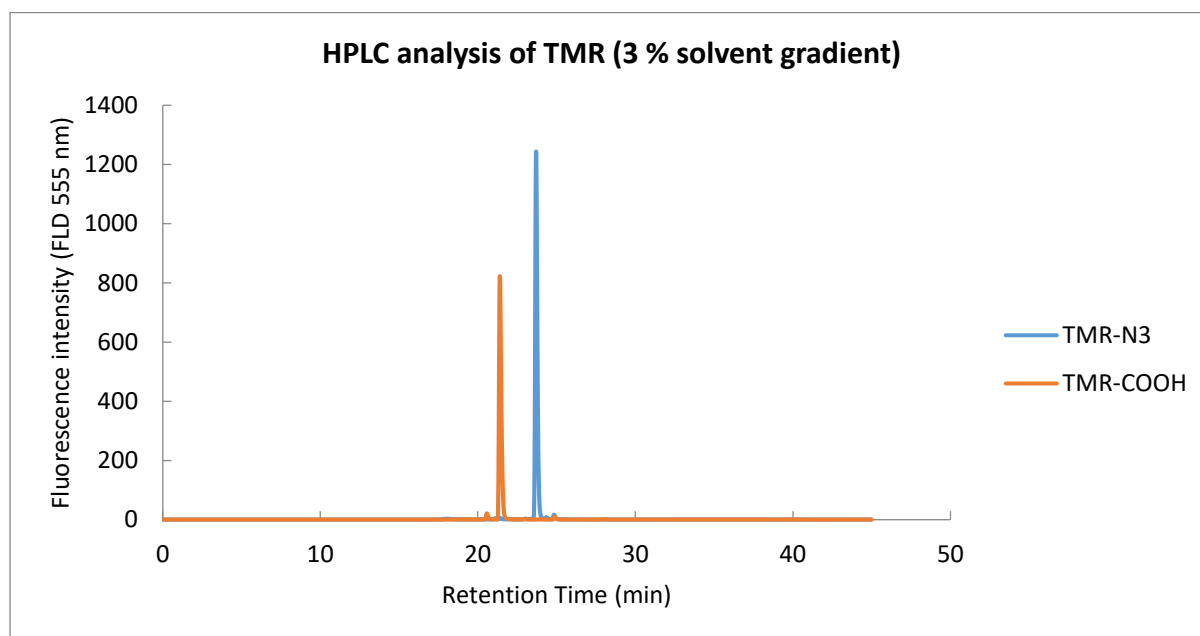
Figure 87 - HPLC analysis of TMR-COOH and TMR-N$_3$

The analysis determined that the two forms of the dye resolved well with retention times of 22 minutes and 24 minutes for TMR-COOH and TMR-N$_3$ respectively.

Having determined the retention time of the reaction product a matrix of reaction times and reagents was tested. TMR-COOH was incubated with the coupling reagents for 5 min before 3-azidopropylamine was introduced to the reaction vessel. Reactions were monitored after 1, 2, 4, 8 and 24 hour intervals.

Table 33 - Results of time course testing of coupling reagents for amide coupling.

% Area of azide peak (HPLC, 555 FLD)

Time (Hr)	HATU	DIC	PyBop	EDC	TSTU
1	68.62	28.02	6.46	16.32	25.06
2	82.56	27.30	5.52	28.75	28.68
4	82.36	28.55	9.47	34.07	28.51
8	82.35	34.03	8.50	42.83	26.60
24	68.49	42.38	-	50.28	26.95

The product was purified by preparative HPLC yielding pure TMR-N$_3$ (40 % yield).

Over the course of this thesis the synthesis of both the TMR-COOH and TMR-N$_3$ were repeated several times with reproducible results.
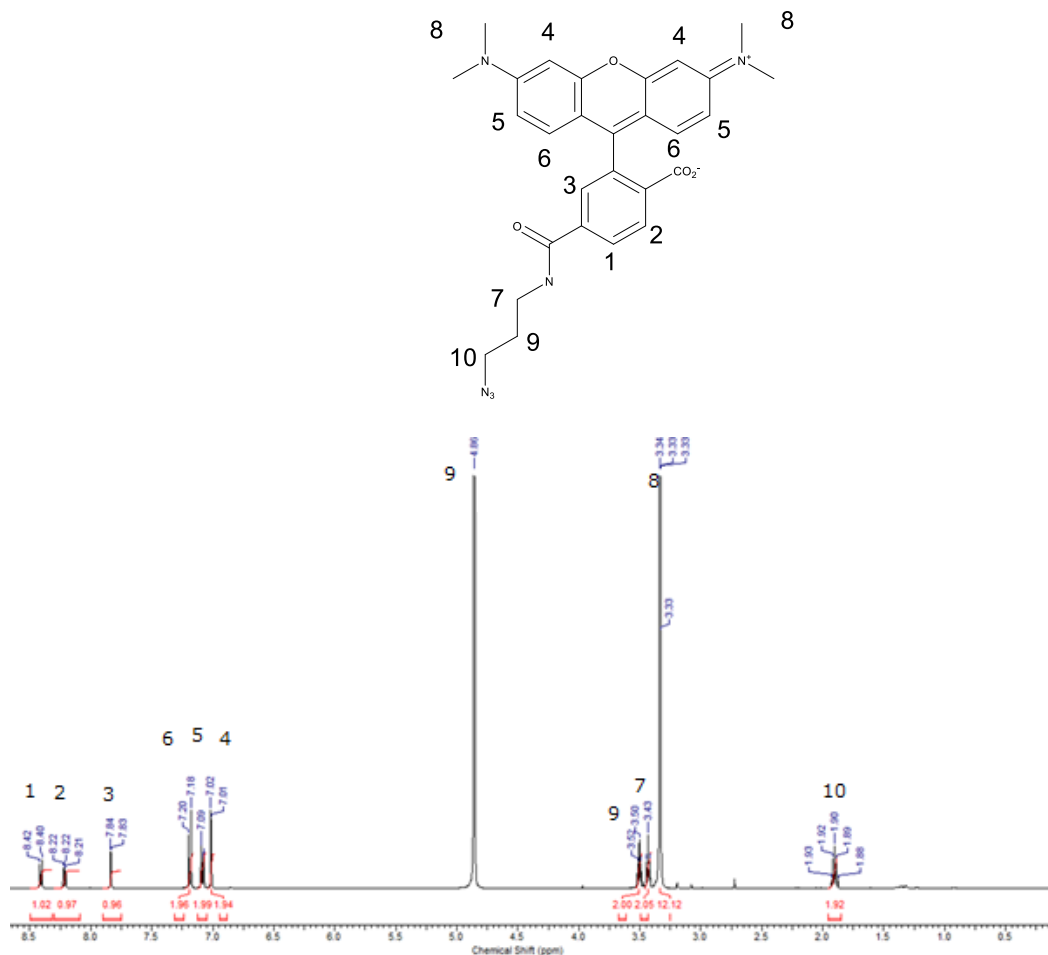
Figure 88- H¹ NMR analysis of TMR-N$_3$

7.3.3. Methionine linker cleavage method testing

Testing of acid activating agents in the synthesis of TMR-N$_3$.

A test peptide TMR*-Ala-Met was synthesised on Tentagel NH$_2$ resin. 4 known literature methods of methionine peptide bond cleavage were tested for single bead cleavage:

Method A: 15 μl of 20 mg/ml CNBr in 0.1 M HCl, 16 hours.[159]

Method B: 15 μl of 20 mg/ml CNBr in 70 % TFA, 30 % water, 16 hours.[160]

Method C: 15 μl of 100 mg/ml CNBr in 70 % formic acid, 24 hours.[161]

Method D: 15 μl of 30 mg/ml CNBr in propionic acid, 24 hours.[28]

Methods were tested on 10 instances of single beads each.

Table 34 - Results of HPLC analysis of methione linker cleavage methods A and B.

Methionine Linker Cleavage Method A			
Sample	RT	Area	%area
1	20.295	4560.9	67.729
2	19.351	4093.7	70.769
3	19.651	6519.0	72.008
4	19.613	7114.7	68.541
5	19.836	5464.3	71.260
6	19.357	3692.2	66.296
7	19.743	4733.0	69.999
8	19.840	4370.8	68.116
9	19.280	3212.6	67.052
10	19.579	6170.1	71.793
avg	19.655	4993.1	69.356
stdev	0.301	1280.2	2.069

Methionine Linker Cleavage Method B			
Sample	RT	Area	%area
1	20.361	2563.3	93.560
2	20.257	2026.8	90.789
3	20.220	1196.4	92.046
4	18.596	1846.9	92.756
5	18.873	1678.3	90.404
6	19.269	2671.3	91.158
7	19.642	817.3	85.884
8	19.284	1481.8	90.128
9	19.599	2345.7	93.338
10	19.451	2911.0	93.203
avg	19.555	1953.9	91.327
stdev	0.591	678.0	2.294

Methionine Linker Cleavage Method C			
Sample	RT	Area	%area
1	19.914	5492.2	84.186
2	20.128	5022.2	85.034
3	19.498	6966.0	95.464
4	18.988	8480.0	85.876
5	19.717	8325.9	87.424
6	19.403	8686.4	83.667
7	18.783	6743.7	85.281
8	19.117	9677.5	93.982
9	19.001	6452.9	93.063
10	21.022	6322.7	93.263
avg	19.557	7217.0	88.724
stdev	0.673	1509.3	4.642

Methionine Linker Cleavage Method D			
Sample	RT	Area	%area
1	19.963	154.4	80.362
2	18.413	37.3	84.055
3	17.716	120.5	79.944
4	19.076	142.9	77.048
5	19.185	114.7	84.621
6	19.214	196.0	88.479
7	18.715	43.9	72.391
8	19.251	119.5	89.554
9	19.168	39.2	69.171
10	19.319	147.0	82.113
avg	19.002	111.5	80.774
stdev	0.604	54.5	6.527

Methods A and C released a higher concentration of material but analysis indicated 2 fluorescently labelled products being released from the resin on cleavage. Analysis of the minor peaks by LCMS gave m/z values of 18 units higher. Method D released a very low concentration of fluorescent label from the resin in comparison to the other methods.

Method B (15 µl of 20 mg/ml CNBr in 70 % TFA, 30 % water, 16 hours.) proved to be the method that provided the highest purity but with a lower cleavage efficiency compared to A

and C. Prioritising purity over quantity, method B was used as a starting template method from which reaction time and temperature were varied. For each set of conditions 10 repetitions of 20 beads was tested. The resulting peptide was dissolved in 3 ml of water and the absorbance at 555 nm was recorded via cuvette.

Table 35 - Results of varying reaction times and temperature, based on Method B of methionine linker cleavage.

Method	Average ABS (555nm)	pmol/bead
SCAL standard	0.3083	809.821
2 hours, room temp	0.0488	128.196
16 hours, room temp	0.1815	476.795
40 hours, room temp	0.1187	311.821
1 hour, 40 ºC	0.1782	468.126
2 hours, 40 ºC	0.2296	603.152
16 hours, 40 ºC	0.0839	220.403
30 mg CNBr, 2 hours, 40 ºC	0.1974	518.564
50 mg CNBr, 2 hours, 40 ºC	0.1484	389.842
100 mg CNBr, 2 hours, 40 ºC	0.1212	318.389

Table 35 shows that 2 hours at 40 °C releases the highest concentration of peptide. A sample of the same resin was used to synthesise the same test peptide on the SCAL linker. The standard linker cleaves with a higher efficiency compared to even the highest yielding methionine cleavage method.

Having confirmed the best methionine cleavage conditions the new method was tested against the 20 natural amino acids coupled to methionine. This experiment aimed to test for undesirable side reactions that could lower the purity of peptides produced on the methionine linker.

Table 36 - Results of MS analysis of amino acids bound to the methionine linker and cleaved using the adjusted Method B.

Amino acid	Expected m/z	Reported m/z	Amino acid	Expected m/z	Reported m/z
Ala	394.15	395.20	Cys	426.12	475.11
Arg	480.22	480.47	Gly	380.14	381.06
His	460.17	461.35	Pro	420.17	421.20
Lys	452.21	452.35	Val	422.18	423.30
Asp	438.14	439.17	Leu	436.20	437.26
Glu	452.16	453.18	Ile	436.20	437.28
Ser	410.15	411.09	Met	-	-
Thr	424.16	425.08	Phe	470.18	471.25
Asn	437.16	438.21	Tyr	486.18	487.24
Gln	451.17	452.24	Trp	509.20	425.17

The resulting masses indicate that the cleavage conditions generally leave the residues untouched with few exceptions. Notably the mass for tryptophan was lower than expected, the m/z matches the expected mass of the C-terminal amide form of the amino acid. Cysteine was also reported to have a significantly higher mass than expected.

7.4. Discussion and Future Work

7.4.1. The synthesis of TMR and azide functionalised TMR

The aim of this project was to synthesise and purify TMR-COOH and TMR-N$_3$ for use in common Auer lab processes. This goal was achieved through the application of a novel synthetic route towards xanthene cored dyes developed in the Auer lab. The novel synthetic route developed by Gemma Mudd was quickly adopted as the new lab standard for rhodamine synthesis. The method has been shown to reproducibly provide isomerically pure dye products and was used to produce all of the TMR labelled peptides that appear in this thesis. Since the publication of this method further improvements to the synthesis of TMR have been published in a method from Dwight *et al* from Promega Biosciences.[156]
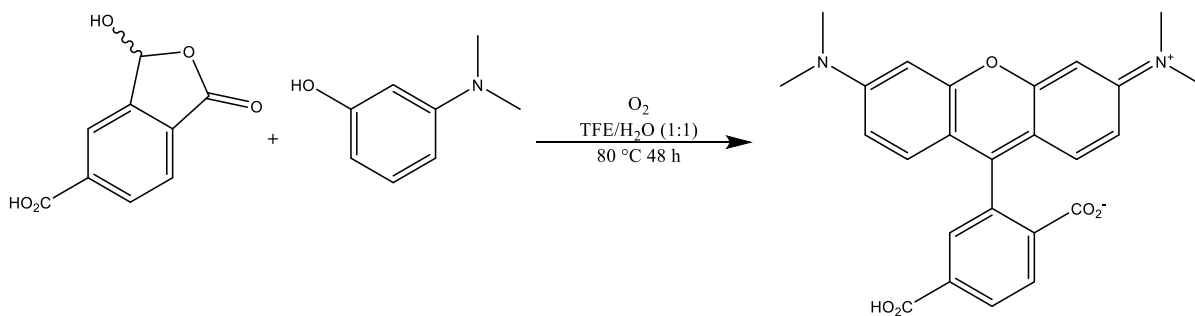
**Figure 89 - Reaction scheme describing the synthesis of 6-TMR-COOH by Dwight *et al.*

The article describes the method as being applicable to the synthesis of a variety of rhodamine and rosamine dyes at a gram scale level. This more recent method avoids the need to synthesise the acid via isolating and saponifying an intermediate methyl ester. Future efforts to synthesise rhodamine dyes would be best carried out trialling this one-pot method to compare the practical differences of the two methods. Between these two methods an investigation into the effects of different groups on the properties of rhodamine dyes.

7.4.2. Methionine linker

The use of methionine as a linker has been reported in literature to have been used for a variety of peptide synthesis projects with general success. This project aimed to specifically characterise and review the applicability of methionine as a linker for use in Auer lab on-bead synthesis projects. Of the 4 methods tested only 2 showed a reasonable degree of labelled product being freed from the resin on cleavage. Method A showed 2 TMR labelled cleavage products when tested by HPLC. LCMS analysis showed that the minor peak had an m/z 18 higher than the expected main peak. This would suggest the addition of water to the structure, possibly indicating that the homoserine lactone has opened to form acid and alcohol groups.
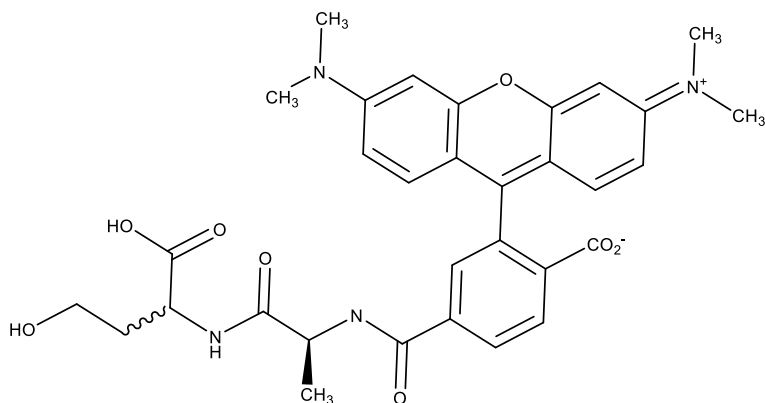
Figure 90 - Suggested structure of minor Method A HPLC peak.

Having 2 cleavage products can lead to confusion during analysis and therefore makes method A unacceptable for common use. Method B showed better purity but a weaker concentration of cleaved product. Variations of the method were tested in an attempt to improve the reaction. Through testing it has been shown that methionine, as expected, is inert to standard Fmoc solid phase peptide synthesis conditions. The literature cleavage method was significantly shortened from 16 hours to 2 hours through heating, making the cleavage method much more appealing from a practical standpoint. Additionally, the testing of the 20 natural amino acids showed that the CNBr containing cleavage conditions generally does not have an effect on the purity of peptides synthesised on the methionine linker with few exceptions. As expected testing of methionine led to multiple peaks in the analysis, however as mentioned previously methionine is generally excluded from library synthesis. The cysteine containing test peptide was reported as having a higher mass than expected.
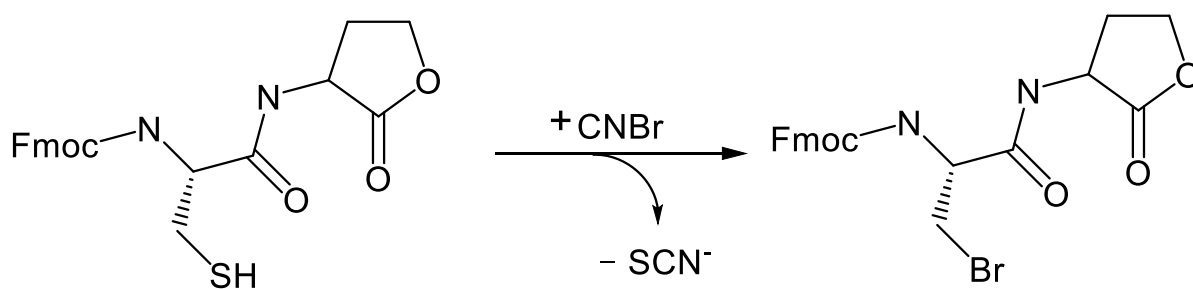
Figure 91 - Scheme of possible bromination of cysteine by CNBr.

Figure 91 offers a possible explanation for the difference in mass. In Figure 82 the mechanism of the methionine linker requires the attack by the sulphur atom to the cyanide group. It is feasible that the thiol of cysteine acts in a similar manner, attacking the cyanide group and forming a thiocyanate. A bromide ion can then attack and displace the thiocyanate ion. Cysteine is omitted as a building block in library synthesis because of its reactivity, its incompatibility with the methionine linker cleavage method is not expected to be a significant hindrance in its expected applications. The lower mass of the tryptophan containing peptide corresponds with the expected mass of the peptide with a C-terminal amide as opposed to the expected homoserine lactone.
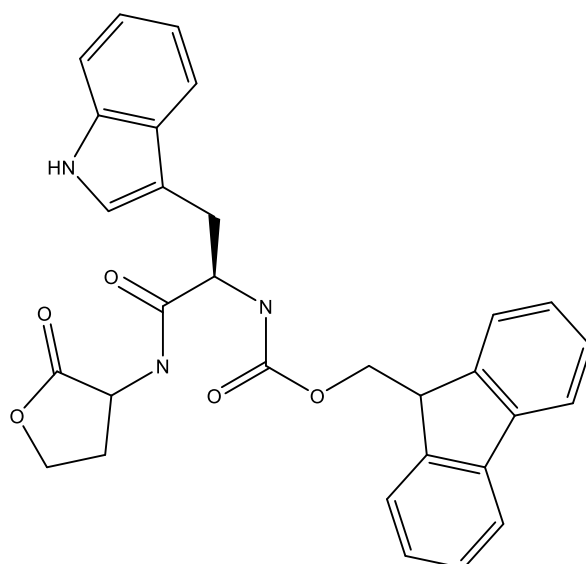
Chemical Formula: $C_{30}H_{27}N_3O_5$
Exact Mass: 509.20
m/z: 509.20 (100.0%), 510.20 (32.4%), 511.20 (2.7%), 511.20 (2.4%), 510.19 (1.1%), 511.20 (1.0%)
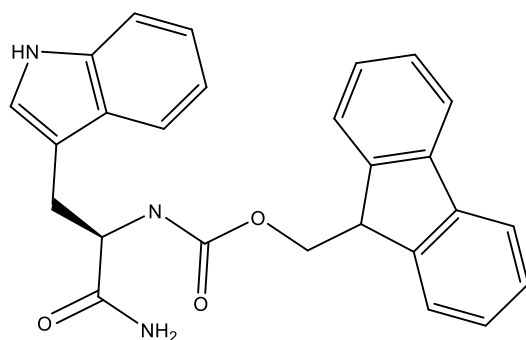
Chemical Formula: $C_{26}H_{23}N_3O_3$
Exact Mass: 425.17
m/z: 425.17 (100.0%), 426.18 (28.1%), 427.18 (2.7%), 426.17 (1.1%), 427.18 (1.1%)

Figure 92 - Structures and masses of Fmoc-tryptophan-M* and Fmoc-trptophan-CONH₂ where M* denotes the homoserine lactone formed by methionine linker cleavage.

A search of the literature finds that CNBr is used in peptide fragmentation at tryptophan residues, however the reagents used suggest that the mechanism requires that the tryptophan residues are oxidised for the cleavage to occur.[162] A microwave method for the cleavage of the methionine linker has been reported that oxidation of the tryptophan residue was avoided by carrying out the reaction under argon.[163] This adjustment to the method would be simple enough to integrate when working with tryptophan containing peptides.

7.4.2.1. Future Work

In the future the Auer lab could consider trialling microwave synthesis methods, they are well documented in their application towards peptide synthesis. Microwave irradiation has also been applied for linker cleavage. Generally, it is accepted that microwave irradiation is an effective method for significantly reducing the reaction times for Fmoc deprotections, building block couplings and cleavage reactions.[164] Microwave-assisted methods also exist for the synthesis of Peptoids (peptides where the R groups are shifted to the nitrogen of the backbone)[165] and for disulphide bonding for application in the synthesis of cyclic peptides[166]

This characterisation project has shown that methionine could be used as a potential alternative to other, more expensive linkers for the purposes of solid phase synthesis of chemical tools or libraries. Given the harsh cleavage conditions it would likely be a reasonable step to test building blocks for compatibility with the methionine linker prior to their use in synthesis.

8. Closing Discussion and Future Work

8.1. WD40 Domain Containing Protein Database and Target Selection

The first piece of work carried out for this thesis was an exercise in information gathering. Collecting data on WD40 domain containing proteins; their structures and their roles in health and disease. This piece of work became an ongoing effort to catalogue all human WD40 domain containing proteins. Through this effort the roles of WD40 domains in nature quickly became clear, WD40 domains are important scaffolds on which protein complexes are formed. Despite having a common structural motif and similar roles different WD40 domains are highly specific in their binding to interaction partners. By investigating the different pathways and interactions WD40 domain were involved in several WD40 domain containing proteins were determined to be promising as research targets. WDR5 stood out as a high value, priority target largely due to the large of literature available on the interactions of WDR5 and roles that the complexes it forms play in cancer.
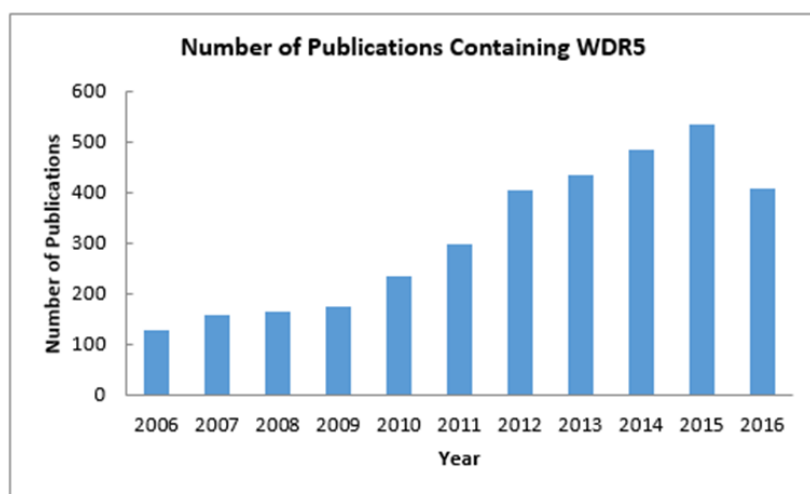
Figure 93 - number of publications containing "WDR5" per year, 2006-2016.

As the research described in this thesis continued the number of publications on the topic of WDR5 began to increase (**Error! Reference source not found.**), validating WDR5 as a target a nd offering valuable input for *in-silico* techniques in the form of protein structures and known binders. However, while research on WDR5 became more and more prevalent in the literature, other WD40 domain containing proteins that could be of interest remain neglected.

For example the protein WD40 domain containing protein 39 (WDR39, also known as Ciao1) is known to bind to Wilms Tumour suppressor protein 1 (WT1). The binding of WDR39 to WT1 is known to modulate its anticancer activity.[167] In comparison to WDR5 which has had hundreds of papers in the past decade WDR39 has only 64. With 283 different WD40 domain containing proteins WDR39 is just one example of many different potentially interesting research targets. Database currently holds information on all human WD40 domain containing proteins. The data collected shows that only a small number of WD40 domain containing proteins have known inhibitors and some even have very limited information available with regards to their role in human biology.

In summary the data collection exercise successfully determined several potentially interesting targets to focus research interests on. Through further research it was determined that WDR5 would offer the highest potential out of the shortlist of proteins. In future work the compiled information could be used to find other WD40 domain containing proteins that could be targeted for the modulation of different pathways.

8.2.   Targeting of WDR5 via **In-silico** Techniques

Two in-silico techniques were used to determine a short list of potential binders to WDR5. The first of these methods was a structure-based virtual screening platform named Q-mol, developed by Anton Cheltsov. Q-mol determined possible binding sites on WDR5, through conformer exploration and molecular dynamics. The binding sites identified by Q-mol also included known protein-protein interaction sites on the top and bottom interface. Anton Cheltsov, using Q-mol, then docked the NCI compound library against these sites and determined several potential binders to these pockets. These compounds were screened against WDR5 in TDA and of the 66 compounds, one was determined to stabilise WDR5 in the assay. This compound, NCI292249, was further characterised by microdialysis where it was determined to have a K_D of 564 μM to WDR5. However, the control experiment with BSA showed a similar affinity (452 μM), suggesting that the compound binds non-specifically to protein. The second in-silico method used was USRCAT, a molecular similarity technique. Molecular similarity compares descriptions of compounds as opposed to structures. This allowed for the comparison of known binders to WDR5 to compounds available from our collaboration partner Vichem, who do not divulge the structures of compounds in their

libraries. USRCAT determined several compounds that had similar molecular descriptions to known binders to WDR5, the top 15 of which were screened in TDA against the target. In screening none of the compounds were determined to have a stabilising effect on WDR5. The results of this screening effort highlight the difficulties in determining a binder to protein-protein interaction targets. The hurdle of determining a starting point from which an optimised binder can be developed proved to be difficult to overcome through small molecule screening without screening a large number of compounds. Further work could possibly be carried out where *In-silico* techniques are used to target WDR5, however a different library of small molecules could be docked against the target.

8.3.   SOBOC and MORPH

Rather than continue investing in small molecule screening efforts it was decided that efforts should be focused on investigating the known interactions of WDR5 and to use these as starting points. The top interface of WDR5 has been targeted by Karatas *et al.*[66] through the development of a binder based on natural interactions at this site, this inspired the targeting the bottom interface (The MYC-interacting interface) in the same manner. The bottom interface has structural information available for a number of interactions, however a publication by Thomas *et.al*[12] had begun probing the contribution of residues through alanine scanning of residues of a peptide sequence conserved in all MYC proteins known in interact with WDR5. This sequence is known as the MYCIIIb sequence.[15] In an effort to build on this research it was determined that the next logical step would be to screen fragments of the MYCIIIb peptide. In parallel with this MorPH software had been developed and trialled against Survivin. During the course of the work to develop MorPH peptides there was discussion on how to best determine what stretch of a long peptide would be best to use as input for a MorPH guided SAR investigation. By synthesising fragments of a known peptide binder we could investigate the contributions of different residues to the MYC-WDR5 interaction and also determine a possible starting point for a MorPH guided SAR project. However the synthetic effort of manually synthesising 45 individual peptides in separate syntheses seemed inefficient. Fmoc solid phase peptide synthesis produces all N-terminal truncates in the course of the synthesis of a full length peptide. These truncates are easily isolated by removing a sample of resin from the reaction vessel after each amino acid is

coupled. By synthesising a series of C-terminal truncates using Fmoc chemistry and isolating each N-terminal truncate as they were produced the synthesis was made significantly and faster. These peptides were then labelled and screened against WDR5 in an on-bead screening format. The original plan was to use labelled WDR5 and labelled peptides in a 2 colour screen to allow for an accurate ranking of the peptide fragments. It was however determined that chemical labelling of WDR5 is not tolerated. A search of the literature shows that other screens involving WDR5 have been carried out using fluorescently labelled antibodies as opposed to chemical labelling or fluorescent-protein conjugates. While this was not an ideal outcome the screening was carried out using unlabelled WDR5 but still allowed for the ranking of hits via the ring intensity of labelled peptides. The data produced from this screen allowed for the evaluation of the contributions of amino acids to the MYC-WDR5 interaction. A previously published alanine scan of the WDR5 binding sequence from MYC found that the "IDVV box" of the sequence was the key contributor, and that the replacement of other residues had a comparable effect, suggesting that all other residues contribute almost equally. The *in-silico* work carried out to supplement this investigation suggested that the IDVV sequence is also at the heart of the MYC-WDR5 interaction. However the results of this investigation suggested that the acidic residues N-terminal to the IDVV box are also important contributors to the interaction. This theory is reinforced by comparing to the sequences of other binders to the bottom interface of WDR5, such as RBBP5. These other binders also have an IDVV box with N-terminal acidic residues. Through this screen it was determined that the sequence EEEIDVV was the highest affinity out of fragments of a size of 7 amino acids. This peptide was carried through to solution phase characterisation using fluorescence anisotropy. The peptide was found to have a K_D of 93 µM, the literature affinity of the full length peptide (DEEEIDVVSVE) is 10 µM.[168] The truncated peptide was considered to be a reasonable starting point for an SAR investigation, as further removal of residues would almost certainly lead to a further decrease of affinity. MorPH software was utilised to determine possible replacements for the natural residues in the WDR5 interaction sequence from MYC. At the time MorPH had been undergoing testing using the protein Survivin as a target. The best amino acid replacements for each position had already been determined, a final resynthesis of exemplar peptides was required to show the effects of iterative replacement of amino acid residues with non-natural replacements from MorPH. It was shown that overall there was an

increase in affinity from using suggestions from MorPH, increasing the affinity from 7.35 μM to 3.03 μM. Plasma stability was also increased, significantly in the case of abu-CarbPhe-E-R. This was a promising result for the first trial of MorPH and encouraged its application towards developing a novel tool compound for the WDR5-MYC interaction. Evaluation of the suggestions from MorPH for the WDR5 interaction sequence from MYC resulted in the determination of several promising replacements. This acts as an outline for a possible future SAR project for the targeting of WDR5, which would also allow for further evaluation of MorPH.

8.4. Characterisation of Methionine as a Solid Phase Linker and Synthesis of TMR-N$_3$

Finally, in chapter 7, two research outputs were described. The first was a contribution towards the determination towards a novel synthetic route to rhodamine dyes. This synthetic route was the outcome from the efforts from several chemists in the research group, the resulting product is used in the majority of screening experiments in the lab. Functionalisation with an azide group allows for the incorporation of the dye in the azide-alkyne Huisgen cycloaddition reaction,[27] as was the case with the labelling of the peptides synthesised for the SOBOC project. The second project was the characterisation of methionine as a linker in solid phase peptide library synthesis. The use of methionine as a linker has been described in literature, however several different cleavage conditions are also described. At the time there was a need for the testing of a different linker suitable for solid phase library synthesis that would allow for selective cleavage from resin. The current lab standard required a synthesis that was 11 steps long and gave a 22 % yield. Methionine by comparison is commonly available from suppliers as an Fmoc protected peptide synthesis building block. Several cleavage methods for the methionine linker were tested, the method that resulted in the highest purity was used as a starting point and further improved. This final method was then tested against the 20 natural amino acids to verify if any undesirable side reactions were likely to occur. As expected methionine and cysteine did not tolerate the cleavage conditions, tryptophan was also noted to have an unexpected side reaction during cleavage. The final cleavage conditions are harsh however, future investigations into possible variations on these conditions could allow for the preservation of tryptophan but likely at the cost of yield.

9. Materials and Methods

9.1.   General Methods

NMR Analysis

^{1}H NMR spectra were recorded on a Bruker AVA500 (500 MHz), Bruker Pro500 (500 MHz), Bruker AVA600 (600 MHz) or a Bruker AVA400 (400 MHz) spectrometer. Chemical shifts (δ) are quoted in parts per million (ppm) downfield of tetramethylsilane, using residual protonated solvent as internal standard ($CDCl_3$ at 7.27 ppm). Abbreviations used in the description of resonances are: s (singlet), d (doublet), t (triplet), q, (quartet), app (apparent), br (broad). Coupling constants (J) are quoted to the nearest 0.1 Hz. Proton-decoupled ^{13}C NMR spectra were recorded on a Bruker AV500 (125.8 MHz) spectrometer or a Bruker AVA400 (100.6 MHz) spectrometer. Chemical shifts (δ) are quoted in parts per million (ppm) downfield of tetramethylsilane, using deuterated solvent as internal standard ($CDCl_3$ at 77.0 ppm).

HPLC-MS Analysis

Samples were analysed on an HPLC-ESI-MS system consisting of a Finnigan Deka XP Plus ESI-MS detector connected to an Agilent 1100 capillary LC with a (G1376A) capillary pump, a (G1377A) Micro-autosampler and a (G1325B) DAD detector with micro flow cell. The system was run with a spray voltage of 5 kV, capillary temperature of 275 °C, a capillary voltage of 15 V in single MS mode. The LC-system contained a Zorbax-SB-C18 0.5x35mm 3.5 μm particle size column and was run with a gradient of solvent A: H_2O containing 0.1% trifluoracetic acid (TFA) and B: MeCN (LCMS grade) containing 0.1% TFA. Standard Gradient: 2.5 min 0% B, 22.5 min 0 - 100% B, 25 min 100 - 0% B, 28 min 0% B at a flow rate of 200 μl/min. HPLC-ESI-MS data were processed using the Xcalibur software package (version 2.0, Thermo Electron Corporation, MA, USA).

HPLC analysis

HPLC analysis was performed on an Agilent 1100 series HPLC system, consisting of a quaternary pump (G1311A), a degasser (G1322A), an FLD detector (G1321A), a DAD detector (G1315B) and a 100 sample well plate autosampler (G1367A); columns: Agilent Zorbax C18 Vydac peptide C18, 4.6 mm x 150 mm, 3.5 μm particle diameter size for peptide and small molecule analysis. Vydac C4 for protein analysis. Analyses were performed using a linear

gradient of A: H$_2$O containing 0.1% TFA and B: MeCN (HPLC grade) containing 0.1% TFA with a flow rate of 0.8 mL/min, Retention times (t$_R$) are denoted in minutes. All analyses use standard gradient unless stated otherwise. Buffers: A: water + 0.1 % TFA, B: acetonitrile + 0.1 % TFA. Standard gradient: 5 min 5% B, 25 min 5 – 95% B, 27.5 min 95 – 5% B, 30 min 5% B. Plasma Stability Gradient: 11 min 5 - 90 %, 5 min 90 % B, 3 min 90 – 5 % B, 3 min 5 % B.

Semi-Preparative HPLC Purification

Purification of labelled peptides and mg quantities of dyes was carried out using a preparative HPLC system (Agilent 1100 prep-HPLC system), equipped with a preparative autosampler (G2260A), preparative scale pumps (G1361A), a fraction collector (G1364B-prep) and a multiwavelength UV detector (G13658 MWD with preparative flow cell). Material was separated at a flow rate of 20 mL/min on an Agilent RP-C-18 column (21.2 x 150 mm, 10 µm particle size), using a H$_2$O/MeCN gradient. Solvent A: H$_2$O, 0.1% TFA; solvent B: MeCN, 0.1% TFA. The size of collected fractions was varied for each purification.

9.2.   Small molecule screening methods

Expression, purification and characterisation of 6xHis-WDR5

Plasmid was prepared by Kris Wilson via a miniprep of Top10 cell stock from Addgene (Structural Genomics Consortium clone: WDR5 (2H9L)). 20 µl of chemically competent BL21 cell stock was incubated with 1 µl of 6xHis WDR5 pet28a plasmid for 10 minutes on ice. Cells were heat shocked at 42 °C for 30 seconds before being cooled on ice for a further 20 minutes. 350 µl of LB broth was added to the cells. Cells were incubated at 180 rpm, 37 °C for 1 hour (Infors-HT shaker unit). 200 µl of the cell culture was plated onto a 50 µg/ml kanamycin LB agar plate and incubated for 16 hours at 37 °C.

A culture from the overnight plate was added to 20 ml 2xTY nutrient broth with 50 µg/ml kanamycin. Culture was incubated at 37 °C, 180 rpm for 16 hours. Culture was scaled up to 1000 ml and incubated for a further 2 hours at 37 °C before the temperature was reduced to 16 °C for 5 hours. Expression was induced with 1 mM IPTG, the culture was incubated for a further 16 hours. Culture was centrifuged at 5000 rcf for 20 minutes (Sorvall RC 5C Plus). Cell pellets were suspended in 40 ml of buffer (20 mM imidazole, 50 mM Phosphate, 300 mM NaCl, pH 7.50) with 1mg/ml lysozyme and 1 tablet of cOmplete™ protease inhibitor. The lysis

solution was stored on ice for 20 minutes. The lysis solution was sonicated (Sonics VibraCell CV18) 2x1 min, 10 seconds of 50% amplitude, 10 seconds off. The resulting lysate was centrifuged to remove cell debris (5000 RCF, 20 minutes). The lysate was incubated for 2 hours with washed Ni-NTA agarose resin beads (Quiagen) from 2 ml of 50 % slurry. Resin was isolated from the lysate via centrifugation (5 min at 800 RCF), the resin was transferred to a sintered column and washed with buffer containing 40 mM imidazole, WDR5 was eluted with buffer containing 250 mM imidazole. The fractions were analysed by SDS-PAGE. SDS-PAGE was carried out with Invitrogen Bolt 12 % gels, samples were incubated with Bolt LDS sample buffers at 72 °C, shaken at 500 rpm for 12 minutes prior to loading onto gel. Seeblue Plus2 prestained standard was also loaded onto gels for evaluation of the molecular weight of the protein content of the samples. Electrophoresis was carried out at 200 V for 32 minutes. After electrophoresis gels were washed with distilled water and stained with SimplyBlue safestain overnight. The following day excess stain was removed and the gels washed with distilled water. Fractions containing significantly higher proportions of WDR5 were combined and concentrated via a 15000 dalton spin column to a volume (Sartorious Vivaspin filter) of <500 µl. The concentrated solution was injected into an AKTA protein purification system directly into the injection valve connected to a 1 ml superloop. Protein was purified by a superdex 200 size exclusion column (Amersham Biosciences) connected to the AKTA protein purification system. The mobile phase was 20 mM HEPES, 250 mM salt, pH 7.00 with a flowrate of 0.5 ml/min. Collected fractions were 0.5 ml in volume.

Fractions containing WDR5 combined, determined by AKTA Unicorn software and SDS-PAGE. Combined fractions were concentrated and characterised by SDS-PAGE and reverse-phase HPLC (C4 column). Repeat expressions of 6xHis WDR5 used a glycerol stock from the initial expression that was stored at -80 °C. On average expression produced 5.5 mg of 6xHIS-WDR5.

Protein concentration determination

Calculation of estimated molar extinction coefficient

Tryptophan and tyrosine residues, as well as cysteine-cysteine disulphide bonds offer considerable contribution to the absorption of a protein at 280 nm. However the absorption of these residues is dependent on the local environment, in turn this environment is dependent

on the folded structure of the protein. Therefore it is necessary to determine the concentration of protein under denaturing conditions. By using guanidine hydrochloride the error is reduced to an approximation of the contributions of N-acetyl-L-tryptophanamide, gly-tyr-gly and cysteine dimers, as opposed to the ± 11 % error expected through the calculation of the molar extinction coefficient using values determined for folded proteins. The molar extinction coefficient of a protein in 6M guanidine hydrochloride is calculated as the sum of the molar extinction coefficient of the absorbing residues as established by Gill *et al.*[169] (Equation 2 below).

Equation 2

$$\varepsilon_{280\ nm} = 5690 \times \#Trp + 1280 \times \#Tyr + 120 \times \#Cys$$

Where $\varepsilon_{280\ nm}$ is the molar extinction coefficient at 280 nm (units = M^{-1} cm^{-1}) and #Trp, #Tyr and #Cys correspond to the number of tryptophan, tyrosine and cysteine residues in the protein sequence. Substituting in the frequencies of these residues from WDR5 gives Equation 3.

Equation 3

$$\varepsilon_{280\ nm} = 5690 \times 10 + 1280 \times 9 + 120 \times 10$$
$$\varepsilon_{280\ nm} = 69620\ M^{-1}cm^{-1}$$

Concentration determination

Having calculated the $\varepsilon_{280\ nm}$ for WDR5 in 6 M guanidine HCL the protein concentration was determined via UV/VIS absorption spectroscopy using an Agilent 8453 spectrophotometer. The buffer subtracted spectra was corrected for Rayleigh scattered light according to Equation 4.[170]

Equation 4

$$A_{280\ nm} = E_{280\ nm} - 1.95 \times E_{330\ nm}$$

Where A$_{280\ nm}$ refers to the protein absorption at 280 nm, and E$_{280}$ and E$_{330}$ refer to the measured absorbance at 280 and 330 nm respectively. The values for A$_{280\ nm}$ and $\varepsilon_{280\ nm}$ are used in the Beer-Lambert law (Equation 5).

Equation 5

$$A_{280\ nm} = \varepsilon_{280\ nm} \times [protein] \times l$$

Where $A_{280\ nm}$ is the corrected absorbance of the sample, [protein] is the concentration of protein in the sample (units = M) and l is the path length of the sample, the length of the sample that the light beam passes through (units = cm).

Compound Preparation

Compounds from the NCI library were received as solids. These compounds were weighed into MS vials and made into 20 mM stock solutions with anhydrous dimethylsulphoxide (DMSO). Compounds from Vichem were received as 5 mM DMSO stocks. Stock solutions were stored at – 20 °C

All compounds underwent quality control; analysis via HPLC and LCMS.

Synthesis of Peptide Control (sequence: ARA)

All reactions were carried out in solid phase extraction vessels (Supelco), dimethylformaldehyde (DMF) used was of peptide grade. Prior to use the solid phase resin was washed thoroughly with DMF.

Amino acid couplings: Amino acid building blocks, Fmoc-alanine and Fmoc-Arginine(PBF), (3 equiv.), were dissolved in 1.5 ml of DMF and DIPEA. HATU (2.8 equiv) was added to the solution, the solution changed from colourless to yellow. The solution was added to the washed resin. Resin was shaken at room temperature for 40 minutes on a platform shaker at 30 oscillations per minute (OPM). Resin was then washed (see resin washing below).

Fmoc deprotections: Resin was incubated with 3 ml 20% piperidine DMF for 20 minutes. The resin was then washed. The deprotection and following wash were then repeated.

N-Terminal Capping: 40 equivalents of acetic anhydride and 40 equivalents of DIPEA in 1.5 ml of DMF was added to the resin. Resin was incubated for 30 minutes with shaking at room temperature, then washed.

Resin Washing: Washing was carried out after all reaction steps. Resin vessel was drained of solvent by vacuum manifold. Once drained the resin was washed with 5 x 3 ml DMF, 5 x 3 ml DCM, 5 x 3 ml DMF.

Testing for Primary Amines: A small sample of solid phase resin was transferred into an insert vial using a needle. These were suspended in 10% DIPEA in DMF to which a drop of 5 % solution of trinitrobenzenesulfonic acid (TNBS) was added. The vial was then sealed and

shaken, if free amines were present on the solid phase resin the bead turned deep orange. This result indicated either that an amide coupling reaction had not successfully reacted with all amines or that Fmoc groups had been removed from the N-terminal amines.

Cleavage: Peptides were cleaved from 100 mg of the resin in using a solution of 564 µl TFA, 18 µl triisopropylsilane (TIS), 18 µl water. Resin was incubated for 3 hours in the solution at room temperature. Cleavage solution was added dropwise into a solution of ice cold diethyl ether where the peptides precipitated as a white solid. The white solid was collected via centrifugation and dried in vacuo. To ensure no moisture affected the calculation of the yield of peptide the peptide was dissolved in water and lyophilised (MartinChrist Alpha 1-2 LDPlus). The fluffy white solid was transferred to a pre-weighed Eppendorf tube, flushed gently with nitrogen and sealed. The sealed Eppendorf was then sealed in a small plastic bag filled with self-indicating silica beads (Sigma Aldrich) to ensure long term dryness.

Ac-ARA-NH$_2$: Peptide was synthesised using the standard methods described above on 134 mg of 90 µm Tentagel resin, functionalised with Rink amide linker (loading of 0.26 mmol/g). (12.85 mg, 0.027 mmol, 77%), Peptide mass: 357.21. Seen: 358.77 [M+H]$^+$

The peptide was dissolved into a 10 mM stock using dry DMSO. A sample of this stock was further diluted for use as a stock for this series of experiments.

Thermal Denaturation Assay

Protein and compounds were incubated together at concentrations of 5 µM and 100 µM respectively in 135 µl of buffer (20mM HEPES, 250 mM NaCl, pH 7.50) for 45 minutes in a sealed 96 well plate at room temperature. 15 µl of 50x Sypro orange dye was added and the solutions agitated. The 150 µl solutions were split into three 50 µl aliquots in a 96 well PCR plate. The plate was sealed and briefly centrifuged (500 RCF for 5 min) to remove air bubbles. The plate was transferred to a quantitative Polymerase Chain Reaction unit, a Biorad IQ5 ICycler, for analysis. The melting temperature of the protein was determined using a temperature range of 20 – 80 °C with intervals of 0.5 °C and a hold time of 30 seconds for each temperature increment. The data was exported using Biorad IQ5 software.

Microdialysis

Concentration gradient for Calibration Curve

A concentration series of 100, 50, 25, 12.5 µM of NCI292249 were analysed by HPLC (injection volume: 15 µl method: 3 % gradient described in general methods section 9.1). Plotting the area under the HPLC curve vs the concentration of the solution provided a calibration curve for concentration determination.

Characterisation of the diffusion of NCI292249

200 µM NCI292249 in 100 µl of buffer (20 mM Hepes, 250 mM NaCl, 5 % DMSO, pH 7.50) were injected by autopipette into the incubation chamber of a microdialysis cartridge in a microdialysis plate. To the connected chamber, 300 µl of buffer was added. The plate holding the cartridge was sealed and shaken for 16 hours at room temperature at 180 rpm on an orbital shaker. A sample of the solution from the diffusion chamber was removed and submitted for HPLC analysis to determine the concentration of the sample (injection volume: 15 µl, method: 3 % gradient described in general methods section 9.1).

Microdialysis of NCI292249 with WDR5 and BSA

80 µM WDR5 and 200 µM NCI292249 in 100 µl of buffer (20 mM Hepes, 250 mM NaCl, 5 % DMSO pH 7.50) were injected into the incubation chamber of a microdialysis cartridge in a microdialysis plate. To the connected chamber 300 µl of buffer was added. The plate holding the cartridge was sealed and shaken for 16 hours at 25 °C at 180 rpm. A sample of the solution from the diffusion chamber was removed and submitted for HPLC analysis to determine the concentration of NCI292249 in the diffusion well (injection volume: 15 µl).

In parallel the experiment was repeated with bovine serum albumin (BSA) in place of WDR5 as a control experiment to test whether NCI292249 binds non-specifically with protein.

The concentration of the ligand in the incubation well was calculated based on the total concentration determined from the control experiment that was carried out in parallel From this the ratio [ligand]$_{incubation}$/[ligand]$_{diffusion}$ could be determined, referred to as the partition coefficient or Pt value. The Pt value was converted into a K_D using a calculator tool developed by Steven Shave.

9.3. SOBOC Methods

Protein labelling determination

Computational Methods

The following commands were input into a PyMol (ver 18.4) session of 2H9M, a PDB file containing WDR5 and a segment of an ART containing peptide. The following commands were used to highlight lysine residues in blue and cysteine residues in yellow on the surface of WDR5 (coloured white).

remove resn hoh	# remove water
h_add	# add hydrogens
as surface	# shows surface
color grey90	# colour surface in grey
select sulf_cys, (resn cys and (elem S))	# select the sulfur atom of cysteine residues
color yellow, sulf_cys	# sulphur of cys are shown in yellow
select nitro_lys, (resn lys and name NZ)	# select the nitrogens of free amines ("NZ" in PDB file)
select hydro_lys, (elem H and (neighbor nitro_lys))	# select the neighboring H atoms
select amine_lys, (nitro_lys or hydro_lys)	
color tv_blue, amine_lys	# amines of lys residues are shown in blue

This presents the functional groups of the residues as colours as opposed to the entire residue.

To further clarify the functional groups the command below was entered.

as spheres, amine_lys + sulf_cys

In theory the more a sphere presents itself through the surface the more readily able to react it will be.

Surface residues were counted by eye, 27 lysine and 6 cysteine residues appear on the surface of WDR5.

WDR5 protein labelling and purification

100 μM of WDR5 and 1 mM Cy5 maleimide were incubated in 200 μl buffer (20 mM HEPES, 250 mM NaCl, 5 % DMSO, pH 7.00) under argon in a sealed glass vial at 4 °C for 16 hours. The resulting solution was purified by size exclusion using a NAP-5 column (Sephadx G-25) eluted using 20 mM HEPES, 250 mM salt, pH 7.00. Fractions were eluted by gravity with a size of around 0.5 ml per fraction, decreasing to 0.1 ml per fraction as label containing solution eluted. The resulting fractions were analysed by reverse-phase HPLC using a C-4 column. Fractions containing pure labelled protein were combined and concentrated using a 15000 Dalton spin column.

Protein labelling quantification

Using an Agilent 8453 spectrophotometer in single beam mode and 150 μl glass cuvette the absorbance at 280, 330 and 649 nm were determined for a solution of Cy5 maleimide. The average 280 and 330 nm absorbance values were each divided by the average 649 nm absorbance. This gives the correction factors for 280 and 330 nm, allowing for protein concentration determination. A sample of the combined WDR5-Cy5 fractions was analysed using an Agilent 8453 spectrophotometer in single beam mode and 150 μl glass cuvette the absorbance at 280, 330 and 649 nm. The average 280 and 330 nm absorbance values were adjusted using the correction factors determined for Cy5. These Cy5 corrected absorbances were then used to determine the corrected 280 nm absorbance (Equation 6).

Equation 6

$$A_{280\,nm} = E_{280\,(corr)} - 1.95 \times E_{330\,(corr)}$$

Where $A_{280\,nm}$ refers to the protein absorbance at 280 nm, $E_{280\,(corr)}$ and $E_{330\,(corr)}$ refer to the Cy5 corrected experimentally observed 280 and 330 nm absorbance.

The concentration of protein and Cy5 in the sample were then calculated using the Beer-Lambert law, the previously calculated molar extinction coefficient for WDR5 (69,620 M^{-1} cm^{-1}) and the known literature molar extinction coefficient of Cy5 (250,000 M^{-1} cm^{-1})

$$A_{280\,nm} = \varepsilon_{280\,nm} \times [protein] \times l \quad and \quad A_{649\,nm} = \varepsilon_{649\,nm} \times [Cy5] \times l$$

Where A_X corresponds to the absorbance at wavelength x (280 nm for protein and 649 nm for Cy5), ε_X corresponds to the molar extinction coefficient of the analyte at wavelength X, l cm is the distance that the light beam travels through the sample solution, [protein] and [Cy5] correspond to the solution concentrations of the protein and Cy5 respectively.

SOBOC Library Synthesis

The scanning peptide fragment library was synthesised on Tentagel 90 μm Rink Amide resin (Rapp Pulymere) with a loading of 0.23 mmol/g (Rapp Polymere). All reactions were carried out in DMF used was peptide grade. Resin was swollen in DMF for at least 20 minutes prior to use. Peptides were synthesised with a C-terminal propargylglycine (PRA) residue for labelling and an 8-amino-3 6-dioxaoctanoic acid (DOA) spacer between this labelling site and the peptide itself. Reactions were carried out in solid phase extraction vessels

Amino acid couplings: Amino acid building blocks (3 equiv.), were dissolved in 1.5 ml of DMF and DIPEA. HATU (2.8 equiv) was added to the solution, the solution changed from colourless to yellow. The solution was added to the washed resin. Resin was agitated on a platform shaker at 30 oscillations per minute (opm) at room temperature for 40 minutes. Resin was then washed

Fmoc deprotections: Resin was incubated with 3 ml 20% piperidine DMF, agitated for 20 minutes on a platform shaker at 30 rpm. The resin was then washed. The deprotection and following wash were then repeated.

Testing for Primary Amines: A small sample of solid phase resin was transferred into an insert vial using a needle. These were suspended in 10% DIPEA in DMF to which a drop of 5 % solution of trinitrobenzenesulfonic acid (TNBS) was added. The vial was then sealed and shaken, if free amines were present on the solid phase resin the bead turned deep orange. This result indicated either that an amide coupling reaction had not successfully reacted with all amines or that Fmoc groups had been removed from the N-terminal amines.

Washing: Washing was carried out after all reaction steps. Resin vessel was drained of solvent by vacuum manifold. Once drained the resin was washed with 5 volumes of DMF, 5 volumes of DCM, 5 volumes of DMF.

Resin and vessel preparation: 9 sintered solid phase extraction vessels of varying size were filled with resin, 1 ml – 10 ml with size selected to allow for 1 ml of solvent per 100 mg of resin. The resin added to each vessel was dependent on the number of peptides to be produced by the reaction series. 100 mg of resin per final peptide was used per vessel, 4.5 g of resin in total. Each sample of resin was reacted to produce the C-terminal trimer with a DOA spacer unit and a Pra labelling site (AA$_1$-AA$_2$-AA$_3$-DOA-Pra) using the Fmoc synthesis methods described above. 100 mg of resin was removed from each reaction vessel, dried in a fresh sintered SPE and stored under vacuum. The remaining bulk resin underwent Fmoc deprotection and couplings according to the methods described above. After each coupling the resin was thoroughly dried and weighed, ~100 mg of resin was removed and stored under vacuum. The cycle of coupling amino acids and removing resin continued until all fragments had been produced.

Cleavage: The labelled peptides were cleaved from the resin using trifluoroacetic acid: water: triisopropylsilane in the ratio of 376:12:12 μl. The cleavage mixture was added to the resin in a sintered solid phase extraction vessel, the vessel was sealed and agitated on a platform shaker at 30 rpmin the absence of light for 3 hours at room temperature. The cleavage solution was drained through the sinter and dried in vacuo, resulting in a purple solid or oil.

Labelling: 10 mg of Fmoc deprotected resin was transferred into an insert vial. 6.4 mg of CuSO$_4$ was dissolved in 250 μl water, 5.1 mg of sodium ascorbate was dissolved in 250 μl water. These two solutions were combined prior to use to give a yellow solution. 4 mg of TMR azide was dissolved in 150 μl tBuOH in H$_2$O (1:2). 100 μl of the copper containing catalyst solution and 100 μl of the TMR azide solution was added to the insert vial. Insert vial was sealed and heated in a water bath at 40 °C for 16 hours. Resin was washed with water and methanol until solutions showed no trace of fluorescent label by eye, then dried. Labelled peptides were cleaved from the resin using a solution consisting of 188 μl TFA, 6 μl TIS and 6 μl water for 3 hours at room temperature in a sealed insert vial, agitated on a platform shaker at 30 rpm. Labelled peptides were purified via preparative HPLC. Peptides were concentrated via lyophilisation.

Purification: Peptides were purified by preparative HPLC. Typically dry samples were dissolved in 450 μl of 1:1 ACN and water solution. The solution was filtered through glass

wool packed into a Pasteur pipette to ensure no possible contamination from fragments of the solid phase support into the samples. 400 µl of the samples was injected into the preparative HPLC system which was set up to collect on detection of absorbance at 555 nm, fraction sizes were typically 200 µl. Preparative HPLC system details found in section 9.1. Fractions were tested by LCMS. Pure fractions containing the desired product were combined and lyophilised.

Peptide Quantification

Peptides were dissolved in water. The concentration of these peptides was determined by determination of the concentration of TMR in the solution (peptides were 1:1 labelled with TMR with a purity >95 % by HPLC). A 1 in 10 dilution of each peptide was made in water and the absorbance at 550 nm determined in a 1 ml quartz cuvette in an Agilent 8453 spectrophotometer. For each peptide the concentration was determined according to the Beer-Lambert law and the known molar extinction coefficient of TMR, 57100 M^{-1} cm^{-1} in water.

Equation 7

$$Abs = l \times \varepsilon \times conc$$

where Abs, l (cm), ε (cm^{-1} M^{-1}) and conc (M) refer to absorbance, pathlength, molar extinction coefficient of solution at a specific wavelength, and concentration of solution.

On-bead screening of the labelled peptides

Ni-NTA functionalised agarose resin (Quiagen) was filtered using 70-100 µM nylon membrane filters (FischerScientific nylon cell strainer) in order to increase the uniformity of the resin. 3 µl of 50 % slurry of the filtered resin was incubated with 7 µl of 2 µM 6xHis-WDR5 for 20 minutes at 25 °C shaken at 180 rpm in an Eppendorf in an orbital Eppendorf shaker. The resin was washed 3 time with 500 µl of buffer by centrifuging the resin (500 RCF, 2 min using an Eppendorf 5417R centrifuge), removing the supernatant and re-suspending the resin in buffer. This wash was repeated twice more. The resin was then resuspended and aliquoted into separate Eppendorf tubes for incubation with the ligand. The resin was incubated with 200 µl of 500 nM ligand in buffer for 20 min at 25 °C shaken at 180 rpm in an Eppendorf tube placed in an Eppendorf shaker. The resin was suspended in the solution and a 25 µl sample transferred to a glass bottomed 384 well plate (SWISSC, PS384B-G175). The samples were analysed on a Perkin-Elmer OPERA high content screening unit.

Confocal scanning (CONA)

Images of the beads were taken on an Opera® High Content screening system (PerkinElmer) using a 20x air objective, numerical aperture 0.45. Images were acquired by peltier cooled CCD cameras with 1.3 megapixel resolution, each well was imaged as 77 overlapping images (20 % overlap) over each well area. The focal height in all experiments was set to 30 μm above the bottom of the well plate surface.

WDR5-Cy5 and peptide-TMR screening

In order to image the resin beads, the labelled protein and the labelled peptides a three exposure setup was used. Camera 2 was used for imaging in all exposures: TMR imaging: exposure one, cw laser 561 nm, 1000 μW power, 200 ms exposure time, primary dichroic 445/561/640, detection dichroic 568sp, filter 585/40; Bead imaging: exposure two (brightfield), top illumination, 50 % LED, 200 ms exposure time, detection dichroic 650sp, filter 690/70; Cy5 imaging: exposure three, cw laser 640 nm, 500 μW power, 120 ms exposure time, primary dichroic 445/561/640, detection dichroic 650sp, filter 690/70.

Unlabelled WDR5 and peptide-TMR screening

In the absence of labelled protein only 2 exposures were used in the bulk of screening experiments. Camera 2 was used to for imaging in all exposures: TMR imaging: exposure one, cw laser 561 nm, 1000 μW power, 200 ms exposure time, primary dichroic 445/561/640, detection dichroic 568sp, filter 585/40; Bead imaging: exposure two (brightfield), top illumination, 50 % LED, 200 ms exposure time, detection dichroic 650sp, filter 690/70.

Image Stitching

Images from the CONA screen (77 per well) were stitched using BREAD, software delivered in the Auer lab that includes a stitching plugin for ImageJ (Fiji Software). The macro utilises plugin Grid/Mosaic stitching plugin.[171]

Analysis of Bead Intensities Using Bead Ring Evaluation and Analysis of Data (BREAD)

Stitched images were then analysed using proprietary image analysis software, Bead Ring Evaluation and Analysis of Data (BREAD). BREAD determined the locations of all resin beads by detecting circular structures of a specified size (60 – 100 pixels in this series of experiments. Detection is carried out in both the fluorescent and brightfield channels, allowing for the

detection of all beads in the stitched images. Once the location of all beads in a stitched well image had been determined fluorescence intensity profiles of each of the resin beads was recorded. In this study for each resin bead in each well 5 profiles were taken. Each intensity profile is observed as 2 peaks, representing the brighter edges of the bead, between these peaks values are expected to be significantly lower as this corresponds to the bead interior. The maxima of these two peaks were averaged and the centre percentiles (20-80th) subtracted in order to correct for any background intensity seen in the bead interior, the two resulting values are averaged. This is repeated 5 times to determine the bead ring intensity of the bead. For each well the mean of the bead ring intensity of all beads is taken to give the mean bead ring intensity.

Fluorescence anisotropy

Experimental planning

A 1:1 binding interaction was expected between WDR5 and the fluorescently labelled ligand. This relationship can be defined via a quadratic equation.

A binding interaction between two partners can be described as

$$P + L \underset{\Leftarrow}{\overset{k_D}{\Rightarrow}} PL$$
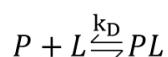
Where the dissociation constant (K_D) is defined as:

Equation 8

$$k_D = \frac{[P]_f [L]_f}{[PL]}$$

Where $[P]_f$ and $[L]_f$ refer to the concentration of free protein and ligand respectively, $[PL]$ refers to the concentration of protein-ligand complex.

The mass balances that describe the amounts of protein and ligand are defined as:

Equation 9

$$[P]_0 = [P]_f + [PL]$$

$$[L]_0 = [L]_f + [PL]$$

Where $[P]_0$ and $[L]_0$ correspond to the starting or total concentration of protein and ligand respectively. $[PL]$ refers to the concentration of protein-ligand complex. $[P]_f$ and $[L]_f$ correspond to the free concentration of protein and ligand respectively.

These are rearranged to put each equation into the context of the free concentrations of protein and ligand.

Equation 10

$$[P]_f = [P]_0 - [PL]$$

$$[L]_f = [L]_0 - [PL]$$

The expressions from Equation 10 are substituted into Equation 8 to give Equation 11.

Equation 11

$$k_D = \frac{([P]_0 - [PL])([L]_0 - [PL])}{[PL]}$$

This is then rearranged into the format of a quadratic equation (Equation 12).

Equation 12

$$0 = [PL]^2 - ([L]_0 + [P]_0 + k_D)[PL] + [P]_0[L]_0$$

Which is solved as the quadratic equation (Equation 13).

Equation 13

$$[PL] = \frac{[L]_0 + [P]_0 + k_D}{2} - \sqrt{\frac{([L]_0 + [P]_0 + k_D)^2}{4} - [L]_0 \times [P]_0}$$

Dividing Equation 13 by the concentration of ligand gives the proportion of complexation relative to the amount of ligand (Equation 14).

Equation 14

$$\text{Proportion of complexation} \left(\frac{[PL]}{[L]_0}\right) = \frac{\frac{[L]_0 + [P]_0 + k_D}{2} - \sqrt{\frac{([L]_0 + [P]_0 + k_D)^2}{4} - [L]_0 \times [P]_0}}{[L]_0}$$

A simulated binding curve was plotted using Origin 9.0 software. The plot assumed a K_D of 100 μM (the literature affinity of the MYC MIIIb peptide[12]) and a ligand concentration of 50 nM. This was used to determine the range of concentrations of protein should be used in a titration experiment.

Fluorescence Anisotropy

Fluorescence anisotropy was determined on a SPEX Fluorolog τ-3 spectrofluorometer (Horiba) in an L set up, polarisers were in place and set to 55° (magic angle). Excitation 552 nm, emission 580 nm with slits set to 4 and 6 nm bandwidth respectively. The starting sample volume was 175 µl of 50 nM EEEIDVV-DOA-PRA(TMR) peptide in a 200 µl cuvette (Sub-Micro Quartz cuvette). Protein was titrated into the cuvette from a highly concentrated stock solution from 175 µl up to 200 µl end volume. Titration was carried out by pipette, the sample agitated by pipette and allowed to equilibrate for 3 minutes after agitation prior to analysis. Data was plotted using Graphit (Erithacus) and fitted using equation Equation 15 below.

Equation 15

$$b = [Ligand]_0 + [Protein] + Kd$$

$$a = (b - (b^2 - 4 \times [Ligand]_0 \times [Protein])^{\frac{1}{2}}/(2 \times [Ligand]_0)$$

$$Q = \frac{qbound}{qfree}$$

$$(rmin + ((rmax \times Q - rmin) \times a))/(1 - (1 - Q) \times a)$$

Where rmax is the expected maximum anisotropy, rmin is the startpoint (ligand anisotropy), [Ligand]$_0$ refers to the starting concentration of labelled ligand, [protein] refers to the protein concentration. Q is the quenching factor, qbound is the molecular brightness of the bound ligand, qfree is the molecular brightness of the free ligand. GraFit 6.0 was used to plot the experimental data, rmax and K_D variable fitting parameters in a nonlinear least square regression fit.

9.4.   MorPH Peptide Examples

Peptide synthesis and Purification

Peptides were synthesised using Fmoc solid phase synthesis methods and labelled with TMR-N$_3$ as described in section 9.3. All peptides were synthesised on a 200 mg scale using Rink

amide functionalised Tentagel resin (Rapp Polymere) as the solid phase in 2 ml solid phase extraction vessels (Supelco).

Reverse phase HPLC analysis of unlabelled peptides: Due to the short retention time of these peptides in reverse phase HPLC the HPLC analysis is of the Fmoc-protected forms of the peptides. HPLC analysis was repeated after Fmoc protecting group removal to verify the deprotection was complete.

H₂N-Abu-CarbPhe-E-R-CONH₂: Peptide was synthesised using the standard methods described above on 217 mg of Tentagel resin, functionalised with Rink amide linker (loading of 0.23 mmol/g). (16.3 mg, 0.028 mmol, 56%), Peptide mass: 577.64. Seen: 578.55 [M+H]⁺

H₂N-Abu-KER-CONH₂: Peptide was synthesised using the standard methods described above on 242 mg of Tentagel resin, functionalised with Rink amide linker (loading of 0.23 mmol/g).(16.0 mg, 0.031 mmol, 56%). MS (ESI) Exact mass calcd: 515.62. Seen: 516.38 [M+H]⁺

H₂N-AKER-CONH₂: Peptide was synthesised using the standard methods described above on 290 mg of Tentagel resin, functionalised with Rink amide linker (loading of 0.23 mmol/g). (8.0mg, 0.016 mmol, 24%) MS (ESI) Exact mass calcd: 501.59. Seen: 502.57 [M+H]⁺

Labelled peptides

H₂N-Abu-CarbPhe-E-R-DOA-PRA(TMR)-CONH₂: Peptide was synthesised using the standard methods described above on Tentagel resin, functionalised with Rink amide linker. A sample of the resin was then removed for labelling according to the method described above. MS (ESI) Exact mass calcd: 1331.62. Seen:1331.21 [M+H]⁺ and 666.44 [M+2H]²⁺

H₂N-Abu-KER-DOA-PRA(TMR)-CONH₂: Peptide was synthesised using the standard methods described above on Tentagel resin, functionalised with Rink amide linker. A sample of the resin was then removed for labelling according to the method described above. MS (ESI) Exact mass calcd: 1268.63. Seen: 1269.26 [M+H]⁺ and 635.29 [M+2H]²⁺

2D-FIDA Fluorescence Anisotropy Measurement – Carried out by Nhan Pham. Peptides synthesised by Olivier Barbeau

Samples were prepared in 20 mM HEPES, 100 mM NaCl, pH 7.5, 0.1 % Pluronic, 5 % DMSO with a ligand concentration of 2 nM. Measurements were carried out using proprietary Pick-

o-Screen instruments with 9 measurements per concentration of protein. Measurements were then averaged to determine anisotropy.

Stability analysis – Carried out by Nhan Pham

Mouse plasma, strain BALB/C, with Li-Heparin ordered from Harlan Sera-Lab Ltd., Loughborough, UK.

The 3 TMR labelled peptides were prepared in a stock of water/methanol 8:2. Each of these stock solutions was used to spike 3 separate 600 μl samples of plasma to a final concentration of 10 μM of peptide. The plasma samples were incubated at 37 °C and shaken. Time points were 0, 1, 2, 4, 6, 23, 30, 47, 57 and 71 hours. At each time point 40 μl of plasma was removed and mixed thoroughly with 120 μl of acetonitrile to precipitate the protein content. The mixture was centrifuged at 12000 rpm for 2 min. The supernatant was transferred to a vial and concentrated in a vacuum centrifuge. The residue was dissolved in water/acetonitrile 8:2 (50 μl), sonicated and centrifuged for 1 min at 12000 rpm. 25 μl of the supernatant was submitted for HPLC analysis. A 4 x 4 mm Merk Lichrospher pre-column was put in place to ensure no proteins reached the C18 column, the solvent gradient is described as 'plasma stability gradient' in the HPLC gradient section above. Compound stability was monitored by detecting HPLC chromatographic peak areas of compound fluorescence emission and maximum absorbance of attached dye in relation to total peak area at time 0 h.

Degradation products appeared as earlier eluting peaks, visible in both fluorescence and absorbance traces.

9.5.   Methionine linker and TMR Synthesis

Synthesis of test sequence: TMR-Ala-Met

Peptide was synthesised using the standard methods described above in section 9.2 on NH_2 functionalised Tentagel resin (Rapp polymere) with a loading of 0.30 mmol/g.

Preparation and cleavage of single bead samples

A needle tip of Fmoc-deprotected resin was transferred from the bulk to an insert vial. The insert vial was filled with methanol and the slurry agitated by pipette. Using a light microscope and pipette single beads in methanol were isolated from the slurry and transferred

to clean insert vials. These insert vials were filled with methanol and dried in a speed vac. Once dry the insert vials were inspected by eye to verify that a single bead was present at the bottom of each vial.

Method A: 15 µl of 20 mg/ml CNBr in 0.1 M HCl, 16 hours.[159]

Method B: 15 µl of 20 mg/ml CNBr in 70 % TFA, 30 % water, 16 hours.[160]

Method C: 15 µl of 100 mg/ml CNBr in 70 % formic acid, 24 hours.[161]

Method D: 15 µl of 30 mg/ml CNBr in propionic acid, 24 hours.[28]

In all test methods cleavage solutions were prepared beforehand and carefully mixed prior to addition to the single beads. Insert vials were sealed with snap-cap lids and wrapped in parafilm. All cleavage reactions were carried out at room temperature, shaken at 180 rpm.

Cleavage mixtures were dried in vacuo. 30 µl of acetonitrile and water (1:1) was added to each vial. The mixture was agitated and 25 µl of the solution was transferred to a fresh insert vial, with care to avoid removing the resin bead. Samples were analysed by HPLC and LCMS.

Development of CNBr Cleavage Method

Cleavage methods were each trialled on 20 single solid phase resin beads that held the methionine linker test resin as described above. 150 µl of 20 mg/ml CNBr in 70 % TFA, 30 % water was added to 2 insert vials containing the test resin. The vials were sealed with snap lids and covered in foil. The cleavage reactions were shaken on a platform shaker at 30 rpm before the cleavage solutions were isolated from the resin after 16 and 40 hour intervals.

150 µl of 20 mg/ml CNBr in 70 % TFA, 30 % water was added to 3 insert vials containing the test resin. The cleavage solutions were agitated by pipette before the vials were sealed with snap lids and covered in foil. The reactions were incubated in a 40 °C water bath for 1, 2 and 16 hours before the cleavage solutions were isolated from the resin.

150 µl of 30, 50 and 100 mg/ml CNBr in 70 % TFA, 30 % water was added to 3 insert vials containing the test resin. The cleavage solutions were agitated by pipette before the vials were sealed with snap lids and covered in foil. The reactions were incubated in a 40 °C water bath for 2 hours before the cleavage solutions were isolated from the resin.

The cleaved peptides were suspended in water and transferred to a 5 ml vial with care to avoid removing any resin. The volume of the peptide solution was made up to 3 ml and the

concentration determined by the concentration of TMR present in the solution using Equation 7.

TMR synthesis

TMR methyl ester

2-Formyl-4-methoxycarbonyl-benzoic acid (604 mg, 2.91 mmol) was stirred in propionic acid (15 ml) and 3-dimethylaminophenol (837 mg, 6.10 mmol) was added, followed by *p*-TsOH (113 mg, 0.58 mmol). The mixture was heated to 80 °C for 16 hours. Chloranil (714 mg, 2.91 mmol) was added, maintaining the temperature at 80 °C and the mixture stirred for 6 hours. The mixture was cooled to room temperature then solvent was removed under reduced pressure. The crude residue was purified via chromatography. The resulting product was a dark purple solid (452 mg, 1.02 mmol, 35 %). Eluent for chromatography: gradient from 5% H_2O/MeCN to 20% H_2O/MeCN, R_f = 0.27 (15% H_2O/MeCN). ^{1}H NMR (500 MHz, CD$_3$OD) δ 8.25 (1H, dd, *J* = 8.2, 1.7 Hz, Ar**H**), 8.14 (1H, d, *J* = 8.2 Hz, Ar**H**), 7.85 (1H, d, *J* = 1.4 Hz, Ar**H**), 7.23 (2H, d, *J* = 9.5 Hz, Ar**H**), 6.98 (2H, dd, *J* = 9.5, 2.5 Hz, Ar**H**), 6.87 (2H, d, *J* = 2.5 Hz, Ar**H**), 3.90 (1H, s, OC**H**$_3$), 3.25 (12H, s, 3 x NC**H**$_3$); ^{13}C NMR (125 MHz, CD$_3$OD) δ 172.5 (C), 167.5 (C), 161.9 (C), 159.2 (C), 158.8 (C), 146.3 (C), 133.9 (C), 132.8 (CH), 132.2 (C), 131.9 (CH), 131.6 (CH), 131.2 (CH), 115.2 (CH), 115.1 (C), 97.5 (CH), 53.1 (CH$_3$), 41.0 (CH$_3$); MS (ESI) Exact mass calcd for C$_{26}$H$_{25}$N$_2$O$_5$ [M+H]$^+$: 445.18, found: 445.38.

TMR acid

TMR-OMe (202 mg, 0.455 mmol) was dissolved in 30 ml of THF/water (1:1). NaOH (45.48 mg, 1.137 mmol) was added to the reaction solution with stirring. The reaction vessel was covered in foil and stirred at room temperature overnight. Concentrated HCl was added to acidify the solution. The solution was stored at 4 °C, a solid precipitated over the course of 3 days. The solid was isolated by filtration and washed with cold water, giving the product as a purple crystalline solid (84 mg, 0.195 mmol, 43 %).^{1}H NMR (500 MHz, MeOD) δ 8.48 - 8.36 (2H, m, ArH), 7.99 (1H, s, ArH), 7.15 (2H, d, *J* = 9.4 Hz, ArH), 7.10 (2H, dd, *J* = 9.4, 1.7 Hz, ArH), 6.97 (2H, s, ArH), 3.33 (12H, s, 3 x CH$_3$); ^{13}C NMR (125 MHz, CD$_3$OD) δ 167.7 (C), 167.5 (C), 160.2 (C), 159.0 (C), 158.8 (C), 136.4 (C), 136.2 (C), 135.4 (C), 133.0 (CH), 132.5 (CH), 132.0 (CH), 115.6

(CH), 114.8 (C), 97.4 (CH), 41.0 (CH$_3$); MS (ESI) Exact mass calcd for C$_{25}$H$_{23}$N$_2$O$_5$ [M+H]$^+$: 431.16, found: 431.33.

Time course experiments

Reverse-Phase HPLC Standards

Stock samples of TMR-N$_3$ and TMR-COOH were analysed by reverse phase HPLC using a 3%/minute gradient to determine their retention times. Column information: Agilent Zorbax C18 Vydac C18, 4.6 mm x 150 mm, 3.5 μm particle diameter size. The retention times of the acid and the azide were 21 and 23 minutes respectively.

Time Course Experiments

Five 5 ml reactivials were prepared, each holding TMR-COOH (20 mg, 0.046 mmol) in DMF, with the exception of test 5 which had a 2:2:1 mixture of DMF/dioxane/water. To vial 1 DIPEA (16.03 μl, 0.092 mmol) was added, followed by HATU (17.49 mg, 0.046 mmol). To vial 2 DIC (5.81 mg, 0.046 mmol) and HOBt (6.22 mg, 0.046 mmol) were added. To vial 3 DIPEA (16.03 μl, 0.092 mmol) was added followed by PyBOP (23.94 mg, 0.046 mmol). To vial 4 DIPEA (16.03 μl, 0.092 mmol) was added, followed by EDC (8.82 mg, 0.046 mmol) and HOBt (6.22 mg, 0.046 mmol). To vial 5 DIPEA (24.04 μl, 0.138 mmol) was added, followed by TSTU (18.01 mg, 0.060 mmol). The reaction solutions were stirred at room temperature for 5 minutes in the absence of light. Then to each vial 3-azidopropylamine (46.40 mg, 0.46 mmol) was added, the vials sealed. At 1, 2, 4, 8 and 16 hour intervals a small sample of reaction solution was removed from each vial, the sample was diluted 1:1000 with 1:1 acetonitrile/water to significantly slow the reaction. Samples were submitted for analysis via reverse phase HPLC. Conversion of the acid to the amide was determined by percentage peak area of the peak with a retention time corresponding to that of the TMR-N$_3$ in the fluorescence detector channel (Ex 555 nm, EM 575 nm).

TMR-N$_3$ Synthesis

TMR-COOH (90.0 mg, 0.209 mmol) and DIPEA (73 μl, 0.418 mmol) were dissolved in DMF (10ml) with stirring. HATU (79.6 mg, 0.209 mmol) was added to the reaction vessel, after a few minutes of stirring 3-azidopropylamine (104.737 mg, 1.046 mmol) was added to the reaction vessel. The vessel was sealed and the reaction mixture stirred at room temperature

for 2 hours. Solvent was removed in vacuo using toluene to form an azeotrope. The resulting solid was purified by preparative reverse-phase HPLC yielding the product as a purple solid (102 mg, 0.199 mmol, 95 %). 1H NMR (500 MHz, METHANOL-d4) ⊚ 8.43 - 8.45 (Ar$\mathbf{H}^1$ dd, 1H), 8.23 (Ar$\mathbf{H}^2$ d, J = 6.38 Hz, 1H), 7.86 (Ar$\mathbf{H}^3$ s, 1H), 7.19 (Ar$\mathbf{H}^6$ d, J = 9.46 Hz, 2H), 7.09 (Ar$\mathbf{H}^5$ dd, J = 2.48, 9.50 Hz, 2H), 7.02 (Ar$\mathbf{H}^4$ d, J = 2.52 Hz, 2H), 3.47 - 3.57 (C$\mathbf{H}_2{}^9$ m, 2H), 3.44 (C$\mathbf{H}_2{}^7$ t, J = 6.66 Hz, 2H) , 3.34 (4 x NC$\mathbf{H}3^8$ s, 12H), 1.91 (C$\mathbf{H}_2{}^{10}$ t, J = 6.78 Hz, 2H). 13C NMR (126 MHz, METHANOL-d4) ⊚ 168.7, 168.0, 161.3, 159.7, 159.6, 140.0, 136.2, 133.5, 132.7, 131.0, 130.6, 116.2, 115.5, 98.1, 50.9, 41.6, 39.4, 31.3, 30.3, 24.9. MS (ESI) Exact mass calculated for $C_{28}H_{29}N_4O_4$ [M+H]+: 513.22, found: 513.19.

10. References

1.   Vanhee, P.; van der Sloot, A. M.; Verschueren, E.; Serrano, L.; Rousseau, F.; Schymkowitz, J., Computational design of peptide ligands. *Trends in Biotechnology* **2011,** *29* (5), 231-239.

2.   Ofran, Y.; Rost, B., Protein–Protein Interaction Hotspots Carved into Sequences. *PLOS Computational Biology* **2007,** *3* (7), e119.

3.   Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A., In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science* **2004,** *303* (5659), 844.

4.   Higueruelo, A. P.; Jubb, H.; Blundell, T. L., TIMBAL v2: Update of a database holding small molecules modulating protein-protein interactions. *Database* **2013,** *2013*, 1-5.

5.   Bourgeas, R.; Basse, M.-J.; Morelli, X.; Roche, P., Atomic Analysis of Protein-Protein Interfaces with Known Inhibitors: The 2P2I Database. *PLOS ONE* **2010,** *5* (3), e9598.

6.   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1. *Advanced Drug Delivery Reviews* **2001,** *46* (1–3), 3-26.

7.   Bakail, M.; Ochsenbein, F., Targeting protein–protein interactions, a wide open field for drug design. *Comptes Rendus Chimie* **2016,** *19* (1–2), 19-27.

8.   Stirnimann, C. U.; Petsalaki, E.; Russell, R. B.; Müller, C. W., WD40 proteins propel cellular networks. *Trends in Biochemical Sciences* **2010,** *35* (10), 565-574.

9.   Wu, X.-H.; Chen, R.-C.; Gao, Y.; Wu, Y.-D., The effect of Asp-His-Ser/Thr-Trp tetrad on the thermostability of WD40-repeat proteins. *Biochemistry* **2010,** *49* (47), 10237-10245.

10.   UniProt: a hub for protein information. *Nucleic Acids Research* **2014,** *43* (D1), D204-D212.

11.   Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P., The ChEMBL bioactivity database: An update. *Nucleic Acids Research* **2014,** *42* (D1), 1083-1090.

12.   Thomas, Lance R.; Wang, Q.; Grieb, Brian C.; Phan, J.; Foshage, Audra M.; Sun, Q.; Olejniczak, Edward T.; Clark, T.; Dey, S.; Lorey, S.; Alicie, B.; Howard, Gregory C.; Cawthon, B.; Ess, Kevin C.; Eischen, Christine M.; Zhao, Z.; Fesik, Stephen W.; Tansey, William P., Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. *Molecular Cell* **2015,** *58* (3), 440-452.

13.   Shiryaev, S. a.; Cheltsov, A. V.; Gawlik, K.; Ratnikov, B. I.; Strongin, A. Y., Virtual ligand screening of the National Cancer Institute (NCI) compound library leads to the allosteric inhibitory scaffolds of the West Nile Virus NS3 proteinase. *Assay and drug development technologies* **2011,** *9* (00), 69-78.

14.   Dharmarajan, V.; Lee, J. H.; Patel, A.; Skalnik, D. G.; Cosgrove, M. S., Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *Journal of Biological Chemistry* **2012,** *287* (33), 27275-27289.

15.   Tansey, W. P., Mammalian MYC Proteins and Cancer. *New Journal of Science* **2014,** *2014*, 1--27.

16.   Thomas, L. R.; Foshage, A. M.; Weissmiller, A. M.; Tansey, W. P., The MYC-WDR5 nexus and cancer. *Cancer Research* **2015,** *75* (19), 4012-4015.

17.   Tuncbag, N.; Keskin, O.; Gursoy, A., HotPoint: Hot spot prediction server for protein interfaces. *Nucleic Acids Research* **2010,** *38* (SUPPL. 2), 402-406.

18.   Shin, S.; Sung, B. J.; Cho, Y. S.; Kim, H. J.; Ha, N. C.; Hwang, J. I.; Chung, C. W.; Jung, Y. K.; Oh, B. H., An anti-apoptotic protein human survivin is a direct inhibitor of caspase-3 and -7. *Biochemistry* **2001,** *40* (4), 1117-1123.

19.   Kato, J.; Kuwabara, Y.; Mitani, M.; Shinoda, N.; Sato, A.; Toyama, T.; Mitsui, A.; Nishiwaki, T.; Moriyama, S.; Kudo, J.; Fujii, Y., Expression of survivin in esophageal cancer: Correlation with the prognosis and response to chemotherapy. *International Journal of Cancer* **2001,** *95* (2), 92-95.

20.   Kawasaki, H.; Altieri, D. C.; Lu, C.-d.; Altieri, D.-o., Inhibition of Apoptosis by Survivin Predicts Shorter Survival Rates in Colorectal Cancer Advances in Brief Inhibition of Apoptosis by Survivin Predicts Shorter Survival Rates in Colorectal Cancer. *Cancer research* **1998,** *54*, 5071-5074.

21.   Tamm, I.; Wang, Y.; Sausville, E.; Scudiero, D. A.; Vigna, N.; Oltersdorf, T.; Reed, J. C., IAP-family protein Survivin inhibits caspase activity and apoptosis induced by Fas (CD95), bax, caspases, and anticancer drugs. *Cancer Research* **1998,** *58* (23), 5315-5320.

22.   Jeyaprakash, A. A.; Basquin, C.; Jayachandran, U.; Conti, E., Structural basis for the recognition of phosphorylated histone H3 by the Survivin subunit of the chromosomal passenger complex. *Structure* **2011,** *19* (11), 1625-1634.

23.   Merrifield, R. B., Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **1963,** *85* (14), 2149--2154.

24.   Wuts, P.; Greene, T., *Greene's Protective Groups in Organic Synthesis*. 2006.

25.   Kvach, M. V.; Stepanova, I. a.; Prokhorenko, I. a.; Stupak, A. P.; Bolibrukh, D. a.; Korshun, V. a.; Shmanai, V. V., Practical synthesis of isomerically pure 5- and 6-carboxytetramethylrhodamines, useful dyes for DNA probes. *Bioconjugate chemistry* **2009,** *20* (8), 1673-82.

26.   Mudd, G.; Pi, I. P.; Fethers, N.; Dodd, P. G.; Barbeau, O. R.; Auer, M., A general synthetic route to isomerically pure functionalized rhodamine dyes. *Methods and Applications in Fluorescence* **2015,** *3* (4), 045002-045002.

27.   Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie (International ed. in English)* **2002,** *41* (14), 2596-9.

28.   Hancock, W. S.; Marshall, G. R., Letter: Cyanogen bromide as a cleavage procedure in solid phase peptide synthesis. *Journal of the American Chemical Society* **1975,** *97* (26), 7488-9.

29.   Gross, E.; Witkop, B., Nonenzymatic Cleavage of Peptide Bonds : The Methionine Residues in Bovine Pancreatic Ribonuclease Nonenzymatic Cleavage of Peptide Bonds : The Methionine Residues in Bovine Pancreatic Ribonuclease. *Journal of Biological Chemistry* **1962,** *237*, 1856-1856.

30.   Xu, C.; Min, J., Structure and function of WD40 domain proteins. *Protein & cell* **2011,** *2* (3), 202-214.

31.   Chatr-aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N. K.; O'Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; Stark, C.; Breitkreutz, B.-J.; Dolinski, K.; Tyers, M., The BioGRID interaction database: 2017 update. *Nucleic Acids Research* **2017,** *45* (Database issue), D369-D379.

32.   Fabregat, A.; Sidiropoulos, K.; Garapati, P.; Gillespie, M.; Hausmann, K.; Haw, R.; Jassal, B.; Jupe, S.; Korninger, F.; McKay, S.; Matthews, L.; May, B.; Milacic, M.; Rothfels, K.; Shamovsky, V.; Webber, M.; Weiser, J.; Williams, M.; Wu, G.; Stein, L.; Hermjakob, H.; D'Eustachio, P., The Reactome pathway Knowledgebase. *Nucleic Acids Research* **2016,** *44* (Database issue), D481-D487.

33.   Su, W.-Y.; Li, J.-T.; Cui, Y.; Hong, J.; Du, W.; Wang, Y.-C.; Lin, Y.-W.; Xiong, H.; Wang, J.-L.; Kong, X.; Gao, Q.-Y.; Wei, L.-P.; Fang, J.-Y., Bidirectional regulation between WDR83 and its natural antisense transcript DHPS in gastric cancer. *Cell Res* **2012,** *22* (9), 1374-1389.

34.   Inamdar, G. S.; Madhunapantula, S. V.; Robertson, G. P., Targeting the MAPK Pathway in Melanoma: Why some approaches succeed and other fail. *Biochemical pharmacology* **2010**, *80* (5), 624-637.

35.   Aster, J.; Blacklow, S.; Pear, W., Notch signalling in T cell lymphoblastic leukaemia/lymphoma and other haematological malignancies. *Journal of pathology* **2011**, *223* (2), 262-273.

36.   Hao, B.; Oehlmann, S.; Sowa, M. E.; Harper, J. W.; Pavletich, N. P., Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Molecular Cell* **2007,** *26* (1), 131-143.

37.   Gross, S.; Rahal, R.; Stransky, N.; Lengauer, C.; Hoeflish, K. P., Targeting cancer with small molecule kinase inhibitors. *The Journal of clinical investigation* **2015**, *125* (5), 1780-1789.

38.   Puca, R.; Nardinocchi, L.; Givol, D.; D'Orazi, G., Regulation of p53 activity by HIPK2: molecular mechanisms and therapeutical implications in human cancer cells. *Oncogene* **2010,** *29*, 4378-4387.

39.   Chauhan, D.; Hideshima, T.; Rosen, S.; Reed, J. C.; Kharbanda, S.; Anderson, K. C., Apaf-1/Cytochrome c-independent and Smac-dependent Induction of Apoptosis in Multiple Myeloma (MM) Cells. *Journal of Biological Chemistry* **2001,** *276* (27), 24453-24456.

40.   Christoph, F.; Kempkensteffen, C.; Weikert, S.; Köllermann, J.; Krause, H.; Miller, K.; Schostak, M.; Schrader, M., Methylation of tumour suppressor genes APAF-1 and DAPK-1 and in vitro effects of demethylating agents in bladder and kidney cancer. *British journal of cancer* **2006,** *95* (12), 1701-7.

41.   Sanchis, D.; Mayorga, M.; Ballester, M.; Comella, J. X., Lack of Apaf-1 expression confers resistance to cytochrome c-driven apoptosis in cardiomyocytes. *Cell death and differentiation* **2003,** *10* (9), 977-986.

42.   Harlan, J.; Chen, Y.; Gubbins, E.; Mueller, R.; Roch, J. M.; Walter, K.; Lake, M.; Olsen, T.; Metzger, P.; Dorwin, S.; Ladror, U.; Egan, D. a.; Severin, J.; Johnson, R. W.; Holzman, T. F.; Voelp, K.; Davenport, C.; Beck, a.; Potter, J.; Gopalakrishnan, M.; Hahn, a.; Spear, B. B.; Halbert, D. N.; Sullivan, J. P.; Abkevich, V.; Neff, C. D.; Skolnick, M. H.; Shattuck, D.; Katz, D. a., Variants in Apaf-1 segregating with major depression promote apoptosome function. *Molecular psychiatry* **2006,** *11* (1), 76-85.

43.   Orlicky, S.; Tang, X.; Neduva, V.; Elowe, N.; Brown, E. D.; Sicheri, F.; Tyers, M., An allosteric inhibitor of substrate recognition by the SCF(Cdc4) ubiquitin ligase. *Nature biotechnology* **2010,** *28* (7), 733-7.

44.   Takeishi, S.; Matsumoto, A.; Onoyama, I.; Naka, K.; Hirao, A.; Nakayama, K. I., Ablation of Fbxw7 Eliminates Leukemia-Initiating Cells by Preventing Quiescence. *Cancer Cell* **2013,** *23* (3), 347-361.

45.   Wu, X.-H.; Wang, Y.; Zhuo, Z.; Jiang, F.; Wu, Y.-D., Identifying the hotspots on the top faces of WD40-repeat proteins from their primary sequences by β-bulges and DHSW tetrads. *PloS one* **2012,** *7*, e43005.

46.   Wu, X.-H.; Wang, Y.; Zhuo, Z.; Jiang, F.; Wu, Y.-D., Identifying the hotspots on the top faces of WD40-repeat proteins from their primary sequences by β-bulges and DHSW tetrads. *PloS one* **2012,** *7* (8), e43005-e43005.

47.   He, J.; Chao, William C. H.; Zhang, Z.; Yang, J.; Cronin, N.; Barford, D., Insights into Degron Recognition by APC/C Coactivators from the Structure of an Acm1-Cdh1 Complex. *Molecular Cell* **2013,** *50* (5), 649-660.

48.   Tian, W.; Li, B.; Warrington, R.; Tomchick, D. R.; Yu, H.; Luo, X., Structural analysis of human Cdc20 supports multisite degron recognition by APC/C. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (45), 18419-24.

49.   Kidokoro, T.; Tanikawa, C.; Furukawa, Y.; Katagiri, T.; Nakamura, Y.; Matsuda, K., CDC20, a potential cancer therapeutic target, is negatively regulated by p53. *Oncogene* **2008,** *27* (11), 1562-71.

50.   Wu, W.-j.; Hu, K.-s.; Wang, D.-s.; Zeng, Z.-l.; Zhang, D.-s.; Chen, D.-l.; Bai, L.; Xu, R.-h., CDC20 overexpression predicts a poor prognosis for patients with colorectal cancer. *Journal of Translational Medicine* **2013,** *11* (1), 142-142.

51.   Sackton, K. L.; Dimova, N.; Zeng, X.; Tian, W.; Zhang, M.; Sackton, T. B.; Meaders, J.; Pfaff, K. L.; Sigoillot, F.; Yu, H.; Luo, X.; King, R. W., Synergistic blockade of mitotic exit by two chemical inhibitors of the APC/C. *Nature* **2014**.

52.   Zeng, X.; King, R. W., An APC/C inhibitor stabilizes cyclin B1 by prematurely terminating ubiquitination. *Nature chemical biology* **2012,** *8*, 383-92.

53.   Abdel-Magid, A. F., Inhibitors of LRRK2 as treatment for parkinson's disease. *ACS Medicinal Chemistry Letters* **2012,** *3*, 701-702.

54.   Guaitoli, G.; Raimondi, F.; Gilsbach, B. K.; Gómez-Llorente, Y.; Deyaert, E.; Renzi, F.; Li, X.; Schaffner, A.; Jagtap, P. K. A.; Boldt, K.; von Zweydorf, F.; Gotthardt, K.; Lorimer, D. D.; Yue, Z.; Burgin, A.; Janjic, N.; Sattler, M.; Versées, W.; Ueffing, M.; Ubarretxena-Belandia, I.; Kortholt, A.; Gloeckner, C. J., Structural model of the dimeric Parkinson's protein LRRK2 reveals a compact architecture involving distant interdomain contacts. *Proceedings of the National Academy of Sciences of the United States of America* **2016,** *113* (30), E4357-E4366.

55.   Dharmarajan, V.; Lee, J. H.; Patel, A.; Skalnik, D. G.; Cosgrove, M. S., Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *Journal of Biological Chemistry* **2012,** *287*, 27275-27289.

56.   Dou, Y.; Hess, J. L., Mechanisms of transcriptional regulation by MLL and its disruption in acute leukemia. *International Journal of Hematology* **2008,** *87* (1), 10-18.

57.   Marschalek, R., Classification of mixed-lineage leukemia fusion partners predicts additional cancer pathways. *Annals of Laboratory Medicine* **2016,** *36* (2), 85-100.

58.   Krivtsov, A. V.; Armstrong, S. a., MLL translocations, histone modifications and leukaemia stem-cell development. *Nature reviews. Cancer* **2007,** *7* (11), 823-833.

59.   Felix, C. A., Secondary leukemias induced by topoisomerase-targeted drugs. *Biochimica et Biophysica Acta - Gene Structure and Expression* **1998,** *1400* (1-3), 233-255.

60.   Senisterra, G.; Wu, H.; Allali-Hassani, A.; Wasney, G. a.; Barsyte-Lovejoy, D.; Dombrovski, L.; Dong, A.; Nguyen, K. T.; Smil, D.; Bolshan, Y.; Hajian, T.; He, H.; Seitova, A.; Chau, I.; Li, F.; Poda, G.; Couture, J.-F.; Brown, P. J.; Al-Awar, R.; Schapira, M.; Arrowsmith, C. H.; Vedadi, M., Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. *The Biochemical journal* **2013,** *449*, 151-9.

61.   Bolshan, Y.; Kuznetsova, E.; Wasney, G. a.; Hajian, T.; Poda, G.; Nguyen, K. T.; Wu, H.; Dombrovski, L.; Dong, A.; Senisterra, G.; Schapira, M.; Arrowsmith, C. H.; Brown, P. J.; Al-awar, R.; Vedadi, M.; Smil, D., Synthesis, Optimization, and Evaluation of Novel Small Molecules as Antagonists of WDR5-MLL Interaction. **2013**.

62.   Li, D.-D.; Chen, W.-L.; Xu, X.-L.; Jiang, F.; Wang, L.; Xie, Y.-Y.; Zhang, X.-J.; Guo, X.-K.; You, Q.-D.; Sun, H.-P., Structure-based design and synthesis of small molecular inhibitors disturbing the interaction of MLL1-WDR5. *European Journal of Medicinal Chemistry* **2016,** *118*, 1-8.

63.   Grebien, F.; Vedadi, M.; Getlik, M.; Giambruno, R.; Avellino, R.; Skucha, A.; Vittori, S.; Kuznetsova, E.; Barsyte-lovejoy, D.; Li, F.; Poda, G.; Schapira, M.; Dong, A.; Senisterra, G.; Stukalov, A.; Huber, K. V. M.; Marcellus, R.; Bilban, M.; Bock, C.; Brown, P. J.; Zuber, J.; Bennett, K. L.; Al-awar, R.;

Delwel, R.; Nerlov, C., Pharmacological targeting of the Wdr5-MLL interaction in C / EBP α N-terminal leukemia. *Nature Chemical Biology* **2015,** *11* (8), 571-578.

64.   Grebien, F.; Vedadi, M.; Getlik, M.; Giambruno, R.; Grover, A.; Avellino, R.; Skucha, A.; Vittori, S.; Kuznetsova, E.; Smil, D.; Barsyte-Lovejoy, D.; Li, F.; Poda, G.; Schapira, M.; Wu, H.; Dong, A.; Senisterra, G.; Stukalov, A.; Huber, K. V. M.; Schönegger, A.; Marcellus, R.; Bilban, M.; Bock, C.; Brown, P. J.; Zuber, J.; Bennett, K. L.; Al-awar, R.; Delwel, R.; Nerlov, C.; Arrowsmith, C. H.; Superti-Furga, G., Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. *Nature Chemical Biology* **2015,** *11*, 571-580.

65.   Karatas, H.; Townsend, E. C.; Bernard, D.; Dou, Y.; Wang, S., Analysis of the binding of mixed lineage leukemia 1 (MLL1) and histone 3 peptides to WD repeat domain 5 (WDR5) for the design of inhibitors of the MLL1-WDR5 interaction. *Journal of medicinal chemistry* **2010,** *53*, 5179-85.

66.   Karatas, H.; Townsend, E. C.; Cao, F., High-Affinity, Small-Molecule Peptidomimetic Inhibitors of MLL1/WDR5 Protein–Protein Interaction. *Journal of the American Chemical Society* **2012,** *2* (135), 669-682.

67.   Cao, F.; Townsend, E. C.; Karatas, H.; Xu, J.; Li, L.; Lee, S.; Liu, L.; Chen, Y.; Ouillette, P.; Zhu, J.; Hess, J. L.; Atadja, P.; Lei, M.; Qin, Z. S.; Malek, S.; Wang, S.; Dou, Y., Targeting MLL1 H3K4 Methyltransferase Activity in Mixed-Lineage Leukemia. *Molecular cell* **2013**, 1-15.

68.   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. *Journal of Medicinal Chemistry* **2010,** *53* (7), 2719-2740.

69.   DeLano, W. *The Pymol Molecular Graphics System*, 2002.

70.   P., M., Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994). *Pure and Applied Chemistry* **2009,** *66* (5), 1077-1184.

71.   Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C. K.; Glick, M.; Davies, J. W., How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space. *Journal of Chemical Information and Modeling* **2009,** *49* (1), 108-119.

72.   Sheridan, R. P.; Kearsley, S. K., Why do we need so many chemical simialrity search methods? *Drug Discovery Today* **2002,** *7* (17), 903-911.

73.   Nikolova, N.; Jaworska, J., Approaches to Measure Chemical Similarity– a Review. *QSAR Combinatorial Science* **2003,** *22* (910), 1006-1026.

74.   Livingstone, D., The characterization of chemical structures using molecular properties. A survey. *Journal of chemical information and computer sciences* **2000,** *40* (2), 195-209.

75.   Yap, C. W., PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *Journal of computational chemistry* **2011,** *32* (7), 1466-1474.

76.   Bajusz, D.; Rácz, A.; Héberger, K., Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics* **2015,** *7* (1), 1-13.

77.   Whittle, M.; Gillet, V. J.; Willett, P.; Alex, A.; Loesel, J., Enhancing the Effectiveness of Virtual Screening by Fusing Nearest Neighbor Lists : A Comparison of Similarity Coefficients Enhancing the Effectiveness of Virtual Screening by Fusing Nearest Neighbor Lists : A Comparison of Similarity Coefficients. *journal of chemical information and computer science* **2004,** *44* (September), 1840-1848.

78.   Willett, P., Combination of similarity rankings using data fusion. *Journal of Chemical Information and Modeling* **2013,** *53* (1), 1-10.

79.   Dixon, S. L.; Koehler, R. T., The hidden component of size in two-dimensional fragment descriptors: Side effects on sampling in bioactive libraries. *Journal of Medicinal Chemistry* **1999,** *42* (15), 2887-2900.

80.   Holliday, J. D.; Salim, N.; Whittle, M.; Willett, P., Analysis and display of the size dependence of chemical similarity coefficients. *Journal of Chemical Information and Computer Sciences* **2003,** *43* (3), 819-828.

81.   Maggiora, G. M.; Vogt, M.; Stumpfe, D.; Bajorath, J. J. r., Molecular similarity in medicinal chemistry. *Journal of medicinal chemistry* **2013,** *57* (8), 3186-3204.

82.   Vainio, M. J.; Kogej, T.; Raubacher, F.; Sadowski, J., Scaffold hopping by fragment replacement. *Journal of Chemical Information and Modeling* **2013,** *53* (7), 1825-1835.

83.   Schreyer, A. M.; Blundell, T., USRCAT: Real-time ultrafast shape recognition with pharmacophoric constraints. *Journal of Cheminformatics* **2012,** *4* (27).

84.   Ballester, P. J.; Richards, W. G., Ultrafast shape recognition for similarity search in molecular databases. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2007,** *463* (2081), 1307-1321.

85.   Shave, S.; Blackburn, E. A.; Adie, J.; Houston, D. R.; Auer, M.; Webster, S. P.; Taylor, P.; Walkinshaw, M. D., UFSRAT: Ultra-fast shape recognition with atom types -The discovery of novel bioactive small molecular scaffolds for FKBP12 and 11βHSD1. *PLoS ONE* **2015,** *10* (2), 1-15.

86.   Zhou, H.; Skolnick, J., FINDSITEcomb: A threading/structure-based, proteomic-scale virtual ligand screening approach. *J Chem Inf Model* **2012**.

87.   Skolnick, J.; Zhou, H.; Gao, M., Are predicted protein structures of any value for binding site prediction and virtual ligand screening? *Current Opinion in Structural Biology* **2013,** *23* (2), 191-197.

88.   Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G., A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology* **1996,** *261* (3), 470-89.

89.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T., Glide: a new approach for rapid, accurate docking and scoring. *J Med Chem* **2004,** *47*, 1739-1749.

90.   Lorber, D. M.; Shoichet, B. K., Flexible ligand docking using conformational ensembles. *Protein Science* **1998,** *7* (4), 938-950.

91.   Schnecke, V.; Kuhn, L. A., Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design* **2000,** *20* (1), 171-190.

92.   Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M., Molecular docking to ensembles of protein structures. *Journal of molecular biology* **1997,** *266*, 424-440.

93.   Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D., DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design* **2001,** *15* (5), 411-428.

94.   Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design* **1997,** *11* (5), 425-445.

95.   Böhm, H.-J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design* **1994,** *8* (3), 243-256.

96.   Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OLPS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118* (15), 11225-11236.

97.   Cheltsov, A. V.; Aoyagi, M.; Aleshin, A.; Chi-wang, Y. E.; Zhai, D.; Bobkov, A. A.; Reed, J. C.; Liddington, R. C.; Abagyan, R., Vaccina Virus Virulence Factor N1L is a Novel Promising Target for Antiviral Therapeutic Intervention. *Journal of medicinal chemistry* **2011,** *53* (10), 3899-3906.

98.   Müller, K. H.; Kakkola, L.; Nagaraj, A. S.; Cheltsov, A. V.; Anastasina, M.; Kainov, D. E., Emerging cellular targets for influenza antiviral agents. *Trends in Pharmacological Sciences* **2012,** *33* (2), 89-99.

99.   Kretschmann, E.; Raether, H., Radiative decay of nonradiative surface plasmons excited by light. *Z. Naturforsch. A* **1968,** *23*, 2135.

100.   Tang, Y.; Zeng, X.; Liang, J., Surface Plasmon Resonance: An Introduction to a Surface Spectroscopy Technique. *Journal of Chemical Education* **2011,** *87* (7), 742-746.

101.   Cimmperman, P.; Baranauskiene, L.; Jachimoviciūte, S.; Jachno, J.; Torresan, J.; Michailoviene, V.; Matuliene, J.; Sereikaite, J.; Bumelis, V.; Matulis, D., A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophysical journal* **2008,** *95* (7), 3222-31.

102.   Weidemann, T.; Seifert, J.-M.; Hintersteiner, M.; Auer, M., Analysis of Protein–Small Molecule Interactions by Microscale Equilibrium Dialysis and Its Application As a Secondary Confirmation Method for on-Bead Screening. *Journal of Combinatorial Chemistry* **2010,** *12* (5), 647-654.

103.   Bolshan, Y.; Getlik, M.; Kuznetsova, E.; Wasney, G. a.; Hajian, T.; Poda, G.; Nguyen, K. T.; Wu, H.; Dombrovski, L.; Dong, A.; Senisterra, G.; Schapira, M.; Arrowsmith, C. H.; Brown, P. J.; Al-awar, R.; Vedadi, M.; Smil, D., Synthesis, Optimization, and Evaluation of Novel Small Molecules as Antagonists of WDR5-MLL Interaction. *ACS Medicinal Chemistry Letters* **2013,** *4* (3), 353-357.

104.   Karatas, H.; Townsend, E. C.; Bernard, D.; Dou, Y.; Wang, S., Analysis of the binding of mixed lineage leukemia 1 (MLL1) and histone 3 peptides to WD repeat domain 5 (WDR5) for the design of inhibitors of the MLL1-WDR5 interaction. *Journal of medicinal chemistry* **2010,** *53* (14), 5179-85.

105.   Senisterra, G.; Chau, I.; Vedadi, M., Thermal denaturation assays in chemical biology. *Assay and drug development technologies* **2012,** *10* (2), 128-136.

106.   McDonnell, P. A.; Yanchunas, J.; Newitt, J. A.; Tao, L.; Kiefer, S. E.; Ortega, M.; Kut, S.; Burford, N.; Goldfarb, V.; Duke, G. J.; Shen, H.; Metzler, W.; Doyle, M.; Chen, Z.; Tarby, C.; Borzilleri, R.; Vaccaro, W.; Gottardis, M.; Lu, S.; Crews, D.; Kim, K.; Lombardo, L.; Roussell, D. L., Assessing compound binding to the Eg5 motor domain using a thermal shift assay. *Analytical Biochemistry* **2009,** *392* (1), 59-69.

107.   Layton, C. J.; Hellinga, H. W., Quantitation of protein-protein interactions by thermal stability shift analysis. *Protein Science* **2011,** *20* (8), 1439-1450.

108.   Pantoliano, M.; Petrella, E.; Kwasnoski, J.; Lobanov, V.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B.; Lane; Salemme, F., High-density miniaturized thermal shift assays as a general strategy for drug discovery. **2001,** *6* (6), 249-40.

109.   Fischer, G.; Rossmann, M.; Hyvönen, M., Alternative modulation of protein-protein interactions by small molecules. *Current Opinion in Biotechnology* **2015,** *35*, 78-85.

110.   Corral, J.; Lavenir, I.; Impey, H.; Warren, A. J.; Forster, A.; Larson, T. A.; Bell, S.; McKenzie, A. N. J.; King, G.; Rabbitts, T. H., An Mll-AF9 Fusion Gene Made by Homologous Recombination Causes Acute Leukemia in Chimeric Mice: A Method to Create Fusion Oncogenes. *Cell* **1996,** *85* (6), 853-861.

111.   Cao, F.; Townsend, E. C.; Karatas, H.; Xu, J.; Li, L.; Lee, S.; Liu, L.; Chen, Y.; Ouillette, P.; Zhu, J.; Hess, J. L.; Atadja, P.; Lei, M.; Qin, Z. S.; Malek, S.; Wang, S.; Dou, Y., Targeting MLL1 H3K4 Methyltransferase Activity in Mixed-Lineage Leukemia. *Molecular Cell* **2014,** *53* (2), 247-261.

112.   Odho, Z. a. S. S. M. a. W. J. R., Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *The Journal of biological chemistry* **2010,** *285* (43), 32967--76.

113.   Dias, J.; Van Nguyen, N.; Georgiev, P.; Gaub, A.; Brettschneider, J.; Cusack, S.; Kadlec, J.; Akhtar, A., Structural analysis of the KANSL1/WDR5/ KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. *Genes and Development* **2014,** *28* (9), 929-942.

114.   Spencer, C. A.; Groudine, M., Control of c-myc Regulation in Normal and Neoplastic Cells. In *Advances in Cancer Research*, George, F. V. W. a. G. K., Ed. Academic Press: 1991; Vol. Volume 56, pp 1-48.

115.   Harada, Y.; Katagiri, T.; Ito, I.; Nakamura, Y.; Emi, M.; Akiyama, F.; Sakamoto, G.; Kasumi, F., Genetic studies of 457 breast cancers. Clinicopathologic parameters compared with genetic alterations. *Cancer* **1994,** *74* (8), 2281-2286.

116.   Escot, C.; Theillet, C.; Lidereau, R.; Spyratos, F.; Champeme, M. H.; Gest, J.; Callahan, R., Genetic alteration of the c-myc protooncogene (MYC) in human primary breast carcinomas. *Proceedings of the National Academy of Sciences of the United States of America* **1986,** *83* (13), 4834-4838.

117.   Erisman, M. D.; Rothberg, P. G.; Diehl, R. E.; Morse, C. C.; Spandorfer, J. M.; Astrin, S. M., Deregulation of c-myc gene expression in human colon carcinoma is not accompanied by amplification or rearrangement of the gene. *Molecular and Cellular Biology* **1985,** *5* (8), 1969-1976.

118.   Brodeur, G. M.; Seeger, R. C.; Schwab, M.; Varmus, H. E.; Bishop, J. M., Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **1984,** *224* (4653), 1121.

119.   Little, C.; Nau, M.; Carney, D.; Gazdar, A.; J, M., Amplification and expression of the c-myc oncogene in human lung cancer cell lines. *Nature* **1983,** *306*, 194-196.

120.   Nau, M.; Brooks, B.; Battey, J.; Sausville, E.; Gazdar, A.; Kirsch, I.; McBride, O.; Bertness, V.; Hollis, G.; Minna, J., L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. *Nature* **1985,** *7* (13), 69-73.

121.   Nesbit, C. E.; Tersak, J. M.; Prochownik, E. V., MYC oncogenes and human neoplastic disease. *Oncogene* **1999,** *18* (19), 3004-3016.

122.   Leimbach, A.; Hacker, J.; Dobrindt, U., Small-Molecule Modulators of c-Myc/Max and Max/Max interactions. 2013; Vol. 358, pp 3-32.

123.   Cowling, V. H.; Chandriani, S.; Whitfield, M. L.; Cole, M. D., A Conserved Myc Protein Domain, MBIV, Regulates DNA Binding, Apoptosis, Transformation, and G(2) Arrest. *Molecular and Cellular Biology* **2006,** *26* (11), 4226-4239.

124.   Hintersteiner, M.; Buehler, C.; Uhl, V.; Schmied, M.; Muller, J.; Kottig, K.; Auer, M., Confocal nanoscanning, bead picking (CONA): PickoScreen microscopes for automated and quantitative screening of one-bead one-compound libraries. *Journal of combinatorial chemistry* **2009,** *11*, 886-894.

125.   Arpad, F., *Combinatorial Chemistry Principles and Techniques*. 2007.

126.   Meisner, N.-C.; Hintersteiner, M.; Seifert, J.-M.; Bauer, R.; Benoit, R. M.; Widmer, A.; Schindler, T.; Uhl, V.; Lang, M.; Gstach, H.; Auer, M., Terminal Adenosyl Transferase Activity of Posttranscriptional Regulator HuR Revealed by Confocal On-Bead Screening. *Journal of Molecular Biology* **2009,** *386* (2), 435-450.

127.   Lim, H.-S.; Archer, C. T.; Kodadek, T., Identification of a Peptoid Inhibitor of the Proteasome 19S Regulatory Particle. *Journal of the American Chemical Society* **2007,** *129* (25), 7750-7751.

128.   Gocke, A. R.; Udugamasooriya, D. G.; Archer, C. T.; Lee, J.; Kodadek, T., Isolation of Antagonists of Antigen-Specific Autoimmune T Cell Proliferation. *Chemistry & biology* **2009,** *16* (11), 1133-1139.

129.   Peng, L.; Liu, R.; Marik, J.; Wang, X.; Takada, Y.; Lam, K. S., Combinatorial chemistry identifies high-affinity peptidomimetics against [alpha]4[beta]1 integrin for in vivo tumor imaging. *Nat Chem Biol* **2006,** *2* (7), 381-389.

130.   Hintersteiner, M.; Buehler, C.; Auer, M., On-bead screens sample narrower affinity ranges of protein-ligand interactions compared to equivalent solution assays. *ChemPhysChem* **2012,** *13* (15), 3472-3480.

131.   Helmer, D.; Brahm, K.; Helmer, C.; Wack, J. S.; Brenner-Weiss, G.; Schmitz, K., Two-channel image analysis method for the screening of OBOC libraries. *Anal. Methods* **2016,** *00,* 1-11.

132.   Hintersteiner, M.; Kimmerlin, T.; Kalthoff, F.; Stoeckli, M.; Garavel, G.; Seifert, J. M.; Meisner, N. C.; Uhl, V.; Buehler, C.; Weidemann, T.; Auer, M., Single Bead Labeling Method for Combining Confocal Fluorescence On-Bead Screening and Solution Validation of Tagged One-Bead One-Compound Libraries. *Chemistry and Biology* **2009,** *16* (7), 724-735.

133.   McGann, M., FRED and HYBRID docking performance on standardized datasets. *Journal of Computer-Aided Molecular Design* **2012,** *26* (8), 897-906.

134.   Naylor, E.; Arredouani, A.; Vasudevan, S. R.; Lewis, A. M.; Parkesh, R.; Mizote, A.; Rosen, D.; Thomas, J. M.; Izumi, M.; Ganesan, A.; Galione, A.; Churchill, G. C., Identification of a chemical probe for NAADP by virtual screening. *Nat Chem Biol* **2009,** *5* (4), 220-226.

135.   Reynolds, C. H.; Bembenek, S. D.; Tounge, B. a., The role of molecular size in ligand efficiency. *Bioorganic & medicinal chemistry letters* **2007,** *17* (15), 4258-61.

136.   Ung, P.; Winkler, D. a., Tripeptide motifs in biology: targets for peptidomimetic design. *Journal of medicinal chemistry* **2011,** *54* (5), 1111-25.

137.   Neduva, V.; Russell, R. B., DILIMOT: discovery of linear motifs in proteins. *Nucleic acids research* **2006,** *34* (Web Server issue), W350-5.

138.   Zuckermann, R. N.; Martin, E. J.; Spellmeyer, D. C.; Stauber, G. B.; Shoemaker, K. R.; Kerr, J. M.; Figliozzi, G. M.; Goff, D. a.; Siani, M. a.; Simon, R. J., Discovery of nanomolar ligands for 7-transmembrane G-protein-coupled receptors from a diverse N-(substituted)glycine peptoid library. *Journal of medicinal chemistry* **1994,** *37* (17), 2678-85.

139.   Kraft, C.; Vodermaier, H. C.; Maurer-Stroh, S.; Eisenhaber, F.; Peters, J.-M., The WD40 propeller domain of Cdh1 functions as a destruction box receptor for APC/C substrates. *Molecular cell* **2005,** *18* (5), 543-53.

140.   Ambrosini, G.; Adida, C.; Altieri, D., A Novel Anti-apoptosis gene, Survivin, expressed in Cancer and Lymphoma. *Nature medicine* **1997,** *3* (8), 917--921.

141.   Fortugno, P.; Wall, N. R.; Giodini, A.; O'Connor, D. S.; Plescia, J.; Padgett, K. M.; Tognin, S.; Marchisio, P. C.; Altieri, D. C., Survivin exists in immunochemically distinct subcellular pools and is involved in spindle microtubule function. *Journal of cell science* **2002,** *115* (Pt 3), 575-585.

142.   Verdecia, M. a.; Huang, H.; Dutil, E.; Kaiser, D. a.; Hunter, T.; Noel, J. P., Structure of the human anti-apoptotic protein survivin reveals a dimeric arrangement. *Nature structural biology* **2000,** *7* (7), 602-608.

143.   Xia, H.; Chen, S.; Huang, H.; Ma, H., Survivin over-expression is correlated with a poor prognosis in esophageal cancer patients. *Clinica Chimica Acta* **2015,** *446*, 82-85.

144.   Shinohara, E. T.; Gonzalez, A.; Massion, P. P.; Chen, H.; Li, M.; Freyer, A. S.; Olson, S. J.; Andersen, J. J.; Shyr, Y.; Carbone, D. P.; Johnson, D. H.; Hallahan, D. E.; Lu, B., Nuclear survivin predicts recurrence and poor survival in patients with resected nonsmall cell lung carcinoma. *Cancer* **2005,** *103* (8), 1685-1692.

145.   Fukuda, S.; Pelus, L. M., Survivin, a cancer target with an emerging role in normal adult tissues. *Molecular cancer therapeutics* **2006,** *5* (5), 1087-1098.

146.   Li, F., Role of survivin and its splice variants in tumorigenesis. *British journal of cancer* **2005,** *92* (2), 212-6.

147.   Li, F.; Ackermann, E. J.; Bennett, C. F.; Rothermel, a. L.; Plescia, J.; Tognin, S.; Villa, a.; Marchisio, P. C.; Altieri, D. C., Pleiotropic cell-division defects and apoptosis induced by interference with survivin function. *Nature cell biology* **1999,** *1* (8), 461-466.

148.   Plescia, J.; Salz, W.; Xia, F.; Pennati, M.; Zaffaroni, N.; Daidone, M. G.; Meli, M.; Dohi, T.; Fortugno, P.; Nefedova, Y.; Gabrilovich, D. I.; Colombo, G.; Altieri, D. C., Rational design of shepherdin, a novel anticancer agent. *Cancer Cell* **2005,** *7* (5), 457-468.

149.   Carrasco, R. a.; Stamm, N. B.; Marcusson, E.; Sandusky, G.; Iversen, P.; Patel, B. K. R., Antisense inhibition of survivin expression as a cancer therapeutic. *Molecular cancer therapeutics* **2011,** *10* (2), 221-232.

150.   Wendt, M. D.; Sun, C.; Kunzer, A.; Sauer, D.; Sarris, K.; Hoff, E.; Yu, L.; Nettesheim, D. G.; Chen, J.; Jin, S.; Comess, K. M.; Fan, Y.; Anderson, S. N.; Isaac, B.; Olejniczak, E. T.; Hajduk, P. J.; Rosenberg, S. H.; Elmore, S. W., Discovery of a novel small molecule binding site of human survivin. *Bioorganic & medicinal chemistry letters* **2007,** *17* (11), 3122-3129.

151.   Chettiar, S. N.; Cooley, J. V.; Park, I. H.; Bhasin, D.; Chakravarti, A.; Li, P. K.; Li, C.; Jacob, N. K., Design, synthesis and biological studies of Survivin Dimerization Modulators that prolong mitotic cycle. *Bioorganic and Medicinal Chemistry Letters* **2013,** *23* (19), 5429-5433.

152.   Mahotka, C.; Liebmann, J.; Wenzel, M.; Suschek, C. V.; Schmitt, M.; Gabbert, H. E.; Gerharz, C. D., Differential subcellular localization of functionally divergent survivin splice variants. *Cell death and differentiation* **2002,** *9* (12), 1334-42.

153.   Click2Drug Directory of computer-aided Drug Design tools. https://www.click2drug.org/directory_StructureBasedScreening.html.

154.   Wuts, P.; Greene, T., *Protective Groups in Organic Synthesis*. 2006.

155.   Detraz, H., Maurice Ceresole. 1860-1936. *Helvetica Chimica Acta* **1937,** *20* (1), 999-1008.

156.   Dwight, S. J.; Levin, S., Scalable Regioselective Synthesis of Rhodamine Dyes. *Organic Letters* **2016,** *18* (20), 5316-5319.

157.   Menchen, S. M.; Fung, S. 5- and 6- Succinimidyl - carboxylate isomers of rhodamine dyes. 1988.

158.   Stagge, F.; Mitronova, G. Y.; Belov, V. N.; Wurm, C. A.; Jakobs, S., Snap-, CLIP- and Halo-Tag Labelling of Budding Yeast Cells. *PLoS ONE* **2013,** *8* (10), 1-9.

159.   Gross, E., The cyanogen bromide reaction. *Methods in enzymology* **1967,** *13* (1877), 238--255.

160.   Upadhyaya, P.; Qian, Z.; Selner, N. G.; Clippinger, S. R.; Wu, Z.; Briesewitz, R.; Pei, D., Inhibition of Ras Signaling by Blocking Ras-Effector Interactions with Cyclic Peptides. *Angewandte Chemie International Edition* **2015,** n/a-n/a.

161.   Yu, Z.; Chu, Y.-h., Combinatorial Epitope Search: Pitfalls of Library Design. *Bioorganic & Medicinal Chemistry Letters* **1997,** *7* (1), 95-98.

162.   Huang, H. V.; Bond, M. W.; Hunkapiller, M. W.; Hood, L. E., Cleavage at tryptophanyl residues with dimethyl sulfoxide-hydrochloric acid and cyanogen bromide. *Methods in Enzymology* **1983,** *91* (C), 318-324.

163.   Lee, S. S.; Lim, J.; Cha, J., Rapid microwave-assisted CNBr cleavage of bead-bound peptides. *Journal of combinatorial …* **2008,** *10* (6), 807-809.

164.   Pedersen, S. L.; Tofteng, A. P.; Malik, L.; Jensen, K. J., Microwave heating in solid-phase peptide synthesis. *Chemical Society reviews* **2012,** *41* (5), 1826-44.

165.   Fara, M. A.; Díaz-Mochón, J. J.; Bradley, M., Microwave-assisted coupling with DIC/HOBt for the synthesis of difficult peptoids and fluorescently labelled peptides - A gentle heat goes a long way. *Tetrahedron Letters* **2006,** *47* (6), 1011-1014.

166.   Galanis, A. S.; Albericio, F.; Grøtli, M., Enhanced microwave-assisted method for on-bead disulfide bond formation: Synthesis of α-comotoxin MII. *Biopolymers* **2009,** *92* (1), 23-34.

167.   Johnstone, R. W.; Wang, J.; Tommerup, N.; Vissing, H.; Roberts, T.; Shi, Y., Ciao 1 Is a Novel WD40 Protein That Interacts with the Tumor Suppressor Protein WT1. *Journal of Biological Chemistry* **1998,** *273* (18), 10880-10887.

168.   Thomas, L. R.; Wang, Q.; Fesik, S. W.; Tansey, W. P.; Thomas, L. R.; Wang, Q.; Grieb, B. C.; Phan, J.; Foshage, A. M.; Sun, Q.; Olejniczak, E. T.; Clark, T.; Dey, S.; Lorey, S.; Alicie, B.; Howard, G. C.; Cawthon, B.; Ess, K. C.; Eischen, C. M.; Zhao, Z.; Fesik, S. W.; Tansey, W. P., Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. *Molecular Cell* **2015**, 1-13.

169.   Gill, S. C.; von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* **1989,** *182* (2), 319-326.

170.   Heirwegh, K. P. M.; Meuwissen, J. A. T. P.; Lontie, R., Selective absorption and scattering of light by solutions of macromolecules and by particulate suspensions. *Journal of Biochemical and Biophysical Methods* **1987,** *14* (6), 303-322.

171.   Preibisch, S.; Saalfeld, S.; Tomancak, P., Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **2009,** *25* (11), 1463-1465.