Exploring RGD-integrin mimics as a novel therapeutic approach in diabetes

Daniel Gugan

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds

School of Chemistry

Leeds Institute of Cardiovascular and Metabolic Medicine

May 2025

I confirm that the work submitted is my own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

My thanks to my supervisors; Dr Katie Simmons, Dr Martin McPhillie, Professor Colin Fishwick and Professor Stephen Wheatcroft.

Thanks to The Dr Simon Waterworth Studentship for funding the PhD studies.

Thanks to Dr Jeanine Williams for running the HPLC purity analysis and Dr Iain Manfield for the assistance with the SPR facilities.

Finally, thanks to Dr Scott Grossman, the Fishwick/McPhillie group and the IGFBP-1 team.

This research has been carried out by a team which has included Dr Lia De Faveri and Emily Rolfe. My own contributions, fully and explicitly indicated in the thesis, have been the modelling and synthesis of the compounds, and the SPR analysis of the compounds.

The other members of the group and their contributions have been as follows: Dr Lia de Faveri and Emily Rolfe developed and ran the live/dead, phosphorylation of Akt and scratch-wound assays, and developed the SPR assay.

Contents

1. Abstract

Type 2 Diabetes (T2D) was estimated in 2021 to affect over 480 million people and can lead to the development of cardiovascular disease (CVD), which is the cause of death for 50.3% of people with T2D. Insulin-like growth factor binding protein 1 (IGFBP-1) overexpression has been shown to improve insulin sensitivity, promote nitric oxide production, reduce blood pressure and protect against atherosclerosis. IGFBP-1 contains an Arginine-Glycine-Aspartic acid (RGD) peptide sequence, used in many proteins to bind to integrins, which are transmembrane proteins, primarily involved in the extracellular matrix. IGFBP-1 specifically binds to the $\alpha_5\beta_1$ integrin via a flexible loop which contains the RGD sequence. It has been demonstrated that hexapeptides containing RGD have the same beneficial effects of IGFBP-1. The research presented investigates if small molecules could be used to mimic the effects of IGFBP-1 overexpression. From preliminary ROCS virtual shape screening and subsequent hit validation by a phosphorylation of Akt assay, two series of compounds were designed from the hit matter using docking in Schrodinger Maestro. The first series was zwitterion based, with a carboxylic acid mimicking the aspartate in the endogenous ligand, and a glycine, histidine or arginine at the other terminus. Phenyl, piperidine, pyrazole and alkyl chains were explored as linking groups. High hydrophilicity and poor solubility limited the synthesis; however, seven compounds were submitted for biological analysis. None of the compounds showed increased cell death but none showed an increase from the baseline in a limited phosphorylation of Akt assay. Direct binding to $\alpha_5\beta_1$ was measured with surface plasmon resonance (SPR), with only one compound, **5.6.1**, showing a measurable K_D of 59 μM ± 0.64-fold. A second series from another hit molecule was designed, replacing the carboxylic acid with a hydantoin group, which is neutral charge at physiological pH but has precedent for metal binding. After the amide linkage of the hydantoin in the hit molecule could not be replicated, an imine linked molecule was produced with a piperazine terminus, compound **6.3.6**. No toxicity was detected in a live/dead assay and the molecule and the Boc-protected intermediate had response in the limited phosphorylation of Akt assay. Modifications to the basic terminus and linker were designed and docked, incorporating examples used in literature. Sixteen of these targets were successfully synthesised and analysed using SPR. Five compounds had K_D values measured, with the best two compounds **6.5.10** and **6.7.3** showing K_D values of 70 μM ± 0.65-fold and 71 μM ± 0.52-fold respectively.

2. Abbreviations and figures

2.1 Abbreviations

ADMIDAS	Adjacent metal ion-dependent adhesion site
Akt	Protein kinase B
AMP	Adenosine monophosphate
AMPK	AMP activated protein kinase
aq.	Aqueous
BCA	Bicinchoninic acid
BMI	Body-mass index
Boc	*Tert*-butyloxycarbonyl
br	Broad
CADD	Computer aided drug design
Cbz	Benzyloxycarbonyl
C/EBP	CCAAT-enhancer-binding protein
CCDC	Cambridge crystallographic data centre
CVD	Cardiovascular disease
d	Doublet
DCM	Dichloromethane
DIPEA	Di-*iso*propylethylamine
DKD	Diabetic kidney disease
DMAP	4-Dimethylaminopyridine
DME	Diabetic macular edema
DMEM	Dulbecco's modified eagle's medium
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DPP-4	Dipeptidyl peptidase 4
EC_{50}	Half maximal effective concentration
ECM	Extracellular matrix
EDC	*N*-[3-(Dimethylamino)propyl]-*N*'-ethylcarbodiimide
ELISA	Enzyme-linked immunosorbent assay
equiv.	Number of equivalents
ERK1/2	Extracellular signal-regulated kinase 1 and 2
ES	Electrospray ionisation
ESC	European society of cardiology
Et	Ethyl
FAK	Focal adhesion kinase
FDA	Food and drug administration
Fmoc	Fluorenylmethyloxycarbonyl
GI	Gastrointestinal
GIP	Gastric inhibitory polypeptide
GLP-1	Glucagon-like peptide-1
GLUT4	Glucose transporter 4
HbA1c	Percentage of glycated haemoglobin in the blood
HPC	High performance computing
HPLC	High performance liquid chromatography
HTS	High throughput screening

I-EGF	Integrin epidermal growth factor
IC_{50}	Half maximal inhibitory concentration
IGF	Insulin-like growth factor
IGFBP	Insulin-like growth factor binding protein
ILK	Integrin-linked kinase
IRE	Insulin response element
IU	International unit
IUGR	Intrauterine growth restriction
IV	Intravenous
K_D	Equilibrium dissociation constant
LCMS	Liquid chromatography mass spectrometry
m	Multiplet
MACE	Major adverse cardiovascular events
Me	Methyl
MIDAS	Metal ion-dependant adhesion site
MDAP	Mass directed automated purification
MS	Mass spectrometry
NHS	National health service
NMR	Nuclear magnetic resonance
NO	Nitric oxide
NOESY	Nuclear Overhauser effect spectroscopy
Pbf	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
PBS	Phosphate buffered saline
PDB	Protein data bank
PI3-K	Phosphatidylinositol-3-kinase
pK_a	Acid dissociation constant
PTEN	Phosphatase and tension homolog
q	Quartet
RA	Receptor agonist
ReCET	Re-cellularization by electroporation therapy
ROCS	Rapid overlay of chemical structures
rt	Room temperature
RT	Retention time
s	Singlet
SAR	Structure-activity relationship
SGLT2	Sodium-glucose cotransporter-2
SPR	Surface plasmon resonance
Su	Succinamide
t	Triplet
T2D	Type 2 diabetes
T3p	Propanephosphonic acid anhydride
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TLC	Thin layer chromatography
Trt	Trityl
UKPDS	UK Prospective Diabetes Study
UV	Ultraviolet
WT	Wild type

2.2 Amino acid abbreviations

Amino acid name	Three letter abbreviation	Single letter abbreviation
Alanine	ALA	A
Arginine	ARG	R
Asparagine	ASN	N
Aspartic acid	ASP	D
Cysteine	CYS	C
Glutamic acid	GLU	E
Glutamine	GLN	Q
Glycine	GLY	G
Histidine	HIS	H
Isoleucine	ILE	I
Leucine	LEU	L
Lysine	LYS	K
Methionine	MET	M
Phenylalanine	PHE	F
Proline	PRO	P
Serine	SER	S
Threonine	THR	T
Tryptophan	TRP	W
Tyrosine	TYR	Y
Valine	VAL	V

2.3 List of figures

Figure 4: Promoter region of IGFBP-1, where factors that regulate the levels of the protein are effective. RE: response element for the various proteins. Adapted from J. H. Bae *et al.*

Figure 5: Integrin signalling pathway following activation of the α5β1 integrin by an RGD binding ligand. Activation of Src and Fak lead to the phosphorylation of PI3K, which leads to the phosphorylation of Akt. Adapted from P. Moreno-Layseca and C. H. Streuli.[41] Figure made with BioRender.

Figure 6: The possible combinations of α (blue) and β (green) subunits in integrins. 24 integrins are known to form with eight recognising the RGD binding motif.

Figure 7: Amino acids adjacent to the RGD binding motif, known to enhance selectivity for binding to α5β1 integrin.

Figure 8: The three proposed conformations of the α5β1 integrin. The inactive, bent-closed conformation is estimated to be occupied 99.76% of the time. It is theorised that by intracellular mechanical force applied to the integrin tail, the extended-closed conformation is adopted, which stabilises to the extended-open conformation which is maintained by extracellular force from the ECM. Adapted from Z. Sun *et al.*[62] Figure made with BioRender.

Figure 9: Crystal structures rendered in PyMol. Each RGD sequence is displayed with spherical atoms. The α5 monomers are shown in green and β1 monomers are shown in blue (B and C) A. C-terminal domain of IGFBP-1 (1ZT3) with the RGD sequence, which resides on a flexible loop. B. The Nagae crystal structure of the α5β1 integrin (3VI4), with the RGD motif bound to the active site. C. The Xia crystal structure of the α5β1 integrin (4WK0), with the RGD motif bound to the active site.

Figure 10: The Nagae crystal structure (3VI4) with the RGD motif bound to the active site. The key residues for RGD binding have been highlighted. The arginine (R) residue forms hydrogen bonds with GLN221 and ASP227. The glycine (G) residue interacts with SER227. The aspartate (D) residue interacts with TYR133, ASN224 and the Mg502 ion.

Figure 11: Approved small molecule therapeutics that target the RGD site of integrins, eptifibatide and tirofiban. Both are used to treat acute coronary syndrome and thrombotic cardiovascular events.

Figure 12: Examples of methods to improve integrin specificity in literature. Methylation of the chain linked nitrogen of guanidine significantly reduces binding potency to αvβ3

integrin, while slightly reducing binding potency to α5β1 integrin. Large functional groups adjacent to the metal binding carboxylate can be manipulated for specificity to a range of integrins.

Figure 13: Examples of RGD mimetic integrin inhibitors in literature. Structural similarities between these compounds include a carboxylate group for metal binding, a bulky aromatic group adjacent to the carboxylate for integrin specificity and a terminal basic group mimicking the guanidine of the RGD endogenous ligand.

Figure 14: Structures of small-molecule antagonists of the α5β1 integrin with the associated IC50 of each. There are close structural similarities to the RGD ligand in most of the compounds, including the entire RGD sequence in cilengitide (**3.4.17**). ATN-161 (**3.4.16**) does not share the similarities as it does not target the RGD binding site.

Figure 15: Functional groups used to mimic the terminal guanidine of the RGD ligand in literature. Heckmann *et al.* found that functional groups with a lower basicity reduced the binding affinity with the integrin.

Figure 16: RGD site integrin ligands. The compounds with the highest activity for α5β1 from each publication are displayed, along with the IC50/EC50 values. A sample of **GM263** was shared by Prof. Daria Giacomini as a positive control.

Figure 17: An example of the stages of a molecular docking workflow using Schrodinger Maestro and Glide. (A) The protein is prepared for docking by adding hydrogen atoms, resolving clashes and optimising conformations, amongst other functions. (B) Small molecules are prepared, accounting for possible charges, chirality and conformations. (C) A grid is generated in which the ligands will be docked. Grid settings can be finely adjusted, such as requiring or preventing interactions with certain atoms or residues. (D) Ligands are docked into the grid and energetic outputs can be assessed.

Figure 18: A selection of compounds identified using the virtual screening platform ROCS to match the shape and electrostatic properties of the RGD sequence in IGFBP-1.

Figure 19: Activity profile of the initial set of compounds, measured by the insulin-stimulated phosphorylation of Akt with 1 μM of compound (24 hours treatment) compared to the GRGDSP and GRADSP hexapeptides, IGFBP-1 and negative controls. n=3; Akt: protein kinase B.

dosed with 100 nM of the controls and compounds and incubated for 40 mins, before exposure to insulin. Both graphs are n=1; Akt: protein kinase B.

Figure 30: Cell migration of HUVEC cells, 24 hours after a scratch-wound while exposed to a selection of proteins, controls and zwitterion compounds (n=4, mean ± SEM)

Figure 31: Experimental setup of surface plasmon resonance (SPR). A derivatised gold film surface on a semi-circular prism reflects a beam to a detector. The refractive index the surface is modified by the molecular weight of the derivatised surface, meaning the association and dissociation of an analyte can be detected, allowing measurements of the analyte-surface affinity. Figure made with BioRender.

Figure 32: Theoretical and experimental responses from the titration SPR experiments. Rligand (shown on the left Y axis) is the protein binding level which decreases over time due degradation of the chip surface. Rmax (shown on the right Y axis) is the theoretical maximum binding of the analytes, based on Rligand, the analyte molecular weight, the ligand molecular weight and assuming a single binding site. Ranalyte (shown on the right Y axis) is the experimentally measured response. Separate batches of experiments are distinguished with vertical grey lines on the X axis.

Figure 33: Theoretical and experimental responses and binding to the reference cell from the titration SPR experiments. Rmax is the theoretical maximum binding of the analytes, based on Rligand, the analyte molecular weight, the ligand molecular weight and assuming a single binding site. Ranalyte is the experimentally measured response. The regularly spaced peaks in the binding to reference are at the beginning of each batch of 123 samples, where the buffer is run and shows a large reference response. Separate batches of experiments are distinguished with vertical grey lines on the X axis.

Figure 34: Binding to immobilised biotinylated α5β1 integrin by control compounds (A) Positive hexapeptide GRGDTP, (B) Negative hexapeptide AAAAAA, (C) Literature compound **GM263,** at six concentrations from 0-100 μM (n=3, mean ± SEM)

Figure 35: Binding to immobilised biotinylated α5β1 integrin by zwitterion compounds (A) **5.3.7** (B) **5.3.9** (C) **5.4.9** (D) **5.6.1,** at six concentrations from 0-100 μM (n=3, mean ± SEM)

Figure 37: Binding to immobilised biotinylated α5β1 integrin by controls and zwitterion compounds; (A) positive hexapeptide GRGDTP (B) negative hexapeptide AAAAAA (C)

Cell migration of HUVEC cells after a scratch-wound while exposed to a selection of proteins, controls and compounds **6.3.5** and **6.3.6.** (n=4, mean ± SEM)

Figure 46: Compound designs from compound **6.3.6,** separated by modification type. (A) Simple modifications to compound **6.3.6** to interrogate the binding mode. (B) Related amine termini in the group inventory for rapid production of analogues. (C) Diamines, designed as intermediates for derivatisation to functional groups such as urea and amidine. (D) Analogues with an aminopyridine terminus, as seen in literature compounds.

Figure 47: Molecular docking of the top scoring, phenyl linked hydantoin modifications to the RGD-binding domain of the α5β1 integrin (PDB: 4WK0) with close residues and hydrogen bonds highlighted. All but compound **6.5.3** form interactions with GLU126 and LYS182, as the original hit compound was modelled to, whereas compound **6.5.3** forms a bidentate hydrogen bond interaction with ASP227, similar to the endogenous RGD ligand.

Figure 48: Constrained molecular docking of the top scoring, phenyl linked hydantoin modifications to the RGD-binding domain of the α5β1 integrin (PDB: 4WK4) with close residues and hydrogen bonds highlighted. Key hydrogen bonds are made with TYR133, ASN224, ASP227, ALA222 and GLN221. The top two scoring structures reach a better score than the unconstrained docking, indicating that the constrained model may better represent the best possible compound interactions.

Figure 49: Derivatives of compound **6.3.6,** with the additional modification of a piperidine replacing the phenyl linker between the hydantoin imine and the basic terminus. (A) Modifications resembling compound **6.3.6.** (B) Modifications with a guanidine terminus, as is present on the RGD peptide.

Figure 50: Molecular docking of the top scoring, piperidine linked hydantoin modifications to the RGD-binding domain of the α5β1 integrin (PDB: 4WK0) with close residues and hydrogen bonds highlighted. As with the phenyl linked modifications, the top scoring structures, **6.5.5** and **6.5.6,** form interactions with GLU126, while other structures, **6.5.7** and **6.5.8,** form interactions with ASP227, ASP228 and GLU320.

Figure 51: Constrained molecular docking of the top scoring, piperidine linked hydantoin modifications to the RGD-binding domain of the α5β1 integrin (PDB: 4WK4) with close residues and hydrogen bonds highlighted. The highest scoring structures, **6.5.9** and **6.5.7** had the highest scores seen in the molecular modelling in the project thus far, both opting to form hydrogen bonds between the terminal guanidine and

GLU320. The next best scoring structures had lower scores, posed in a linear shape similar to the RGD peptide.

Figure 52: Molecular docking of the top scoring, phenylalanine linked hydantoin modifications to the RGD-binding domain of the α5β1 integrin (PDB: 4WK0) with close residues and hydrogen bonds highlighted. Both structures achieved high scores by adopting a contorted pose to form interactions with ASP227, ASP228 and GLU320.

Figure 53: Constrained molecular docking of the top scoring, phenylalanine linked hydantoin modifications to the RGD-binding domain of the α5β1 integrin (PDB: 4WK4) with close residues and hydrogen bonds highlighted. The docking scores were reduced compared to the unconstrained docking, although the binding pose matches the crystal structure of the RGD peptide, forming interactions with GLN221 and ASP227 with the guanidine group.

Figure 54: Cell viability of C2C12 (n=3, mean ± SEM) and HUVEC (n=4, mean ± SEM) cell lines after exposure to a selection of hydantoin compounds (100 μM).

Figure 55: Binding to immobilised biotinylated α5β1 integrin by control compounds (A) Positive hexapeptide GRGDTP, (B) Negative hexapeptide AAAAAA, (C) Literature compound **GM263,** at six concentrations from 0-100 μM (n=3, mean ± SEM)

Figure 56: Binding to immobilised biotinylated α5β1 integrin by hydantoin compounds (A) **6.3.5** (B) **6.3.6** (C) **6.6.2** (D) **6.6.4** (E) **6.7.4** (F) **6.8.3,** at six concentrations from 0-100 μM (n=3, mean ± SEM)

Figure 57: Structures of six compounds that returned weak or no dose dependant activity from the SPR narrow titration experiment, shown in Figure 56.

Figure 58: Binding to immobilised biotinylated α5β1 integrin by hydantoin compounds (A) **6.5.2** (B) **6.5.5** (C) **6.5.10** (D) **6.6.1** (E) **6.6.3** (F) **6.7.3,** at six concentrations from 0-100 μM (n=3, mean ± SEM)

Figure 59: Structures of six compounds that returned dose dependant activity from the SPR narrow titration experiment, shown in Figure 58.

Figure 60: Binding to immobilised biotinylated α5β1 integrin by hydantoin compound **6.5.1** (structure shown) at six concentrations from 0-100 μM (n=3, mean ± SEM)

Figure 61: Binding to immobilised biotinylated α5β1 integrin by hydantoin compounds (A) **6.5.2** (B) **6.5.10** (C) **6.6.3** (D) **6.7.3,** at eight concentrations from 0-400 μM, with measured KD values (n=4, mean ± SEM)

2.4 List of tables

2.5 List of schemes

2.6 List of equations

3. Introduction

3.1. Diabetes: cost, treatment and failings

Type 2 Diabetes (T2D) is defined as a condition in which the body's tissues become resistant to the effects of insulin, combined with insufficient production of insulin by the pancreas to compensate for the insulin resistance.[1] It represents approximately 90% of diabetes cases and is associated with a shortening of life of up to 12 years (for very obese men aged 20-39) and years of healthy life reduced by 22 years, with lower years lost for those diagnosed over 39 with a lower BMI.[2] The direct cost of treating T2D in 2010/11 was estimated to be £8.8 billion, representing 8.9% of the NHS spending.[3,4] In comparison, cancer is modelled to represent 5.0% of the UK healthcare budget in 2018.[5] Worldwide cases of T2D are estimated to be over 500 million and are projected to rise to over 700 million by 2045.[6] The factors that cause T2D are complex. Half of individuals with T2D are obese and almost 90% are overweight, but factors including hereditary and environmental factors, in addition to lifestyle choices (diet, smoking, inactivity) are implicated.[7-9]

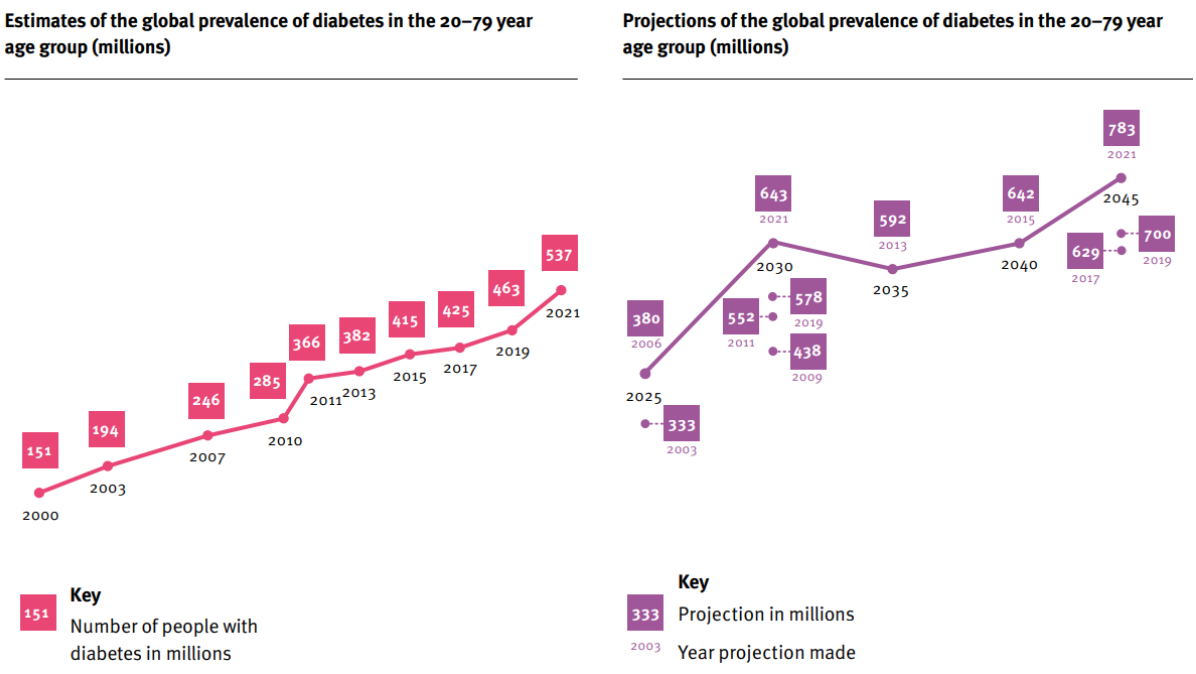


Figure 1: Estimates and projection of the prevalence of diabetes from the International Diabetes Federation. Type 2 Diabetes accounts for over 90% of cases. Figure taken from the IDF Diabetes Atlas 2021.[6]

Clinically, T2D is diagnosed when blood glucose measurements reach a defined threshold. A standard measure for blood glucose is the amount of glycated haemoglobin in the blood, known as HbA1c. A person with a HbA1c consistently above 6.5% or 48 mmol/mol is defined as having diabetes.[10]

One of the most common complications for people with diabetes is the development of cardiovascular disease (CVD). The combined cost of treating cardiovascular complications of T2D, including ischaemic heart disease, myocardial infarction and heart failure was estimated in 2010/11 to be £2.8 billion, and CVD represented 50.3% of deaths of people with T2D in a meta-analysis of worldwide studies from 2007-2017.[3,11] This indicates there is a unmet need for treatments that can prevent CVD.

The first-line treatment of T2D and prediabetes is diet and exercise changes. The European Society of Cardiology (ESC) recommends a reduced calorie intake to lower excessive body weight, adoption of a mediterranean or plant-based diet and regular moderate-to-vigorous exercise.[12] A study of primary care led weight management for T2D patients found that 46% (n=68) of the intervention group achieved remission after 12 months, defined by HbA1c of < 6.5%.[13] Where control of T2D exclusively through diet and exercise is insufficient, use of medical intervention is necessary, but it is important to recognise that effective support of patients to manage the condition via diet and exercise is the best treatment. The current medicines used to treat T2D are described and discussed below.

Metformin (**3.1.1**), shown in Figure 2, is used as a first-line glucose control medication, with a reliable reduction of HbA1c levels by about 1.5%, a good safety profile and low cost.[10,14] The mechanism of action of metformin is complex, involving inhibition of the mitochondrial respiratory chain in the liver and increasing the glucose utilisation in the intestine amongst other effects, all with an unknown hierarchy of importance.[15] There are gastrointestinal (GI) side effects such as nausea and diarrhoea in 20-30% of patients, with about 5% discontinuing the drug as a result.[14, 15] In addition, metformin is contraindicated with chronic kidney disease as it increases the risk of lactic acidosis which is rare, but potentially fatal.[16]

Few studies have analysed the cardiovascular effects of metformin with meaningful control groups, as the overall safety and glycaemic control benefits are well established. The UK Prospective Diabetes Study (UKPDS) is the most comprehensive study, comparing use of metformin to dietary control only.[17] A meta-analysis, which was heavily weighted from the UKPDS, showed that all outcomes aside from stroke favoured metformin, but none showed statistical significance and concluded that definitive trials to establish this are unlikely to occur.[18]

Weight loss has been shown to drastically improve outcomes for T2D patients. Analysis of the Look AHEAD trial, 16.7 years after the start of the study showed that participants who lost ≥10% of weight after 1 year of intervention had a 21% reduced risk of mortality.[19]

As such, the glucose control agents that induce weight loss, glucagon-like peptide-1 receptor agonists (GLP-1 RAs) and sodium-glucose cotransporter-2 (SGLT2) inhibitors, are recommended by the European Society of Cardiology (ESC) for overweight and obese patients.[20]

SGLT2 inhibitors, such as empagliflozin (**3.1.2**), increase glucose excretion via urine by inhibiting the transporters that recover glucose in the kidneys. This reduces HbA1c by 0.5-0.8%, as well as reducing patient weight and blood pressure. The main side effect of empagliflozin is an increased risk of genital infections, with rarer effects reported to include ketoacidosis, fractures and foot amputations.[21] The other approved SGLT2 inhibitors are

GLP-1 RAs, such as liraglutide (**3.1.3**), function by agonising GLP-1 receptors to promote glucose-dependent insulin secretion, reducing HbA1c by 0.5-1%. The drawbacks of GLP-1 RAs are that, with the exception of one oral medication, they require injection dosing and up to 50% of patients experience GI side effects, although typically transient, which leads to 5-10% of patients discontinuing the treatment. The GLP-1 RAs approved for use by the FDA are semaglutide, tirzepatide, liraglutide, albiglutide, exenatide, dulaglutide and lixisenatide.[22] An oral formulation of selemglutide is in use under the name Rybelsus, which represents an accessible alternative for drug administration. A non-peptide, orally active GLP-1 RA, orforglipron has shown promising results in phase III trials for the treatment of T2D and weight loss.[23]

Tirzepatide agonises both GLP-1 and gastric inhibitory polypeptide (GIP), which also increases insulin secretion. This duel-effect have shown increased weight loss and glycemic control when compared to typical GLP-1 agonists.[22] Retatrutide, a triple targeting therapy (GLP-1, GIP and glucagon receptors) in clinical trials for weight loss, may represent the future direction of drugs in this class.[24]

In meta-analyses of cardiovascular outcome trials in people with T2D, only GLP-1 RAs and SGLT2 inhibitors have been shown to moderately reduce major adverse cardiovascular events (MACE) hazard ratios for patients with established atherosclerotic CVD (15% and 11% reduction respectively) and are recommended for cardiovascular risk management independent of glucose control considerations.

Another treatment option is thiazolidinediones, such as pioglitazone (**3.1.4**). They activate peroxisome proliferator-activated receptor-γ, which increases insulin sensitivity by upregulating protein synthesis in lipid homeostasis, which increases glucose usage for energy.[14] They are effective at reducing blood glucose levels and also have evidence of modest cardiovascular benefit for patients with CVD. In the PROactive trial,

pioglitazone had a 16% reduction in MACE, showing similar effects to GLP-1 RAs and SGLT2 inhibitors.[25] The side effects of pioglitazone include fluid retention, an annual 0.4% increased risk of heart failure and weight gain which are of significant concern and must be considered.

Other therapeutics for glycaemic control include sulfonylureas and dipeptidyl peptidase-4 (DPP-4) inhibitors.[10] Second and third generation sulfonylureas are an alternative when metformin is not tolerated or contraindicated. Their mechanism of action comprises binding to the sulfonylurea receptor-1 in the pancreatic β-cells which increases insulin secretion. DPP-4 inhibitors are also used as second line to metformin. Inhibiting DPP-4 prevents breakdown of GLP-1, acting in a similar manner to GLP-1 RAs, although there are unsuitable for patients with heart failure. The studies that have looked at medicines in these classes have shown no difference in MACE outcomes compared to a control.[20]

Another medication used for T2D, especially in China, is the α-glucosidase inhibitor, acarbose (**3.1.5**). Two major studies (STOP-NIDDM and ACE) have analysed its effects and shown that for patients with impaired glucose tolerance, progression to diabetes diagnosis was reduced by 25% and 18% respectively.[26, 27] A 22.5% reduction in cardiovascular effects was seen in the STOP-NIDDM trial (n=1429) however no significant benefit was seen in the ACE trial (n=6522). An analysis of the ACE trial suggests this difference could be due to chance as the number of participants having ≥1 adverse cardiovascular events in the STOP-NIDDM trial were only 47. Other suggestions include differences in dosing, age of participants, ethnicity of participants and cardiovascular prevention measures.[27]

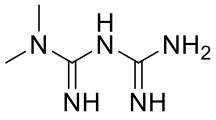
Aspirin has also been used to treat CVD in cases of diabetes when risk of bleeding is not elevated due to the medication's anticoagulant effect. Meta-analyses of trials analysing the CVD events of patients with and without diabetes showed a modest (approximately 9%) improvement in MACE outcomes but with a ≥2-fold increase in relative risk of bleeding.[28]

Bile acid sequestrants, such as colesevelam, are an alternative method of treating T2D. They remove bile acids from the intestine, forcing the body to increase synthesis from cholesterol, whilst also significantly reducing HbA1c levels.[29]

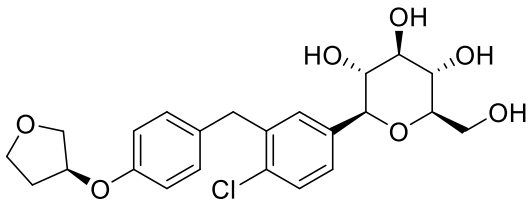
Another alternative treatment method is re-cellularization by electroporation therapy (ReCET), where cells in the duodenal mucosa (mucus membrane of the first section of the small intestine) are exposed to repeated high voltage pulses which induces pores opening in the cell membrane, thus killing the cells.[30] The duodenal mucosa cells have important roles in glucose homeostasis, and the regrown cells have been found to reduce

insulin resistance and improve glycemic control. 86% of patients who required basal insulin for glycemic control before treatment only required use of semaglutide, 12 months after ReCET treatment (with 14%, n=2 who returned to insulin use after receiving a lower voltage treatment of ReCET).[30]

Exercise remains a critical facet of treating T2D, with a 0.6% reduction in HbA1c levels found in a review of exercise in 2006.[31] Intensive treatment methods have a limited risk reduction in major macrovascular and microvascular events which, given the high risk of cardiovascular disease for T2D patients, needs improving.[10] New therapeutics need to be developed to improve the cardiovascular outcomes of T2D, with insulin-like growth factor 1 (IGFBP-1) a possible target for this.



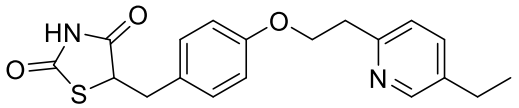
Metformin
3.1.1



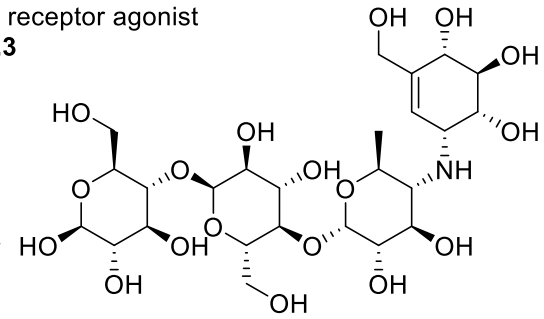
Empagliflozin - SGLT2 inhibitor
3.1.2

H_2N-His-Ala-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val-Ser-Ser-Tyr-Leu-Glu-Gly-Gln-Ala-Ala-Lys(Glu-C-16 fatty acid)-Glu-Phe-Ile-Ala-Trp-Leu-Val-Arg-Gly-Arg-Gly-OH

Liraglutide - GLP-2 receptor agonist
3.1.3



Pioglitazone - peroxisome proliferator-activated receptor-gamma agonist
3.1.4



Acarbose - alpha-glucosidase inhibitor
3.1.5

Figure 2: Structures of commonly prescribed therapeutics for type 2 diabetes.

Table 1: List of FDA approved medicines commonly prescribed for T2D in the drug classes detailed above.[32] Therapies without similar comparison, such as metfomin, are not included.

Drug class	**Approved drugs**
SGLT2 inhibitors	Bexaglifozin, canagliflozin, dapagliflozin, empagliflozin, ertugliflozin
GLP-1 receptor agonists	Albiglutide, dulaglutide, exenatide, liraglutide, lixisenatide, semaglutide, tirzepatide
Thiazolidinediones	Rosiglitazone, pioglitazone
Sulfonylureas	Glimepiride, glipizide, glyburide
DDP-4 inhibitors	Alogliptin, linagliptin, saxagliptin, sitagliptin
α-carbose inhibitors	Acarbose, miglitol

3.2 Metabolic homeostasis and IGFBP-1

Insulin and glucagon maintain blood sugar homeostasis. When blood sugar levels are high, the pancreas secretes additional insulin; when blood sugar levels drop, the pancreas releases glucagon. This system is controlled by a complex network of proteins including insulin-like growth factors (IGFs) and insulin-like growth factor binding proteins (IGFBPs). IGFs are growth-promoting, circulating proteins that act as both endocrine hormones and autocrine/paracrine growth factors. IGF-1 stimulates glucose uptake into cells and increases the sensitivity of tissues to insulin. IGF-1 concentrations are relatively constant, with activity modulated by interactions with circulating IGFBPs.[33] In addition, IGFBPs are known to have independent action of IGFs, with IGFBP-1 and IGFBP-2 of particular interest for treatment of T2D.[33]

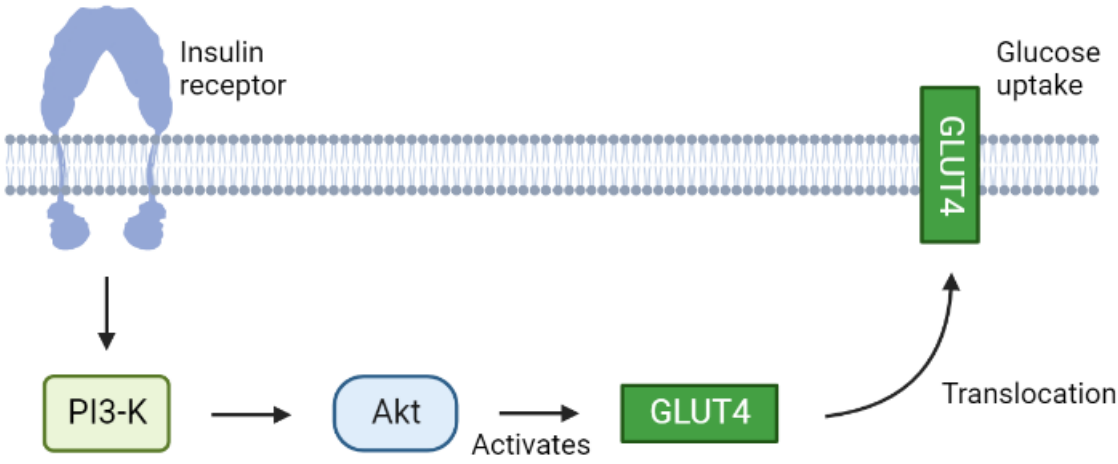


Figure 3: The glucose uptake biomechanism of the insulin receptor upon binding of insulin. PI3-K: phosphatidylinositol-3-kinase; Akt: protein kinase B; GLUT4: glucose transporter 4. Adapted from S. B. Wheatcroft and M. T. Kearney.[33]

Insulin modulates blood glucose concentration by binding to the insulin receptor, which activates a phosphorylation cascade of phosphatidylinositol-3-kinase (PI3-K), then protein kinase B (Akt), which causes translocation of glucose transporter 4 (GLUT4) to the cell membrane, allowing uptake of glucose from the blood to the cells (Figure 3). IGF-1 has also been shown to activate these cascades via the IGF-1 receptor.[34] IGFBPs, particularly IGFBP-1 and IGFBP-3, form a complex with IGF-1 that renders it biologically inert, regulating the IGF signalling.[35] IGFBP-1 has been identified as the main bioactivity control of IGF-1, controlled by three phosphorylation sites that when phosphorylated, forms a higher affinity complex with IGF-1 than the IGF-1 receptor.[35] The kinases that phosphorylate IGFBP-1 have been suggested to be part of the casein family, although this has not been confirmed.[36-38]

3.3 Structure and function of IGFBP-1

IGFBP-1 has a mass of about 30 kDa and is composed of structured N and C-terminal domains with a less structured linker domain. No full structure of the protein has been solved but the C-terminal domain, which is involved in IGF binding, has two crystal structures (1ZT3, 1ZT5) (Figure 9).[33, 39, 40] Both structures were collected by X-ray diffraction and have resolutions of 1.80 and 1.82 Å respectively, with the second structure complexed with iron. IGFBP-1 contains an RGD sequence on a flexible loop in the C-terminal domain which has IGF independent effects.[33]

In their 2013 review of IGFBP-1 regulation, Bae *et al.* described the various factors that affect IGFBP-1 transcription.[35] IGFBP-1 is mainly secreted in the liver, with a number of

factors stimulating secretion and insulin inhibiting gene expression. In the gene promotor region for IGFBP-1, shown in Figure 4, the insulin response element (IRE) can be blocked by CCAAT-enhancer-binding proteins (C/EBPs), where C/EBP *β* forms a complex with other factors and blocks insulin from binding to the IRE site, preventing the inhibition effect. Factors that have been indicated to stimulate IGFBP-1 secretion include AMP activated protein kinase (AMPK), cAMP, progesterone, relaxin, glucagon, peroxisome proliferator-activated receptor (PPAR)-γ agonists, glucocorticoids and hepatocyte nuclear factor 1 (HNF-1).[35]

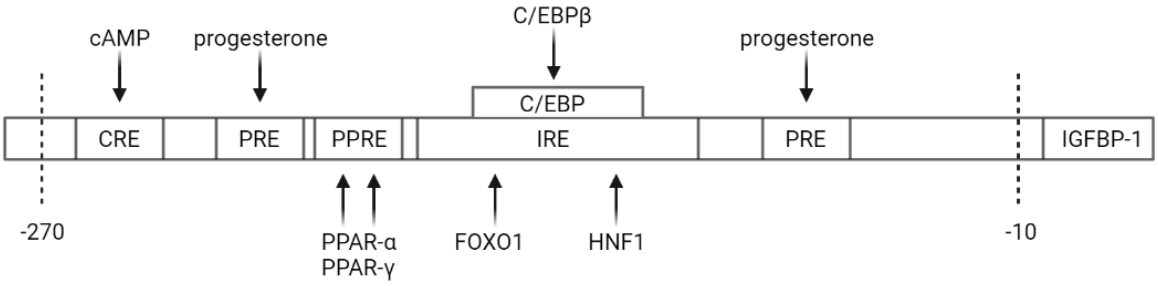


Figure 4: Promoter region of IGFBP-1, where factors that regulate the levels of the protein are effective. RE: response element for the various proteins. Adapted from J. H. Bae *et al.*[35]

When IGFBP-1 is not phosphorylated and no longer binds IGFs with a strong affinity, it has IGF independent interactions. After a meal, when blood glucose is high, it is beneficial for the IGFs to be released by IGFBPs and for the IGF-1 receptors to be activated and stimulate glucose uptake, much like insulin itself. The free IGFBP-1 can then bind to the $\alpha_5\beta_1$ integrin via its RGD motif, presented on a flexible loop in the C-terminal domain.[33] The RGD motif (arginine, glycine and aspartate) is recognised by several integrins and starts a phosphorylation cascade which activates the focal adhesion kinase (FAK) and proto-oncogene tyrosine-protein kinase (Src) which activate a phosphoinositide 3-kinase (PI3K), which subsequently activate Akt (Figure 5).[33, 41, 42]

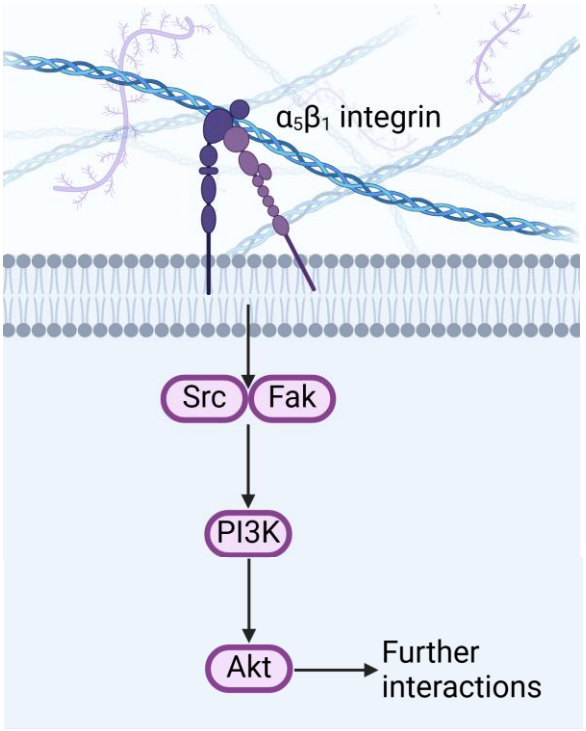


Figure 5: Integrin signalling pathway following activation of the $\alpha_5\beta_1$ integrin by an RGD binding ligand. Activation of Src and Fak lead to the phosphorylation of PI3K, which leads to the phosphorylation of Akt. Adapted from P. Moreno-Layseca and C. H. Streuli.[41] Figure made with BioRender.

Haywood *et al.* found that a dose of 0.75 IU/kg recombinant IGFBP-1 to mice improved whole-body insulin sensitivity and glucose tolerance. Overexpression of IGFBP-1 in mice was associated with reduced blood pressure, reduced susceptibility to atherosclerosis and had enhanced vascular nitric oxide (NO) production.[43] In relation to this, circulating IGFBP-1 levels have shown a strong, positive correlation to insulin sensitivity in a diverse range of study groups.[44, 45] Low levels of circulating IGFBP-1 have been shown to be a strong predictor of development of type 2 diabetes, although during onset of T2D, IGFBP-1 concentrations tend to increase.[46] In agreement is the finding that a high IGFBP-1 level is associated with a lower risk of diabetes in older adults.[47]

Further establishing the association with T2D, IGFBP-1 levels were found to significantly negatively correlate with several established cardiovascular risk factors, including blood pressure, BMI, waist:hip ratio and fasting insulin and triglyceride concentration.[48] In people with T2D, Leinonen *et al.* found that reduced IGFBP-1 concentrations correlated with an increase in the thickness of the carotid wall, which predicts clinical cardiovascular events.[49]

Other rodent studies have had differing results when overexpressing IGFBP-1, showing glucose intolerance, hyperglycaemia and hyperinsulinemia.[44] An explanation for these mixed results by Haywood *et al.* is differing phosphorylation statuses of IGFBP-1, with

the native protein likely to be phosphorylated and bind to IGFs, therefore reducing the metabolic response to glucose and interfering with the typical mechanisms.[44] In comparison, recombinant IGFBP-1 will have varying phosphorylation levels based on the production method, meaning that it may follow the IGF-independent mechanism of activating the insulin receptor cascade and having effects beneficial for hyperglycaemia.

In related rodent trials, overexpression of IGFBP-2 has been shown to protect against obesity and insulin resistance, however the mechanisms by which this occurs appear unrelated to IGFBP-1 and integrin signalling.[50]

IGFBP-1 has been linked to diabetic kidney disease (DKD), with reduced IGFBP-1 expression in early stage DKD for patients with T2D.[51] IGFBP-1 has also been linked to intrauterine growth restriction (IUGR) as a measure of nutrition and oxygenation. Increased phosphorylation of IGFBP-1 has been found to correlate to limited oxygen and nutrient availability and could be used as a biomarker for IUGR in the first trimester.[52] Finally, increased serum IGFBP-1 levels three months after an ischemic stroke were found to predict poor long-term outcomes and post-stroke mortality.[53]

3.3. Integrin structure and conformation

Integrins are proteins that operate across the cell membrane, with extracellular binding sites which mediate intercellular coordination and signalling. The primary function of integrins is as cell-adhesion receptors, allowing proteins such as fibronectin to form the extracellular matrix (ECM) to hold cells together. All integrins are heterodimers, composed of an α and β subunit. Integrins have an extracellular headpiece domain, a transmembrane domain and an intracellular cytoplasmic tail. Eighteen α and eight β subunits have been identified which assemble into 24 integrins (Figure 6). Eight of the integrins recognise the RGD (Arg-Gly-Asp) motif present on fibronectin, vitronectin and fibrinogen as well as IGFBP-1.[54, 55]

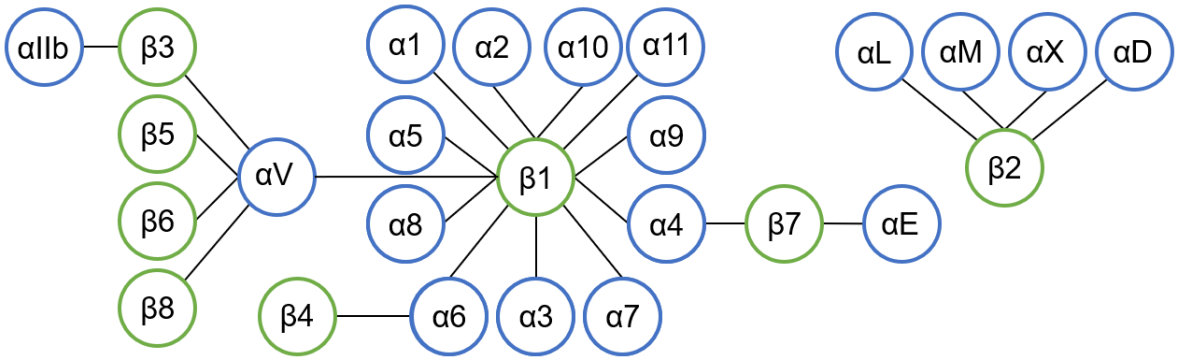


Figure 6: The possible combinations of α (blue) and β (green) subunits in integrins. 24 integrins are known to form with eight recognising the RGD binding motif.

Fibronectin is involved in several processes including cell adhesion, migration, growth and differentiation.[56] As with IGFBP-1, the RGD motif of fibronectin is located on a flexible loop which allows for optimal binding at the metal ion-dependent adhesion site (MIDAS, Figure 9). In addition to the RGD interaction, fibronectin binds selectively to $\alpha_5\beta_1$ and $\alpha_{IIb}\beta_3$ via a DVRPHSRN sequence which is located approximately 32 Å away from the RGD sequence. This enables fibronectin a forty-fold selectivity over other RGD binding integrins, but does not affect the RGD binding site.[57, 58] Recent research has uncovered an additional synergy site that fibronectin engages with when binding to $\alpha_5\beta_1$ on the β_1 monomer.[58] This is proposed as an explanation to why isolated RGD containing peptides induce only partial conformation change in the integrin.[59, 60]

Integrins and their ligands interact with specificity via the conformation of the RGD sequence and neighbouring residues. For $\alpha_5\beta_1$, flanking residues that improve ligand recognition include P -1 to the N terminus, M/N/G +1 to the C terminus and P/F/W +2 to the C terminus (Figure 7).[57] The W +2 neighbouring residue has been described to improve ligand specificity by forming hydrophobic interactions with TRP157 on the α_5 monomer.[61]

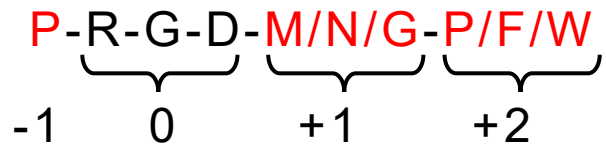


Figure 7: Amino acids adjacent to the RGD binding motif, known to enhance selectivity for binding to $\alpha_5\beta_1$ integrin.

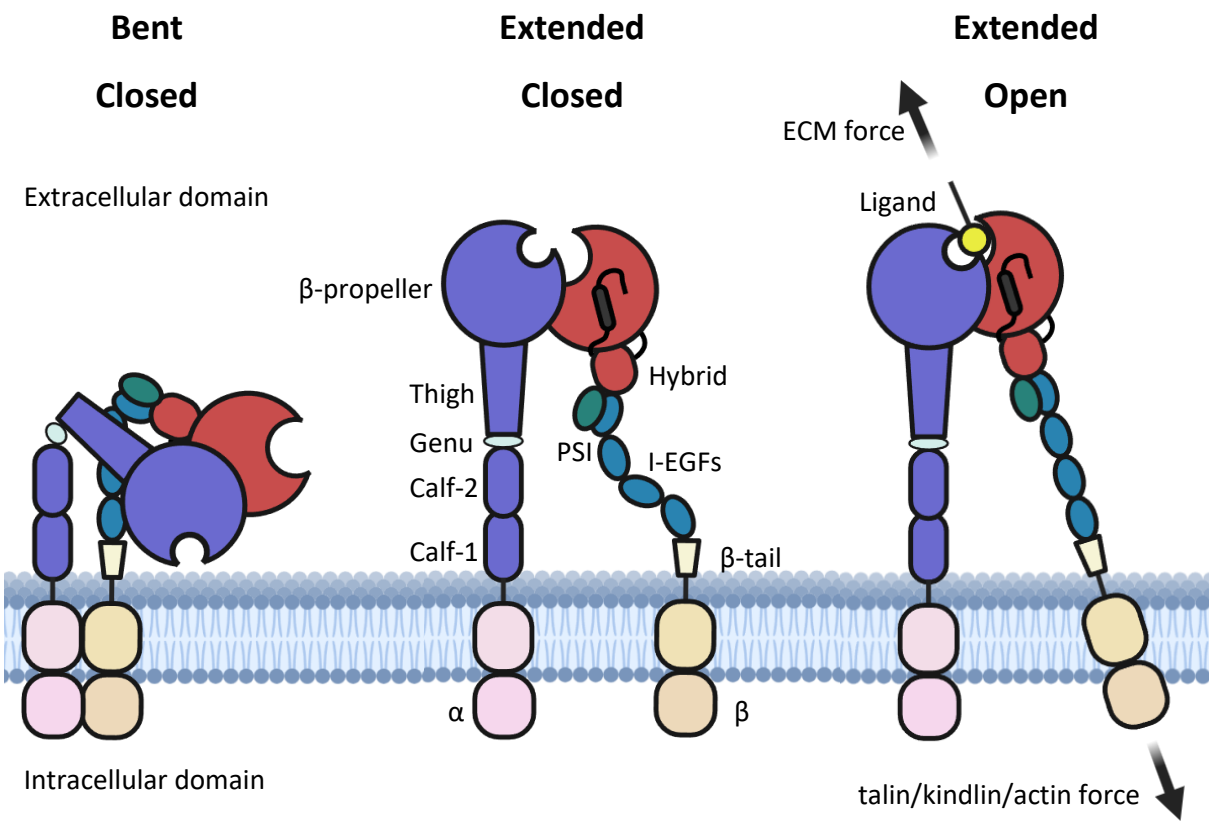


Figure 8: The three proposed conformations of the $\alpha_5\beta_1$ integrin. The inactive, bent-closed conformation is estimated to be occupied 99.76% of the time. It is theorised that by intracellular mechanical force applied to the integrin tail, the extended-closed conformation is adopted, which stabilises to the extended-open conformation which is maintained by extracellular force from the ECM. Adapted from Z. Sun *et al*.[62] Figure made with BioRender.

The standard model of integrin activation indicates three conformational states: bent-closed, extended-closed and extended-open (Figure 8).[63] The integrin headpiece opens and the 'leg' regions (extracellular components excluding the binding regions) of the headpiece move from 45 Å apart to 120 Å apart. Li *et al.* found that for $\alpha_5\beta_1$, the extended-open confirmation is directly associated with a 5000-fold increase in ligand binding affinity over the closed confirmation.[60, 63-65] Springer and Dustin describe that for RGD binding integrins, the conformational change at the MIDAS is limited to the β subunit where the β1-α1 loop, which coordinates to the MIDAS ion with Asp-X-Ser-X-Ser side chains and to the adjacent metal ion-dependent adhesion site (ADMIDAS) Ca^{2+} ion with the amide oxygen backbone, is shifted 2.3 Å to increase binding affinity of RGD-containing ligands.

The free energy of the conformational states of $\alpha_5\beta_1$ has been measured to be -3.8 kcal mol^{-1}, +0.4 kcal mol^{-1} and 0 kcal mol^{-1} for the bent-closed, extended-closed and extended open-states respectively.[65] With this energetic favourability, an estimated 99.76% of the integrins would exist in the bent-closed conformation. Two critical proteins in the

activation of $\alpha_5\beta_1$ are talin and kindlin, which interact with the intracellular tails of the integrin. This has been referred to as inside-out signalling. It is theorised that these proteins also rely on mechanical force, applied to the intracellular tail, to maintain the thermodynamically unfavourable extended-open conformation once adopted.[62] This mechanical force is matched by the ECM to hold the cell in place. ECM rigidity is encouraged by clustering extended-open integrins together, with many localised interactions of the ECM proteins and dimerisation and crosslinking of the intracellular proteins. Cilengitide (**3.4.17**, Figure 14), an RGD mimetic integrin inhibitor has been shown to induce conformational change to the extended-open state, which at low doses induces an agonistic effect.[66]

Conversely, recent research by Schumacher *et al.* found that $\alpha_5\beta_1$ does not adopt the typical integrin conformations, and instead exists in half-bent or extended-open states.[58] They found that fibronectin enabled the adoption of the extended-open state by binding to the half-bent state and engaging all three known binding sites between fibronectin and $\alpha_5\beta_1$. Isolated RGD-containing peptides were not found to promote the extended-open state which was justified by their lack of synergistic site engagement. Finally, they also found that in the absence of fibronectin, an approximately equal equilibrium between the half-closed and extended-open conformations formed. These conclusions are disputed by Li, Yan and Springer, stating that the nanodisc membrane scaffold used by Schumacher *et al.* frustrated opening by isolated RGD sequences and that the half-bent conformation fits within the bent-closed classification.[67]

Published research agrees that Mn^{2+} is critical to adoption of the extended-open conformation. When the ADMIDAS site is occupied by a Ca^{2+} ion, the bent-closed conformation is stabilised, inhibiting conformation change as a result of RGD binding.[68] As part of the integrin activation, the Ca^{2+} is replaced with Mn^{2+} in the extended-open state. A fluorescence binding assay measuring the binding of the cyclic peptide AC**RGD**GWCG, cyclised by a disulfide bond, found a 10-fold decrease in affinity to a six-domain headpiece of $\alpha_5\beta_1$ when 1 nM Ca^{2+} was included in the buffer.[68]

The MIDAS metal ion has typically been assigned as Mg^{2+} in all conformations but there have been suggestions that it may be replaced by Mn^{2+} during integrin activation. Schumacher *et al.* assigned the MIDAS metal ion as Mn^{2+} in their 'open' crystal structure complex with fibronectin (7NWL) whereas Xia *et al.* and Nagae *et al.* assigned the ion as Mg^{2+}.[58, 59, 68] In agreement with the majority of the literature, the MIDAS ion has been referred to as Mg^{2+} in this work.

Xia *et al.* also found that significant conformational change was induced when the cyclic peptide was co-crystallised with $\alpha_5\beta_1$.[68] This was similar to the conformational change with the linear GRGDSP peptide with depleted Ca^{2+}, despite the 10-fold activity dependence on Ca^{2+}.

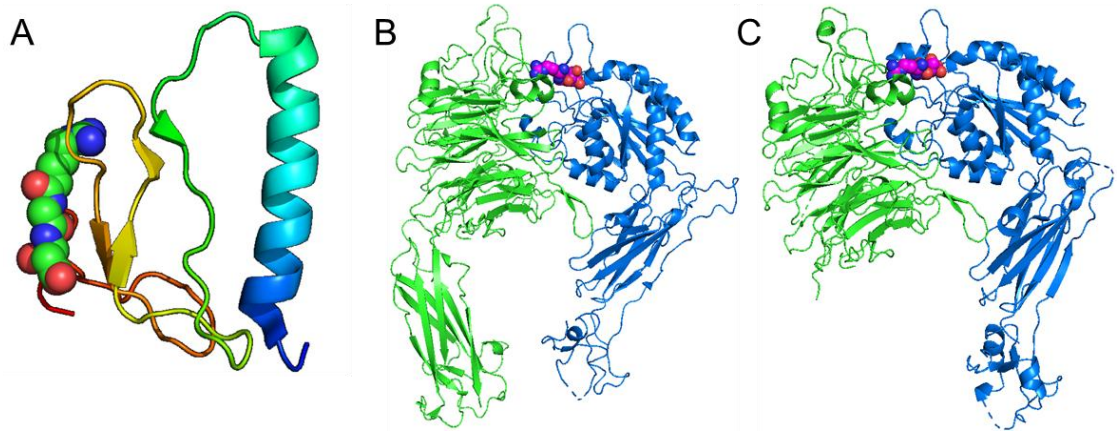


Figure 9: Crystal structures rendered in PyMol. Each RGD sequence is displayed with spherical atoms. The α_5 monomers are shown in green and β_1 monomers are shown in blue (B and C) A. C-terminal domain of IGFBP-1 (1ZT3) with the RGD sequence, which resides on a flexible loop. B. The Nagae crystal structure of the $\alpha_5\beta_1$ integrin (3VI4), with the RGD motif bound to the active site. C. The Xia crystal structure of the $\alpha_5\beta_1$ integrin (4WK0), with the RGD motif bound to the active site.

Structure determination of integrins can prove challenging due to the transmembrane nature of the proteins, however truncating the proteins to their extracellular regions has yielded crystal structures that can be used in ligand development. There are two published X-ray crystal structures of $\alpha_5\beta_1$ integrin with an RGD peptide bound (4WK0 and 3VI4), shown in Figure 9.[59, 68] The structures only contain the truncated, extracellular headpieces and both show a conformational shift of the protein in response to RGD site occupation. Nagae *et al.* could not achieve full occupancy by the RGD ligand, explaining that they believe the RGD bound structure is a mean of the ligand-bound and ligand-unbound forms.[59] Xia *et al.* achieved a complete occupancy of the ligand and resolved a higher resolution (1.78 Å vs 2.90 Å).[68] RGD binding is mediated by bidentate hydrogen bonds from the arginine guanidine group to GLN221 and ASP227, a hydrogen bond between the glycine carbonyl and SER227 and the carboxylate from the aspartate interacts with TYR133, ASN224 and the Mg^{2+} ion (Figure 10).

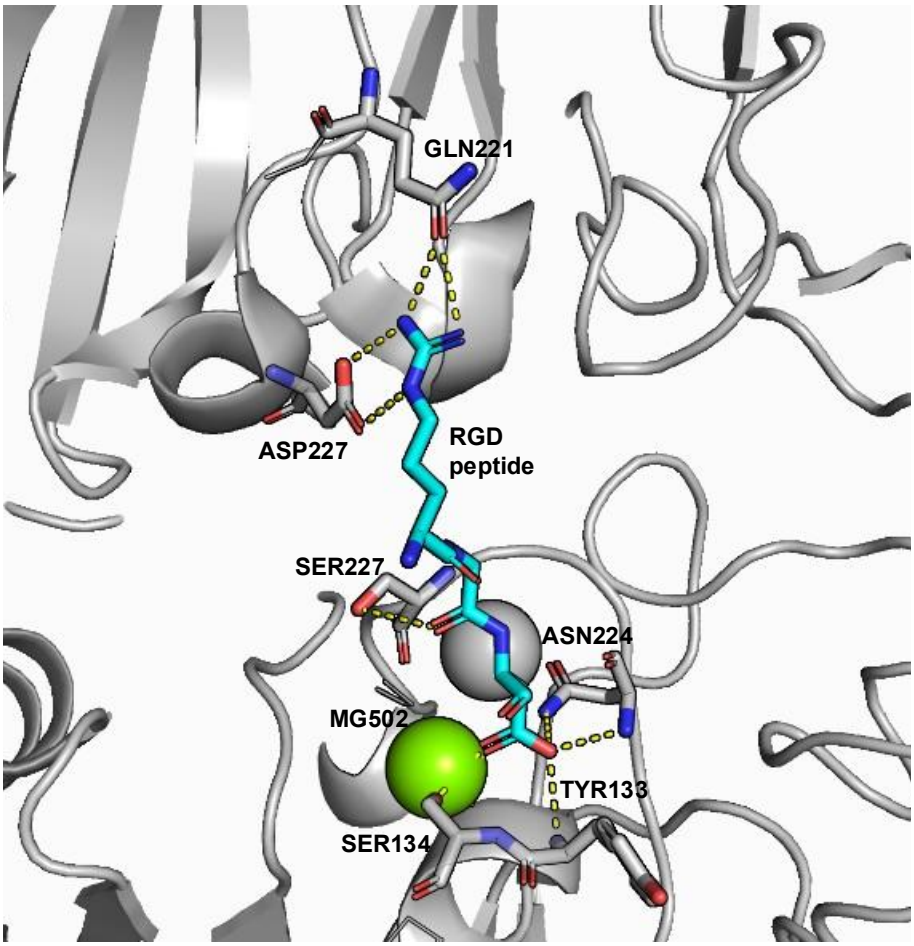
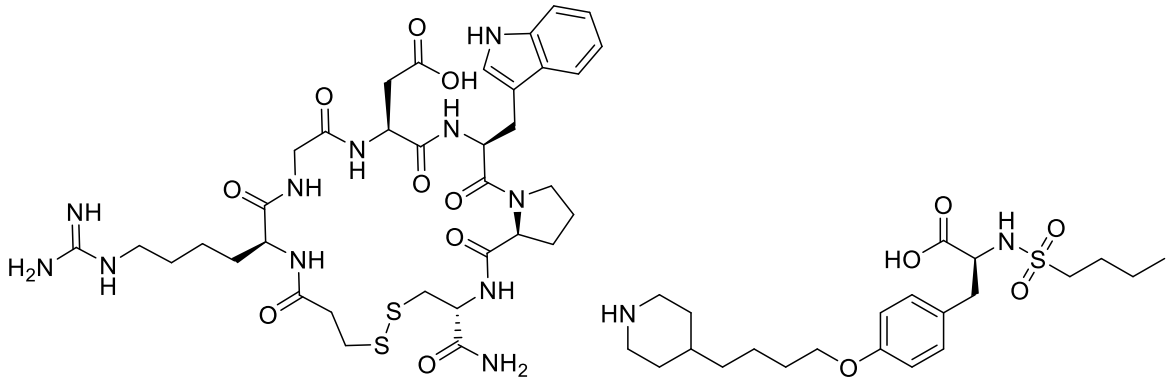


Figure 10: The Nagae crystal structure (3VI4) with the RGD motif bound to the active site. The key residues for RGD binding have been highlighted. The arginine (R) residue forms hydrogen bonds with GLN221 and ASP227. The glycine (G) residue interacts with SER227. The aspartate (D) residue interacts with TYR133, ASN224 and the Mg502 ion.

In addition to fibronectin and IGFBP-1, $\alpha_5\beta_1$ has a wide range of ligands including fibrinogen and fibrillin-1 (regulating cell adhesion and migration), VEGFRI (affecting angiogenesis), CD97 (mediating migration and angiogenesis), CD154 (inducing intracellular signalling), CD87 (inducing migration invasion and angiogenesis) amongst others.[69] Testament to the many functions of $\alpha_5\beta_1$, especially in cell angiogenesis, migration and invasion, aberrant upregulation of the integrin is found in several cancers such as glioblastoma, colon cancer, breast carcinoma, melanoma, node-negative non-small cell lung cancer and cervical cancer.[69] Patients with hyperexpression of $\alpha_5\beta_1$ in the setting of cancer exhibited poor survival.[69] Another ligand of $\alpha_5\beta_1$ integrin is endostatin, a fragment of collagen, which is an inhibitor of angiogenesis, the formation of blood vessels, and has been strongly indicated as a treatment for reducing tumour growth.[70]

3.4. Development of integrin ligands as therapeutic agents

Many therapeutics, including small molecules, biologics and antibodies, have been and are being developed to target integrins. Currently there are six therapeutics approved for clinical use that target integrins; three are small molecules (two of which are RGD-site binding, Figure 11), two are humanised antibodies and one is an antigen-binding fragment.[71] Currently there are no treatments approved which bind to $\alpha_5\beta_1$ integrin, although several have gone through trials to treat diabetic macular edema (DME), dry age-related macular degeneration, glioblastoma and solid tumours.[71] THR-687, targeting DME by inhibiting $\alpha_5\beta_1$ and α_V-containing integrins, entered a phase II clinical trial but development was discontinued due to lack of efficacy in the primary endpoint.[72] Risuteganib, also targeting DME by inhibiting $\alpha_5\beta_1$, $\alpha_V\beta_5$ and $\alpha_V\beta_3$, showed recovery of eyesight for 48% of participants in a phase II trial in 2021.[73] No further trials have been started as yet. GLPG0187, targeting solid tumours by inhibiting $\alpha_5\beta_1$ and α_V-containing integrins, showed no efficacy in a phase I trial.[74]



Eptifibatide
3.4.1

Tirofiban
3.4.2

Figure 11: Approved small molecule therapeutics that target the RGD site of integrins, eptifibatide and tirofiban. Both are used to treat acute coronary syndrome and thrombotic cardiovascular events.

As the RGD sequence is featured in many proteins targeting integrins with high selectivity and potency, there has been significant research into developing short peptides for integrin binding. Linear peptides are metabolised rapidly due to abundant proteases, so strategies including cyclisation and incorporation of D-amino acids have been utilised to generate highly active and metabolically stable ligands.[75] N-methylation has also been used to improve stability, bioavailability and selectivity and masking of charged residues can improve the bioavailability.[75]

Two strategies have been described for optimised $\alpha_5\beta_1$ selectivity for small molecules, illustrated in Figure 12. Methylation of the chain-linked nitrogen on a guanidino group

optimises selectivity for $\alpha_5\beta_1$ by blocking the side hydrogen bonds.[76] This is selective at the cost of binding affinity strength as $\alpha_5\beta_1$ forms both side and end interactions with the guanidine group, whereas $\alpha_v\beta_3$ exclusively forms side interactions. Integrin specificity can also be achieved using large groups, often aromatic, adjacent to the metal binding acid.[75]

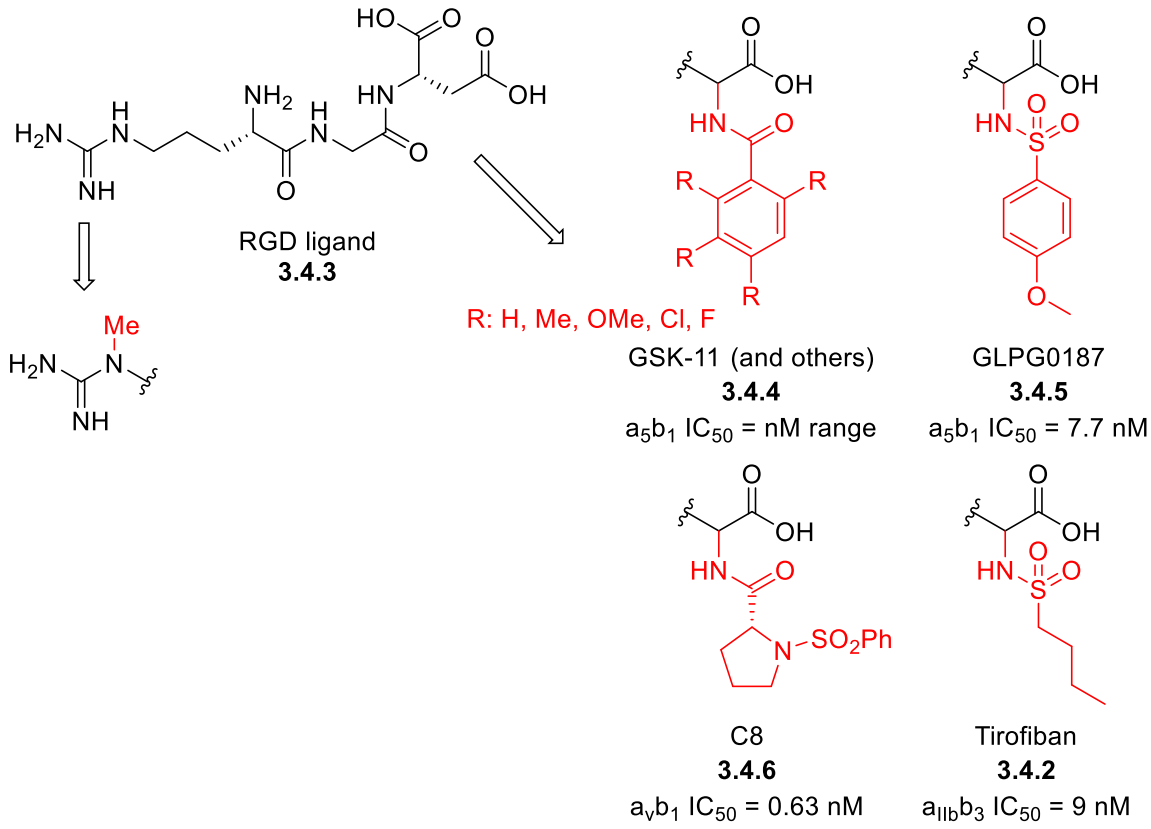


Figure 12: Examples of methods to improve integrin specificity in literature. Methylation of the chain linked nitrogen of guanidine significantly reduces binding potency to $\alpha_v\beta_3$ integrin, while slightly reducing binding potency to $\alpha_5\beta_1$ integrin. Large functional groups adjacent to the metal binding carboxylate can be manipulated for specificity to a range of integrins. IC_{50} values are determined by separate publications and assays and are not directly comparable.

C8 (**3.4.6**) which targets $\alpha_v\beta_1$, utilises a sulfonamidoproline group to produce a 0.63 nM $\alpha_v\beta_1$ inhibitor with all other RGD recognising integrins measured at >100 μM.[77] A number of examples have been shown in Figure 13, illustrating the variety of RGD mimetics in literature.[78-83]

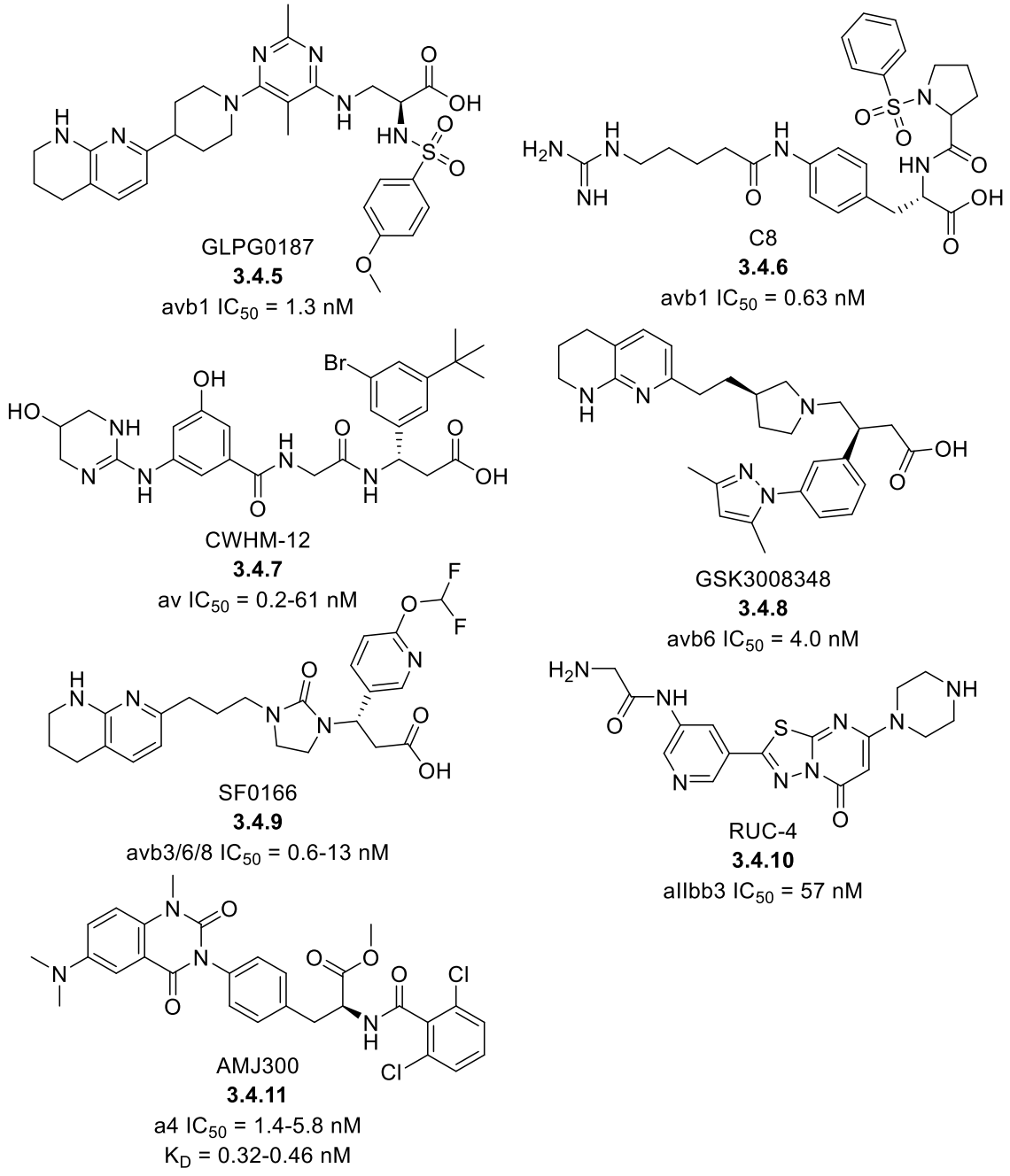


Figure 13: Examples of RGD mimetic integrin inhibitors in literature. Structural similarities between these compounds include a carboxylate group for metal binding, a bulky aromatic group adjacent to the carboxylate for integrin specificity and a terminal basic group mimicking the guanidine of the RGD endogenous ligand. IC_{50} values are determined by separate publications and assays and are not directly comparable.

Several publications and patents have described inhibitors targeting $\alpha_5\beta_1$ integrin, with some in clinical trials. All but one of these compounds target the RGD site and mimic the interactions of the endogenous ligand, with ATN-161 (**3.4.16**), being the exception. A selection of these structures is shown in Figure 14.

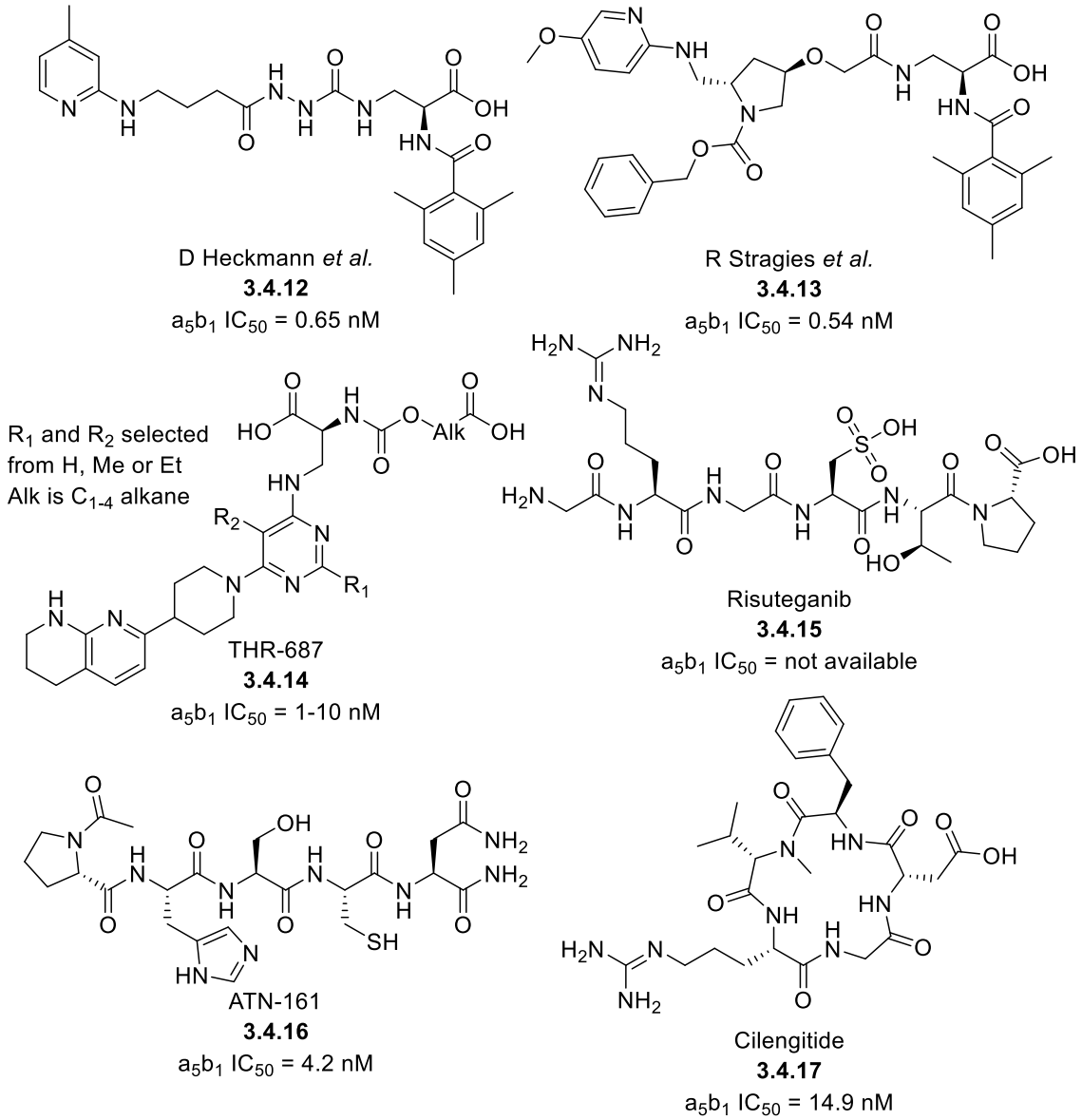


Figure 14: Structures of small-molecule antagonists of the $\alpha_5\beta_1$ integrin with the associated IC_{50} of each. There are close structural similarities to the RGD ligand in most of the compounds, including the entire RGD sequence in cilengitide (**3.4.17**). ATN-161 (**3.4.16**) does not share the similarities as it does not target the RGD binding site. IC_{50} values are determined by separate publications and assays and are not directly comparable.

The compounds targeting the RGD site contain a carboxylic acid to interact with the MIDAS ion and a guanidine or similar group with two nitrogen atoms held adjacent to one another to interact with the Asp-227 as shown by Slack *et al.*[71] Having the acidic group and a basic nitrogen group means the compounds are zwitterionic and that at physiological pH these groups will be charged. This presents significant hurdles for optimisation of pharmacokinetics and as such, clinically approved small molecules targeting RGD integrins (eptifibatide, **3.4.1**, and tirofiban, **3.4.2**) require intravenous (IV) delivery. ATN-161 (**3.4.16**) is not zwitterionic as it targets the synergistic site on $\alpha_5\beta_1$, although as it is a peptide it would still require IV delivery. In addition, ATN-161 shows a

non-linear dose response, with the lowest tested dose having the greatest inhibition in several assays, indicating complexities when targeting the synergistic site.[84]

Molecular design for non-zwitterionic compounds is attractive and can be achieved in a few ways. RUC-4 (**3.4.10**) is a small molecule targeting the RGD-site on $\alpha_{IIb}\beta_3$ integrin that does not contain an acid to interact with the MIDAS ion, instead displacing it and inhibiting the protein.[85] Another approach is to use less basic nitrogen groups to interact with Asp-227. This was exemplified by GSK when designing inhibitors for $\alpha_V\beta_1$, using a phenylurea group.[86] In three of the inhibitors shown in Figure 14, an aminopyridine can be seen in replacement of the guanidine. Heckmann *et al.* observed reduced activity against $\alpha_5\beta_1$ when using less basic groups and the greatest activity when using a cyclic guanidine, meaning this approach may have limited use.[87] The aminopyridine type group appears to be good compromise between these factors, with a carboxylic acid present in all RGD mimics.

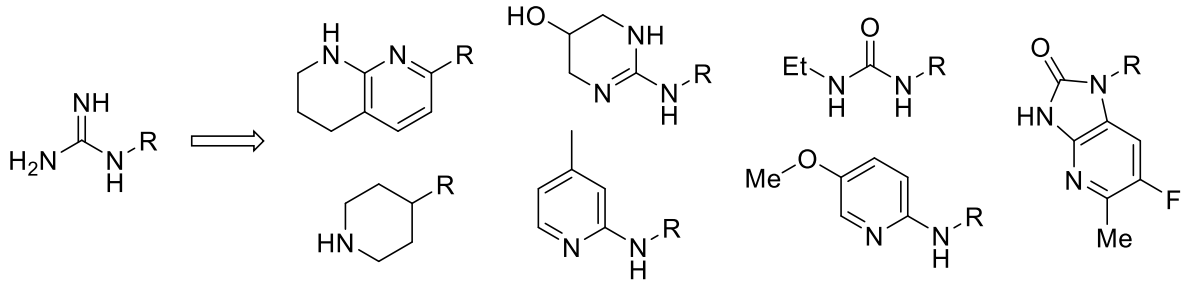


Figure 15: Functional groups used to mimic the terminal guanidine of the RGD ligand in literature. Heckmann *et al.* found that functional groups with a lower basicity reduced the binding affinity with the integrin.[87]

Heckmann *et al*. produced the only known investigation into replacement of the carboxylic acid, using a hydroxamic acid group to adapt $\alpha_5\beta_1$ and $\alpha_V\beta_3$ inhibitors.[88] In most cases activity was significantly reduced for $\alpha_5\beta_1$, although $\alpha_V\beta_3$ activity was increased in three out of six examples.

AJM300 (**3.4.11**) is a pro-drug, small molecule antagonist that targets the α_4 monomer, approved for use in Japan for treatment of ulcerative colitis.[89] The structure contains an ester which is hydrolysed in the active form. Although the drug targets the α_4 monomer and is not expected to interact at the RGD binding site, it does highlight a potential strategy by targeting an integrin subunit specifically.

Cilengitide (**3.4.17**), an integrin antagonist drug candidate, is known to have some agonistic activity at low nanomolar concentrations, stimulating tumour growth in glioblastoma models by shifting the integrin to the extended-open state without total inhibition.[66, 90, 91] The agonism of integrins has had much less research than the

antagonism, although significant compounds have been produced, with some showing specificity for $\alpha_5\beta_1$.

An RGD mimetic strategy was used by Rechenmacher *et al.* in the design of $\alpha_5\beta_1$ specific antagonists.[92, 93] The compounds they produced, shown in Figure 16, are similar to the inhibitors shown in Figure 13 and Figure 14, sharing a guanidine and a substituted aryl group adjacent to a carboxylic acid.[94] In another publication, Bauila *et al.* produced a series of agonistic compounds containing a carboxylic acid with a novel scaffold of a β-lactam. With the paper focusing on cancer treatment and cell-adhesion, it is unclear what other agonistic effects could be occurring on the integrin. As has been shown in the case of cilengitide, the requirements for inhibition and activation are unclear, with several of the compounds in the paper have mixed agonist and antagonist activities for different integrins. Baiula *et al.* also found that their agonistic compounds (**3.4.19-21**) had an concentration-dependent phosphorylation effect on the extracellular signal-regulated kinase 1 and 2 (ERK1/2) mediated by the $\alpha_5\beta_1$ integrin, with up to 4.5x increased phosphorylation compared to the control.[95] This is a highly interesting result with regard to the research presented above on agonists for treatment of diabetes.

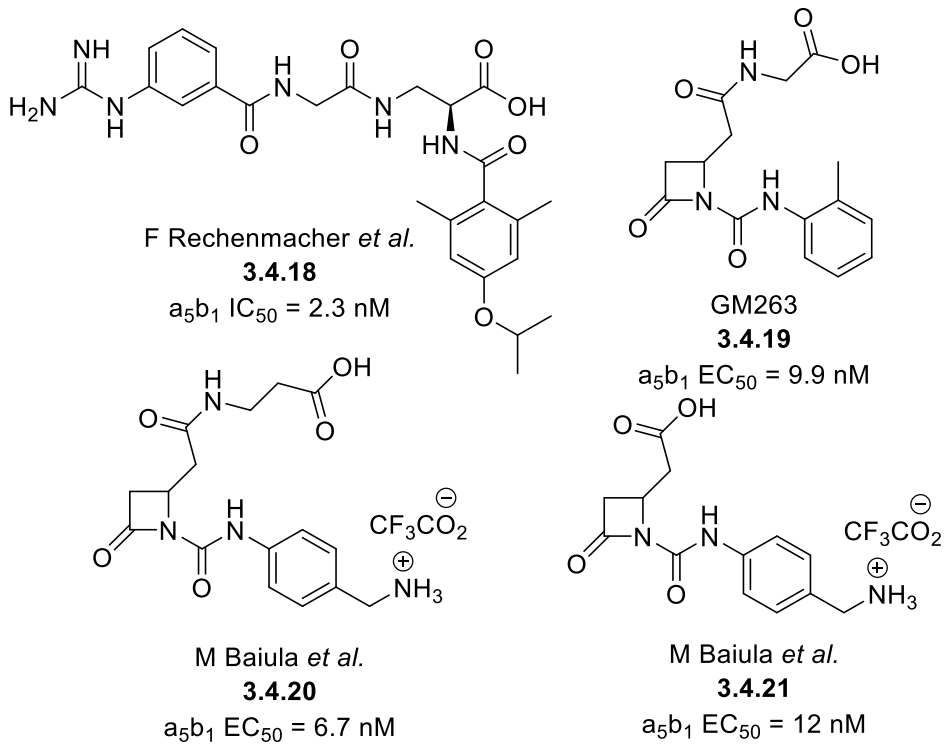


Figure 16: RGD site integrin ligands. The compounds with the highest activity for $\alpha_5\beta_1$ from each publication are displayed, along with the IC_{50}/EC_{50} values. A sample of **GM263** was shared by Prof. Daria Giacomini as a positive control. Stereochemistry of the lactams is not defined.

3.5. Computer aided drug design

Computer aided drug design (CADD) has become ubiquitous in drug development, with many academic and industrial groups using software to direct target design.[96] The first synthetic planning program was described in 1971 by J. B. Hendrickson and the first molecular dynamic calculation of a protein in 1977 by M. Karplus *et al.*, while 3D modelling for drug design was described from 1975 with a variety of systems.[97-99] These discoveries were made possible by advances in structural biology, where advances in solving protein structures unveiled the binding sites of drug molecules, and increasing computational power, which allowed for unprecedented ability to perform complex tasks.[100]

Databases of chemicals can be compared to a structure to find energetic and conformational analogues, known as shape screening, one of the first CADD methods. Properties of these chemicals can be used to predict the properties of novel molecules, directing synthesis for improved drug-like behaviour. These methods are computationally light, limited by predetermined data rather than computational capacity.

More complex is analysing ligands flexibly into a rigid protein environment, typical of docking. Accounting for possible ionisation states, ligand conformations and location in a protein binding site is challenging and requires significant computational resource to complete beyond a small scale. Molecular dynamics, where proteins and ligands can be treated flexibly, are the most resource intensive simulations typically used in CADD and is typically used in small scale on specific model systems.

The ability of CADD to speed up drug programs was demonstrated in the development of AIDS treatments. Crystal structures of HIV PR (key viral protease) in 1989 allowed rapid development of the inhibitors saquinavir (FDA approved in 1995) and ritonavir (FDA approved in 1996), amongst others.[101, 102] Saquinavir and ritonavir were based on mimicking tripeptide sequences using CADD analysis to identify key modalities to form interactions with the protein.[103, 104]

Tirofiban, **3.2.4**, was another of the first drugs to be developed with CADD, using a shape screen of Merck compounds to find analogues of the RGD tripeptide.[105] A lead compound with an IC_{50} value of 27 μM was identified which was improved to 0.011 μM for the final compound, targeting $\alpha_{IIb}\beta_3$ integrin.

Advances in structural biology including X-ray crystallography, NMR and homology modelling have provided accurate protein structures, meaning ligand affinity can be more

reliably modelled *in silico*, enabling researchers to take greater direction from computational modelling.

Similarly, improvements in computational power have allowed screening of increasingly large virtual libraries of compounds. Virtual screening is a fast, cost-effective and widespread technique to generate initial hits for projects with well validated biological targets. Frequently used chemical databases include Enamine (9.6 billion compounds), ZINC (230 million compounds), ChEMBL (2.4 million compounds), eMolecules (20 million compounds) and ChemBridge (300 million compounds).[106-108] This approach requires careful handling of large volumes of data and coordination of computational resource but has proven effective for the development of potent, bioactive compounds.

Using a known ligand, a shape comparison screen can identify compounds with a similar volume and electronics to identify new hit matter.[109] The Rapid Overlay of Chemical Structures (ROCS) program, developed by OpenEye Scientific, is an established method for shape screening. It operates by taking the ligand in a known binding conformation and generating hard-sphere representations of the heavy atoms, which contribute to the overall molecular volume.[110] The same is done for possible conformations of a chemical database and the volume overlap measured using a Gaussian description with approximations for computational efficiency.[111] A colour feature is used to quantify the chemical similarities between the ligand and the test compound. A key feature of this software is that it exclusively compares compounds to the known ligand and not the protein itself. Molecular docking of the highest scoring hits is often used to validate the outputs.[109]

Molecular docking is the typical method of estimating the binding of molecules to a protein. Crystal structures typically provide the protein environment which are minimised as part of the generation of the docking grid. The molecules are prepared to account for the possible ionisation states and conformations they can occupy. Finally, the molecules are fitted into the allocated grid and interactions with the protein are estimated. The Schrodinger package (including Maestro, Epik, and Glide) is an example of software for this type of modelling. The details of these stages using the Schrodinger package is described in detail (Figure 17).

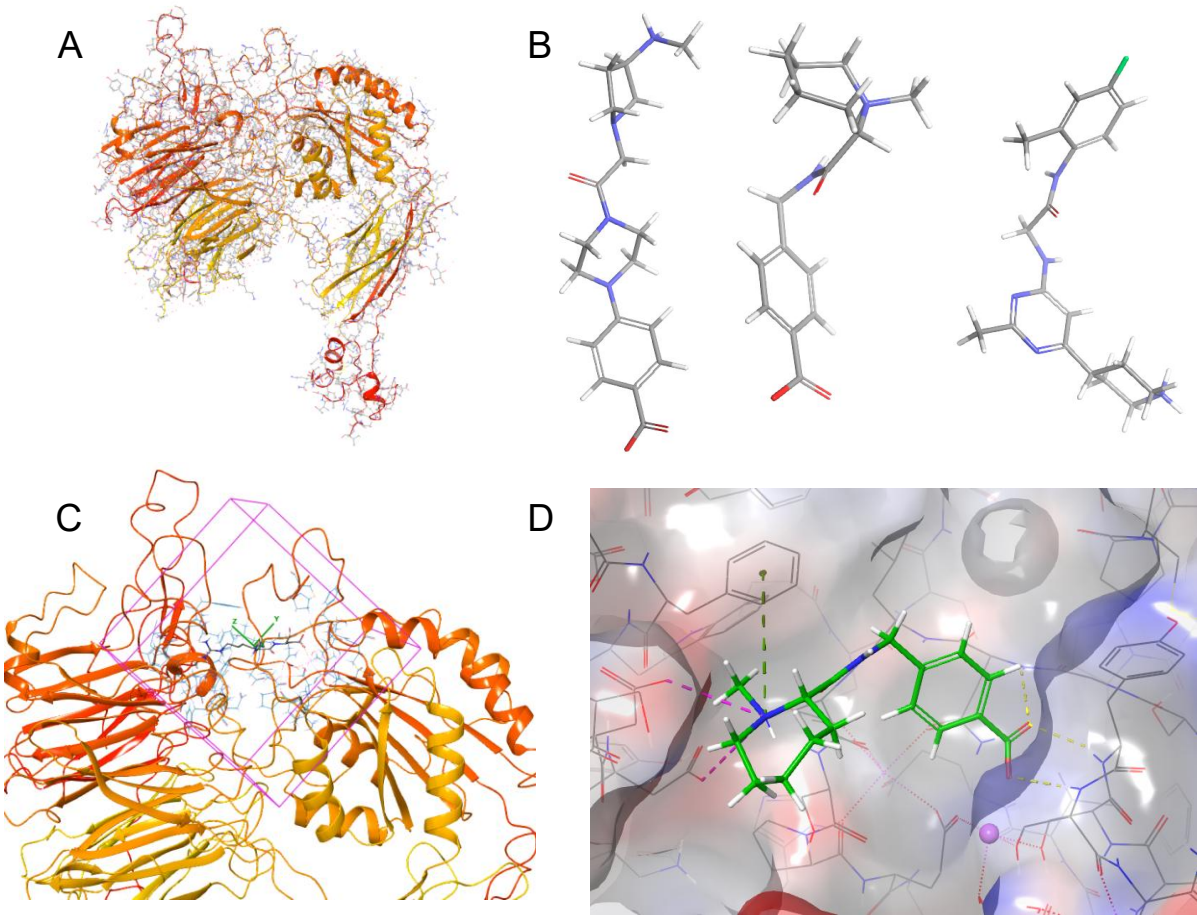


Figure 17: An example of the stages of a molecular docking workflow using Schrodinger Maestro and Glide. (A) The protein is prepared for docking by adding hydrogen atoms, resolving clashes and optimising conformations, amongst other functions. (B) Small molecules are prepared, accounting for possible charges, chirality and conformations. (C) A grid is generated in which the ligands will be docked. Grid settings can be finely adjusted, such as requiring or preventing interactions with certain atoms or residues. (D) Ligands are docked into the grid and energetic outputs can be assessed.

Protein structures represent a structural snapshot of a protein. For molecular docking, the best protein structures are experimentally generated, high-resolution, include a known binding ligand at the active site and can be compared to other published structures. Protein structure files lack information about the protein environment including bond orders, charges and hydrogen atoms which must be added during protein preparation. Conflicts are common and can be ignored depending on the distance to the active site. The minimisation of the generated structure represents a computational model of a protein snapshot. A receptor grid is generated around the active site, limiting the docking of molecules to save computational resource and direct results. Grid generation can include constraints to further control outputs, including excluded volumes and required binding modes. Changing these settings can wildly vary the molecular docking outputs and systematic variation of each variable will indicate which settings are most appropriate for the project.

Preparation of ligands involves creating the likely binding states for each compound, including any ionisation states, tautomers, stereoisomers and minimised conformations. As with protein preparation, the settings of each of these stages will strongly influence the outputs. For example, a narrow pH range centred on physiological pH is the standard setting for ionisation state generation, but during ligand binding a local protein environment will affect the possible ligand states, such as metal coordination stabilising deprotonation of mildly acidic protons. This is one of many settings available in the LigPrep program.

The computational principle of molecular docking varies across the many programs available, with scoring functions categorised into three groups: force-field based, empirical and knowledge-based. As Glide employs an empirical function this is explained in detail. Using the generated receptor grid, a progressively more accurate set of fields are generated that represent the shape and properties of the receptor. The prepared ligands are then screened across the entire surface to locate the promising binding sites. The ligands are then minimised in the receptor field using a molecular mechanics energy function (OPLS-AA force field), in conjunction with a distance-dependent dielectric model. Monte Carlo calculations are then used on the lowest energy poses to further optimise ligand binding conformation. Scoring is estimated by measuring various interactions and effects for the complex, including hydrogen bonding, ionic and hydrophobic interactions. These factors are weighted using a set of ligand-protein complexes with established binding affinities to output a binding affinity.[108, 112]

3.6. Preliminary work

The project follows the results found by Haywood *et al.* and the related work that validated $\alpha_5\beta_1$ integrin as a target for agonistic therapeutics.[113] Previous work by Dr Katie Simmons used the virtual screening tool ROCS to identify ligands in the eMolecules screening library that match the shape and electrostatics of the RGD motif of IGFBP-1.[111] Molecules that had matching electrostatics and conformations to the endogenous ligand were then filtered for structural diversity to yield 30 compounds, a selection of which are shown in Figure 18 (full set shown in appendix 10.1).

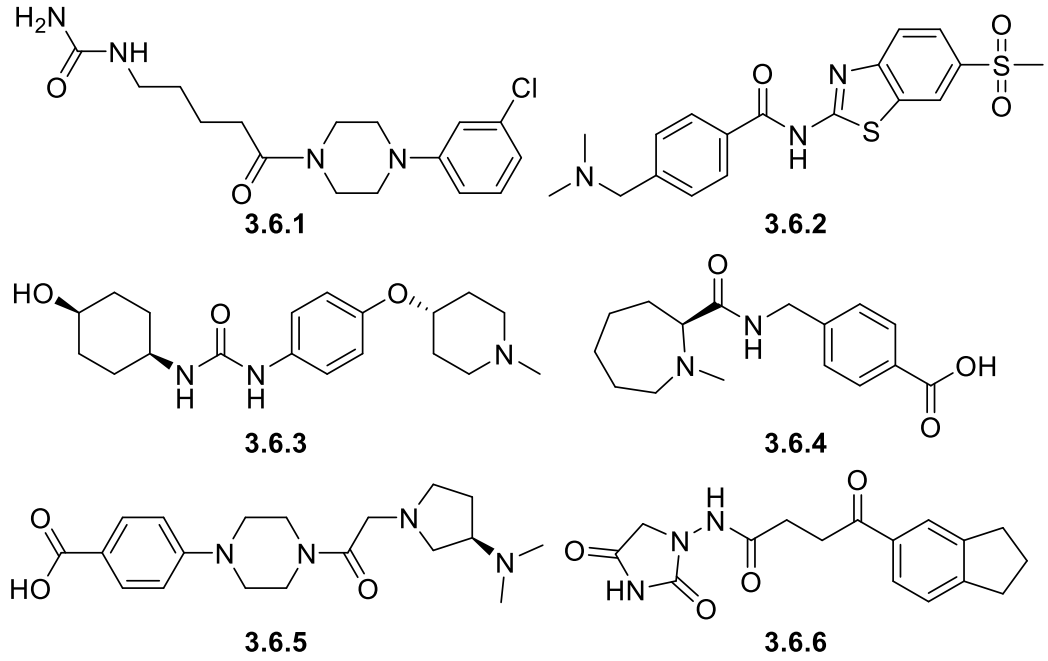


Figure 18: A selection of compounds identified using the virtual screening platform ROCS to match the shape and electrostatic properties of the RGD sequence in IGFBP-1.

The compounds were purchased and tested in the assays established by Haywood *et al.*, by Dr Lia De Faveri. Compounds were tested for their ability to increase insulin-stimulated phosphorylation of Akt at a range of concentrations (western blotting measured pAkt and total Akt in C2C12 mouse myotube cells, 100, 10 and 1 μM in triplicate, 24 hour treatment). The result from the 1 μM assay are shown in Figure 19.

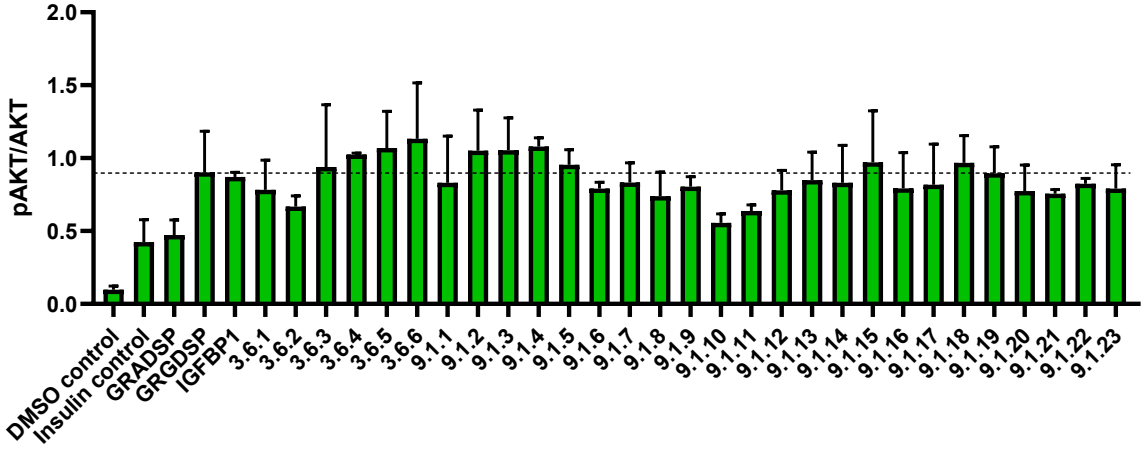


Figure 19: Activity profile of the initial set of compounds, measured by the insulin-stimulated phosphorylation of Akt with 1 μM of compound (24 hours treatment) compared to the GRGDSP and GRADSP hexapeptides, IGFBP-1 and negative controls. n=3; Akt: protein kinase B.

Compound activities were compared to the linear RGD containing hexapeptide-GRGDSP, which Haywood *et. al* showed to have activity comparable to full-length IGFBP-1 protein. When dosed at 100 and 10 μM, all the compounds have activity comparable to the GRGDSP peptide, However, when tested at a lower concentration of 1 μM, compounds **3.6.3-3.6.6**, **9.1.2-9.1.5**, **9.1.15** and **9.1.18** appear to have the ability to enhance phosphorylation of Akt in a manner equivalent to, or better than the RGD-containing peptide.

These results were further validated using a glucose uptake assay. The Glucose-Glo™ Assay (Promega, UK) is a bioluminescent assay which measures glucose in cell lysates to detect altered glucose consumption when cells are treated with the compounds of interest (100 μM, 24 hours incubation). Compounds which enhanced pAkt also showed enhanced glucose uptake. A live/dead cell viability assay (Sigma Aldrich, UK) to measure compound toxicity in the C2C12 skeletal muscles cell line used for all assays showed a lack of toxicity for all but one of the purchased RGD mimics when tested at the maximum concentration of 100 μM. The compounds with promising activity in cellular assays were assessed for their aqueous solubility at pH 7.4 by Dr De Faveri using HPLC, with eight compounds flagged for low solubility.[114]

The number of compounds with promising activity is surprising, especially given the diversity of structures. This may indicate that the assay is not suitably discriminating between binding and non-binding compounds, however, as discussed in section 5.1, there are structural similarities of many compounds to literature binders.

In addition to the assays described above, an external contract research organisation, Cyprotex, conducted assays to measure compound metabolic stability in human and murine liver microsomal fractions. These results are summarised in Table 2. Most of the compounds tested in these assays had very good stability. Although two compounds performed poorly in mouse microsomes, they had much higher values in human microsomes. This data was used to inform compound development and establishes the assay protocols for new compounds.

Table 2: Human and mouse microsomal half-life of select compounds from the initial screening. *Negative values indicate no measurable decrease in concentration.

No.	**3.6.1**	**3.6.2**	**3.6.3**	**3.6.5**	**3.6.6**	**9.1.8**	**9.1.9**	**9.1.13**	**9.1.19**
$T_{1/2}$ (human) (min)	509	144	>300	648	426	186	170	169	165
$T_{1/2}$ (mouse) (min)	180	14	85.8	>300	114	308	29	708	47.5

4. Aims and objectives

The aim of this project is hit validation and optimisation of the compounds identified by virtual ligand-based screening to generate improved agonists of the $\alpha_5\beta_1$ integrin which could provide improved treatment of cardiovascular disease in type 2 diabetes patients. The project will follow iterative cycles of computer aided molecular design, chemical synthesis and biological evaluation, with the individual objectives described below.

Target design:

- Develop understanding of crystal structures and ligand binding modes to predict activity and inform ligand design.

Six crystal structures containing the RGD-binding site of the $\alpha_5\beta_1$ integrin binding site have been selected for analysis, including co-crystalised structures with the RGD tripeptide and RGD-containing cyclic peptides (PDB codes: 3VI3, 3VI4, 4WJK, 4WK0, 4WK2, 4WK4).[59, 68] These will be used for molecular docking and design of ligands, primarily in Schrodinger Maestro and Glide. Biological evaluation of the designed ligands will inform the accuracy of the molecular docking.

Chemical synthesis:

- Design and optimise synthetic routes for generation of hit analogues and derivatise active compounds.

Top scoring targets from molecular docking will be prioritised for synthesis, with analogues made from common intermediates where possible. Synthesised compounds will inform the design hypotheses with a range of functionalities and properties to determine the ligand binding modes.

Biological evaluation:

- Run and analyse results from established biological assays to inform target design and develop understanding of agonist and antagonist compound activity and integrin specificity.

Orthogonal and high throughput assays will be prioritised to reliably inform compound activity, including cell-based and proteogenic systems.

5. Design, synthesis and biological evaluation of the zwitterion series

A zwitterion series was developed from compound **3.6.5**, identified in the ROCS screen which shared structural similarities with the RGD motif and integrin targeting compounds in literature. Further designs were based on compound **3.6.4** and a pyrazole motif identified from virtual screening by Emily Rolfe.

5.1 Zwitterions targeting integrins in literature

A wide variety of zwitterionic compounds have been developed targeting integrins including $\alpha_5\beta_1$, as discussed above. Matching the key carboxylic acid on the endogenous RGD ligand, while carefully adapting the guanidine and varying the linking region appears to be a key design principle for most literature examples of integrin targeting, examples of which are shown in Figure 20.

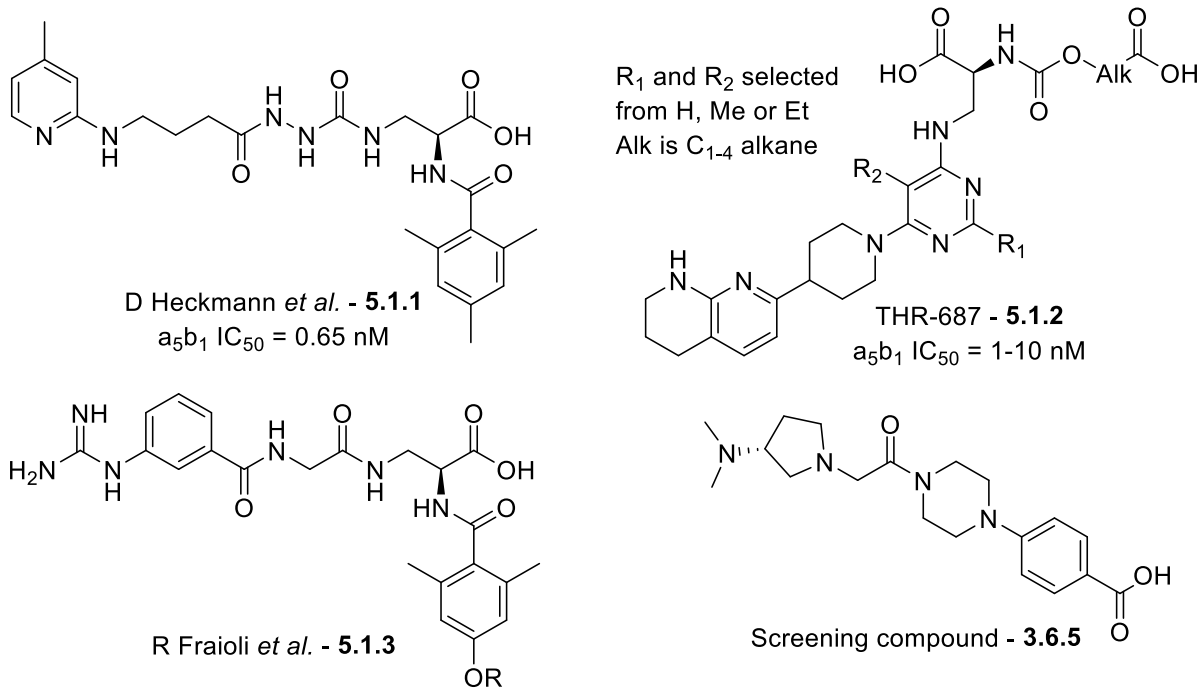


Figure 20: Examples of integrin targeting compounds in literature containing a carboxylic acid and guanidine mimic group. Screening compound **3.6.5** has structural similarities to this trend in literature.[87, 92]

As one would expect, the ROCS shape screen which generated the initial hits for the project identified several compounds with carboxylic acid groups which showed activity in the initial assays. Due to the similarity with the compounds in literature, these were prioritised for development.

5.2 Docking studies and initial zwitterion design

The entire initial set of compounds identified via ligand-based screening were docked using Schrodinger Glide into the available crystal structures of $\alpha_5\beta_1$ integrin to indicate potential expansion points or variation sites. A selection of the best scoring compounds is shown in Figure 21. The compounds including carboxylic acids typically showed higher scores and binding poses that matched the endogenous ligand closer to most of the other compounds.

The Nagae *et al*. crystal structure (3VI4) was used for this process, though later docking was completed predominantly on the Xia *et al*. structure (4WK0) which the authors believe to be in an intermediate of the open and closed integrin conformations, compared to the Nagae *et al.* structure which is described to be in the closed conformation.[87, 92] Using the higher resolution and half-open conformation structure (4WK0) is theorised to yield more accurate docking results. None of the poses directly matched the endogenous ligand, which interacts with ASP227 and GLN221 but later runs using improved grid generation and docking settings showed closer similarity.

Glide reports the affinity of the docked structure and protein with a GlideScore, also referred to as ‘gscore’. These values have been reported for the compounds shown in the docking figures.

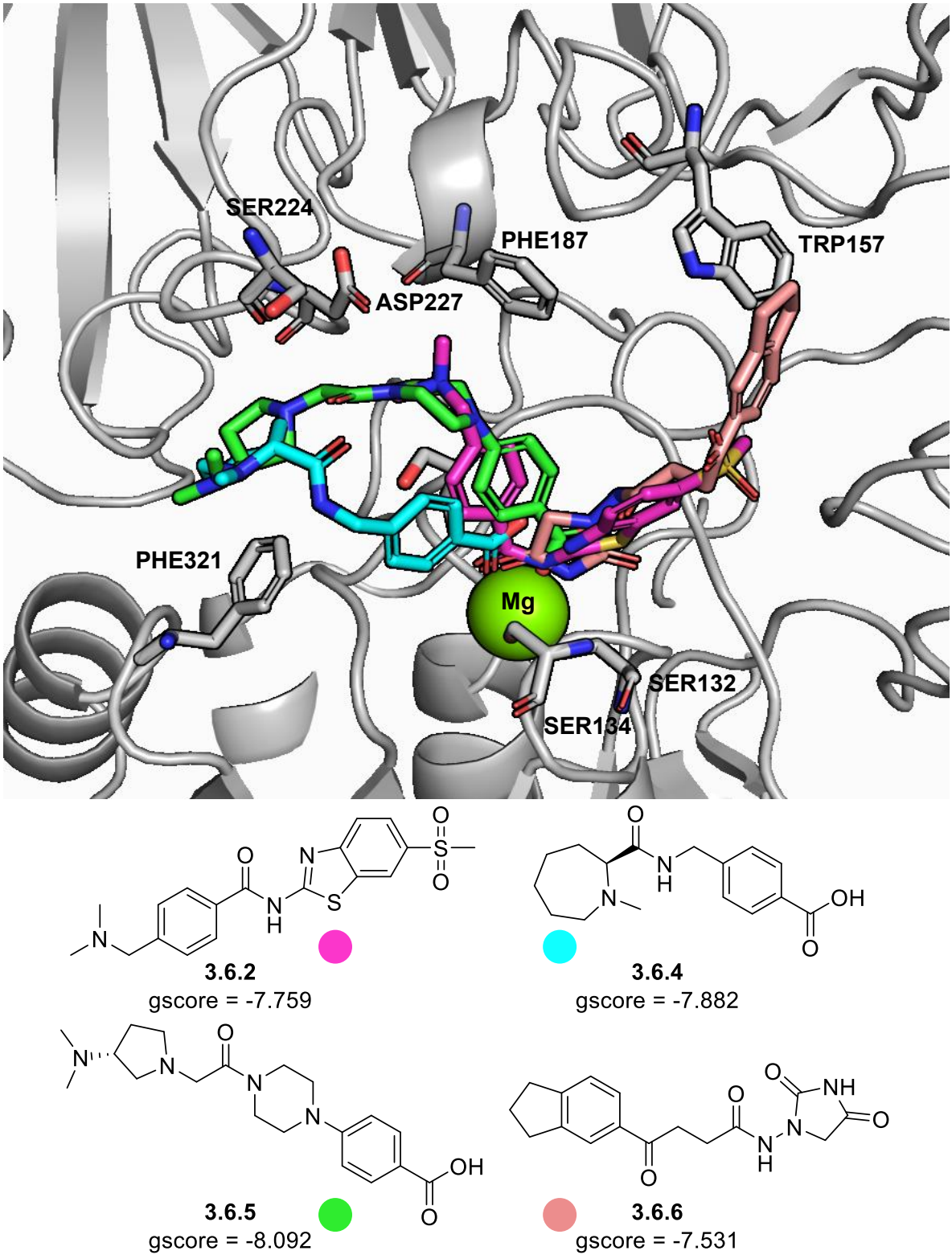


Figure 21: Molecular docking of four screening compounds to the RGD-binding domajn of $\alpha_5\beta_1$ integrin (PDB: 3VI4) with close residues highlighted to design modifications for SAR development.

Compound **3.6.5** showed some of the best activity from the initial assay screen and the top docking score. It was hypothesised that the substituted pyrrolidine group could be

replaced with a basic amino acid residue. Modelling of these changes is shown in Figure 22.

Compound **5.2.1**, where the arginine matches a portion of the endogenous ligand is predicted to bind in an identical pose to the crystal structures of the RGD peptide and has a high gscore as a result.

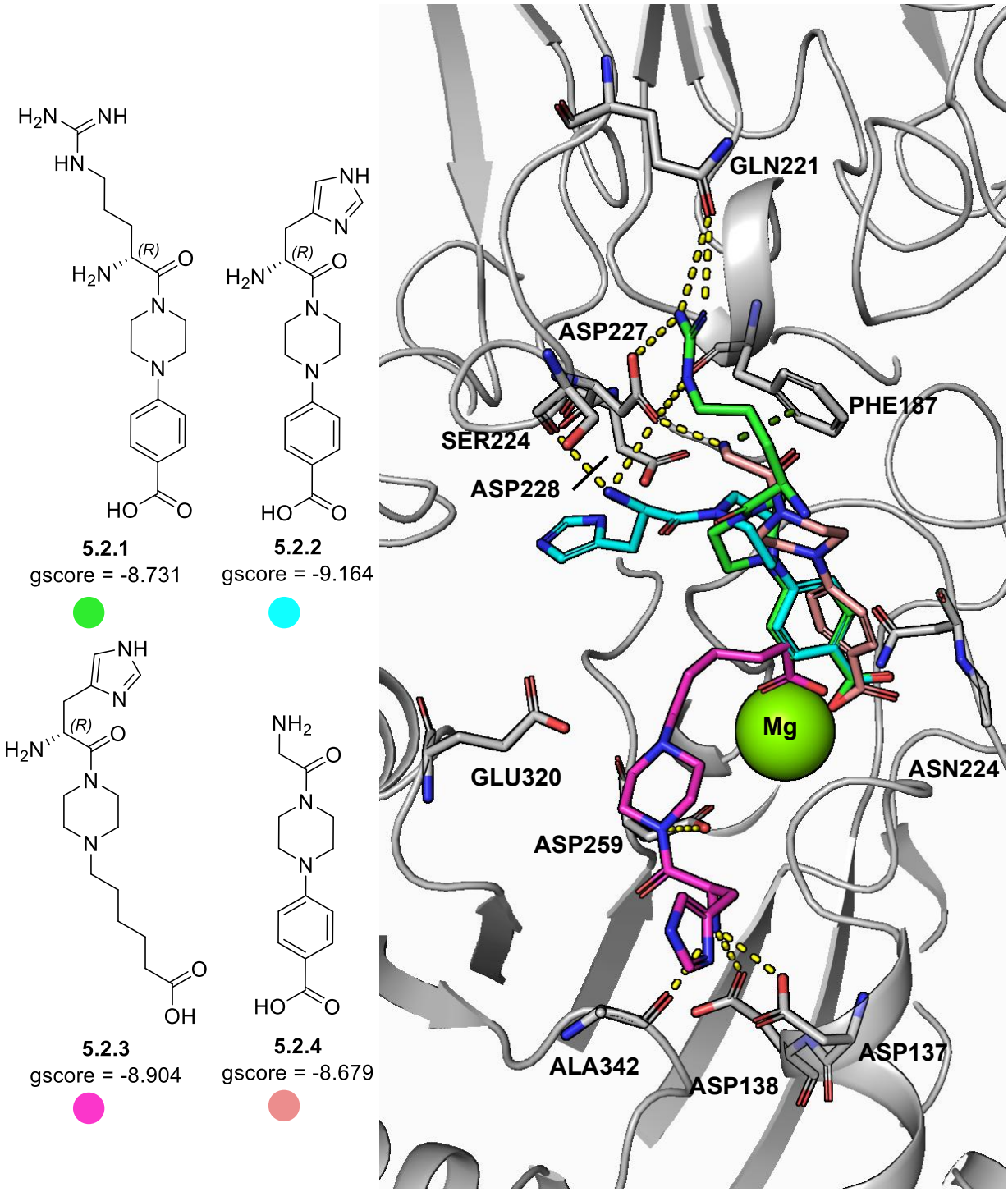


Figure 22: Molecular docking of the top scoring zwitterion alterations to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 3VI4) with key residues and hydrogen bonds highlighted. All the

compounds except **5.2.3** form interactions at the α_5 side of the pocket, mimicking the endogenous ligand. Compound **5.2.3** was instead designed to reach into a highly electronegative pocket on the β_1 monomer.

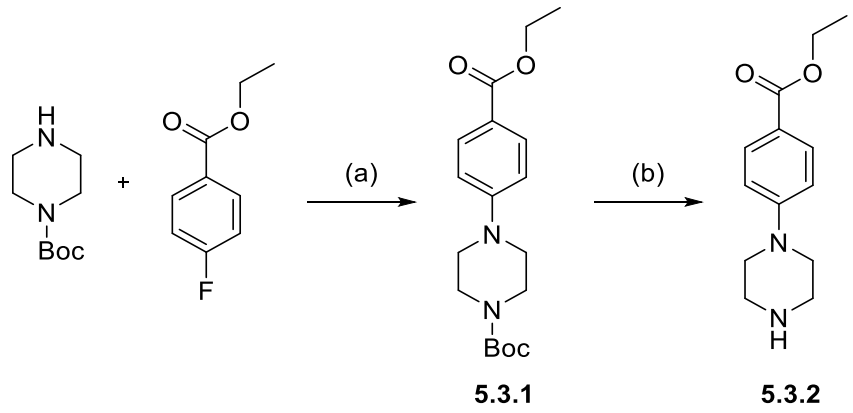
Interestingly, compound **5.2.2**, with a histidine terminus, has a higher gscore despite binding in an alternate conformation by forming interactions with ASP228 and SER224. It is notable that the unnatural R isomers of the amino acids are predicted to bind with higher affinity than the naturally occurring S isomer. Compound **5.2.4**, with a glycine terminus, also scores well but the chain length may be too short to fully form interactions with the residues at the top of the pocket.

Compound **5.2.3**, with a histidine terminus and an alkyl chain in place of the phenyl linker, was designed to reach a highly electronegative pocket on the β_1 monomer, composed of ASP259, ALA342, ASP138 and ASP137. The design received a high docking score, indicating it would be beneficial to explore binding at the β_1 pocket.

Other designs that were not selected for synthesis include halogenation and methylation of the phenyl, replacing the piperazine with a phenyl group, variation of the amine terminus and a variety of carboxylic acid isosteres, including hydroxamic acids, esters and catechols. Most of structures had no significant change in docking scores compared to the hit compound, had significant synthetic challenges to produce, or had concerns of cross activity, such as with the zinc binding of hydroxamic acids.

5.3 Synthesis of initial zwitterion targets

Synthesis of the common intermediate **5.3.2** is shown in Scheme 1 and began with nucleophilic aromatic substitution (S_NAr) of Boc-piperazine and ethyl 4-fluorobenzoate at 130°C in DMSO to obtain compound **5.3.1** with a 20% yield. This was subsequently deprotected using 4N HCl in dioxane to obtain intermediate **5.3.2** in quantitative yield.



5.3.1 **5.3.2**

Scheme 1: Synthesis of intermediates **5.3.1** and **5.3.2**. Reagents and conditions: (a) K_2CO_3, DMSO, 130°C, 24 h, 20%; (b) 4N HCl in dioxane, rt, 2 h, 99%.

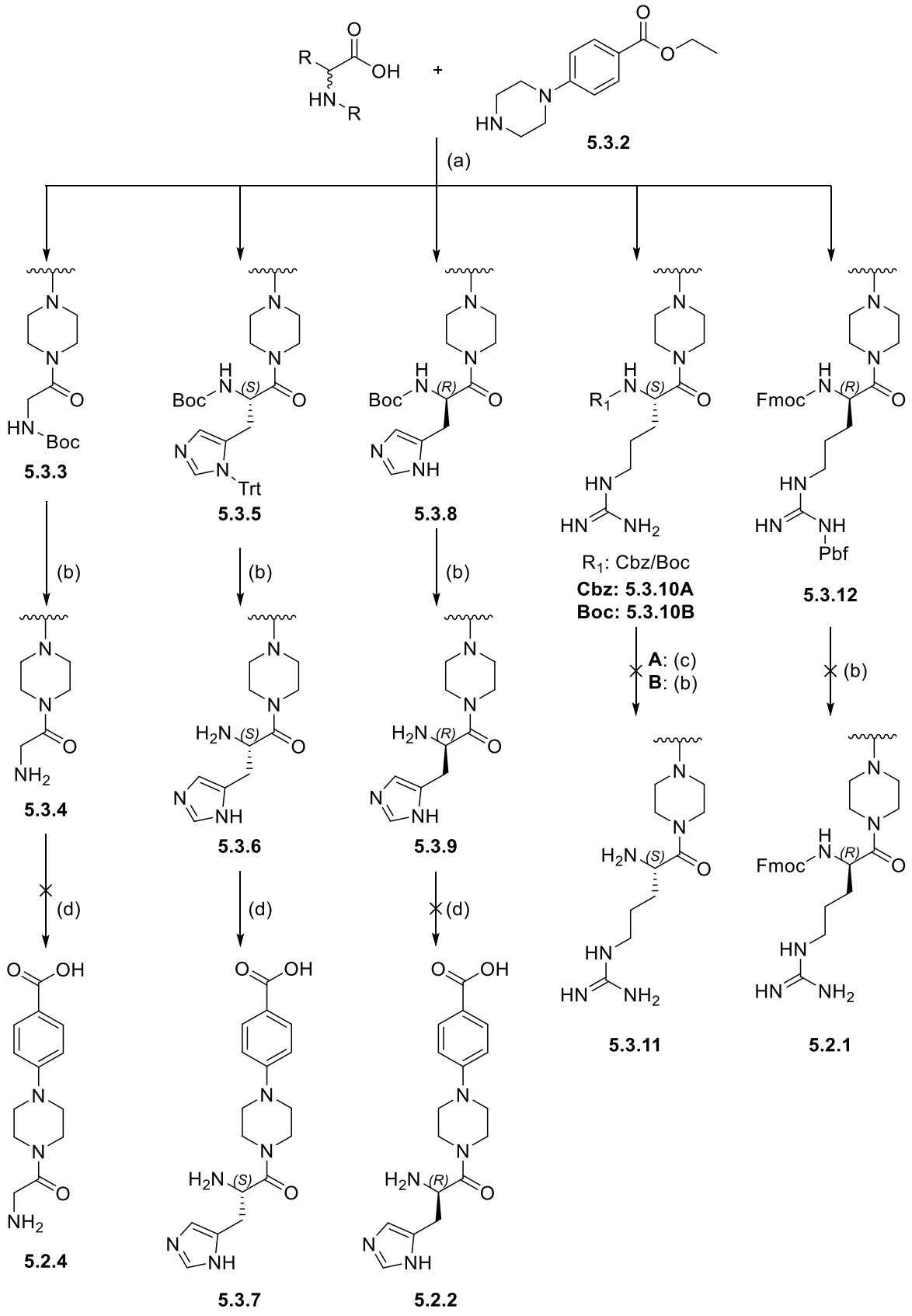
Amide coupling of intermediate **5.3.2** using nitrogen protected glycine, histidine and arginine (Boc-glycine, Boc-D-histidine, Boc-L-(Trt)histidine, Fmoc-D-(Pbf)arginine, Cbz-L-arginine and Boc-D-arginine), using EDC and DMAP as coupling agents in DCM gave yields ranging from 33-87%, as shown in Scheme 2.

The glycine derivatised product **5.3.3** was deprotected using 4N HCl in dioxane to obtain compound **5.3.4** in a 54% yield. Although NMR analysis indicated purification had been successful, HPLC analysis indicated significant, unknown impurities remained in the sample and the de-esterification was not attempted. It is possible that the compound is not stable when dissolved in MeOH with addition of formic acid for HPLC as the NMR is likely to have traces of impurities that are UV active, however this remains unclear.

The L-histidine derivatised product **5.3.5** was reacted initially in 4N HCl in dioxane to remove the Boc group and subsequently in 1% TFA in DCM when the Trt group was not initially removed, obtaining compound **5.3.6** in a 65% yield. The methyl ester was hydrolysed in 35% aqueous ammonium hydroxide and methanol over 10 days to fully complete. Purification using Reverse-phase chromatography yielded no clean elution, but the loading DMSO eluted a small amount of product **5.3.7** which was collected with a 91% purity and a 7% yield.

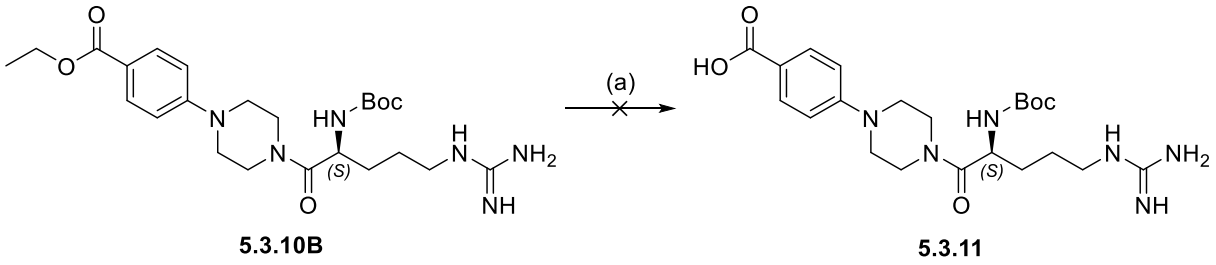
Removal of the N-Boc group of the D-histidine derivative **5.3.8** was successfully achieved using 4N HCl in dioxane to obtain compound **5.3.9** in quantitative yield. The de-esterification was completed using 35% aqueous ammonium hydroxide and methanol. Reverse-phase column chromatography of the product was attempted to remove remaining impurities but was not successful after several attempts. When using DMSO to dissolve the compound and load the column, as with compound **5.3.7**, there was detectable product eluting with the DMSO, but the collected **5.2.2** was too impure for assay submission. No further product peak eluted during the normal gradient, or during elution with the sample dry-loaded on silica.

The synthesis of the L-arginine derivative was attempted with both Cbz and Boc protecting groups (compounds **5.3.10A** and **B**). Removal of the Cbz protecting group to produce compound **5.3.10A** using H_2 and Pd/C was attempted and appeared to proceed to completion when monitored using LCMS. However, purification using reverse-phase column chromatography failed to elute the product.



Scheme 2: Synthesis of the phenyl linked zwitterion targets. Reagents and conditions: (a) EDC, DMAP, DCM, rt, 60 h, 33-87%; (b) 4N HCl in dioxane or TFA and DCM, rt, 4-60 h, 54-99%; (c) H_2, Pd/C, MeOH, rt, 20 h, (d) 35% $NH_4OH_{(aq)}$, MeOH, rt, 230 h, **5.3.7** - 7.6%.

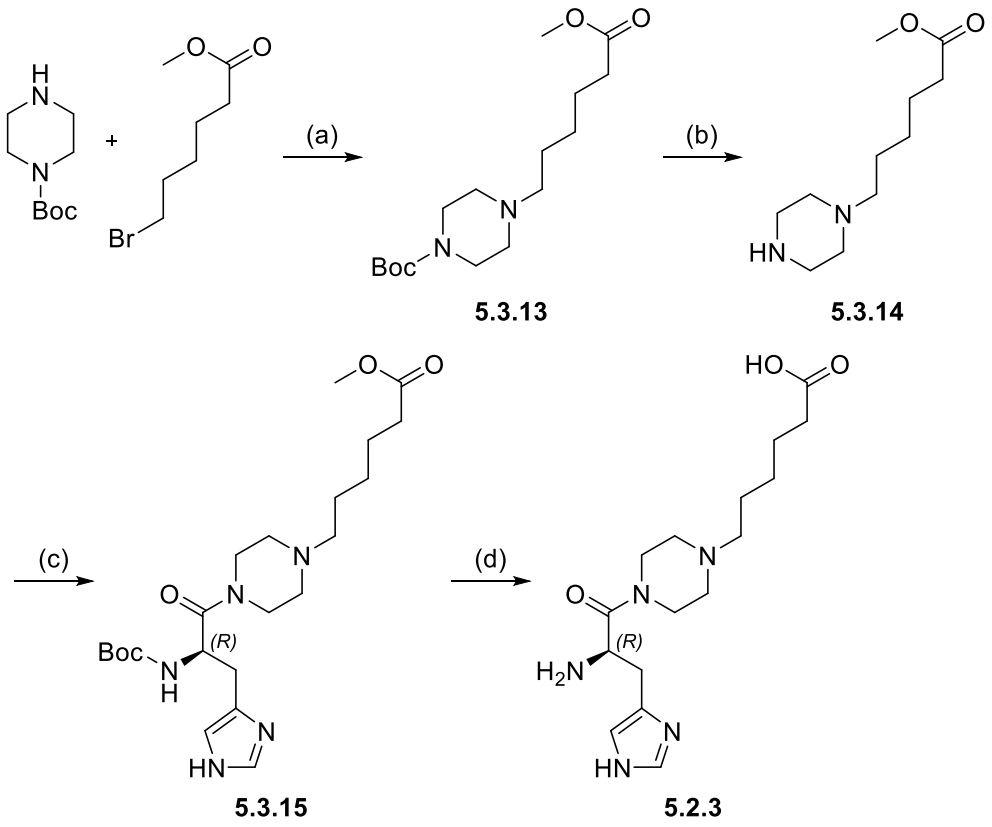
Removal of the N-Boc group of compound **5.3.10B** was attempted using 4N HCl in dioxane. The crude mixture showed significant impurities which could not be removed by purification using reverse-phase column chromatography, despite collecting a cleanly resolved peak. Hydrolysis of the ester, shown in Scheme 3, was also attempted but abandoned after the product could not be extracted into the organic phase during a workup.



5.3.10B **5.3.11**

Scheme 3: Attempted deprotection of **5.3.10B** by reversing the order to hydrolyse the ester first. Reagents and conditions: (a) 35% $NH_4OH_{(aq)}$, MeOH, rt, 40 h.

The D-arginine derivative **5.3.12** was treated with TFA to remove the Pbf group. LCMS showed the desired mass ion, however purification by normal and reverse-phase column chromatography failed to isolate the product. The product could be flushed from the normal-phase column in 20% methanol, 80% DCM but did not elute cleanly.



Scheme 4: Synthesis of the alkyl linked target **5.2.3**. Reagents and conditions: (a) NEt_3, THF, rt, 16 h, 45%; (b) 4N HCl in dioxane, rt, 60 h, 99%; (c) EDC, DMAP, dioxane, rt, 24 h, 35%; (d) 4N HCl in dioxane, MeOH, rt, 18 h, 59%.

Synthesis of compound **5.2.3**, with an alkyl chain replacing the phenyl ring at the metal binding terminus, followed the same general method as the other targets, shown in Scheme 4. N-Boc-piperazine was reacted with methyl 6-bromohexanoate in a substitution reaction using NEt_3 in THF to obtain compound **5.3.13** with a yield of 45%. Removal of the N-Boc group using 4N HCl in dioxane to obtain compound **5.3.14** in quantitative yield. This was coupled with Boc-D-histidine using EDC and DMAP in dioxane to obtain compound **5.3.15** in 35% yield, which was deprotected using 4N HCl in dioxane to remove both the Boc group and the methyl ester. LCMS showed the desired product had formed. Purification using reverse-phase column chromatography gave the desired product **5.2.3** in a 59% yield with high purity.

5.4 Docking studies and design of pyrazole zwitterion series

In a virtual screen of compounds by Emily Rolfe, several high scoring compounds were identified containing a common pyrazole core. The structures were redocked, with a selection of the high scoring compounds are shown in Figure 23.

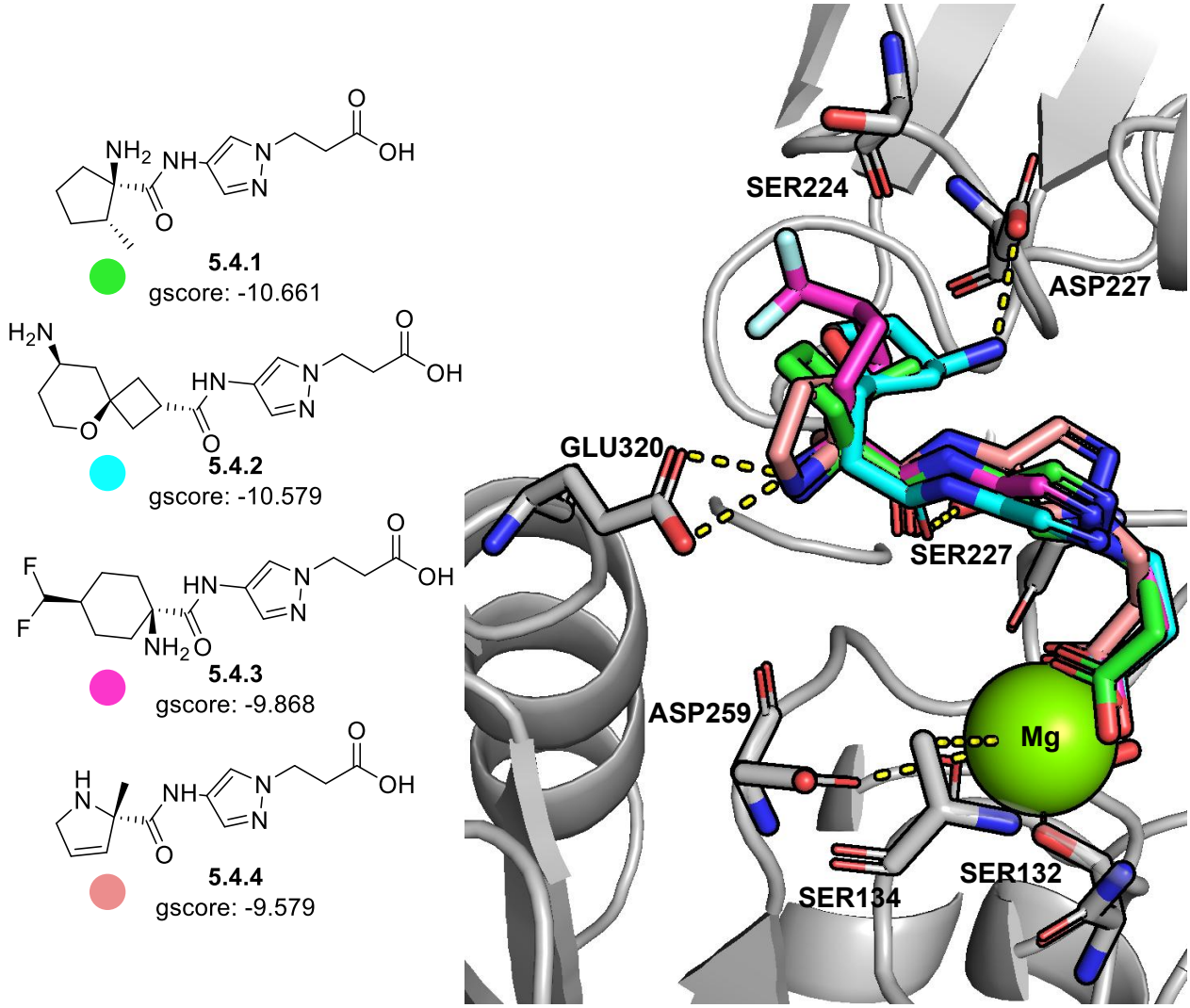


Figure 23: Top scoring pyrazole compounds from molecular docking of virtual HTS hits to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK0).

As the molecules shared a common pyrazole core and carboxylic acid, a synthesis of the common intermediate was an attractive target. As the basic head groups all largely contained amines, a docking run of the protected amines with free carboxylate acids for the linking amide bonds contained in the group inventory was run. The docking results in two different crystal structures (4WK0 and 4WK4) are shown in Figure 24 and Figure 25. The 4WK0 crystal structure, as detailed in section 3.3, is the $\alpha_5\beta_1$ integrin headpiece with the RGD tripeptide co-crystalised. The 4WK4 crystal structure, from the same publication, has a cyclic RGD peptide co-crystalised.[68] Modelling targets with both structures may reveal differences in how the residues in proximity to the RGD core adjust to the ligation of the 20-fold higher affinity cyclic peptide, compared to the undecorated RGD tripeptide.

As found in the design of the initial targets, the top scores were achieved by the amino acid derivatives, **5.4.5-9**. The top non-amino acid result **5.4.10** has been included in the

4WK4 result despite other amino acid derivatives scoring higher. Compound designs that were longer than the screening results in Figure 23 appeared desirable to engage GLN221 and ASP227; key interactions that define the activity of the endogenous ligand. This is highlighted by the arginine derivatives which mimic the interactions by RGD shown in Figure 10. The docking scores appear to reflect this improvement, with compound **5.4.5** reaching a gscore of -11.2 in the 4WK0 docking run, while the screening results had a top score of -10.7.

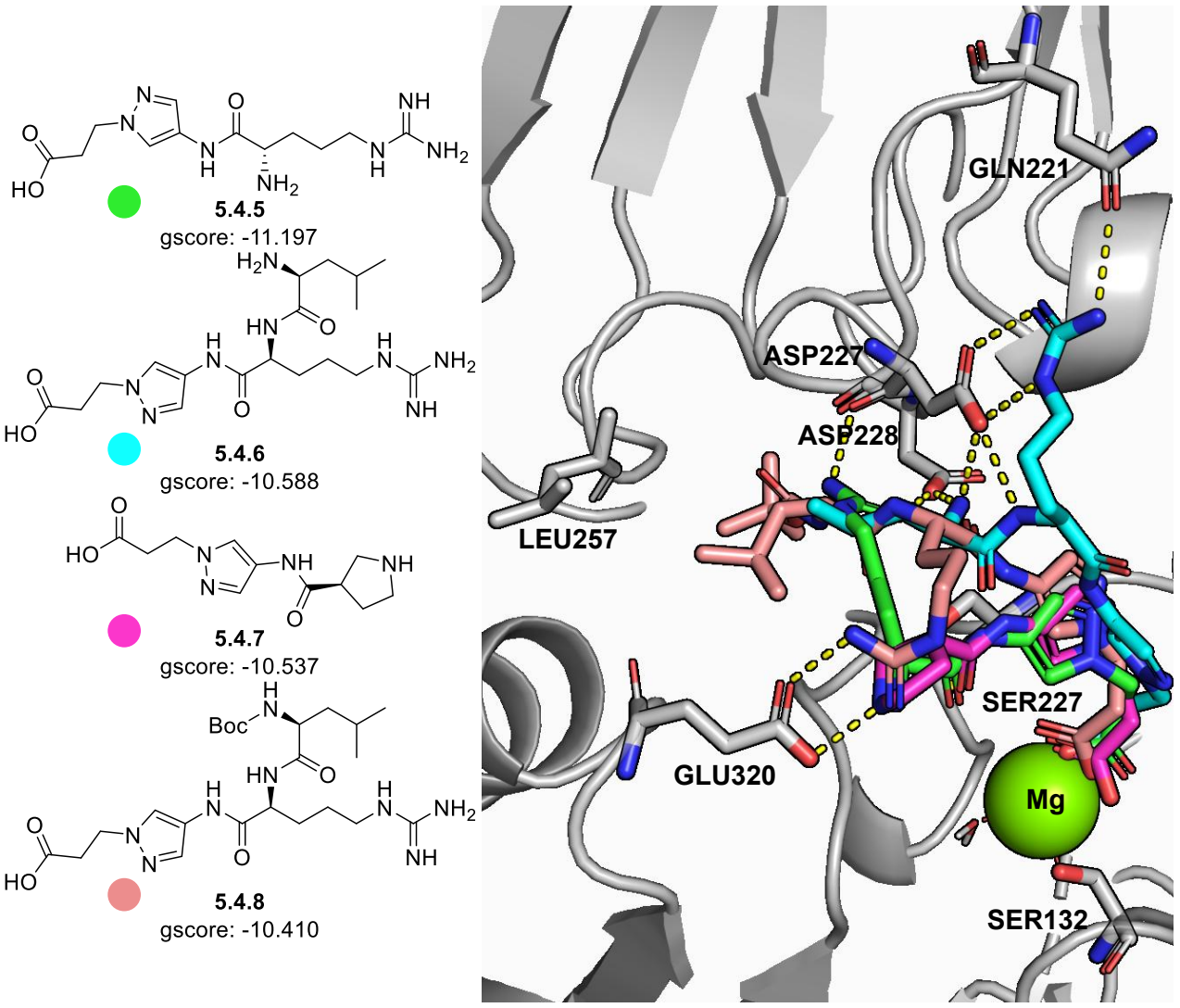


Figure 24: Top hits from molecular docking of potential pyrazole targets to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK0), targeting interactions with residues GLN221 and ASP227 to match the endogenous ligand.

Derivatives with arginine-leucine and arginine-isoleucine attached to the pyrazole were designed to form hydrophobic interactions at the left of the endogenous binding site with residues such as LEU257, GLY255 and ILE225, and to reduce the hydrophilicity of the compounds overall. This approach was also utilised when docking the compounds with Boc protecting groups remaining on the compounds, such as for compounds **5.4.8** and

5.4.9. In the 4WK0 docking model, the deprotected arginine derivative, compound **5.4.5**, scored the highest, forming similar interactions to the docking in Figure 23. This was followed by the arginine-leucine derivative **5.4.6** and the arginine-Boc-leucine derivative **5.4.8**. However, in the 4WK4 docking results the arginine-Boc-leucine derivative **5.4.8** was the highest scoring, followed by the Boc-arginine derivative **5.4.9**, and the arginine-leucine derivative **5.4.6**.

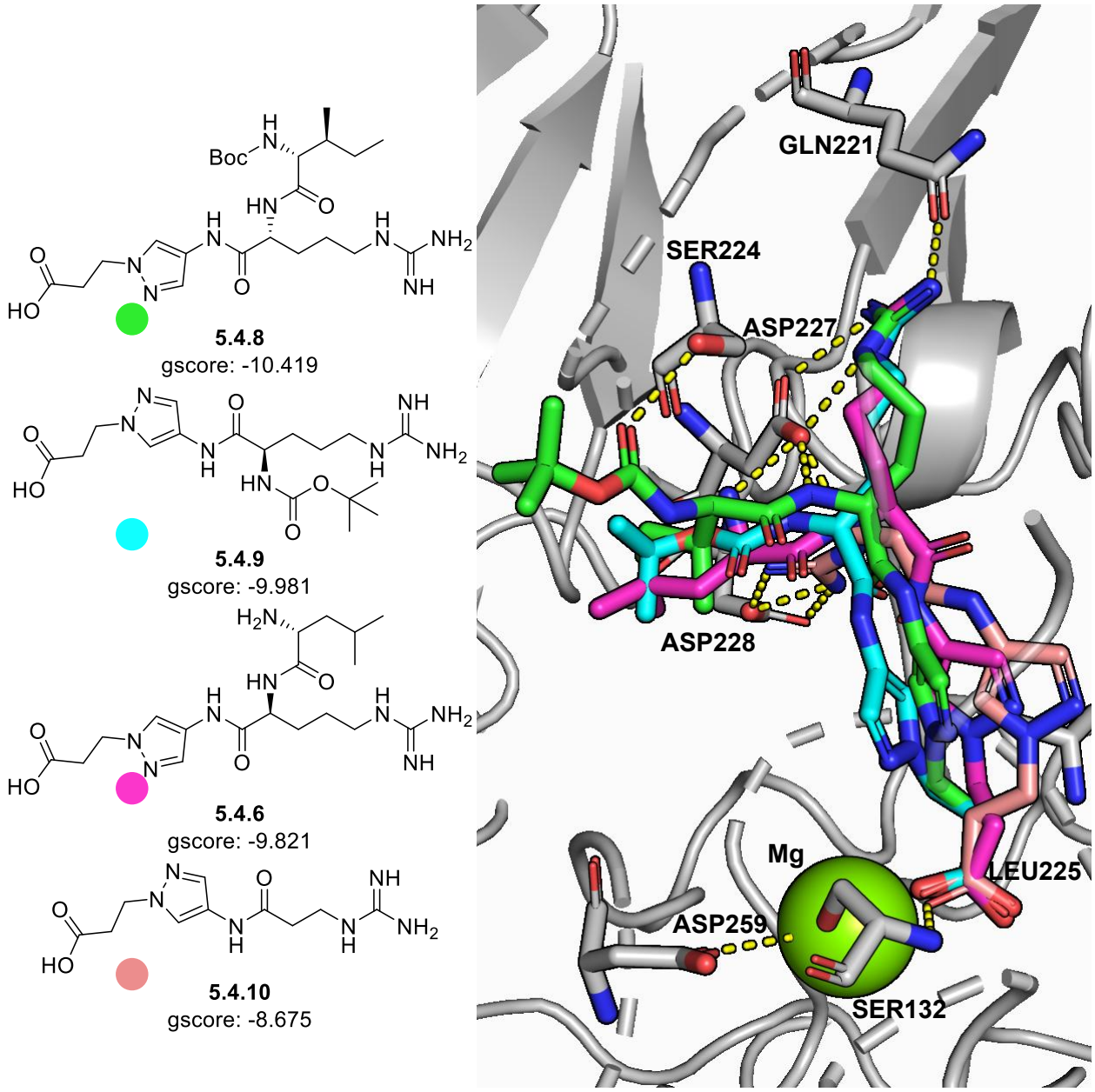


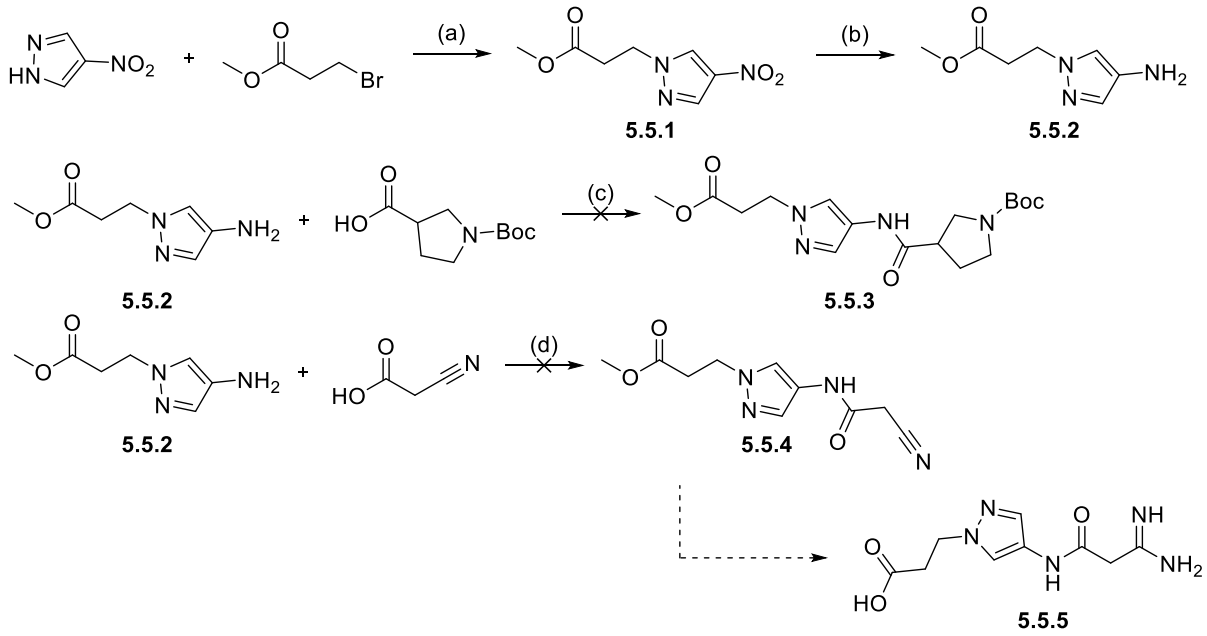
Figure 25: Top hits from molecular docking of potential pyrazole targets to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK4). The alternative docking model formed similar interactions to the protein but had overall lower scores and had different top-scoring structures.

Synthetic targets were selected based on the docking scores, length of synthetic route and potential of late-stage derivatisation. The initial targets were the arginine derivatives

5.4.5 and **5.4.9**, the pyrrolidine target **5.4.7** and a lower scoring amidine derivative via a nitrile intermediate.

5.5 Synthesis of pyrazole zwitterion series

The intermediate **5.5.1** was synthesised by a substitution reaction between 4-nitro-1H-pyrazole and methyl 3-bromopropionate using K_2CO_3 in DMF with a 99% yield. This was hydrogenated with Pd/C to obtain the amine **5.5.2** in 99% yield.

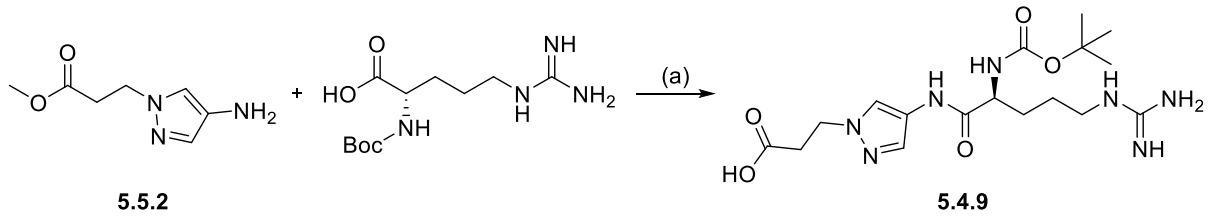


Scheme 5: Attempted synthesis of pyrazole targets **5.4.7** and **5.5.5**. Reagents and conditions: (a) K_2CO_3, DMF, 60°C, 70 h, 99%; (b) H_2, Pd/C, EtOH, rt, 40 h, 90%; (c) Oxalyl chloride, NEt_3, DMF, rt, 40 h; (d) EDC, DMAP, DCM, rt, 60 h.

Reaction of compound **5.5.2** with 1-Boc-pyrrolidine-3-carboxylic acid was initially attempted using oxalyl chloride and DMF, then NEt_3 to produce the amide coupled product. This appeared to co-elute with an unknown impurity during normal-phase column chromatography, observed significantly in the NMR. The ester hydrolysis was attempted with the impure material and purified by reverse-phase column chromatography and MDAP, yielding 7 mg of compound **5.5.3** that was recognisable by NMR, however the Boc deprotection was not attempted due to the low material amount, low purity and changing target priorities.

Amide coupling of compound **5.5.2** was simultaneously attempted with cyanoacetic acid using EDC and DMAP in DCM to produce **5.5.4** to allow for derivatisation of the nitrile group to the amidine. Significant reagent co-elution and poor column performance

prevented a clean elution and only a small amount of impure material was collected. An ester hydrolysis was attempted but no material was collected from purification.



5.5.2 **5.4.9**

Scheme 6: Synthesis of pyrazole target **5.4.9**. Reagents and conditions: (a) EDC, DMAP, DCM, rt, 18 h, 41%.

Boc-L-arginine was also coupled to compound **5.5.2** with EDC and DMAP in DCM. LCMS analysis indicated that the intended, acid protected product was generated, but during workup, while attempting to neutralise the guanidine and remove excess unreacted material, the mixture was sufficiently basified to hydrolyse the ester. Mass-directed automated purification (MDAP, a mass directed HPLC system) was required to isolate **5.4.9** with a 98% purity in an overall yield of 41%. The Boc deprotection was completed using 4N HCl in dioxane but could not be sufficiently purified by a reverse-phase column chromatography due to poor elution that had been observed with many of the late-stage zwitterionic products.

The purification of zwitterions may be improved by using HPLC as a primary purification method, as is the case for peptide synthesis, or by incorporating buffers into the reverse-phase flash column system, which is precedented to add lipophilicity to zwitterionic compounds.[115]

5.6. Design and synthesis of additional zwitterions

Compound **3.6.4**, identified from screening, contains a 7-membered ring which did not appear to form any strong interactions with the protein from modelling, shown in Figure 26. Altering the basic terminus to a glycine derivative (**5.6.1**) yielded an improved docking score and appeared to form improved interactions to the metal ion and to the ASP227.

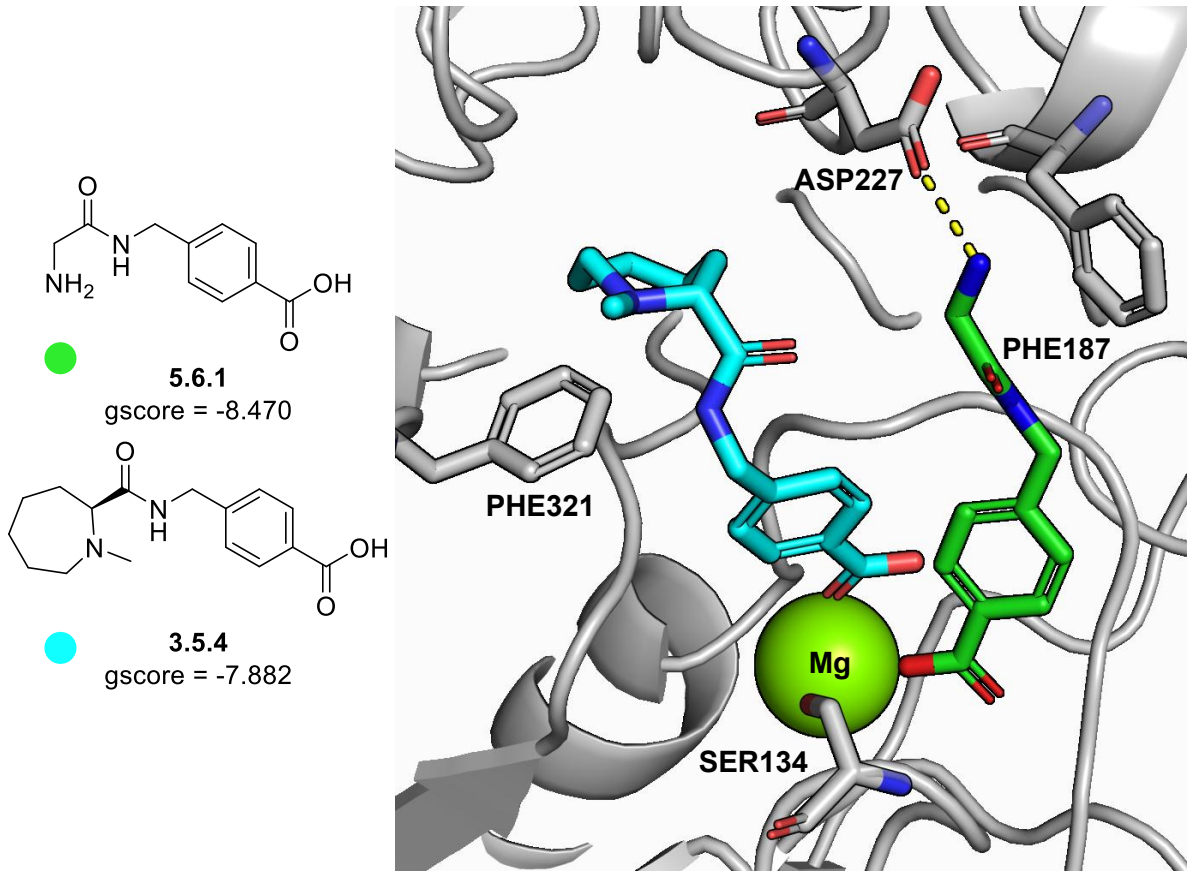
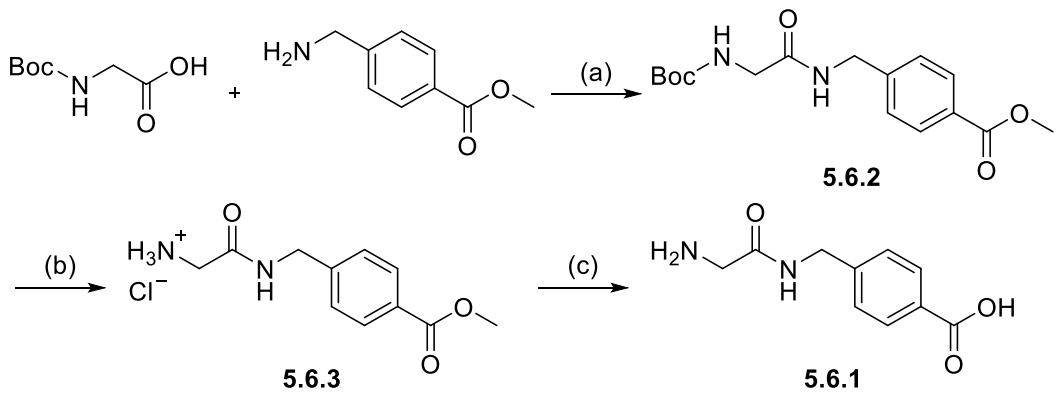


Figure 26: Molecular docking of screening compound **3.4.5** and the modified **5.6.1**, with the 7-membered ring removed, to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 3VI4).

Boc-glycine was coupled with methyl 4-(methylamino)benzoate using EDC and DMAP in DCM to obtain **5.6.2** in an 84% yield. This was subsequently deprotected using 4N HCl in dioxane to obtain **5.6.3** in quantitative yield, before hydrolysing the ester in ammonium hydroxide and methanol. The reaction mixture was evaporated to obtain **5.6.1** with quantitative yield.



Scheme 7: Synthesis of zwitterion target **5.6.1**. Reagents and conditions: (a) EDC, DMAP, DCM, rt, 16 h, 84%; (b) 4N HCl in dioxane, rt, 60 h, 99%; (c) 35% $NH_4OH_{(aq)}$, MeOH, rt, 16 h, 99%.

5.7 Cellular evaluation of zwitterion series

The zwitterion products and some intermediates were submitted to colleagues in the School of Medicine for biological analysis.

A live/dead assay was run using C2C12 mouse myotube cells (also used in the western blot protocol) and in HUVEC cells (Figure 27). This work was completed by Emily Rolfe. No issues were seen for any of the submitted compounds, with no compounds showing more than 10% increased cell death over the control in any repeat.

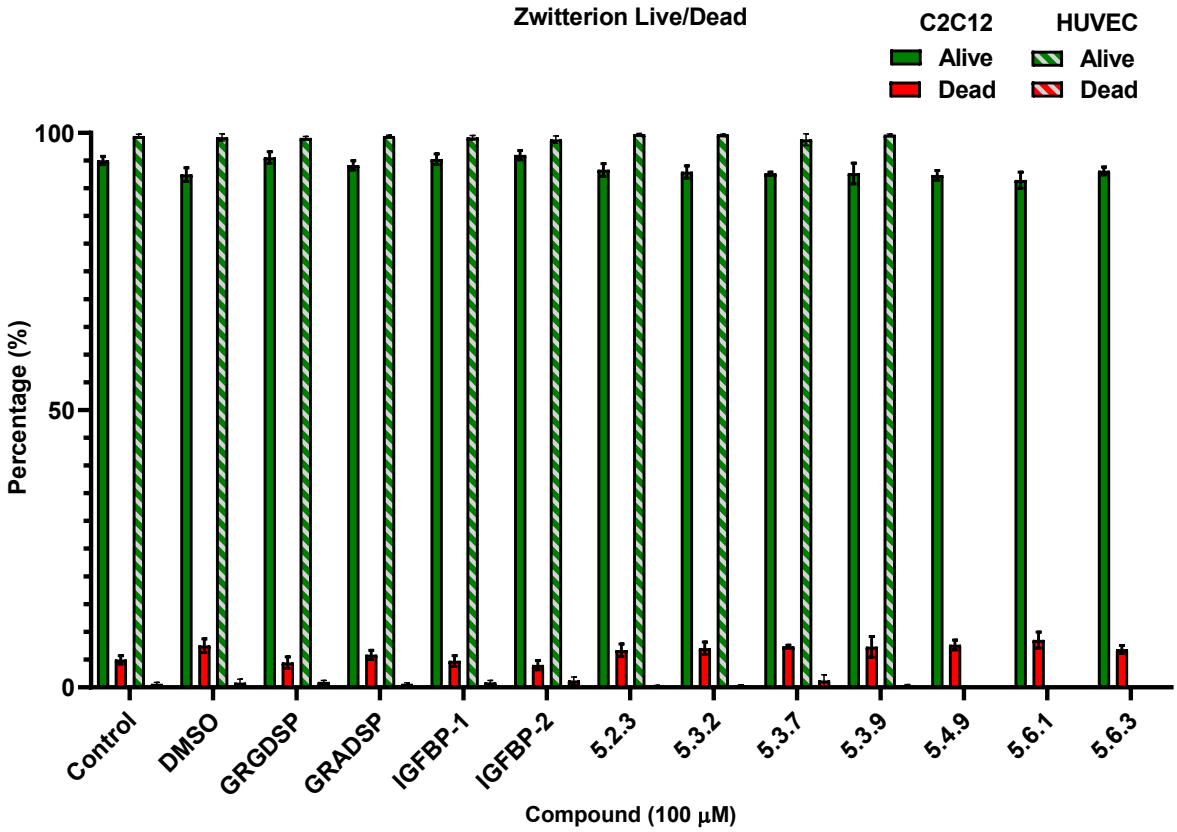
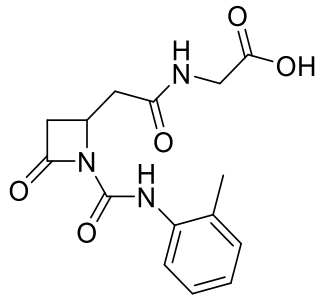


Figure 27: Cell viability of C2C12 and HUVEC cell lines 24 hours after exposure to a selection of zwitterion compounds (100 µM). C2C12 cells (n=3, mean ± SEM) and HUVEC cells (n=4, mean ± SEM). Assay not run in HUVEC cells for compounds **5.4.9**, **5.6.1** and **6.5.3**.

The same phosphorylation of Akt western blot protocol used in the preliminary results was used for activity quantification. All the western blot assays run for this project were completed by Dr Lia De Faveri, a post-doctoral researcher working in the School of Medicine. The controls used were a negative, insulin only, GRADSP (the negative hexapeptide), GRGDSP (the positive hexapeptide), IGFBP-1 and IGFBP-2. At this stage, the insulin control was beginning to show significant variability compared to previous runs of the protocol which prompted two runs of the assay, using 100 nM and 10 nM of insulin. The compounds were dosed at 100 µM in each.

A sample of **GM263**, shown in Figure 28 and measured to increase cell adhesion to fibronectin with an EC_{50} of 9.9 nM, from Baiula *et al.* was generously shared by Prof. Daria Giacomini and was used as a positive control in the assays, specific to $\alpha_5\beta_1$ integrin.[95] The group synthesised a number of agonists with selectivity for several integrins, with activity measured by cell-adhesion and immobilised integrin competition binding against the endogenous ligands, such as fibronectin.



GM263

Figure 28: Compound **GM263**, synthesised and provided by Prof. Daria Giacomini and her research group. It has a reported 9.9 nM agonistic activity, specifically targeting the $\alpha_5\beta_1$ integrin.

While they had not measured the effect of **GM263** on phosphorylation of Akt, they had measured a 4.5x increased phosphorylation of the extracellular signal-regulated kinases 1 and 2 (ERK1/2). The authors found that ERK1/2 phosphorylation was modulated in a concentration dependent manner by $\alpha_5\beta_1$ integrin agonists and antagonists, adding to the complexity of the integrin function.[95]

In a further preliminary assay measuring phosphorylation of Akt, cells treated with 100 nM of insulin (Figure 29, A), the negative controls of insulin and GRADSP have both shown a greater response than expected. In the critical preliminary results (Figure 19), all compounds had a greater response than observed with insulin only, with a mean of approximately 0.9 pAkt/Akt. In this result (Figure 29, A), the insulin control and the weakly responding compounds show a similar response (approximately 0.85 pAkt/Akt) and given the diversity of structures, this is thought to be the insulin baseline. This is further complicated in the results when the cells were treated with 10 nM of insulin (Figure 29, B), as the insulin response is slightly below the positive control hexapeptide GRGDSP and higher than all other positive controls and compounds. The negative control hexapeptide GRADSP in this case is much lower; a striking difference considering the two hexapeptides differ by a single methyl group. In this case, the insulin only control is not expected to represent the assay baseline but indicates that the effect of the insulin on the cells is highly variable. Overall, it is hard to draw robust conclusions from these results without replication when considering the variability of the controls. The assay was

deemed unsuitable for use after these results due to the variability of the insulin control and subsequent issues of cell proliferation, with development of alternative assay conditions prioritised by the team, rather than replication of results. However, as this was the first biological data for the compounds and in lieu of other assay results, some trends were used for target design and synthesis.

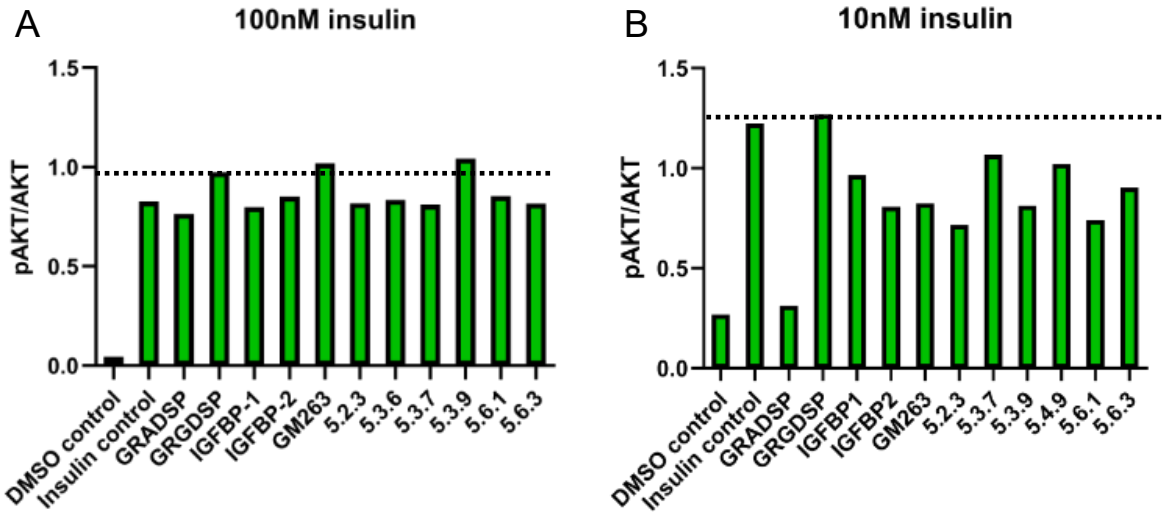


Figure 29: Western blot analysis of insulin-stimulated phosphorylation of Akt by synthesised compounds and controls with C2C12 mouse myotube cells, at 100 (A) and 10 (B) nM concentrations of insulin in each well except the DMSO control. Cells were dosed with 100 nM of the controls and compounds and incubated for 40 mins, before exposure to insulin. Both graphs are n=1; Akt: protein kinase B.

No significant result was seen for any of the zwitterion compounds, with the response from the compounds close to or below the insulin only control. Despite the wide variety of structures, the results cannot be compared due to the lack of replicates, however overall, the activity is low and does not indicate potent agonists.

A scratch-wound assay was developed to measure the migration of endothelial cells, which is impaired in patients with T2D, but is enhanced by IGFBP-1 *in vitro*.[116] The proliferation and migration of the endothelial cells was measured over 24 hours after a displacement of HUVEC cells in a plate (Figure 30). The migration of cells to close the scratch involves interactions between cells and with the extracellular matrix.[117] Compounds that reduce cell migration may indicate an inhibitor of the cell adhesion integrins.[117] These results were collected by Emily Rolfe.

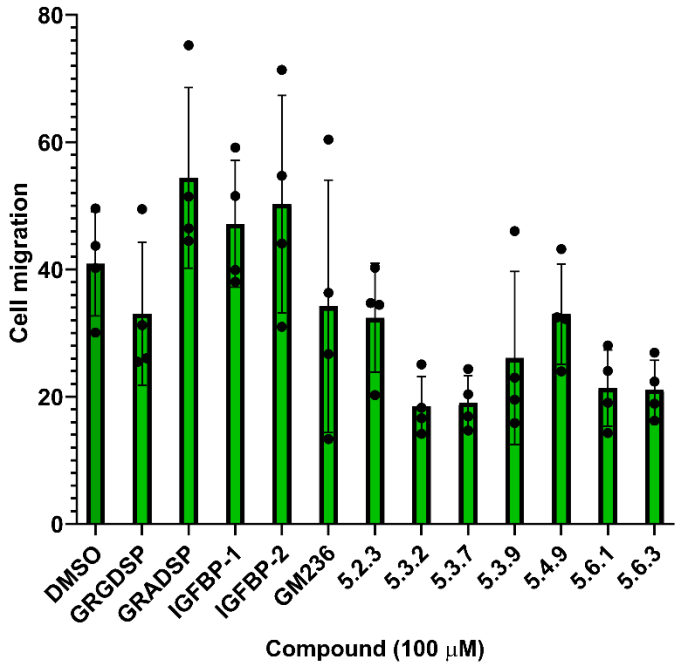


Figure 30: Cell migration of HUVEC cells, 24 hours after a scratch-wound while exposed to a selection of proteins, controls and zwitterion compounds (n=4, mean ± SEM).

The hexapeptide GRGDSP is a binding control, while the GRADSP hexapeptide is the non-binding control. It is theorised that cell migration that matches IGFBP-1 could indicate an $\alpha_5\beta_1$ agonist. Activity matching GRGDSP may also represent success of small molecule binding to the $\alpha_5\beta_1$ integrin, but due to the complex nature of the protein interactions this may be different to the activity of IGFBP-1.

The results show no statistical significance between the positive controls and the compounds, although with four replicates completed some trends can be seen. The GRADSP and IGFBP-1 and 2 controls show the strongest cell migration effect, above the DMSO control and all other compounds. This result does not appear to agree with the preliminary result that showed that dosing GRGDSP had the same effects on phosphorylation of Akt as recombinant IGFBP-1.[113]

Comparison against the GRGDSP hexapeptide shows three compounds are within error of the binding control, including **5.2.3**, **5.3.9** and **5.4.9**. Four compounds, **5.3.6**, **5.3.7**, **5.6.1** and **5.6.3**, fall below the binding control, reducing cell migration. While **5.3.9** and **5.4.9** had previously shown a stronger response in one of the phosphorylation of Akt experiments, **5.2.3** had shown no previous indication of activity and **5.3.7**, which showed a good response in the phosphorylation of Akt using 100 nM insulin, showed reduced cell migration.

Despite the assay measurements not being as thorough as desired, it was determined that the zwitterion series had not shown sufficient promise to justify continued synthetic efforts, and the project moved on to different targets.

5.8 Biophysical evaluation by SPR

Surface plasmon resonance (SPR) is a powerful technique for quantifying direct binding affinities of analytes to an immobilised protein. It provides a real-time measure of analyte binding to the protein surface without fluorescent labels or tags, and can be used to measure association, dissociation and steady state kinetics.

The physical process involves an incident beam of light colliding with a semi-circular prism placed on a gold film. Total internal reflection occurs in the prism and the reflected light reaches a detector which measures the angle of reflection. At specific angles of incidence, the photons from the light source are absorbed by the gold film and convert the energy into surface plasmons, which are excited, oscillating electrons at the metal-dielectric interface (gold-prism interface). Because the incident photons are being absorbed by electrons on the surface, they are not reflected, hence the intensity of the reflected beam is reduced. This creates conditions where the local environment of the gold surface is highly sensitive to change.[118, 119]

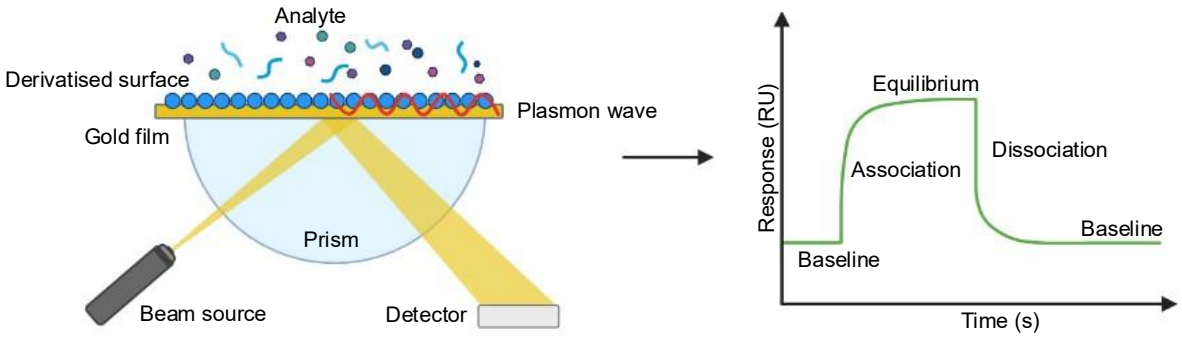


Figure 31: Experimental setup of surface plasmon resonance (SPR). A derivatised gold film surface on a semi-circular prism reflects a beam to a detector. The refractive index the surface is modified by the molecular weight of the derivatised surface, meaning the association and dissociation of an analyte can be detected, allowing measurements of the analyte-surface affinity. Figure made with BioRender.

With a derivatised gold surface, such as with a protein receptor, changes such as analyte binding change the refractive index of the surface, resulting in a detectable change in the angle of refection. In this project, the surface was derivatised with streptavidin which binds to proteins tagged with biotin. The refractive index is also very sensitive to other condition changes, most commonly in the buffer.[118, 119]

SPR analysis was used by X. Pang *et al.* to identify a non-RGD inhibitor of $\alpha_V\beta_3$ integrin with a K_D of 102 nM.[120] Seven integrins, including $\alpha_5\beta_1$, were immobilised on a carboxyl methylated dextran surface via an amine coupling to determine the binding affinities. The best performing compound was tested in orthogonal assays, showing an 85% decreased tumour weight in a mouse model and a dose-dependent reduction of pAkt/Akt in a model epithelial cells (22Rv1).[120]

Similarly, SPR was used by Cheng *et al.* to evaluate the equilibrium dissociation constant (K_D) of macrocyclic RGD-peptides targeting $\alpha_v\beta_3$ for the treatment of cancer.[121] The response of the peptides was measured at 50 μM against $\alpha_v\beta_3$ and $\alpha_v\beta_5$, before the K_D values of the best responding compounds were determined in a nine-point titration against $\alpha_v\beta_3$, finding micromolar and sub-micromolar values.

5.9 Biophysical evaluation of zwitterion series

Subsequent to the previous assay results, an SPR assay to measure direct binding was developed by Dr Lia De Faveri and Emily Rolfe under the guidance of Dr Iain Manfield to directly analyse the interactions between the molecules and the integrin, to prove target engagement. The principles of this assay are detailed in section 5.8.

The protein used in the SPR assay was the $\alpha_5\beta_1$ integrin with a biotin tag on the N-terminus of the α_5 monomer, yielding a molecular weight of 196.1 kDa (112.4 kDa and 83.7 kDa respectively for the α_5 and β_1 monomers).[122] The response of the protein (referred to as the ligand in SPR literature), minus the background response of an underivatised control cell, is defined as R_{ligand}. This is related to the maximum response from the analyte binding, R_{max}, by Equation 1.[123] The other terms are the molecular weight of the protein (Mr_{ligand}), the molecular weight of the compounds being analysed ($Mr_{analyte}$) and the number of binding events for each unit of immobilised protein (V_{ligand}).

Equation 1: Calculation of the maximum analyte response (R_{max}), which is dependent on the ligand response (R_{ligand}), molecular mass of the analyte ($Mr_{analyte}$), molecular mass of the ligand (Mr_{ligand}) and the number of binding events for each immobilised protein (V_{ligand}).

$$R_{max} = R_{ligand} \times \frac{Mr_{analyte}}{Mr_{ligand}} \times V_{ligand}$$

Generally, an ideal R_{max} is 50-100 RU, with a lower R_{max} of 25 also being described as optimal for analysis of small molecule binding.[123, 124] Due to the high molecular weight of the integrin and the expected low affinity binding of the small molecules, the surface was

saturated as much as was possible, using three injections of 100 nM integrin to derivatise the surface.

After the first surface derivatisation, for a point concentration analysis of the compounds at 100 and 500 μM, R_{ligand} was 2247.6 RU, which degraded to 1225.6 RU at the end of the six batches of experiments over a week. After the second surface derivatisation, for a six- and eight-point titration series of the compounds, R_{ligand} was 1553.0 RU, which degraded to 733.2 RU at the end of the eight batches of experiments over a week. As the surface decay is slow, it is not expected to affect the results.

V_{ligand} is expected to be 1, with the compounds only designed to target the RGD binding site and the compound molecular weights range from 200-500 daltons. Using these values, R_{max} can be calculated for each experiment, depending on R_{ligand} during the experiment and Mr_{ligand} for each compound (Figure 32). For the titration experiments, R_{max} varied from 2.376 to 1.122 RU for a molecular weight of 300 and from 3.960 to 1.869 RU for a molecular weight of 500. The surface degradation and corresponding decrease in R_{max} may introduce errors between replicates.

Cytiva, manufacturers of the Biacore T200 used in the experiments, have published a poster regarding the detection of compounds to membrane proteins, with a low R_{max}.[125] They found that the K_D of a test ligand could be evaluated with R_{max} values as low as 1.0 RU. The R_{max} values obtained from the immobilisation are therefore compatible with the machine, although errors are likely to be larger than experiments with a greater R_{max}.

Figure 32 shows the protein binding response, R_{ligand}, the calculated R_{max} and the experimental $R_{analyte}$. Due to the ratio of the low molecular weight analytes to the high molecular weight protein, R_{max} is much lower than optimal, meaning the analyte responses are limited to a small range. Experimental $R_{analyte}$ values were greater than the theoretical R_{max} for several compounds. This is unexpected but could have several explanations.

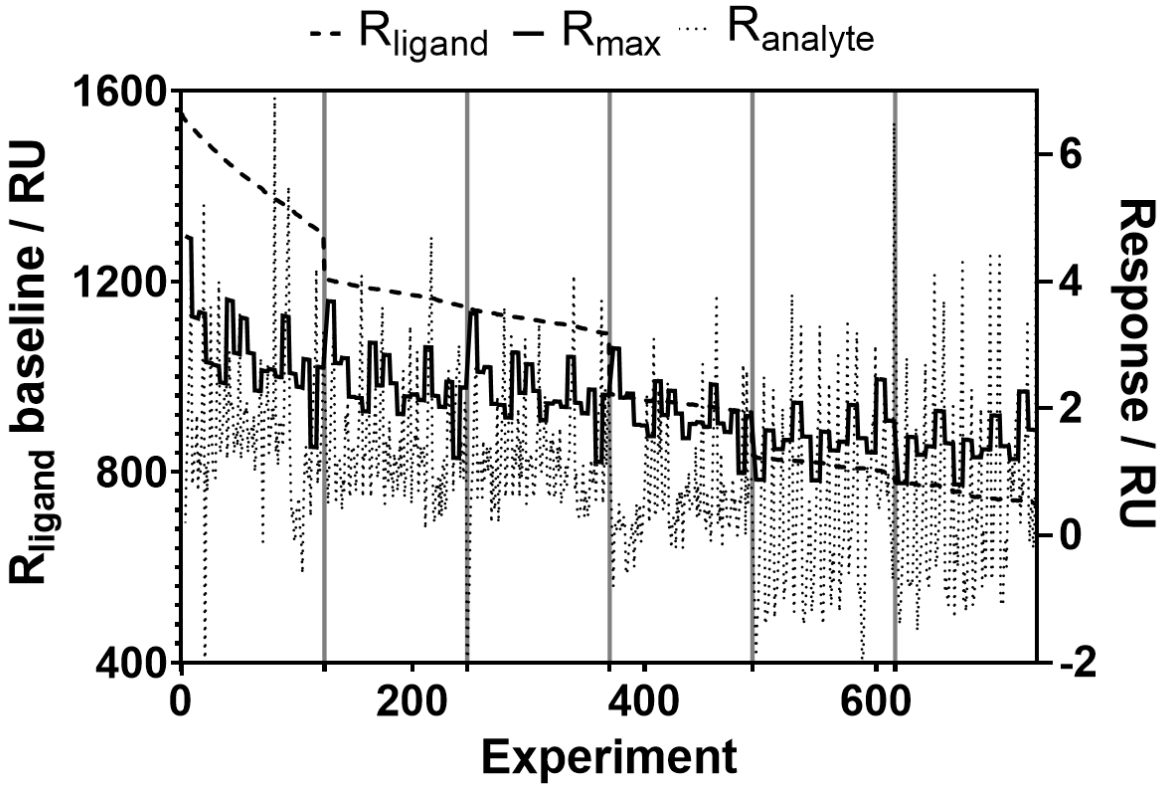


Figure 32: Theoretical and experimental responses from the titration SPR experiments. R_{ligand} (shown on the left Y axis) is the protein binding level which decreases over time due degradation of the chip surface. R_{max} (shown on the right Y axis) is the theoretical maximum binding of the analytes, based on R_{ligand}, the analyte molecular weight, the ligand molecular weight and assuming a single binding site. $R_{analyte}$ (shown on the right Y axis) is the experimentally measured response. Separate batches of experiments are distinguished with vertical grey lines on the X axis.

Non-specific interactions, where the analyte is interacting with the surface, rather than the ligand, will produce an increased response and can be measured by comparing the response to the undecorated control cell. This is shown with the theoretical R_{max} and measured $R_{analyte}$ in Figure 33. The binding to the reference cell is initially low, with expected peaks when the buffer is injected for a new batch of experiments, with additional peaks for compounds with non-specific interactions. However, during the eight-point titration, from experiment 493, the baseline binding to reference increases significantly. In addition, some compounds, such as the hexa-alanine peptide control, exhibit significant non-specific binding at higher concentrations. It is unclear why the overall binding to reference increases significantly for the later experiments, but this may explain why the $R_{analyte}$ responses are reduced for the eight-point titrations, as the response is calculated by subtracting the reference cell response from the derivatised cell response.

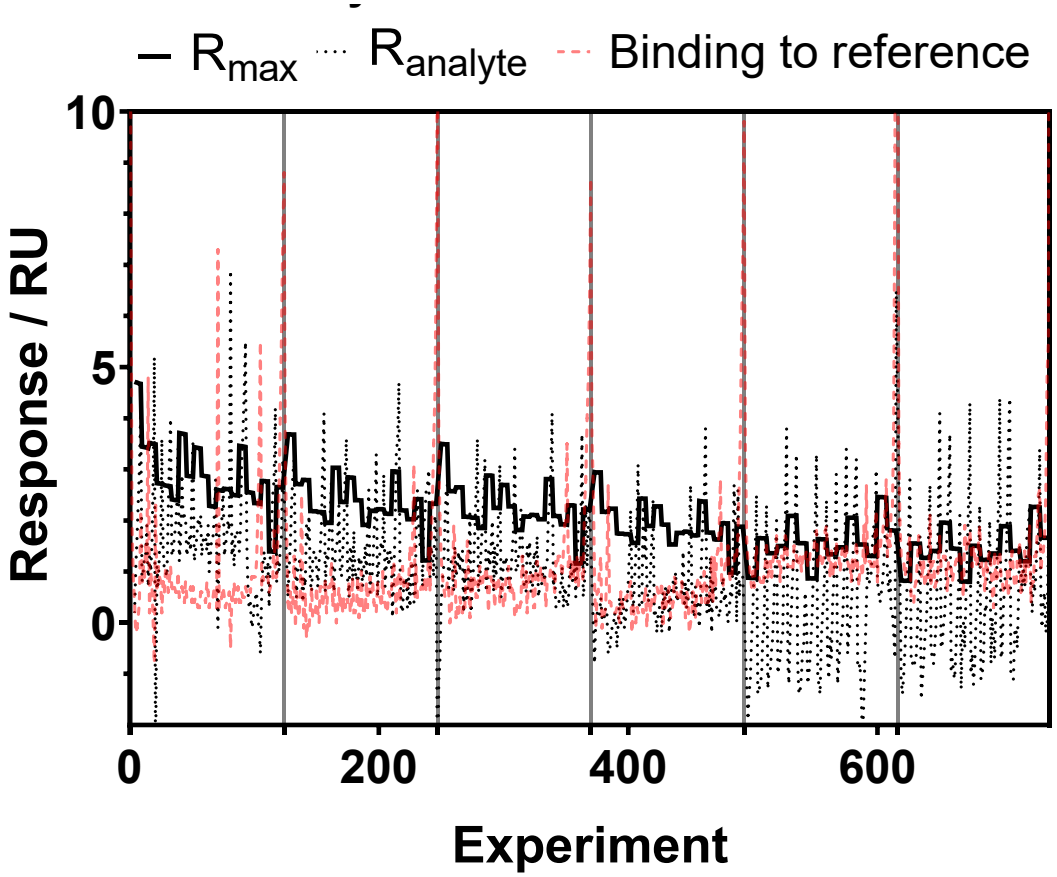


Figure 33: Theoretical and experimental responses and binding to the reference cell from the titration SPR experiments. R_{max} is the theoretical maximum binding of the analytes, based on R_{ligand}, the analyte molecular weight, the ligand molecular weight and assuming a single binding site. $R_{analyte}$ is the experimentally measured response. The regularly spaced peaks in the binding to reference are at the beginning of each batch of 123 samples, where the buffer is run and shows a large reference response. Separate batches of experiments are distinguished with vertical grey lines on the X axis.

One other possible explanation of $R_{analyte}$ being larger than R_{max} is an increase in V_{ligand}, where multiple binding interactions occur with a single immobilised protein dimer. The other possibility is a decrease in Mr_{ligand}, meaning the protein molecular weight has decreased. As the protein is tethered to the surface by the α_5 monomer and the two integrin monomers are not covalently linked, a dissociation of the β_1 monomer could be occurring over time, reducing the protein molecular weight. This could cause the reduction of R_{ligand} over time, rather than dissociation of the streptavidin-biotin bond. Each of the possibilities would approximately double R_{max} which would be greater than many of the increased responses.

Analysis of the waste lines for protein content could identify dimer dissociation if future SPR experiments were undertaken. It is unclear which factor is more likely, and the increased responses may be due to a combination of non-specific binding, multiple binding events and dimer dissociation.

Overall, this may indicate that the protein on the SPR surface is not an accurate reflection of the protein on a cell surface *in vivo*, however this is not possible to confirm with the experiments that were run. Despite this, the results have been analysed to assess the affinity of compounds with the immobilised integrin.

Eight compounds from the zwitterion series, **5.2.3**, **5.3.6**, **5.3.7**, **5.3.9**, **5.4.9**, **5.6.1** and **5.6.3**, were screened in 100 μM and 500 μM solubility tests with no issues identified, although only **5.3.7**, **5.3.9**, **5.4.9** and **5.6.1** were taken forward to short titrations due to plate capacity. The compounds were selected as they were final compounds and late-stage intermediates, and from the results of the phosphorylation of Akt assay and scratch-wound assay. It is important to note that most of these compounds (**5.3.7**, **5.3.9** and **5.4.9**) were taken from previously dissolved aliquots, rather than solid stocks due to the low yields of their synthesis. As SPR is highly sensitive to the buffer, the DMSO in the aliquots was evaporated with nitrogen before redissolution. However, without weighing out the solids, the results are less precise.

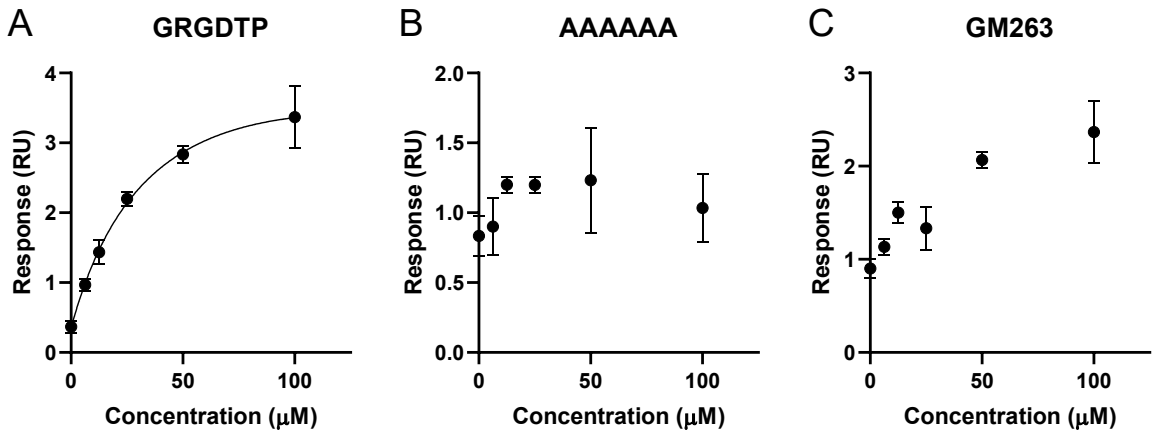


Figure 34: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by control compounds (A) Positive hexapeptide GRGDTP, (B) Negative hexapeptide AAAAAA, (C) Literature compound **GM263**, at six concentrations from 0-100 μM (n=3, mean ± SEM).

The controls used in this assay were the positive hexapeptide GRGDTP, changed from the previously used GRGDSP due to greater use of the former in literature. The negative control peptide was hexa-alanine, selected to show any non-specific interactions. All compounds, including the peptide controls, were dissolved in DMSO before they were diluted into the aqueous buffer and sequentially diluted. However, the solubility of the hexapeptides in the DMSO stock was poor, meaning the results are likely to have greater variation.

GM263 was also included as a positive control, although this was also collected from a pre-dissolved aliquot. Blank buffer samples were run at the start of the assay and

between each sample titration as the zero concentration. The peptide controls showed the expected behaviour of a dose-dependent binding response for the positive control and no dose-dependent response for the negative control. The response to the buffer was consistent throughout the assay and a dose-dependent effect was seen for **GM263**, although given the reported potency of the compound, the size of the response was unexpectedly weak. As the assays used by Baiula *et al*. to assess the activity of **GM263** were based on cell and protein adhesion to a fibronectin surface and differ significantly to the assays used in this project, the results may not be consistent with the reported results.[95]

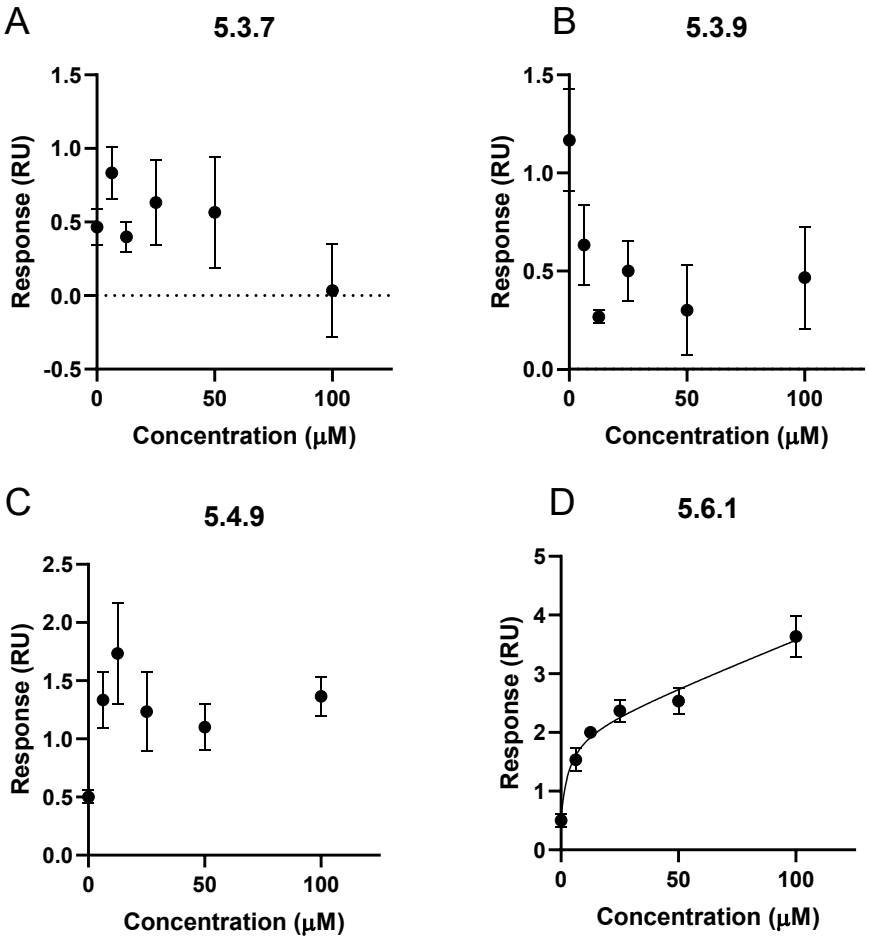


Figure 35: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by zwitterion compounds (A) **5.3.7** (B) **5.3.9** (C) **5.4.9** (D) **5.6.1**, at six concentrations from 0-100 μM (n=3, mean ± SEM).

Compounds **5.3.7** and **5.3.9** unfortunately showed no significant binding, both with responses falling below hexa-alanine and no dose-dependent effect. Compound **5.4.9** also showed no significant binding, with the response at 6.25-100 μM just above the buffer baseline with no dose-dependent effect. Contrary to the results in the previous tests, compound **5.6.1** showed a distinct dose-dependent response, very similar to the

GRGDTP positive control. As such, it was selected for a longer titration which would enable an equilibrium dissociation constant (K_D) measurement.

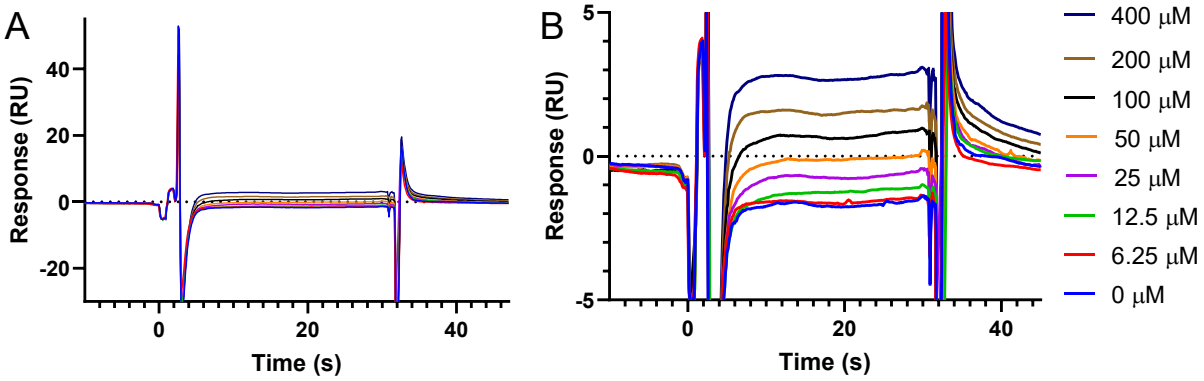


Figure 36: SPR sensorgram of compound **5.6.1** in the long titration experiments, by which the K_D was measured. Measurements are taken at 0.1 second intervals, with the mean response of the replicates (n=4) reported. The time is relative to the sample injection and the response is relative to the baseline response before sample injection. The change in response at 0 seconds and 30 seconds corrospond to the change from the runnning buffer to the sample.

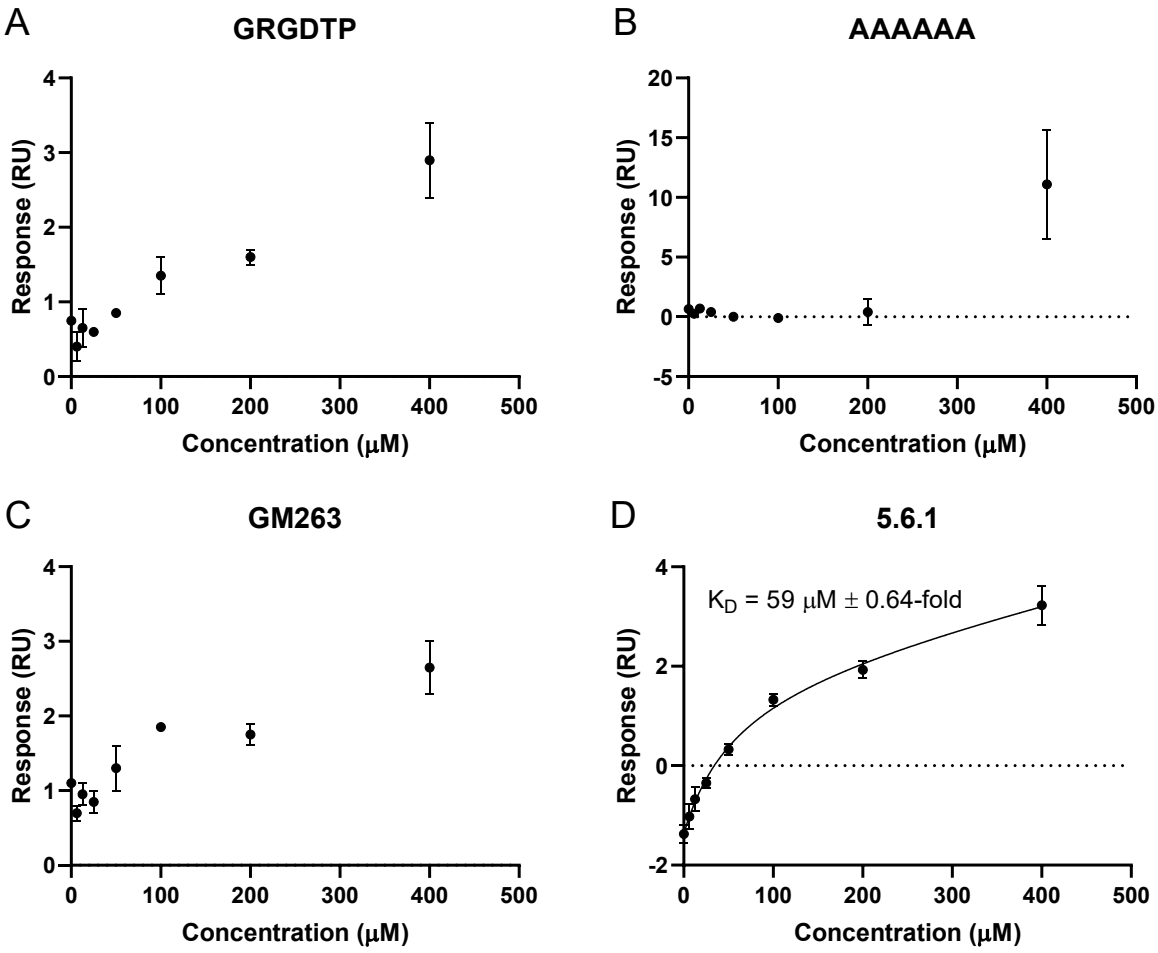


Figure 37: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by controls and zwitterion compounds; (A) positive hexapeptide GRGDTP (B) negative hexapeptide AAAAAA (C) literature compound

GM263 (D) **5.6.1**, at eight concentrations from 0-500 μM. (controls n=2, compound n=4, mean ± SEM).

The serial dilution, 400-6.25 μM, of the compounds (five others from other series are shown in Figure 61 and Figure 63) was completed twice on each plate, with this set up run twice, yielding four replicates. The sensorgram of **5.6.1** is shown in Figure 36 and the K_D measurements from this data is in Figure 37. Due to the number of wells available, each control could only be run once on each plate. Unfortunately, the controls were inconsistent for the long titrations, with the 0-50 μM concentrations of **GM263** and GRGDTP showing an increased response compared to the low concentrations of the other compounds in the assay (compound **5.6.1** had a typical response). The hexa-alanine compound also showed an increased response at the lower concentrations but showed a very large non-specific response at 400 μM. The solubility of the controls may have affected the results, but that does not explain the greatly increased response at low concentrations

As the two replicates generally showed closely correlated results, the cause of the unusual control behaviour is unclear. The compound running before the controls (not shown) had a poor response, with significant variation that may be due to non-specific interactions that were not clear from the screening or the short titration. No evidence of compound sticking, or surface degradation could be seen from the reference cell.

Due to the very fast on and off rates of all the compounds tested in the SPR assay, typical of low affinity molecules, only the equilibrium dissociation constant, K_D, can be calculated, using a non-linear regression analysis (calculated using GraphPad). This is measured by estimating the ligand concentration where half of the receptor sites are occupied at equilibrium.

Despite the behaviour of the controls, compound **5.6.1**, which was run earlier on the plate, showed a good concentration response curve (Figure 37) and was used to calculate a K_D of 59 μM ± 0.64-fold. The confidence interval of the K_D measurement was 18-266 μM. Although the confidence interval is wide, this is a good binding response of the compound

5.10 Summary of zwitterion series

The design of zwitterion-containing compounds to target $\alpha_5\beta_1$ integrin is a well validated approach and the derivatives were designed from the initial hits of the project to attempt to expand on the initial activity. However, the polarity of the compounds presented many issues during the purification of targets which severely limited the series output. These methods were improving with increasing experience, with attempts at deprotecting

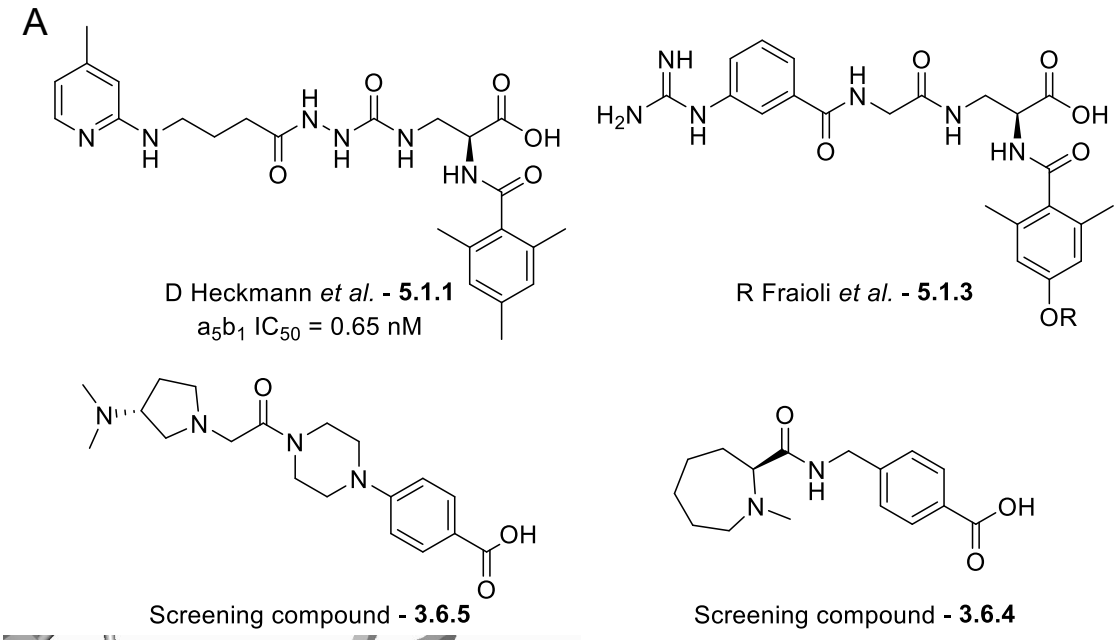
compound **5.3.10B** (Scheme 3) showing more success than previous attempts in the series. With further review of purification methodologies and more time spent on the series, the synthesis may have shown greater success.

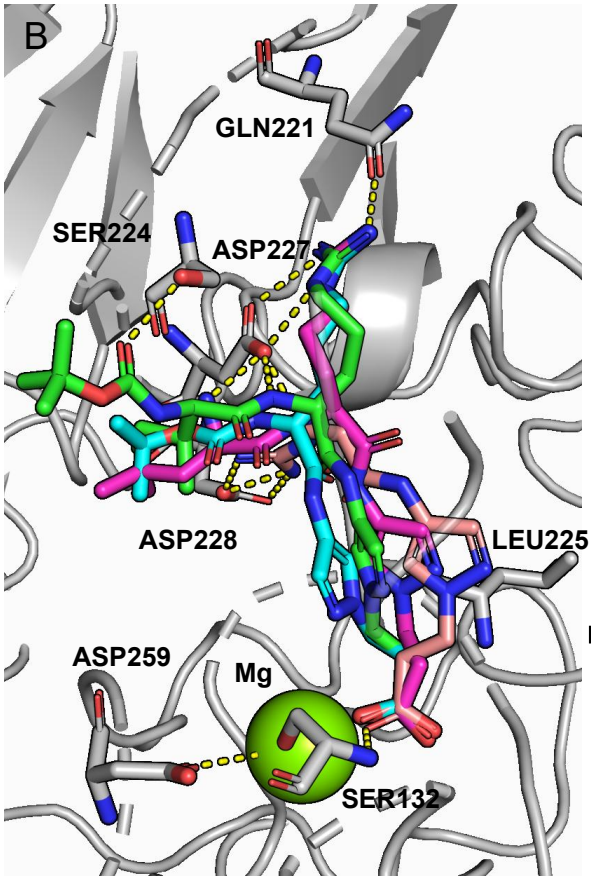
As compounds were produced and synthetic challenges navigated, the issues faced in the biological assays became quite significant. As such, the data collected for the compounds was at first, very limited. Although the compounds showed no cytotoxic effects, no clear conclusions about the design philosophy could be made from the phosphorylation of Akt or scratch-wound assays and the limited data that was available did not indicated potent activity of the molecules. It was decided to focus the investigations on a different series based on these results, which was validated when the SPR assay was established.

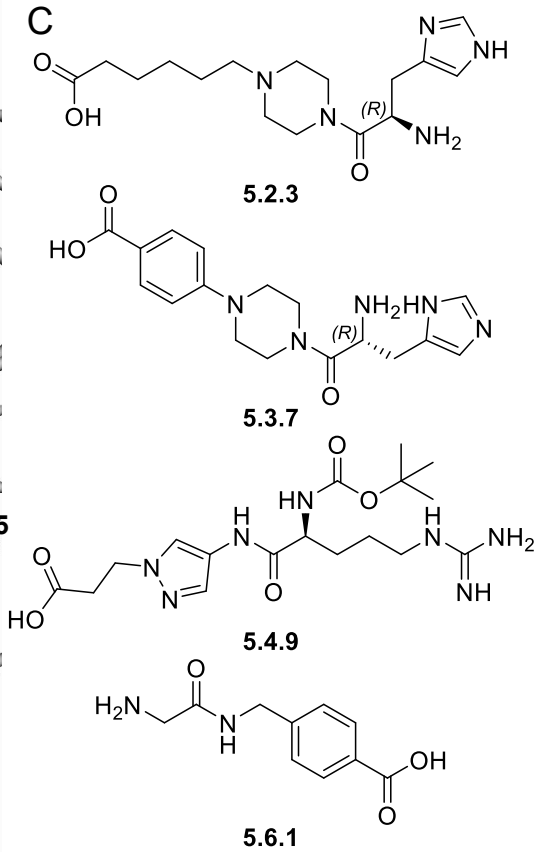
Only compound **5.6.1**, after showing no activity in the previous tests, showed a binding response that justified a binding affinity measurement, yielding a K_D of 59 μM ± 0.64-fold. This could indicate an antagonistic interaction, which is supported by the phosphorylation of Akt assay where it had one of the lowest ratios of pAkt/Akt at both insulin concentrations (Figure 29), and a significant reduction in cell migration 24 hours in the scratch-wound assay (Figure 30).

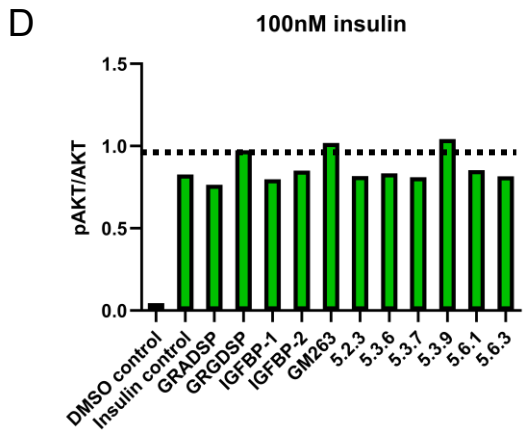
Although compound **5.6.1** employs a simple zwitterion structure, molecular modelling had calculated that the structure would be a potent binder, with a gscore of -8.470. The size of the compound may mean that the compound is forming a small number of strong interactions, similar to a fragment binder. Furthermore, the other compounds with higher molecular weight may not be forming enough strong interactions to overcome any clashes and a greater entropic binding penalty, despite better scores in modelling. This highlights an area where modelling and experimental results are in poor agreement and can be used to inform future modelling.

Notwithstanding that the activity of compounds **5.3.7**, **5.3.9** and **5.4.9** may have been affected by the SPR sample preparation, the results from the docking studies were not validated and the zwitterionic design philosophy had not generated the agonistic activity that the project was aiming for.


A
D Heckmann et al. - 5.1.1
a5b1 IC50 = 0.65 nM
R Fraioli et al. - 5.1.3
Screening compound - 3.6.5
Screening compound - 3.6.4


B
GLN221
SER224
ASP227
ASP228
LEU225
ASP259
Mg
SER132


C
5.2.3
5.3.7
5.4.9
5.6.1


D
100nM insulin
pAKT/AKT
1.5
1.0
0.5
0.0
DMSO control
Insulin control
GRADSP
GRGDSP
IGFBP-1
IGFBP-2
GM263
5.2.3
5.3.6
5.3.7
5.3.9
5.6.1
5.6.3

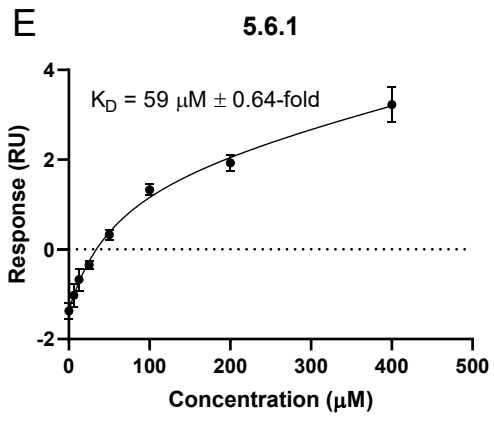

E
5.6.1
KD = 59 µM ± 0.64-fold
Response (RU)
4
2
0
-2
0
100
200
300
400
500
Concentration (µM)

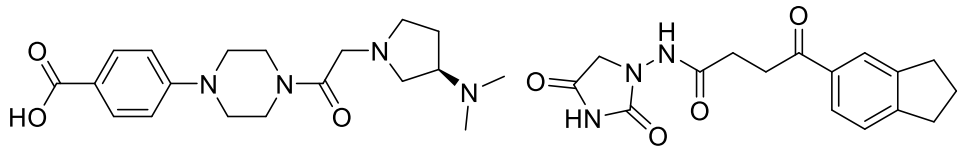
Figure 38: Summary figure of the work completed for the zwitterion series. (A) Literature compounds and screening results provided the source of the compound design. (B) Compounds were docked to optimise targets. (C) Targets were synthesised with a variety of structures sharing a carboxylic acid and basic termini. Compound activity was assessed in phosphorylation of Akt (WB) (D) and direct binding (SPR) (E) assays.

6. Design, synthesis and biological evaluation of the hydantoin series

Compounds in this chapter were designed based on the screening hit **3.6.6**, containing a hydantoin, a potential bioisostere for the carboxylic acid used to interact with the binding site metal ion in literature compounds and chapter 5.[126]

6.1 Hydantoin structure and metal binding alternatives

Unique among the ROCS screening hits was compound **3.6.6**, shown in Figure 39, which contains a hydantoin modality. Hydantoin is a 5 membered ring with an $NHCONHCOCH_2$ structure, synthesised from the coupling of urea and glycolic acid.



3.6.5 **3.6.6**

Figure 39: Compounds identified using the virtual screening platform ROCS to match the shape and electrostatic properties of the RGD domain in IGFBP-1. Compound **3.6.5** was the design source of the zwitterion-based compounds. Compound **3.6.6** was the design source of the hydantoin-based compounds.

Six marketed drugs incorporate the hydantoin group (Figure 40), with three (phenytoin (**6.1.1**), fosphenytoin (**6.1.2**) and ethotoin (**6.1.3**)) targeting epilepsy as anti-convulsants, nilutamide (**6.1.4**) is used for prostate cancers as an androgen receptor antagonist, nitrofurantoin (**6.1.5**) is used as an antibiotic for treatment of UTIs in a selection of Gram positive and negative bacteria, and dantrolene (**6.1.6**) is a muscle relaxant used to treat hypothermia. Several of the compounds regulate ion channels, though none are described as binding to metal ions.[127-130]

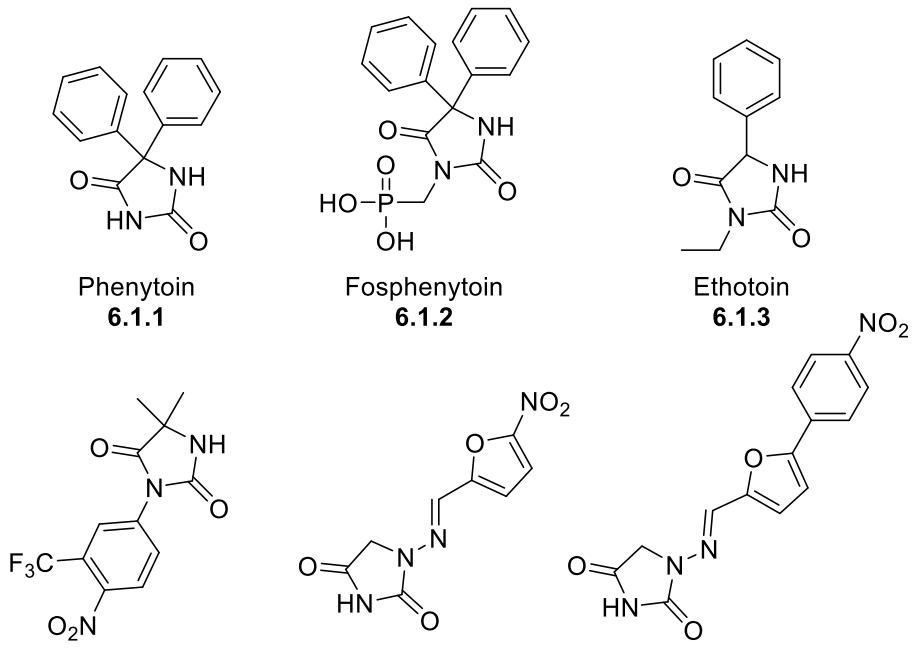


Figure 40: Marketed small molecule drugs incorporating the hydantoin functional group.

Hydantoin containing structures have been used for metal binding in literature, typically to zinc and are used as alternatives to carboxylic acids and hydroxamic acids, most commonly targeting metalloproteinases for the treatment of osteoarthritis (shown in Figure 41).[131-135] Metal binding of the hydantoin has been explained as between the deprotonated imide NH and between the adjacent OH in the enol tautomer.[136] The imide NH is established as the most acidic hydrogen on unsubstituted hydantoin by NMR analysis and DFT calculation.[137, 138] The pK_a of hydantoin and 1-methylhydantoin was measured to be 9.0 and 9.1 respectively by NMR.[139] Crystal structures all show the Zn interaction in plane with the hydantoin, centred on the imide NH (PDB: 4WK7, 6YJM, 3LEA, 4JPA, 4WKI) supporting the deprotonated imide NH binding. At physiological pH in solution, the hydantoin is primarily of neutral charge, however the affinity of metal binding stabilises the deprotonation to form the metal complex.[140, 141] This has been shown extensively for uracil (**6.1.7**), which shares the heteroatomic arrangement of hydantoin in a 6-membered ring.[142-144]

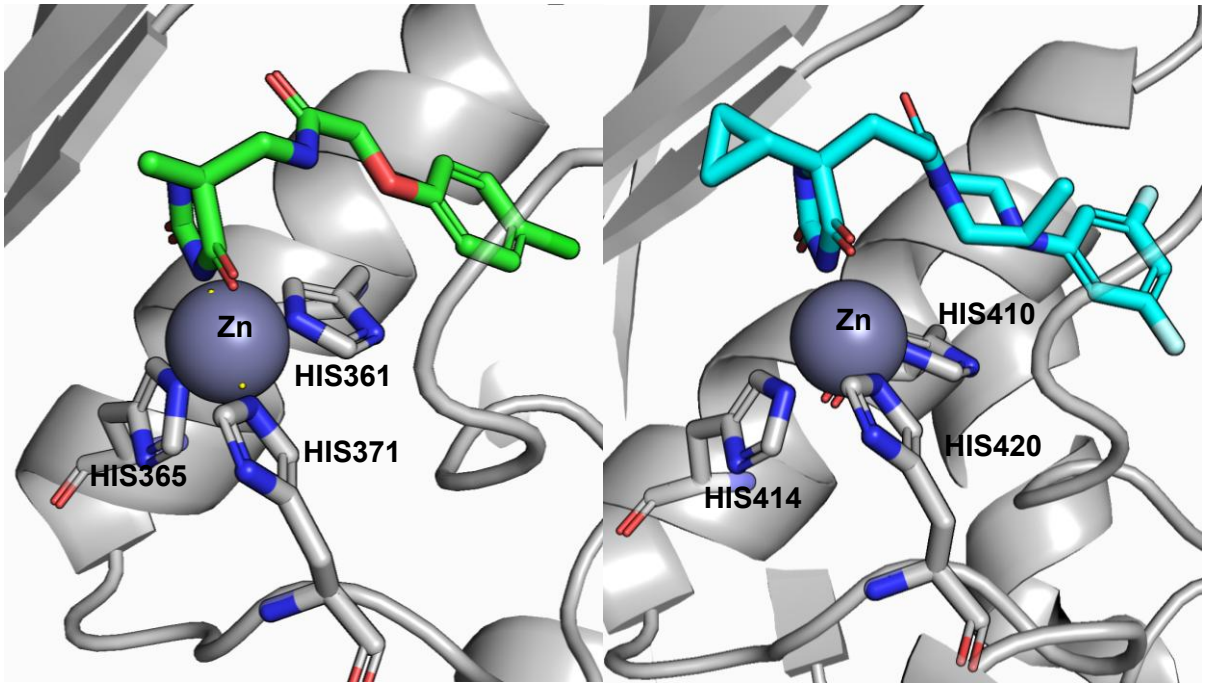


Figure 41: Metalloprotease PDB crystal structures of hydantoin containing small molecules interacting with zinc atoms (PDB: 4WK7 and 6YJM).

Several studies using DFT-based theoretical calculations have indicated that the most stable tautomer of hydantoin in water is the diketo structure (**6.1.8**).[145-147]

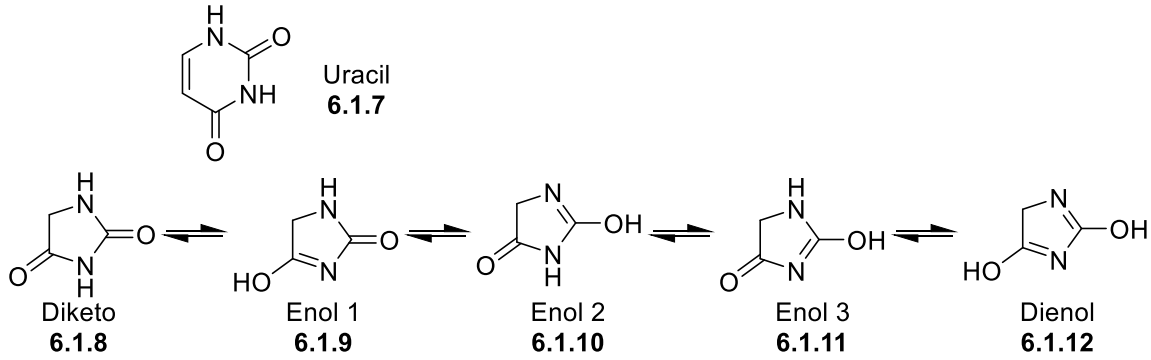


Figure 42: Uracil and the possible tautomers of the hydantoin group. The diketo structure (**6.1.8**) has been calculated to be the most stable.

6.2 Docking studies and initial design

The initial screening compound (**3.6.6**) was modelled to interact with TRP157 and LYS182, off to the side of the endogenous ligand binding pose, whilst retaining interaction with the Mg^{2+} ion (Figure 43). The ketone was substituted for an amide (**6.2.1**), which improved the docking score by 0.785 and the synthetic tractability. Replacement of the indane for the indanone (**6.2.2**), targeting interaction with LYS182, marginally

improved the docking score and was included in the initial targets, along with the indane (**6.2.1**).

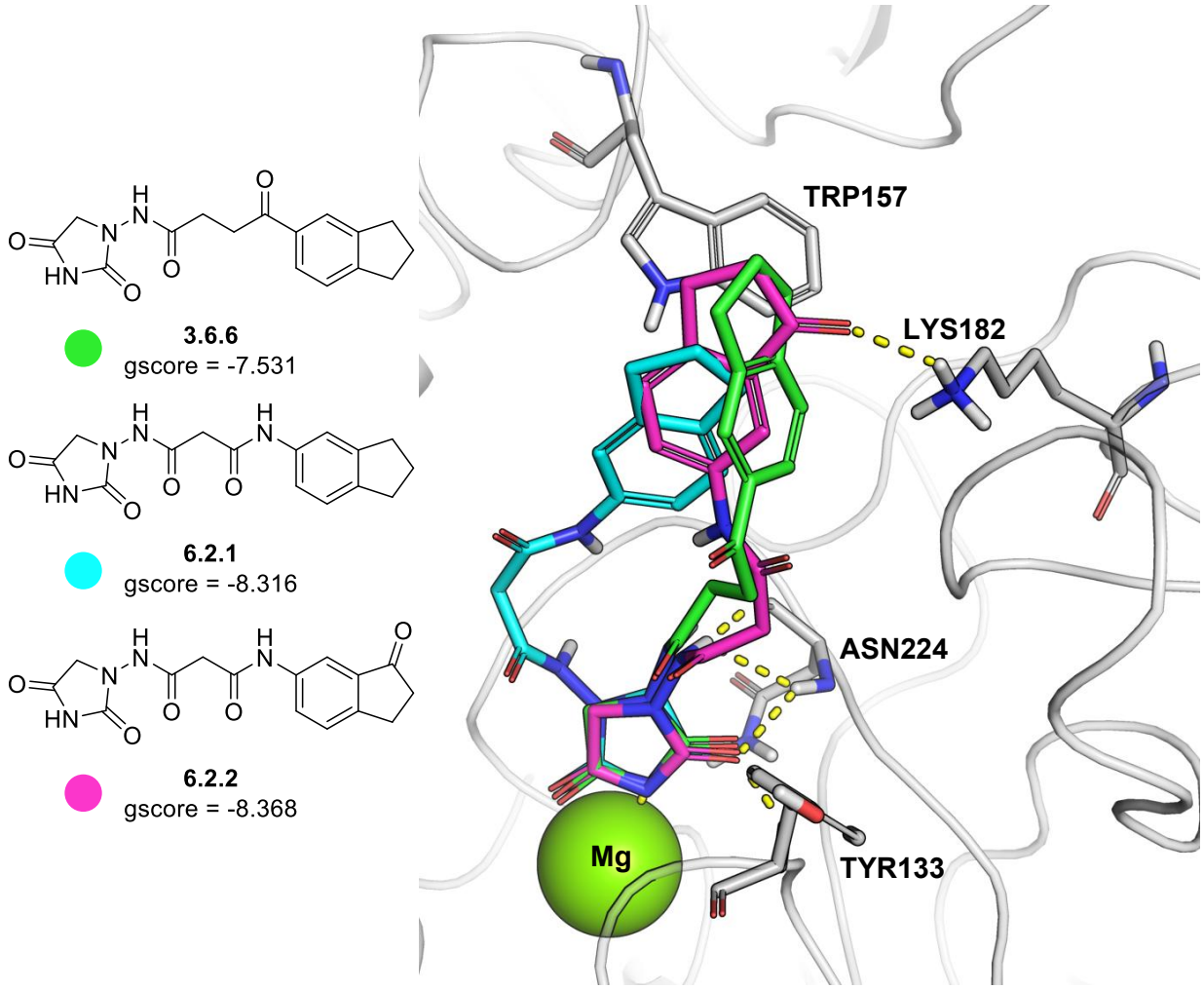


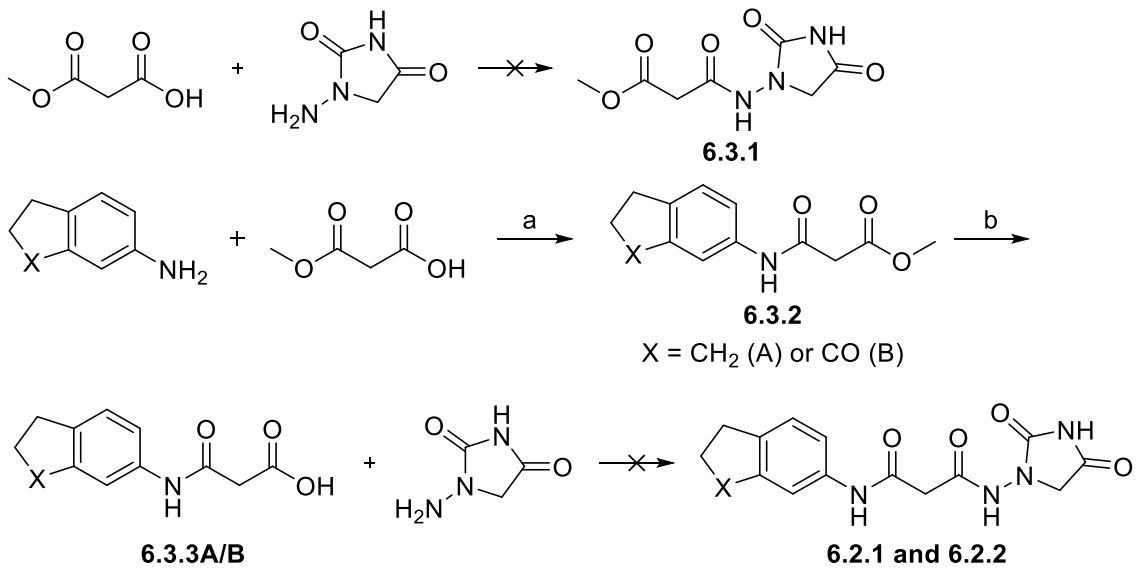
Figure 43: Molecular docking of the hydantoin containing hit compound (**3.6.6**) and two modifications to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 3VI4) with close residues and hydrogen bonds highlighted. The modifications were designed to enable a hydrogen bond to form with LYS182 and to simplify the target synthesis.

6.3 Synthesis of initial targets

Amide coupling of 1-aminohydantoin with 3-methoxy-3-oxopropanoic acid was attempted using EDC/DMAP and $POCl_3$ conditions, shown in Scheme 8. In each case, no trace of product (**6.3.1**) could be detected. These conditions are detailed in Table 3, along with the other conditions and reactants attempted for the amide coupling with 1-aminohydantoin.

Reversing the synthetic route was attempted, coupling 3-methoxy-3-oxopropanoic acid to 5-indanamine and 6-aminoindanone with propanephosphonic acid anhydride (T3p)

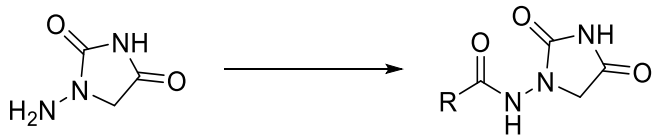
and NEt_3 to obtain compounds **6.3.2A** and **B** in 43% and 31% yield respectively. Hydrolysis of **6.3.2A** and **B** using ammonium hydroxide (35% solution) successfully obtained the acids, **6.3.3A** and **B** in quantitative yields. Amide coupling of the indane product **6.3.3B** was attempted with 1-aminohydantoin in various conditions, detailed in Table 3, but no product formation could be detected.

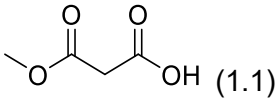
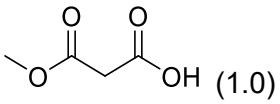
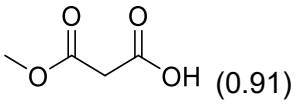
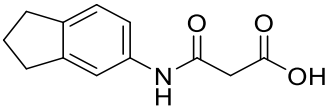
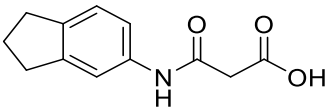
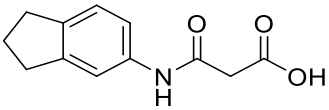
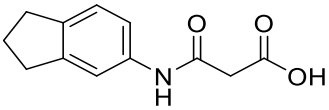
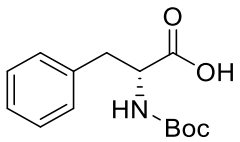
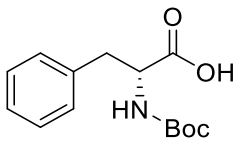
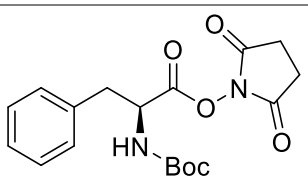


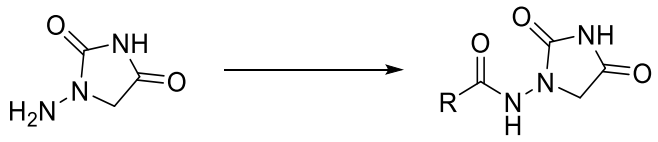
Scheme 8: Attempted synthesis of targets **6.2.1** and **6.2.2**. Reagents and conditions: (a) T3p (50% solution in ethyl acetate), NEt_3, ethyl acetate, rt, 16 h, 31-43%; (b) NH_4OH (35% aqueous solution), MeOH, rt, 60 h, 99%. Failed reaction conditions are detailed in Table 3.

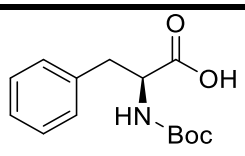
Amide coupling of 1-aminohydantoin with Boc-phenylalanine and Boc-phenylalanine-OSu were attempted, detailed in Table 3, with no significant product formation. A product mass trace was detected by LCMS when using a one-pot $SOCl_2$ and NEt_3 mixture described for unreactive amines by Leggio *et al.* however no material could be isolated by reverse-phase column chromatography.[148] The poor reactivity of 1-aminohydantoin was corroborated by another colleague working on the project and it remains unclear how the initial hit compound **3.6.6** was produced. Three publications have detailed amide bond formation with 1-aminohydantoin, using thionyl chloride and the coupling agent DMTMM, both of which were attempted, or by first derivatising the 3 position which is modelled to be critical to the metal ion interaction.[149-151]

Table 3: Attempted reagents, conditions and results for amide couplings with 1-aminohydantoin.

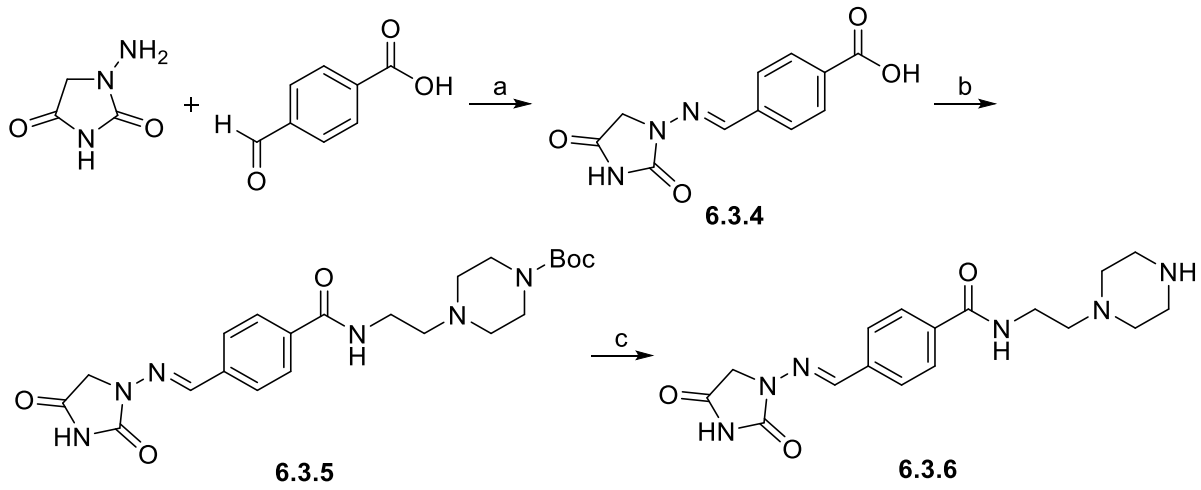


Reactant (eq.)	Reagents (eq.)	Conditions	Yield
 (1.1)	EDC (1.1), DMAP (1.5)	DCM, rt, 18h	N/A
 (1.0)	$POCl_3$ (2.0)	1,2-dichloroethane, 90°C, 2h	N/A
 (0.91)	DMTMM (2.0), N-methylmorpholine (1.2)	MeOH:H_2O (10:1), rt, 22h	N/A
 (1.0)	T3p (2.0), NEt_3 (3.0)	EtOAc, rt, 60h	N/A
 (1.0)	PyBop (1.1), NEt_3 (1.2)	DMF, rt, 22h	N/A
 (0.25)	PyBrop (0.275), DIPEA (1.0)	DMF, rt, 22h	N/A
 (0.91)	$SOCl_2$ (1.1)	Toluene, 100°C, 5h	N/A
 (1.0)	EDC (1.1), DMAP (1.5)	DCM, rt, 120h	N/A
 (1.0)	Oxalyl chloride (1.5), NEt_3 (2.0)	DCM, 60°C, 4h	N/A
 (0.91)	NEt_3 (2.0)	DMSO, rt, 16h	N/A



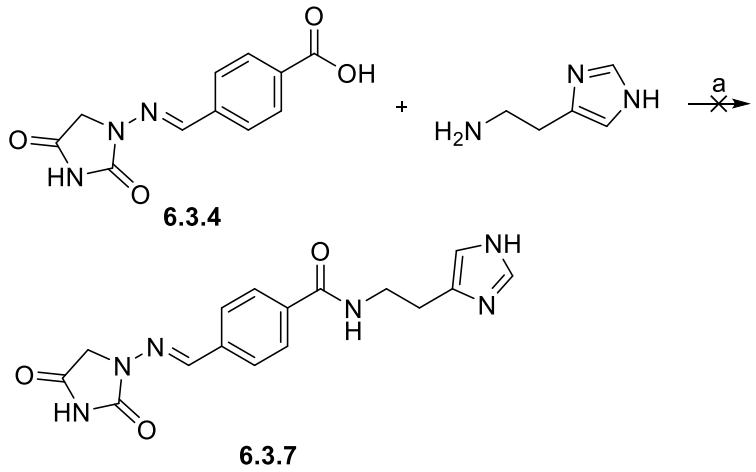
Reactant (eq.)	Reagents (eq.)	Conditions	Yield
 (1.0)	One pot: $SOCl_2$ (1.0), NEt_3 (5.0)	DMF, rt, 24h	Trace

There is much greater literature precedent for imine formation using 1-aminohydantoin. Imine formation between 1-aminohydantoin and 4-formylbenzoic acid using DIPEA, DMF and trimethyl orthoformate at 50°C, yielded a colourless precipitate of compound **6.3.4** in 75% yield as the DIPEA salt. Amide coupling of **6.3.4** with 4-N-(2-aminoethyl)-1-N-Boc-piperazine using $SOCl_2$ in DCM at 60°C obtained compound **6.3.5** in 85% yield. The piperazine was selected as a simple amine terminus from available reagents, expected to have favourable interactions at the basic site. Removal of the N-Boc group of **6.3.5** using 4N HCl in dioxane obtained compound **6.3.6** in 51% yield.



Scheme 9: Synthesis of compounds **6.3.5** and **6.3.6**. Reagents and conditions: (a) NEt_3, DMF, Trimethyl orthoformate, 0-50°C, 16 h, 75%; (b)(i) $SOCl_2$, DCM, 60°C, 18 h; (ii) 4-N-(2-aminoethyl)-1-N-Boc-piperazine, 1,4-dioxane, rt, 2 h, 85%; (c) 4N HCl in dioxane, rt, 18 h, 51%.

Amide coupling of intermediate **6.3.4** was also attempted with histamine, using $SOCl_2$ in DCM at 60°C, before evaporation and addition of histamine in dioxane. A product trace was detected by LCMS but could not be isolated by reverse-phase column chromatography, with significant starting material unreacted.



Scheme 10: Incomplete synthesis of compound **6.3.7** due to remaining impurities after purification attempts.

All the hydantoin compounds show a distinct imine CH peak in ^{1}H NMR and ^{13}C NMR at 7.8 ppm and 144-141 ppm respectively for aromatic linked compounds and 7.0 ppm and 150-148 ppm for aliphatic compounds, such as **6.5.5**, shown in Figure 44. This indicates a single isomer is obtained from the imine formation. The E-isomer assignment agrees with imine formation in literature, but was confirmed by NOESY NMR experiments of the common intermediate compounds **6.3.4** and **6.5.5**.[152]

NOESY shows strong cross peaks between the hydantoin CH_2 and the imine CH for both intermediates (appendix 10.2) although for the aromatic intermediate **6.3.4**, the cross peaks unexpectedly match the phase of the diagonal peaks. No cross peaks are observed between the imine CH and ring hydrogens which would correspond to the Z-isomer.

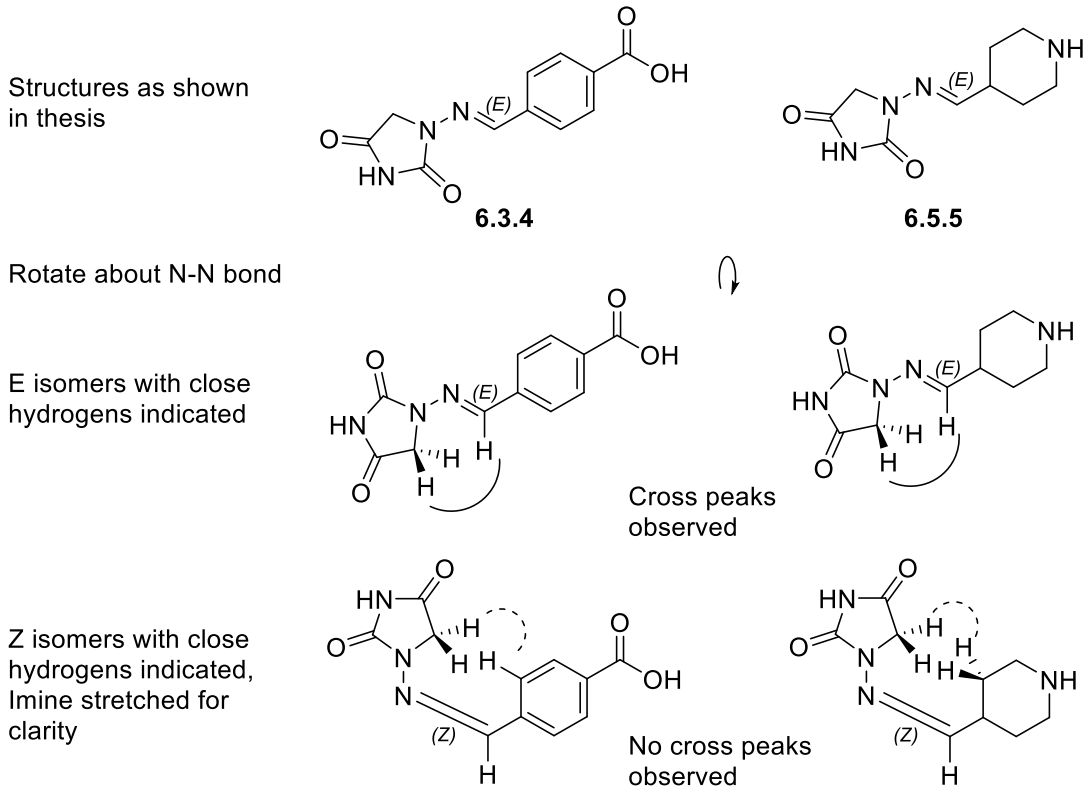


Figure 44: NOESY analysis of the hydantoin key intermediates **6.3.4** and **6.5.5** to show that both are the E-isomers. NOESY spectra are shown in appendix 10.2.

NMR spectra of the synthesised hydantoin compounds show a sharp CH_2 signal at 4.37-4.14 ppm (^1H) and 51-46 ppm (^{13}C) supporting the diketo structure.

Compounds **6.3.5** and **6.3.6** were submitted for testing in the western blot protocol with the compounds in the zwitterion series

6.4 Preliminary biological evaluation of initial hydantoin targets

Hydantoin compounds **6.3.5** and **6.3.6** were submitted for the phosphorylation of Akt assay at the same time as the zwitterionic compounds (Figure 45). Both compounds had a good response in the experiment using 100 nM insulin, finding pAkt/Akt levels above the GRGDSP control, with compound **6.3.6** showing the highest result in the assay.

In the experiment using 10 nM insulin, a similar trend of responses was observed, although the insulin and GRGDSP controls had responses stronger than both the compounds. Despite this, compound **6.3.6** again had the highest response of the compounds, with the Boc protected intermediate **6.3.5** showing a lower response.

The compounds showed a good response in the live/dead assay, with cell death below the DMSO control (Figure 54). In the scratch-wound assay, although there was

significant variation between replicates, compound **6.3.6** had the best response of the compounds, showing similar cell migration compared to the binding control GRGDSP and the literature agonist **GM263**. Compound **6.3.5** also had a good response, with similar migration to the GRGDSP control. As discussed in section 5.7, these results do not conclusively show that the compounds are agonistic or potent, but compound **6.3.6** had shown the best results across the assays of any compound in the project so far. For this reason, further development of the hydantoin series was prioritised.

These compounds were later tested in the SPR assay, with the results shown in section 6.9 with the other compounds in the series.

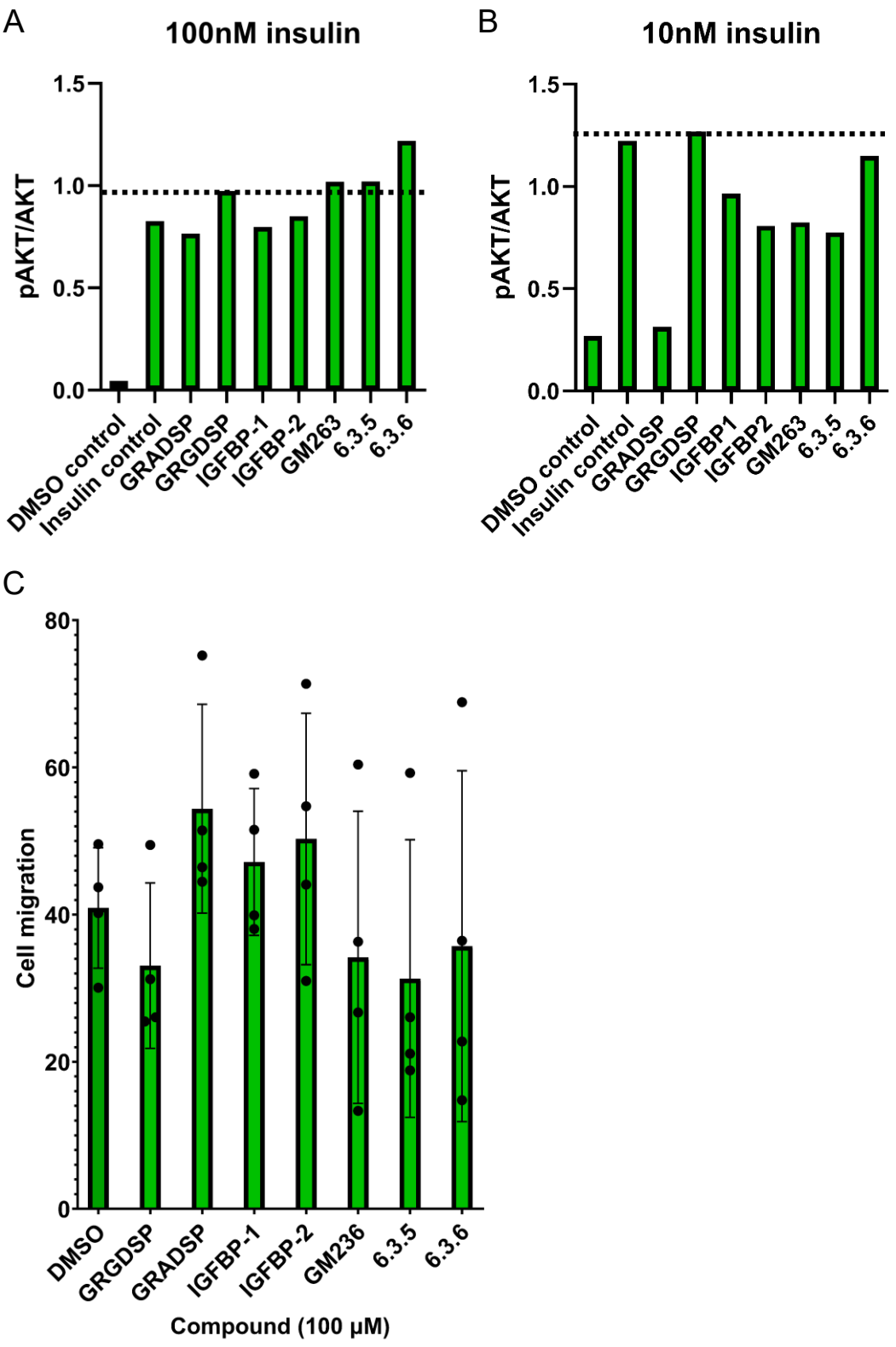


Figure 45: Western blot analysis of phosphorylation of Akt by compounds **6.3.5** and **6.3.6** and controls with C2C12 mouse myotube cells, at 100 (A) and 10 (B) nM concentrations of insulin in each well. Controls and compounds were dosed at 100 nM and incubated for 40 mins before exposure to insulin. (n=1); Akt: protein kinase B. (C) Cell migration of HUVEC cells after a scratch-wound while exposed to a selection of proteins, controls and compounds **6.3.5** and **6.3.6**. (n=4, mean ± SEM).

6.5 Docking studies and analogue design principles

A multi-faceted approach was taken when designing analogues of the hydantoin molecules. Firstly (Figure 46, A), five simple changes to compound **6.3.6** were designed to identify the key binding locations of the compounds which could then be optimised. Then (Figure 46, B), four simple amines were identified that would explore other amine termini. Next (Figure 46, C), two short diamines were selected which could be transformed to explore late-stage derivatisation (ureas, amidines, etc.). Finally (Figure 46, D), analogues with the aminopyridine functionality seen in literature compounds (**5.1.1** and **5.1.2**) were targeted as a guanidine mimic.

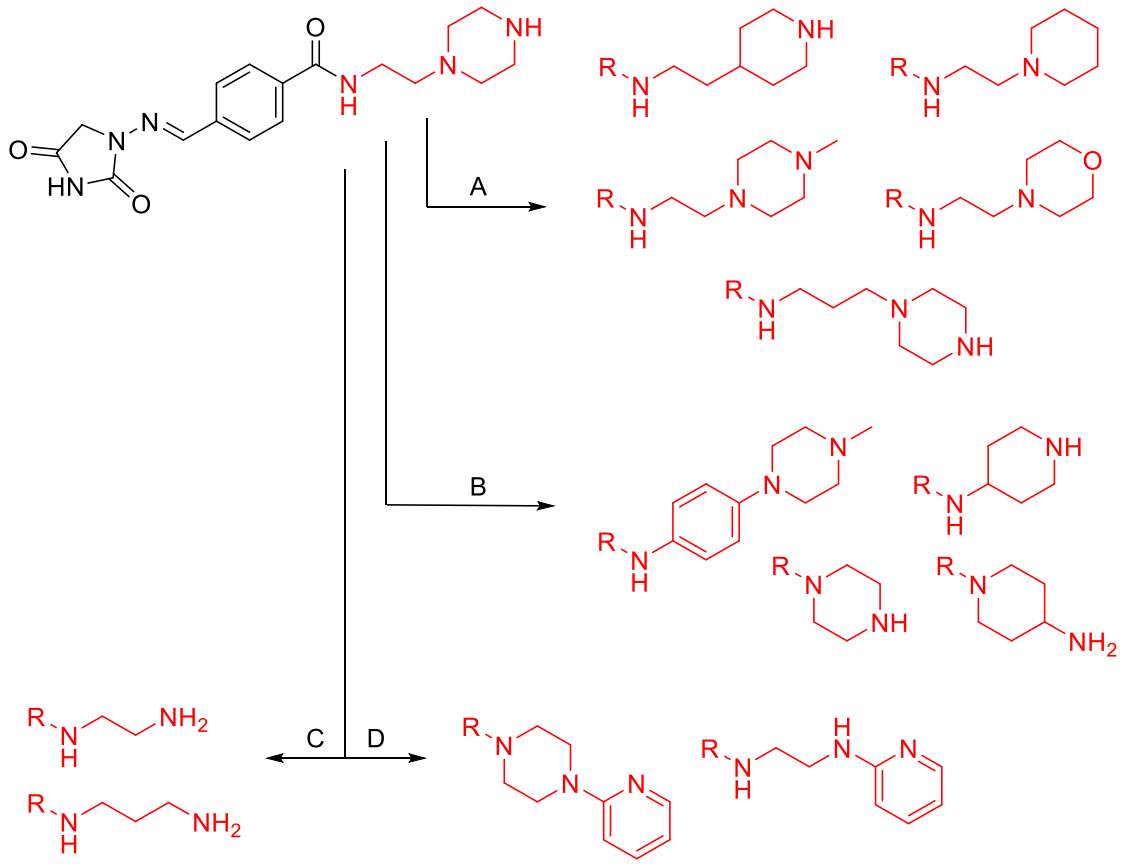


Figure 46: Compound designs from compound **6.3.6**, separated by modification type. (A) Simple modifications to compound **6.3.6** to interrogate the binding mode. (B) Related amine termini in the group inventory for rapid production of analogues. (C) Diamines, designed as intermediates for derivatisation to functional groups such as urea and amidine. (D) Analogues with an aminopyridine terminus, as seen in literature compounds.

As the docking of the zwitterions had not been shown to accurately predict compound activity, the hydantoin targets were not designed from the computational modelling. Instead, the targets were designed first, then docked to examine the expected interactions.

Docking of compound **6.3.6** and the simple alterations, compounds **6.5.1** and **6.5.2**, (Figure 47) favoured the same interactions that had been predicted in the initial hydantoin docking (Figure 43), forming hydrogen bonds with LYS182 and GLU126 and cation-pi interactions with TRP157.

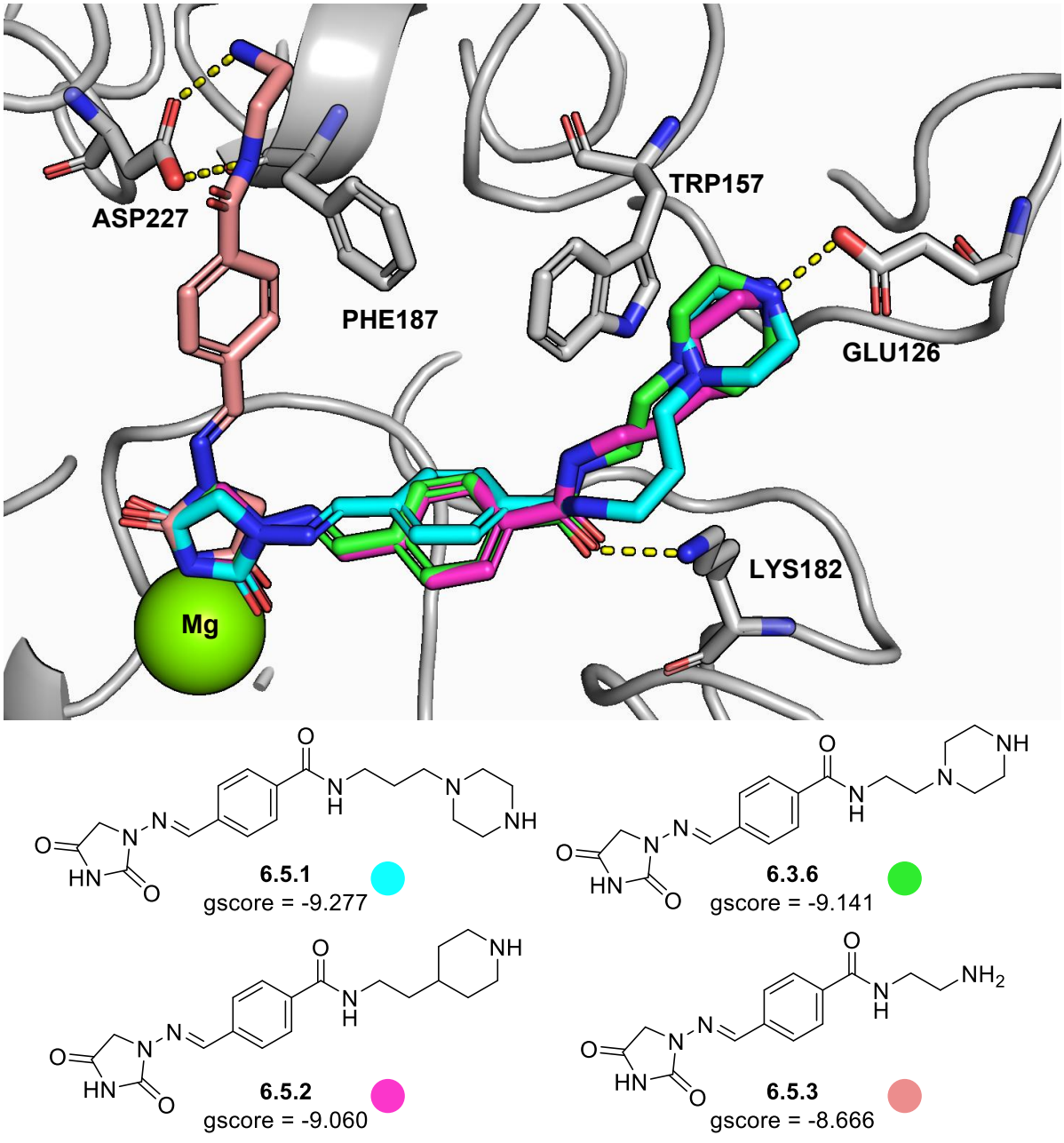


Figure 47: Molecular docking of the top scoring, phenyl linked hydantoin modifications to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK0) with close residues and hydrogen bonds highlighted. All but compound **6.5.3** form interactions with GLU126 and LYS182, as the original hit compound was modelled to, whereas compound **6.5.3** forms a bidentate hydrogen bond interaction with ASP227, similar to the endogenous RGD ligand.

Compound **6.5.3** however, designed with late-stage functionalisation in mind, was modelled to form hydrogen bonds with ASP227 and an edge-face pi-pi interaction with PHE187. Despite the docking score being slightly lower than the other hydantoin ligands

in the modelling, matching the interactions of the endogenous ligand was hypothesised to be the best approach of improving the agonistic activity of the compounds.

A new grid was generated that had constraints to form an interaction with the metal ion and a at least one of ASP227 and GLN221. This was designed to force the ligands to interact to the same key residues as the RGD peptide. The cilengitide co-crystal (4WK4) was used as the environment for this grid. The results are shown in Figure 48.

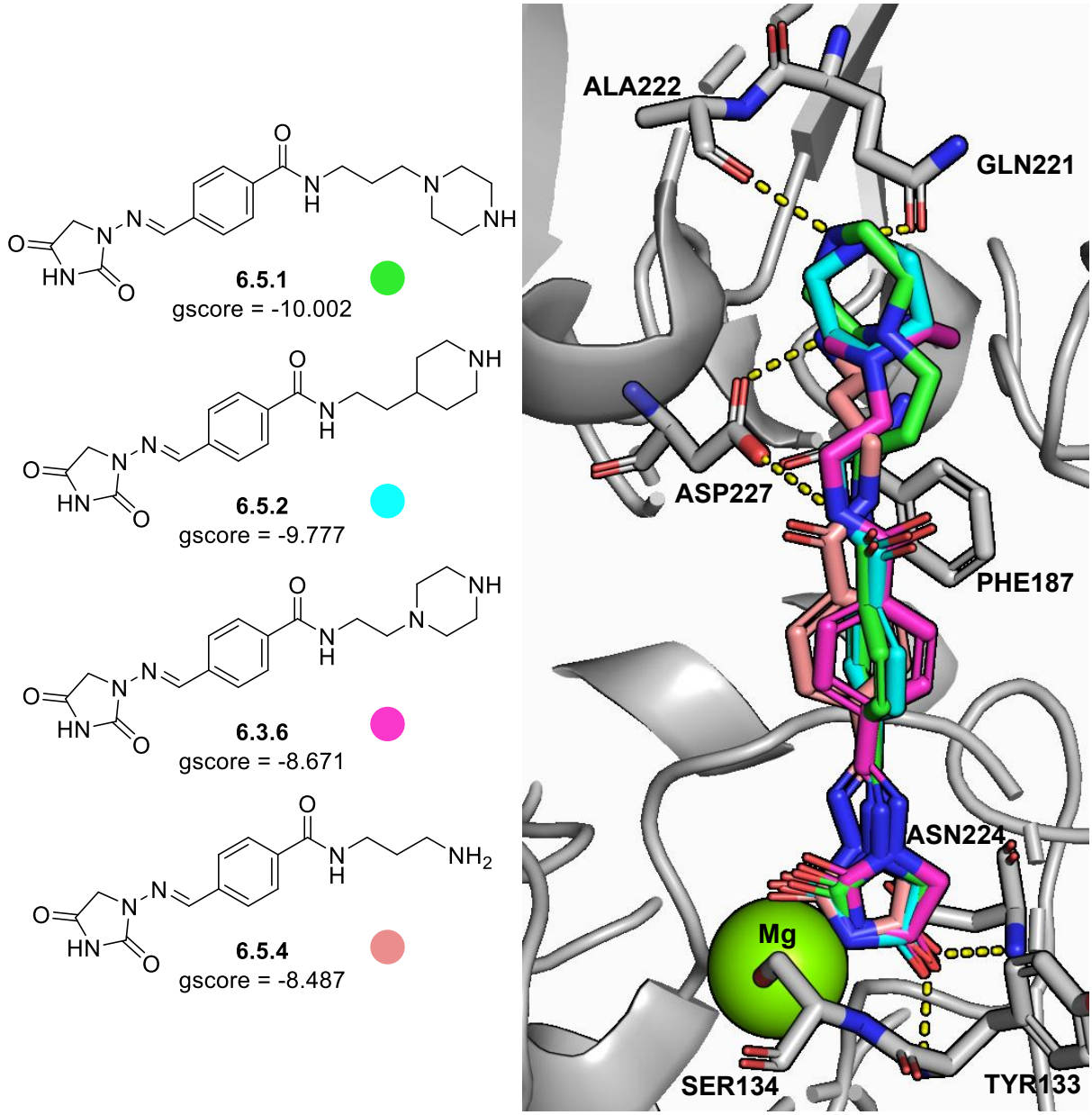


Figure 48: Constrained molecular docking of the top scoring, phenyl linked hydantoin modifications to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK4) with close residues and hydrogen bonds highlighted. Key hydrogen bonds are made with TYR133, ASN224, ASP227, ALA222 and GLN221. The top two scoring structures reach a better score than the unconstrained docking, indicating that the constrained model may better represent the best possible compound interactions.

Compound **6.5.1** had scored the best of the phenyl analogues, with an increase from -9.227 to -10.002 in the constrained docking, forming interactions with GLN221 and ALA222. Similarly, compound **6.5.2** also had an increase in docking score with the constrained run, from -9.060 to -9.777, forming interactions with the same residues.

The docking score of compound **6.3.6** fell from -9.141 to -8.671 under the new constraints. This appears to be because the CH_2 groups adjacent to the chain-linked nitrogen of the piperazine prevent interactions with both GLN221 and ALA222 without a steric clash with SER224 and PHE187. As a result, the top scoring pose avoids these clashes by forming a hydrogen bond exclusively with GLN221. The same steric penalties are encountered by the other compounds with piperidine and piperazine termini but appear to minimise the effect. This is clearest with compound **6.5.1**, which uses the additional carbon in the linking alkyl chain to point the piperazine into the pocket.

The best scoring simple amine was the diaminopropyl, compound **6.5.4**, with a slight reduction in the docking score from -8.666 to -8.487, which is likely due to the minor differences in the crystal structures.

The phenyl linker was also targeted for modification. This would enable examination of the role of the linking group in the binding affinity and allow for further diversity in synthesis. A piperidine was selected as the ring structures are similar and the growth site was a secondary amine, allowing for coupling to a greater variety of termini.

The design principle for the piperidine core molecules was to make simple derivatives of the hit molecule (Figure 49, A) and to investigate other guanidines as the optimised terminus for the endogenous ligand (Figure 49, B). Guanidine containing reagents were easily obtained with a carboxylic acid, so this design principle was much easier to explore with the piperidine intermediate.

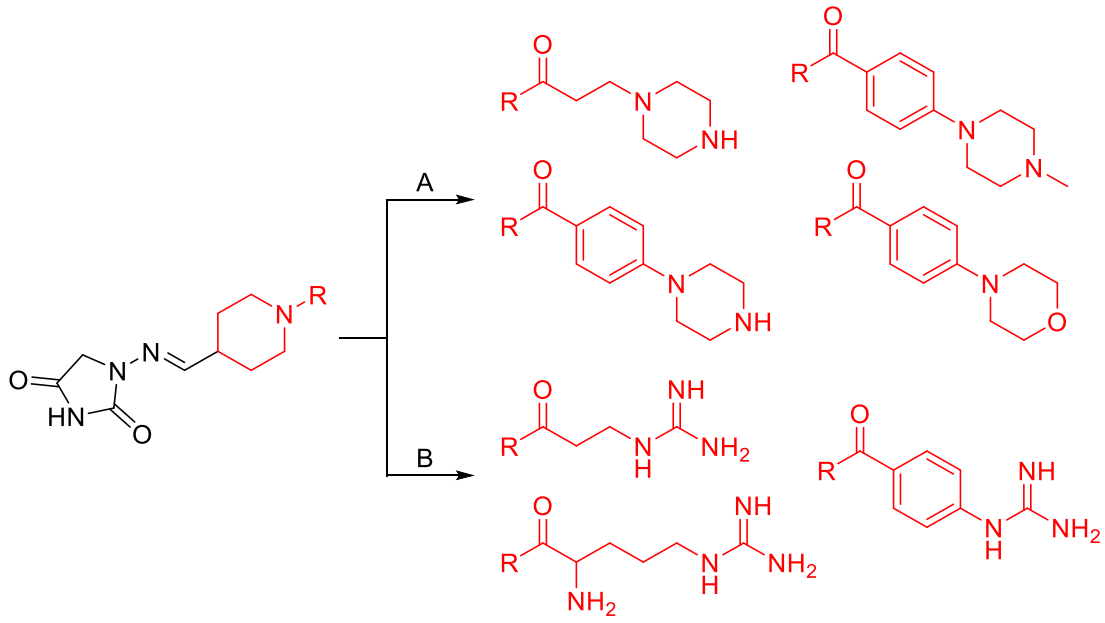


Figure 49: Derivatives of compound **6.3.6**, with the additional modification of a piperidine replacing the phenyl linker between the hydantoin imine and the basic terminus. (A) Modifications resembling compound **6.3.6**. (B) Modifications with a guanidine terminus, as is present on the RGD peptide.

These targets, and intermediates, were then docked in the unconstrained grid (Figure 50), finding that the top scoring compounds, compound **6.5.5** and **6.5.6**, -9.055 and -8.768 respectively, formed interactions with GLU126, TRP157 and LYS182. As seen for compound **6.5.3**, compounds **6.5.7** and **6.5.8** formed hydrogen bonds with ASP227 and ASP228, with the guanidine of compound **6.5.8** reaching further out to form a hydrogen bond with GLU320. The Boc protected amine form of the arginine target had a slightly higher docking score of -8.768 than the deprotected target, with a score of -8.601.

As before, the targets were docked in the constrained grid (Figure 51), with considerable differences in docking scores. The arginine derivatives formed similar interactions with ASP227 and GLU320, with the deprotected amine **6.5.9** returning a docking score of -11.167, while the Boc protected form, **6.5.7**, returned a score of -10.098. In this case, the S isomer of compound **6.5.8** scored the highest, though chirality of the targets was not a priority consideration in the design.

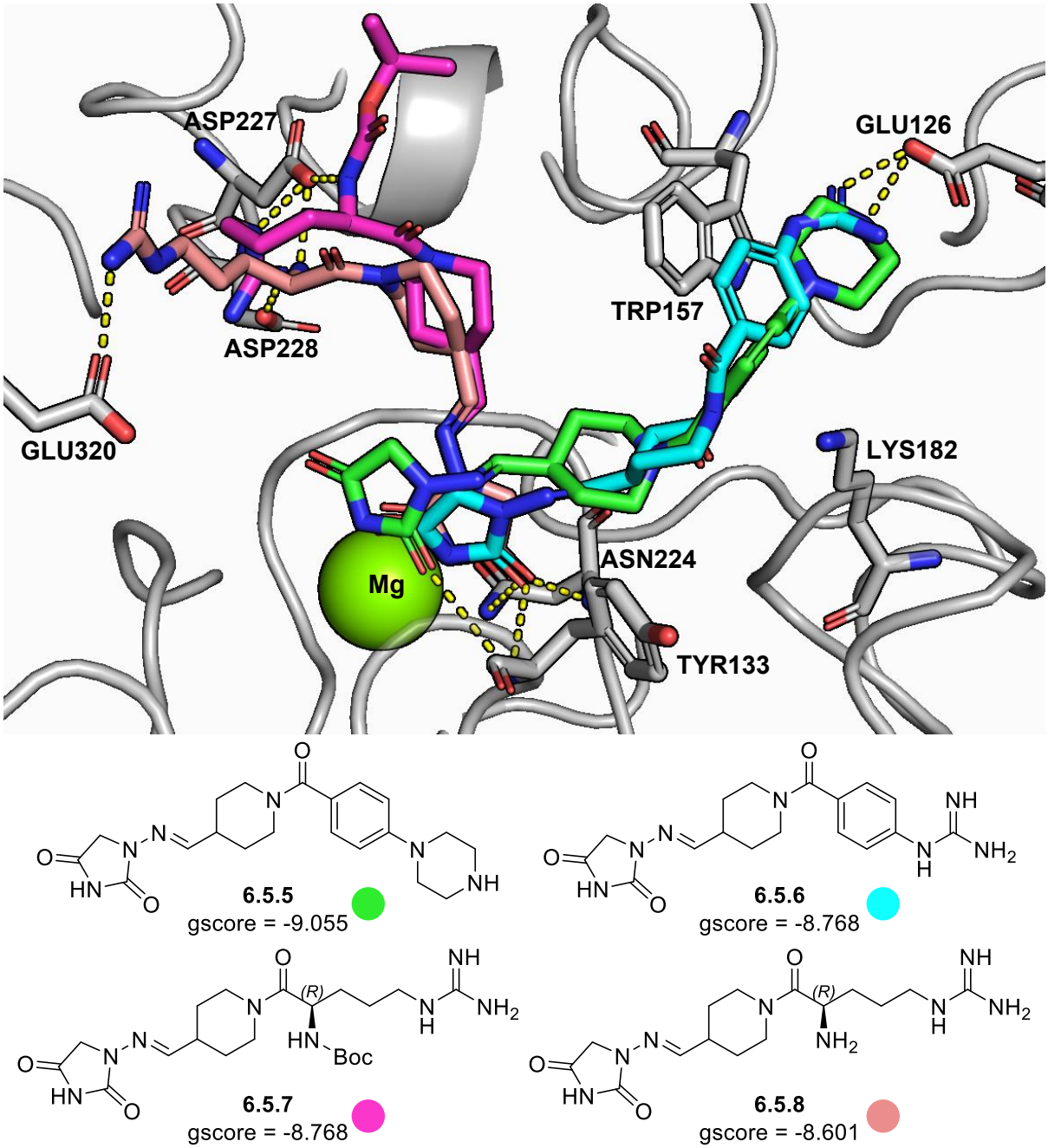
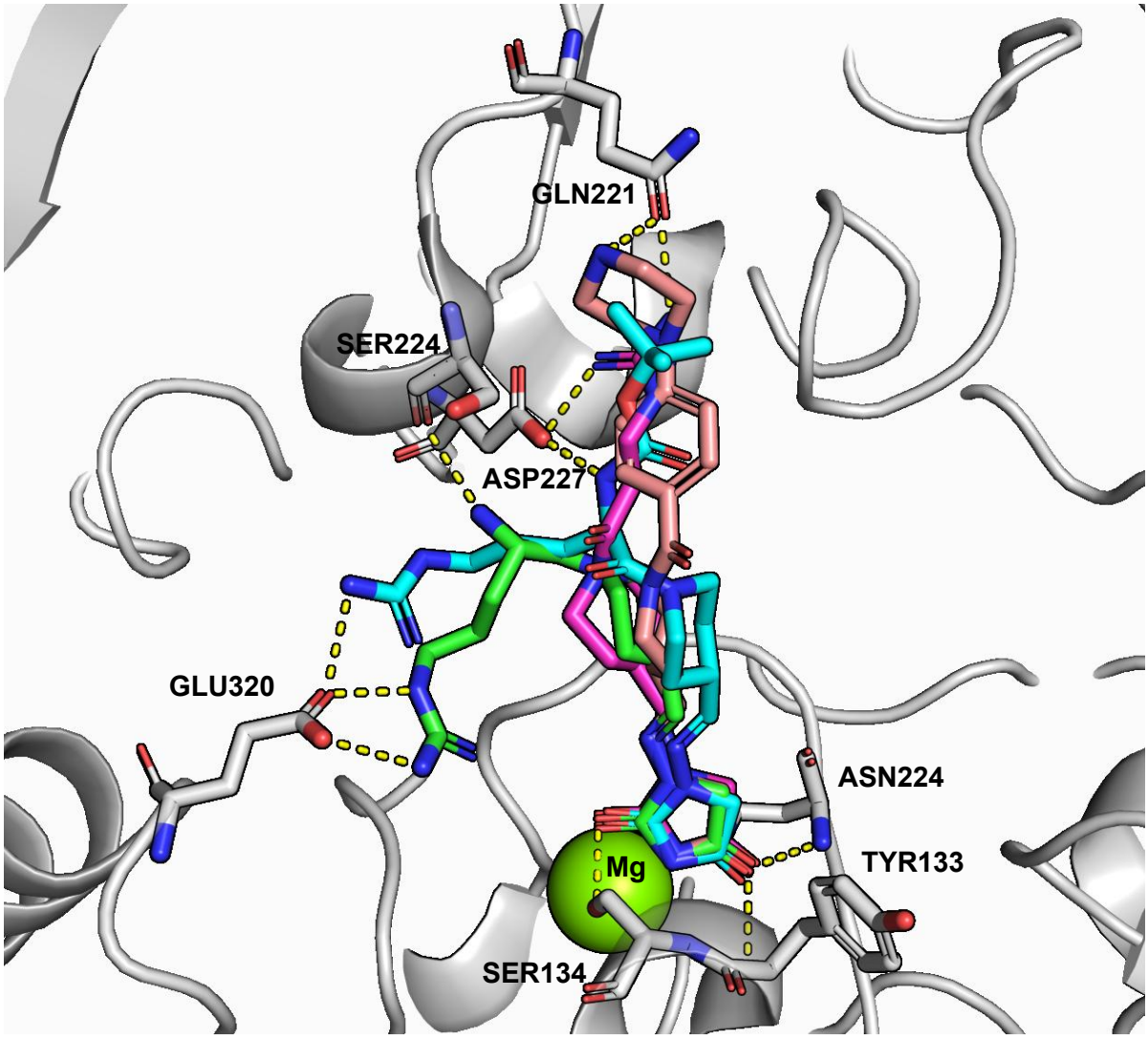


Figure 50: Molecular docking of the top scoring, piperidine linked hydantoin modifications to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK0) with close residues and hydrogen bonds highlighted. As with the phenyl linked modifications, the top scoring structures, **6.5.5** and **6.5.6**, form interactions with GLU126, while other structures, **6.5.7** and **6.5.8,** form interactions with ASP227, ASP228 and GLU320.

The guanidine containing **6.5.10** was the next highest scoring compound with a score of -8.609, far below the scores of the arginine derivatives. This is closely followed by the phenyl piperazine derivative **6.5.5**, which had a docking score of -7.944. Overall, the docking scores were comparable to the phenyl core molecules.



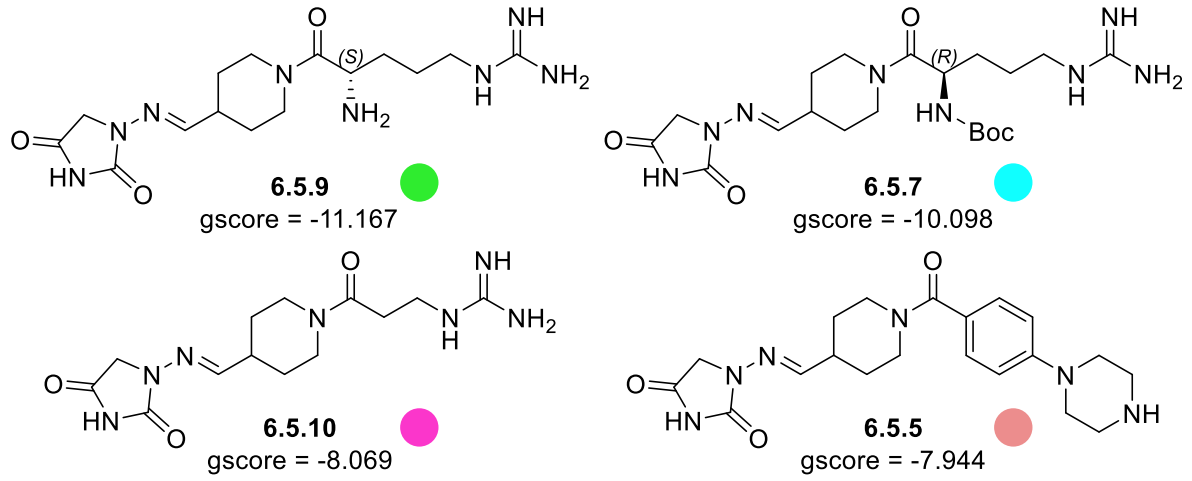


Figure 51: Constrained molecular docking of the top scoring, piperidine linked hydantoin modifications to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK4) with close residues and hydrogen bonds highlighted. The highest scoring structures, **6.5.9** and **6.5.7** had the highest scores seen in the molecular modelling in the project thus far, both opting to form hydrogen bonds between the terminal guanidine and GLU320. The next best scoring structures had lower scores, posed in a linear shape similar to the RGD peptide.

Building from the docking scores of the arginine derivatives, a phenylalanine based core was also investigated specifically with an arginine terminus. These compounds and their Boc protected intermediates were docked in both grids. In the unconstrained model (Figure 52), the middle of each molecule appears to bulge out of the pocket, allowing for interactions to form with ASP227, ASP228, SER227 and GLU320, while retaining the typical interactions at the metal binding site. The Boc deprotected **6.5.11** scored the highest with -10.634, with the Boc protected **6.5.12** returning a score of -9.610.

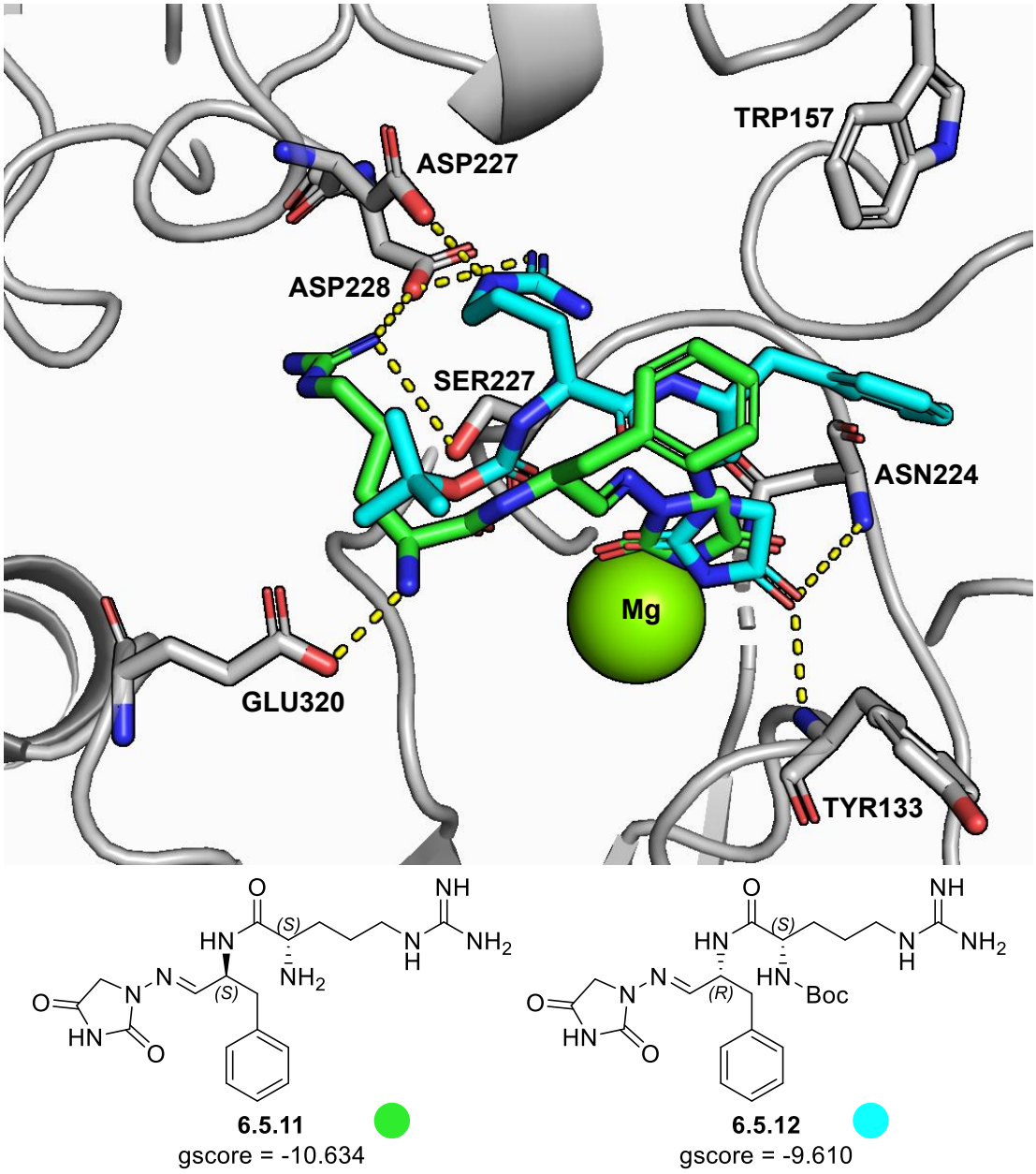
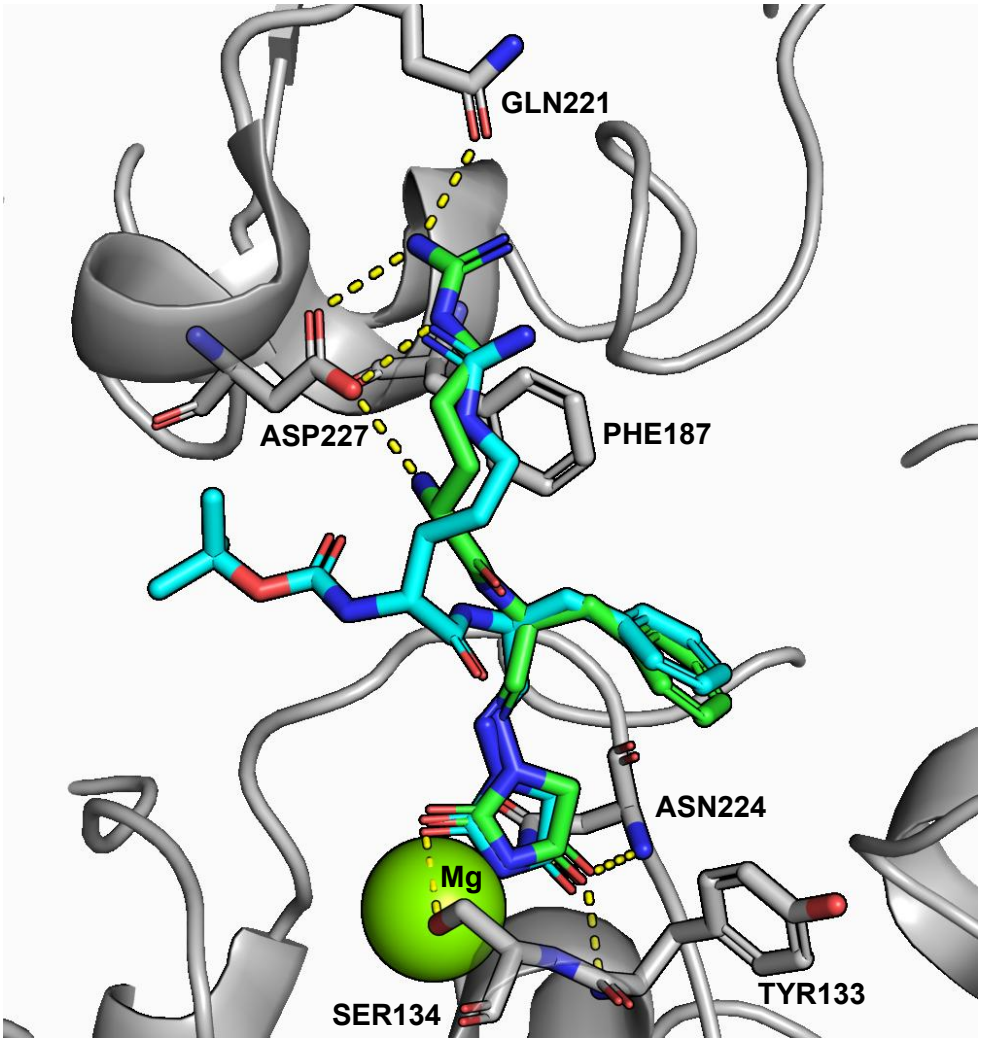


Figure 52: Molecular docking of the top scoring, phenylalanine linked hydantoin modifications to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK0) with close residues and hydrogen bonds highlighted. Both structures achieved high scores by adopting a contorted pose to form interactions with ASP227, ASP228 and GLU320.

For the constrained docking into the 4WK4 crystal (Figure 53), both compounds adopted the pose of the endogenous ligand, although only the deprotected amine **6.5.13** reached up fully to interact with GLN221 and ASP227 with a score of -9.619. The Boc protected **6.5.14**, forming interactions with ASP227, returned a score of -9.123. The phenyl side chain of the linker could fulfil a similar role to the acid adjacent aromatic groups seen in literature (Figure 12), with the group appearing to fit in a similar place in the binding site.



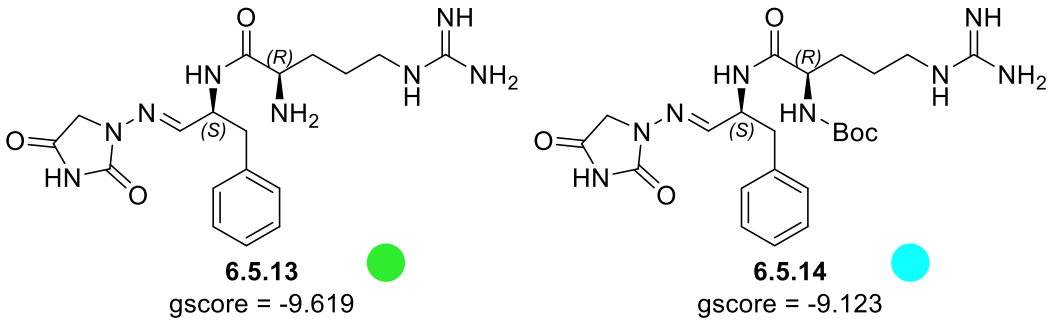


Figure 53: Constrained molecular docking of the top scoring, phenylalanine linked hydantoin modifications to the RGD-binding domain of the $\alpha_5\beta_1$ integrin (PDB: 4WK4) with close residues and hydrogen bonds highlighted. The docking scores were reduced compared to the

unconstrained docking, although the binding pose matches the crystal structure of the RGD peptide, forming interactions with GLN221 and ASP227 with the guanidine group.

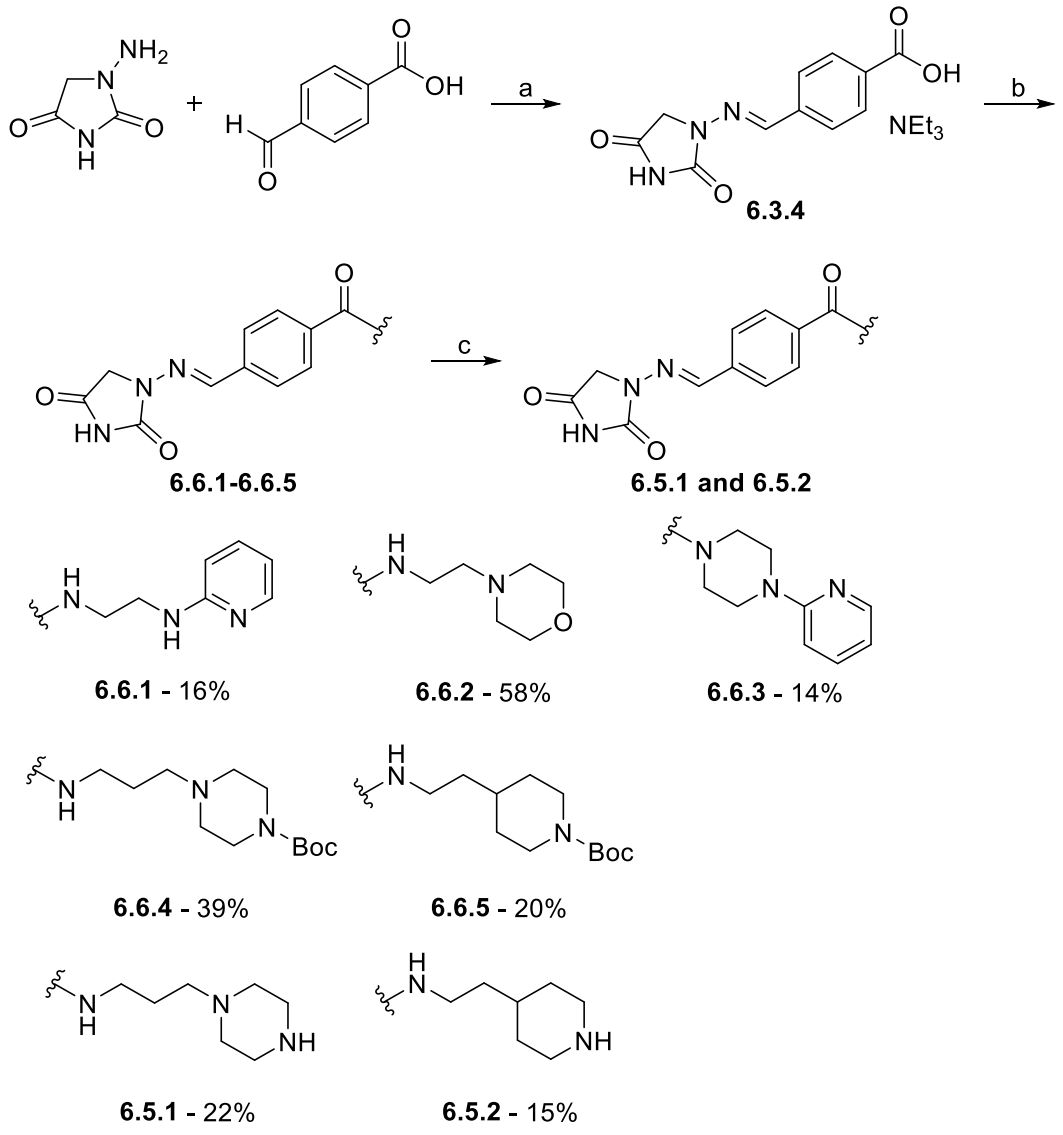
As the phenylalanine linked molecules scored well, an example of these was also prioritised for synthesis.

Overall, the constrained grid settings with the 4WK4 protein structure produced docking poses that more closely matched the endogenous ligand and achieved higher scores in the phenyl and piperidine linked structures. The top scoring compounds from each of the three docking sets were prioritised for synthesis, with a wide range of structures also prioritised to explore the chemical space.

6.6 Synthesis of phenyl core molecules

The synthesis of intermediate **6.3.4** was repeated, forming an imine between 1-aminohydatoin and 4-formylbenzoic acid using NEt_3 and trimethyl orthoformate to obtain the triethylammonium salt in 10 mmol scale. A 1:1.4 ratio of product to NEt_3 was measured by NMR which was adjusted for when measuring reagents.

Amide coupling of intermediate **6.3.4** with N-(2-pyridinyl)-1,2-ethanediamine was attempted to obtain compound **6.6.1**, but was unsuccessful using the $SOCl_2$ conditions used for compound **6.3.5**. Alternate conditions of EDC and DMAP in DCM were used to amide couple intermediate **6.3.4** and 2-(4-morpholinyl)ethanamine, successfully obtaining compound **6.6.2** in a 58% yield. These conditions were repeated for amide coupling of intermediate **6.3.4** and N-(2-pyridinyl)-1,2-ethanediamine, obtaining compound **6.6.1** with a 16% yield after two rounds reverse-phase column chromatography.



Scheme 11: Synthesis of phenyl linked hydantoin targets. Reagents and conditions: (a) NEt_3, DMF, trimethyl orthoformate, 50°C, 16 h, 75%; (b) EDC, DMAP, DCM or DMF, rt-40°C, 16-60 h, 14-58%; (c) 4N HCl in dioxane, DMF, rt, 2-20 h, 15-22%.

As the solubility of intermediate **6.3.4** and products **6.6.1** and **6.6.2** in DCM was poor, so the conditions used during amide coupling of intermediate **6.3.4** and 1-(2-pyridinyl)piperazine were changed to EDC and DMAP in DMF to obtain compound **6.6.3**. Compound **6.6.3** required two rounds of reverse-phase column chromatography (Isolera then MDAP) to obtain the product with 14% yield and high purity. Amide coupling using EDC and DMAP in DMF was subsequently used for 1-Boc-4-(3-aminopropyl)piperazine and intermediate **6.3.4** to obtain compound **6.6.4** in 39% yield, then N-Boc deprotected using 4N HCl in dioxane to obtain compound **6.5.1** in 22% yield. The same steps were used for 4-(aminoethyl)-1-N-Boc-piperidine to obtain compound **6.6.5** in 20% yield, which was N-Boc deprotected to obtain compound **6.5.2** in 15% yield.

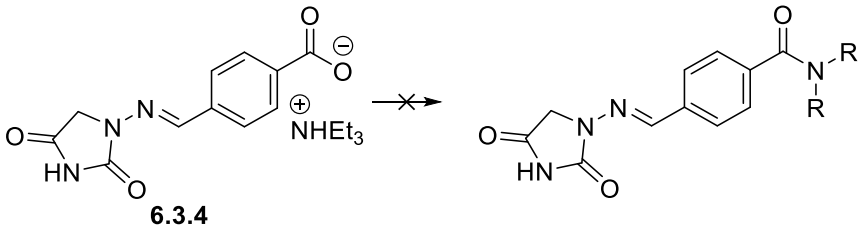
The low yields of these reactions and others in this series reflects the continued difficulty of purification. As the compounds are hydrophilic, they cannot be separated from the coupling agents by workup, so the reaction mixtures were loaded for purification using reverse-phase column chromatography, typically by dry-loading. As most of the reactions were not running to completion, the elution had many components that sometimes-required secondary purification. The N-Boc deprotections required additional purification for the high purity required for submission.

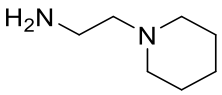
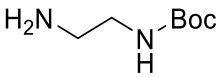
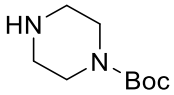
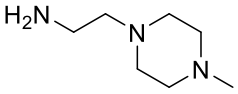
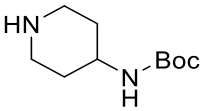
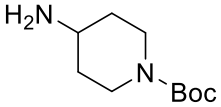
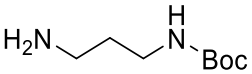
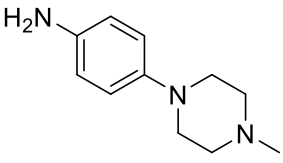
Eight selected targets were not successfully synthesised (Table 4), including with 2-(1-piperidinyl)ethanamine and 2-(4-methyl-1-piperazinyl)ethanamine where the product could be detected but not isolated. This may be due to significant unreacted starting material and further coupling agent optimisation could have pushed the reactions to completion.

Reaction with five amines, N-Boc-1,2-diaminoethane, N-Boc-piperazine, 4-(N-Boc-amino)piperidine, 1-Boc-4-piperidinamine and N-Boc-1,3-diaminopropane showed no indication of product formation, with some attempted in multiple conditions.

Finally, amide coupling of intermediate **6.3.4** and 4-(4-methylpiperazino)aniline showed significant product formation upon further addition of aniline and EDC after LCMS analysis indicated incomplete product formation, expected to be due to degradation of the aniline starting material. Upon purification using reverse-phase column chromatography, intermediate **6.3.4** eluted close to the product which could not be isolated. The reaction was not repeated due to changing project priorities.

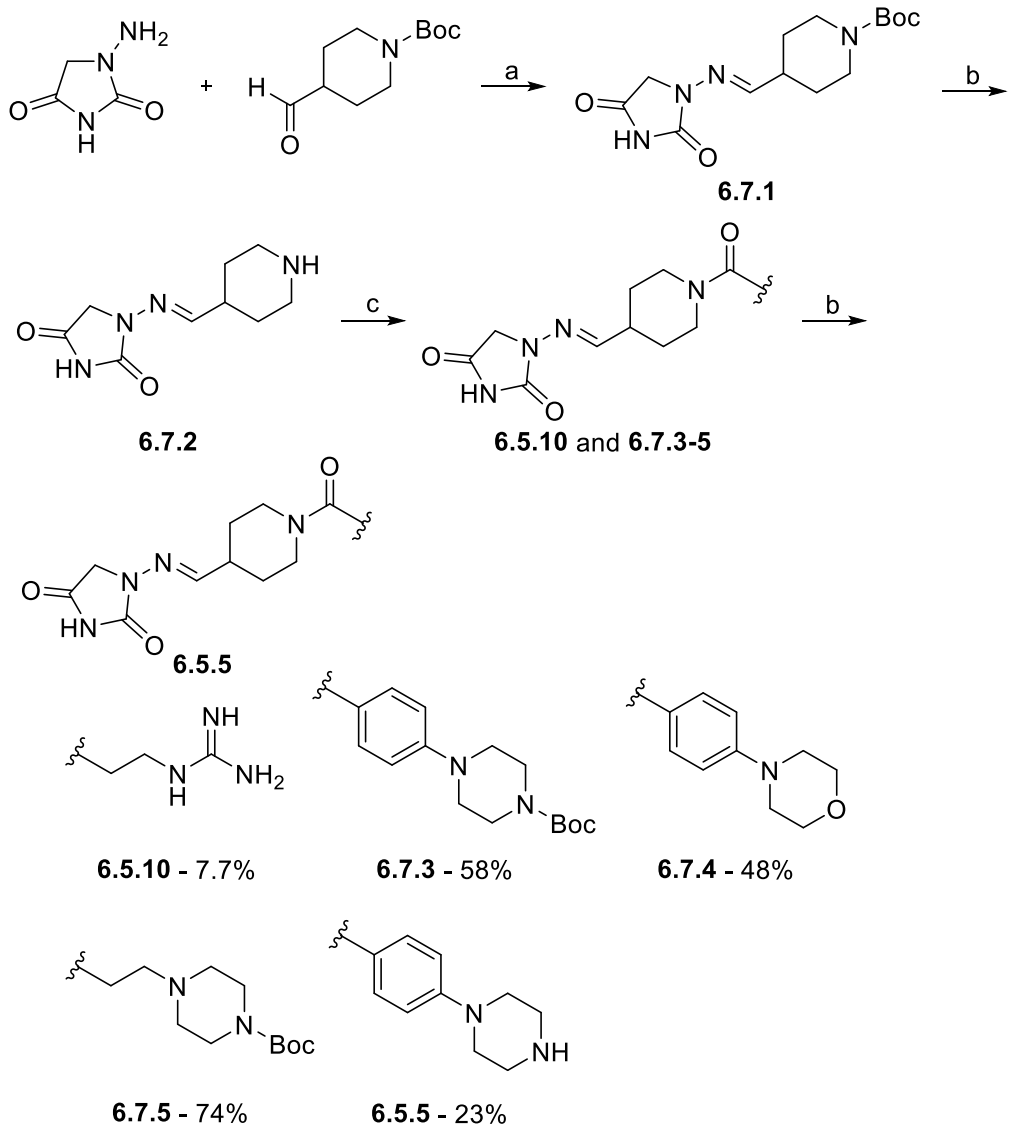
Table 4: Reagents, conditions and results of attempted amide couplings with intermediate **6.3.4**. Reagents: (A) EDC (1.1), DMAP (0.5); (B) PyBop (1.1), DIPEA (3.0). *Not repeated but likely would have been successful



Reactant (1.1 eq.)	Reagents	Conditions	Yield
	A	1.DCM, DMF, rt, 90h 2.DMF, rt, 24h	Product appeared to form but could not be isolated
	A (1,2), B (3)	1.DMF, rt, 40h 2.DMF, 30°C, 120h 3.DMF, 80°C, 40h	No product trace detected
	A	1.DMF, rt, 120h Then 80°C, 4h	No product trace detected
	A, B	1.DMF, rt, 160h Then 40°C, 24h 2.DMF, rt, 20h	B. Product mass detected but could not be isolated
	A	1.DMF, rt, 22h Then 2 eq. NEt_3, 50°C, 24h	No product trace detected
	A	1.DMF, rt, 22h Then 2 eq. NEt_3, 50°C, 24h	No product trace detected
	A	1.DMF, rt, 90h	No product trace detected
	A	1.DMF, rt, 24h Then 2 eq. amine, 1 eq. EDC, rt, 90h	Reaction pushed to completion but could not be isolated*

6.7 Synthesis of piperidine core molecules

The piperidine core intermediate **6.7.2** was synthesised by an imine formation between 1-aminohydantoin and 1-Boc-piperidine-4-carboxaldehyde using DIPEA in trimethyl orthoformate to obtain compound **6.7.1** in 25% yield. Removal of the N-Boc group was completed using 4N HCl in dioxane to obtain **6.7.2** in quantitative yield.

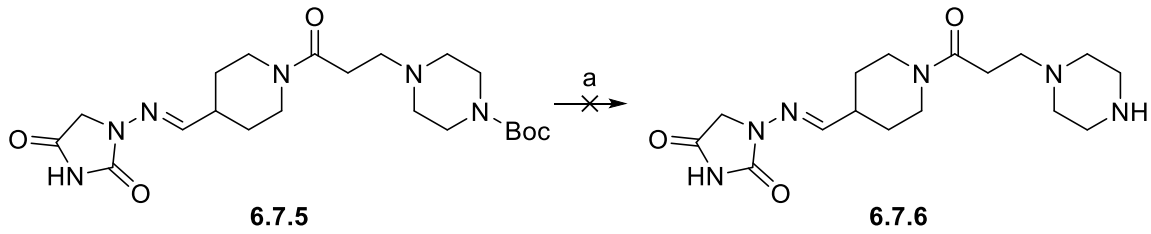


Scheme 12: Synthesis of the piperidine linked hydantoin targets. Reagents and conditions: (a) DIPEA, trimethyl orthoformate, 0°C-rt, 22 h, 25%; (b) 4N HCl in dioxane, DMF, rt, 24-168 h, 23-99%; (c) EDC, DMAP, DMF, rt, 16-60 h, 7-84%.

Amide coupling of intermediate **6.7.2** and 3-carbamimidamidopropanoic acid was completed using EDC and DMAP in DMF to obtain compound **6.5.10** in a 7.7% yield. The same conditions were used for amide coupling of intermediate **6.7.2** and 1-(4-Boc-piperazinyl)benzoic acid to obtain compound **6.7.3** in 58% yield. Analysis by NMR indicated the product was pure, however HPLC analysis showed an approximately 20%

impurity. Analysis using accurate LCMS could not identify the impurity. Removal of the N-Boc group of compound **6.7.3** using 4N HCl in dioxane was successful nonetheless to obtain compound **6.5.5** in 23% yield and 94% purity. As the deprotected product was completed and was expected to be of greater biological interest, repurification of compound **6.7.3** was not prioritised.

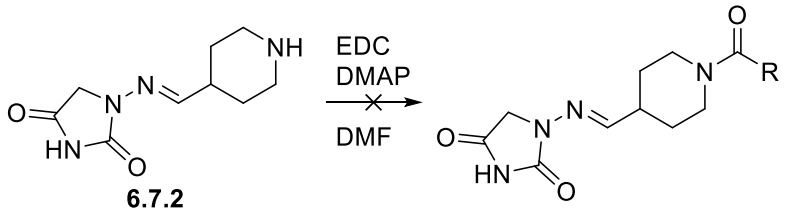
Amide coupling of intermediate **6.7.2** and 4-Morpholinobenzoic acid was completed with the same conditions to obtain compound **6.7.4** in 48% yield. Finally, amide coupling of intermediate **6.7.2** and 3-(1-Boc-piperazin-4-yl)-propionic acid using EDC and DMAP in DMF successfully obtained compound **6.7.5** in 74% yield. Removal of the N-Boc group to obtain compound **6.7.6** was attempted twice in 4N HCl in dioxane, with analysis by LCMS indicated reaction completion, but the product could not be isolated in sufficient purity by reverse-phase column chromatography (Scheme 13).



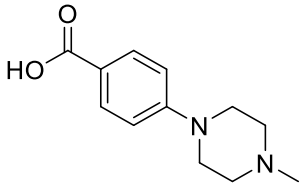
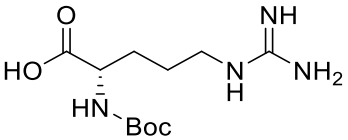
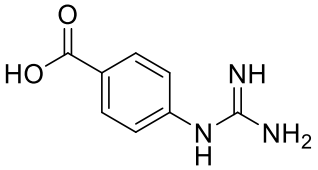
Scheme 13: Attempted deprotection of compound **6.7.5**. The product could not be isolated in a sufficient quantity for analysis.

Amide coupling with three further acids was attempted but were unsuccessful (Table 5). Reaction of intermediate **6.7.2** with 4-(4-methylpiperazin-1-yl)benzoic acid showed clear product formation by LCMS but co-elution of an unknown impurity in two reverse-phase columns meant that the product could not be isolated with sufficient purity. Amide coupling of intermediate **6.7.2** and Boc-L-arginine yielded a cream suspension, so the reaction was heated to 40°C to improve dissolution. The product mass trace was detected but analysis by NMR spectrometry could not clearly identify the product after purification using reverse-phase column chromatography or MDAP. Finally, amide coupling of intermediate **6.7.2** and 4-guanidinobenzoic acid showed clear product formation by LCMS analysis but the product collected from reverse-phase column chromatography contained significant unknown impurities. These reactions weren't repeated as the priority was to generate as many products as possible, rather than these specific targets.

Table 5: Reagents, conditions and results of attempted amide couplings with intermediate **6.7.2**.



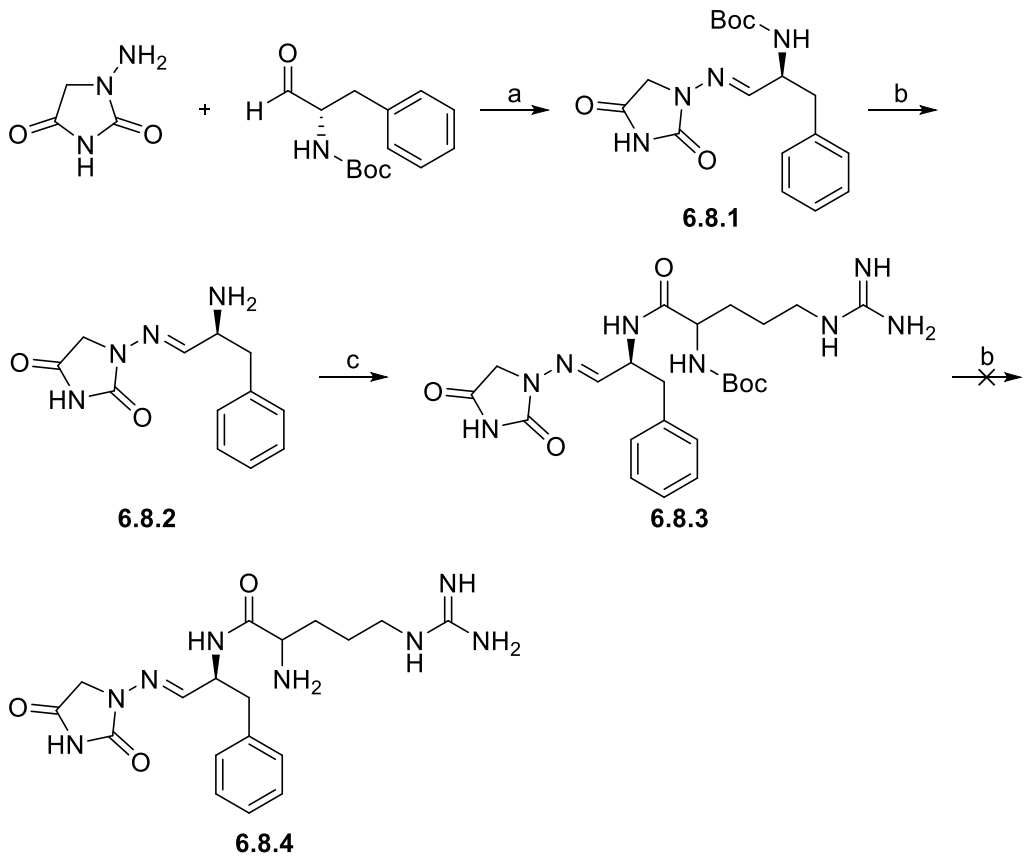
6.7.2

Reactant	Conditions	Yield
	rt, 16h	Clear product formation by LCMS. Co-elution during reverse-phase purification meant product was 76% pure. No improvement from second reverse-phase column.
	40°C, 22h	Product mass detected. Impure material collected from reverse-phase purification. Product still impure after MDAP.
	rt, 18h	Clear product formation by LCMS. Two peaks eluted during reverse-phase purification, both containing significant impurities.

6.8 Synthesis of phenylalanine core molecules

Imine formation between 1-aminohydantoin and N-Boc-L-phenylalaninal using NEt_3 in trimethyl orthoformate successfully provided intermediate **6.8.1** in 65% yield. Removal of the N-Boc group was completed using 4N HCl in dioxane and purified with two reverse-phase columns to obtain compound **6.8.2** in 55% yield. Amide coupling of intermediate **6.8.2** and Boc-L-arginine using EDC and DMAP in DMF successfully led to compound **6.8.3** in 42% yield. HPLC analysis showed two very close eluting peaks, indicating that the chiral centre of the arginine had racemised during the amide coupling. Analysis of **5.4.9** using a chiral HLPC column, coupled using the same conditions, revealed a small additional peak, indicating that some racemisation had occurred in previous reactions with these conditions, however not to the same extent as for **6.8.3** (appendix 10.3). As no racemic standard was used this analysis does not fully clarify if racemisation is occurring.

Removal of the N-Boc group of compound **6.8.3** was attempted using 4N HCl in dioxane. The product was detected in analysis by LCMS and NMR, however significant unknown impurities remained in the product and could not be removed by purification using reverse-phase column chromatography. It is suspected that the increasing polarity throughout the synthesis inhibited removal of the impurities and reduced performance of the reverse-phase column.



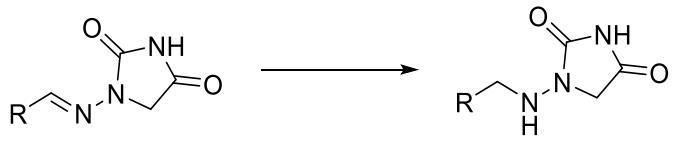
Scheme 14: Synthesis of compound **6.8.3** and attempted synthesis of compound **6.8.4**. Reagents and conditions: (a) NEt_3, DMF, trimethyl orthoformate, rt, 90 h, 65%; (b) 4N HCl in dioxane, MeOH, rt, 35 h, 55%; (c) EDC, DMAP, DMF, rt, 20 h, 42%.

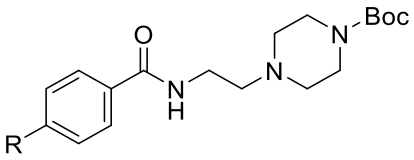
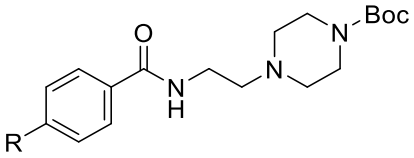
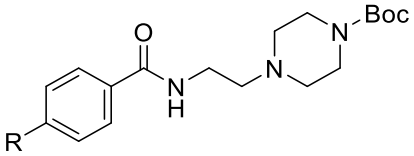
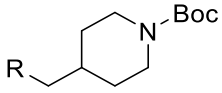
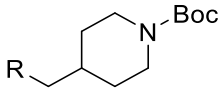
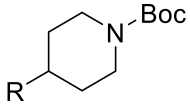
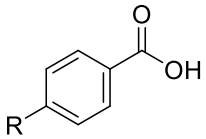
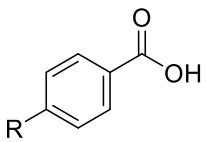
6.9 Reduction of hydantoin imine

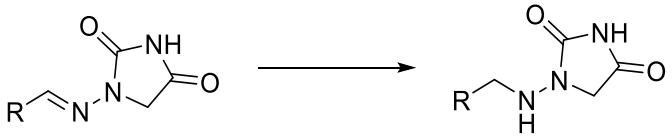
Reduction of the imine, common across all the hydantoin products produced in the project, would provide a simple alteration to the scaffold and provide an interesting SAR analysis of the importance of flexibility of the metal binding terminus. Only two patents have detailed the transformation, so typical imine reduction conditions were attempted, first with compound **6.3.5** from the initial hydantoin synthesis (Table 6).

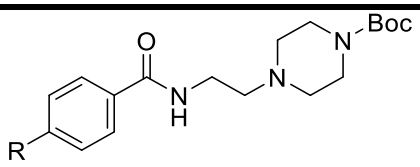
No reaction was observed using $NaBH(OAc)_3$ with excess acetic acid in MeOH, so in the same reaction mixture, $NaBH_4$ was added, followed by $NaBH_3CN$. No reaction progression was observed from either addition.

Table 6: Reagents and conditions for the attempted reduction of the hydantoin imine. *NMR analysis showed characteristic product peaks but could not be fully purified across multiple attempts.



Reactant	Reagents (eq.)	Conditions	Yield
	$NaBH(OAc)_3$ (1.5), acetic acid (excess)	MeOH, rt, 24h	N/A
	$NaBH_4$ (2.0), acetic acid (excess)	MeOH, rt, 22h	N/A
	$NaBH_3CN$ (3.0), acetic acid (excess)	MeOH, rt, 20h	N/A
	$NaBH(OAc)_3$ (2.0), acetic acid (2.0)	DCM, rt, 100h	N/A
	$NaBH_4$ (2.0), acetic acid (2.0)	DCM, rt, 24h	N/A
	H_2, Pd/C (30% w/w)	DMF, MeOH, rt, 6h	N/A
	H_2, Pd/C (5% w/w)	DMF, MeOH, rt, 16h	N/A
	$NaBH_4$ (5.0), $CoCl_2$ (1.5)	MeOH, rt, 23h	Trace



Reactant	Reagents (eq.)	Conditions	Yield
	H_2, PtO_2 (10% w/w)	TFA, rt, 24h	45% by LCMS*

The reduction of two piperazine intermediates was attempted, at first with $NaBH(OAc)_3$ and acetic acid in DCM, then with $NaBH_4$. Then stronger conditions of H_2 and Pd/C were trialled but once again no progression was observed. This was also attempted using the phenyl intermediate **6.3.4** with no success. Using $CoCl_2$ as a means of increasing the strength of the $NaBH_4$ reduction was attempted and showed a weak product mass ion when examined by accurate LCMS but no further progression could be achieved with further additions of $CoCl_2$ and $NaBH_4$.[153]

The stronger reducing conditions of PtO_2 and H_2 in TFA were then trialled on a small, test scale. Analysis by accurate LCMS detected the distinctive 361 m/z mass trace, indicating that the imine had been reduced and the Boc group removed (as anticipated with TFA as the solvent). The product collected from purification using reverse-phase column chromatography could be identified as the reduced product by NMR analysis but contained significant impurities. The reaction was scaled up following the same procedure but once again the product could not be successfully isolated.

6.10 Biological evaluation of hydantoin series

Ten compounds from the hydantoin series were evaluated in the live/dead assay in both C2C12 and HUVEC cells. All responses were within error of the DMSO control aside from compound **6.6.5**, which showed an average cell death of 20% in both cell lines, although this was primarily due to one replicate in each line showing especially high cell death. As compound **6.6.5** is a Boc protected intermediate and an outlier among the other compounds tested there is not a concern for toxicity of the rest of the series.

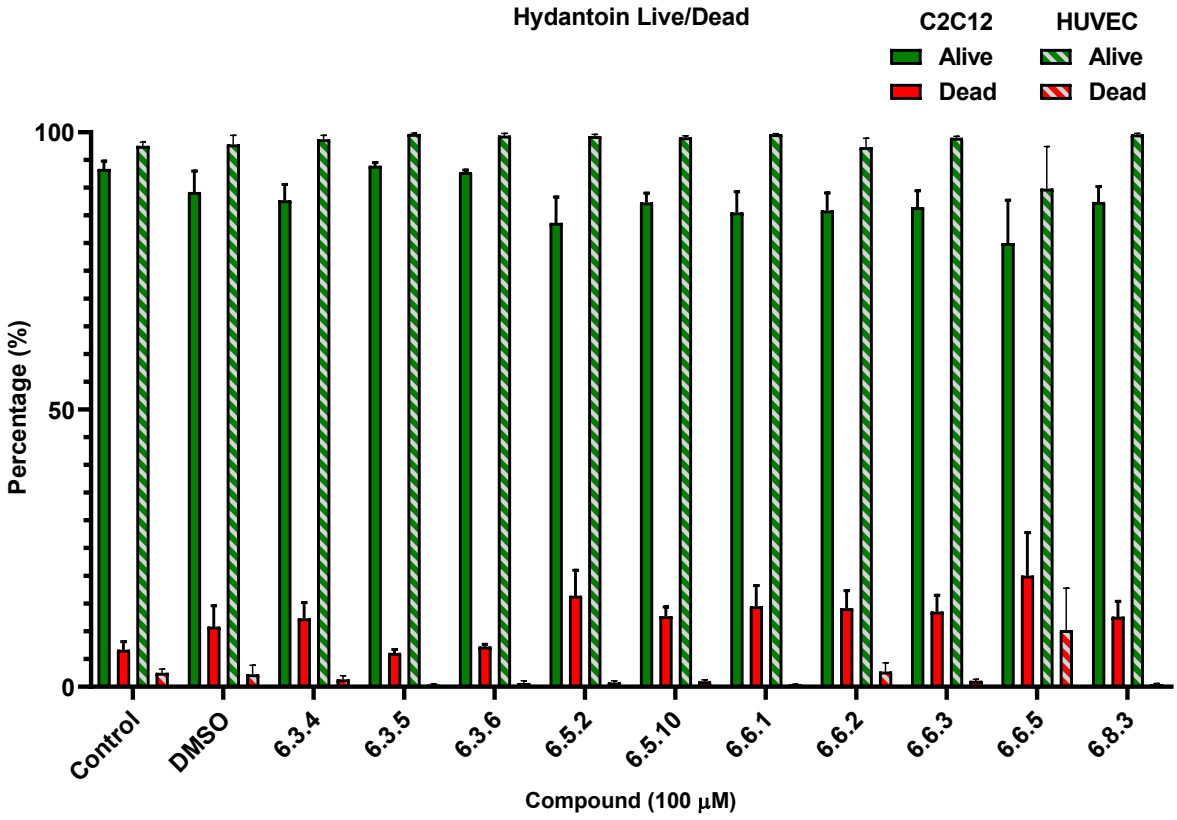


Figure 54: Cell viability of C2C12 (n=3, mean ± SEM) and HUVEC (n=4, mean ± SEM) cell lines after exposure to a selection of hydantoin compounds (100 μM).

The completed compounds and some intermediates from the series were taken forward to SPR analysis. These structures and a summary of the assay results are shown in Table 7.

The background of this assay is detailed in section 5.8 and the results have been discussed in section 5.9 for the zwitterion series. In solubility tests at 100 and 500 μM, detailed previously, compounds **6.5.5** and **6.6.5** showed evidence of non-specific interactions. For compound **6.5.5**, this was inconsistent and since there was additional structural interest in the target, it was included but placed later in the batch of experiments to minimise disruption of the other results. Compound **6.6.5** showed significant and consistent non-specific interactions and was not a priority for testing as an intermediate so was not included for titrations.

One SPR flow-cell was derivatised with integrin $\alpha_v\beta_3$ to test the selectivity of the compounds. The responses in the 100 μM and 500 μM screens showed an equipotent response from both integrins yielding no specificity. As only one cell in the SPR chip was available for derivatisation for the titration experiments, the cell was derivatised with $\alpha_5\beta_1$ and no further integrin specificity could be measured.

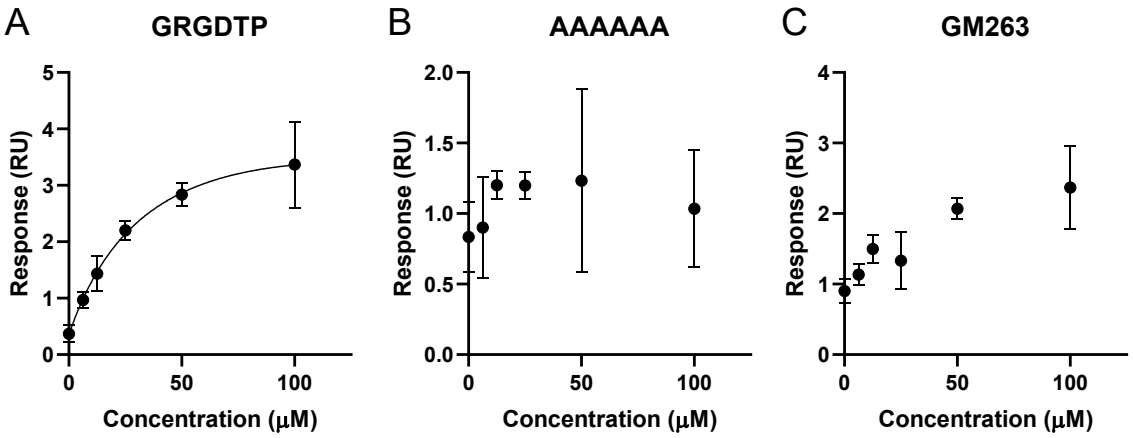


Figure 55: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by control compounds (A) Positive hexapeptide GRGDTP, (B) Negative hexapeptide AAAAAA, (C) Literature compound **GM263**, at six concentrations from 0-100 μM (n=3, mean ± SEM).

As discussed in section 5.9, in the short titrations the peptide controls performed as expected, showing a potent response for the positive control GRGDTP and no dose-dependent activity for the negative control hexa-alanine.

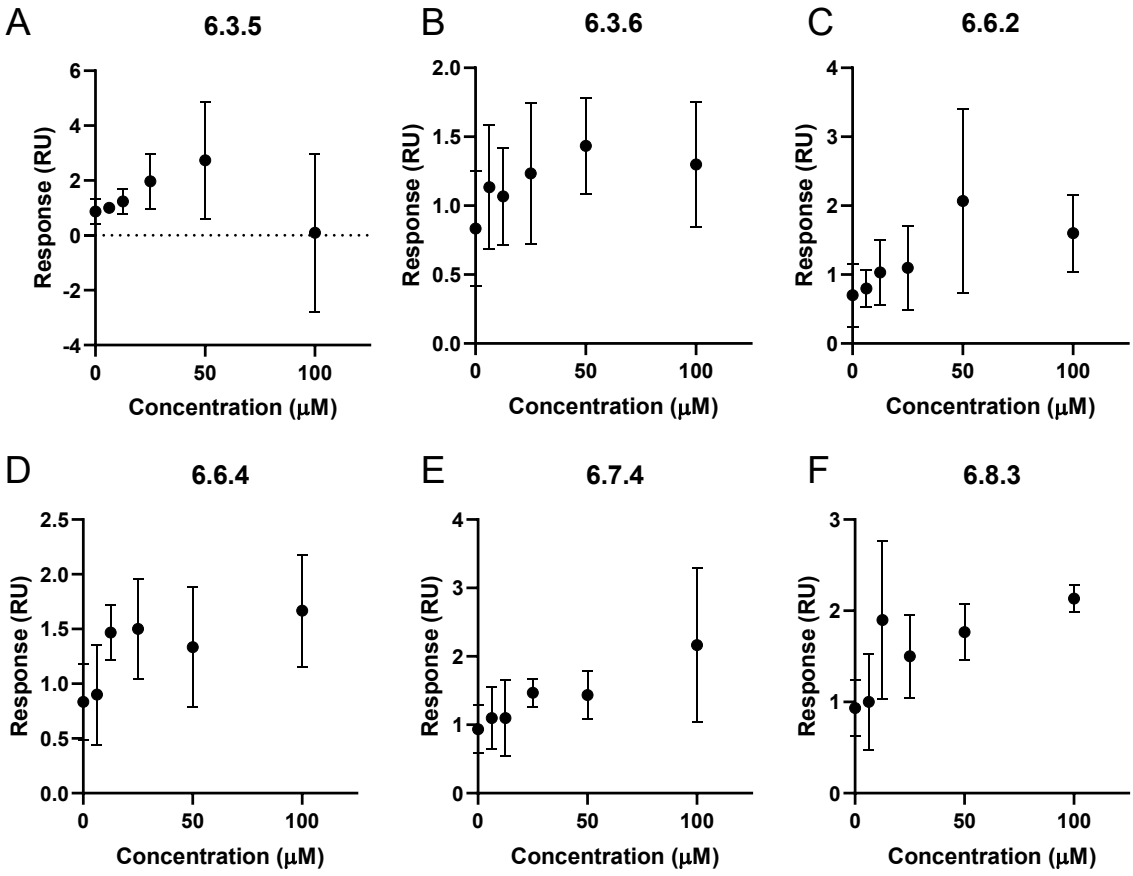


Figure 56: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by hydantoin compounds (A) **6.3.5** (B) **6.3.6** (C) **6.6.2** (D) **6.6.4** (E) **6.7.4** (F) **6.8.3**, at six concentrations from 0-100 μM (n=3, mean ± SEM).

In the six-point titration, 13 compounds from the hydantoin series were tested with most showing a dose-dependent response. Several compounds, **6.3.5**, **6.3.6**, **6.6.2**, **6.6.4**, **6.7.4** and **6.8.3**, had an overall weak response, with any dose-dependency being weak or showing significant errors or inconsistencies, shown in Figure 56. For compound **6.6.4**, a weak response was expected as the compound was an intermediate, not fully deprotected at the terminus thought to be key to binding. Similarly, compounds **6.6.2** and **6.7.4** had a morpholine terminus instead of the piperazine that had shown activity in the previous assays for **6.3.6**. However, in this assay **6.3.6** did not show a strong binding response, with the least dose-dependent activity of the compounds. This response is unexpected, firstly given the promising data seen in the limited phosphorylation of Akt and scratch-wound results and secondly, from the response from the Boc protected form **6.3.5**, which showed an inconsistent but greater response overall. This inconsistency may be due to the sample of **6.3.6**, where the originally produced sample had become insoluble in DMSO, potentially due to degradation, so a remade sample was used that was completely soluble, but did not show the expected activity.

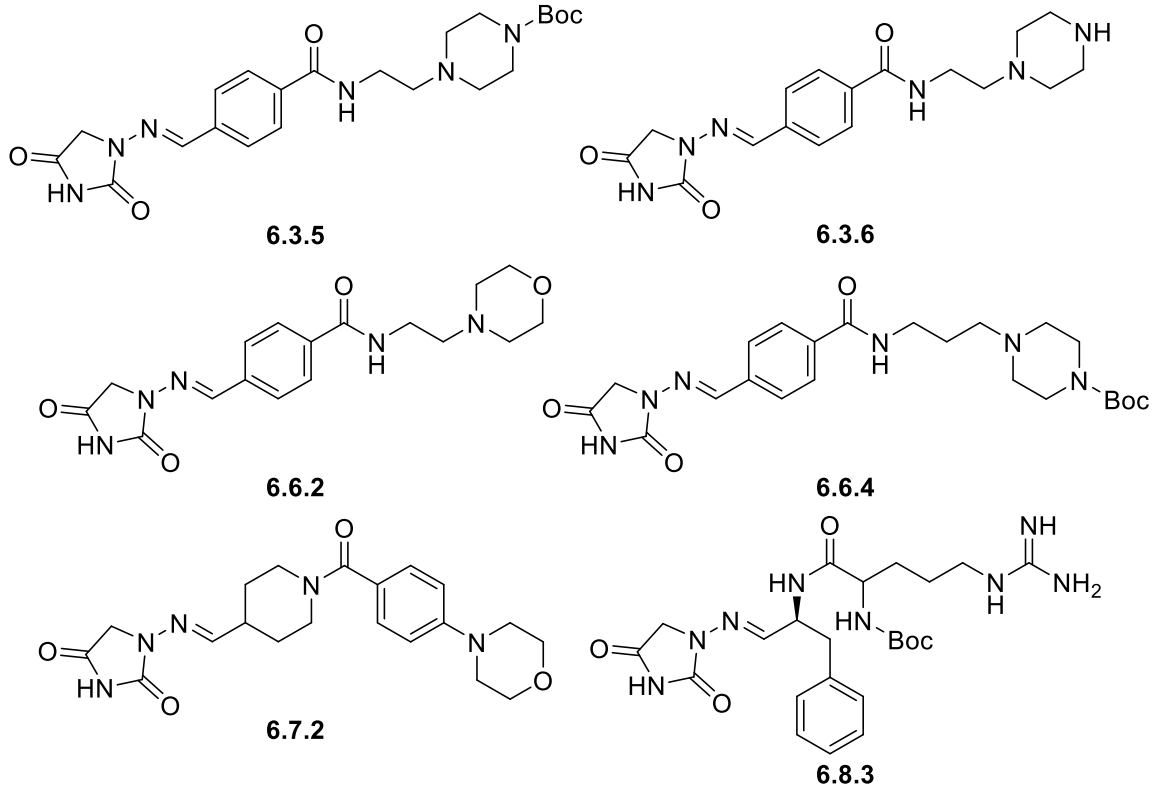


Figure 57: Structures of six compounds that returned weak or no dose dependant activity from the SPR narrow titration experiment, shown in Figure 56.

Compound **6.8.3** had promising docking scores but only a weak response. Further compounds with the phenylalanine core and orthogonal assays would confirm if this was an outlier or if the docking is not a good indicator of activity for this structure. Full

deprotection to compound **6.4.8**, although unsuccessful in previous attempts, would likely be the most informative as hydrophobic groups have little literature precedence near to the basic terminus.

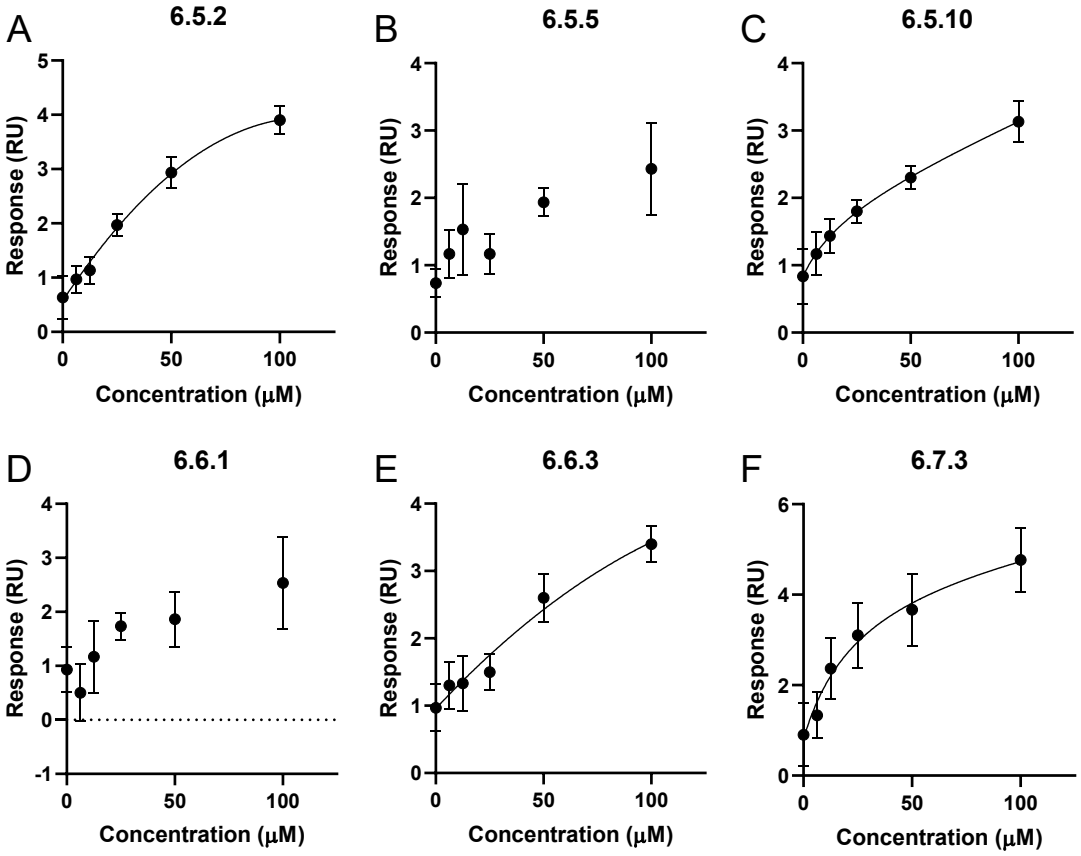


Figure 58: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by hydantoin compounds (A) **6.5.2** (B) **6.5.5** (C) **6.5.10** (D) **6.6.1** (E) **6.6.3** (F) **6.7.3**, at six concentrations from 0-100 μM (n=3, mean ± SEM).

A good dose-dependent response was returned for compounds **6.5.2**, **6.5.5**, **6.5.10**, **6.6.1**, **6.6.3** and **6.7.3**, shown in Figure 58. From the phenyl core targets, compounds **6.6.1** and **6.6.3** had the aminopyridine motif used in literature and returned good responses, despite the fact they had not shown high docking scores. Compound **6.5.2**, simply modified with a piperidine in place of the piperazine in compound **6.3.6** showed a very good response, very similar to the positive control GRGDTP, with low errors between the experiments. While this structure had shown promising docking scores, the gap in response between **6.5.2** and **6.3.6**, which had shown similar docking scores, is difficult to explain.

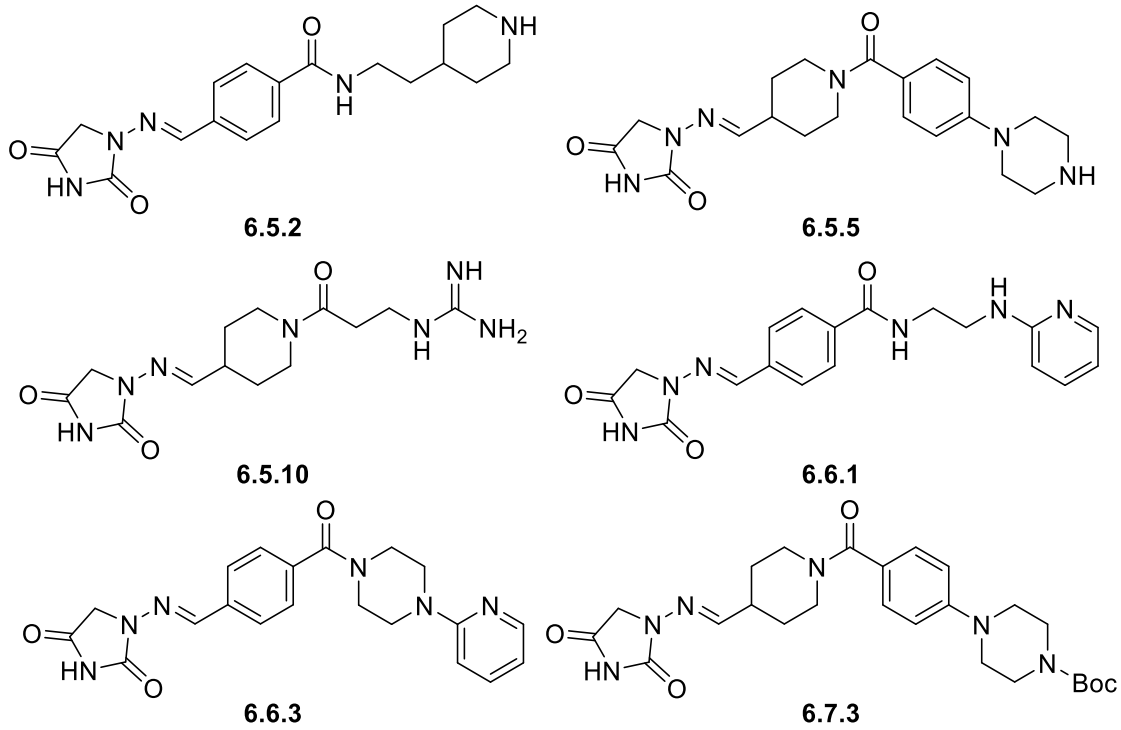


Figure 59: Structures of six compounds that returned dose dependant activity from the SPR narrow titration experiment, shown in Figure 58.

From the piperidine core targets, compound **6.7.3** returned the strongest response of all the compounds tested, considerably better than the GRGDTP positive control. The Boc deprotected form, compound **6.5.5**, also showed a dose-dependent response but had significant non-specific interactions (Figure 58, B) which significantly impacted the measured response. It is unclear how compound **6.7.3** forms the magnitude of the binding response seen in the results given the basic terminus will be significantly hindered in forming interactions by the Boc group.

Compound **6.5.10**, with a guanidine basic terminus, showed a slightly weaker response than GRGDTP but had very low errors between repeats. The constrained docking had the compound as the best structure of the piperidine core targets that was able to be synthesised which is a promising result.

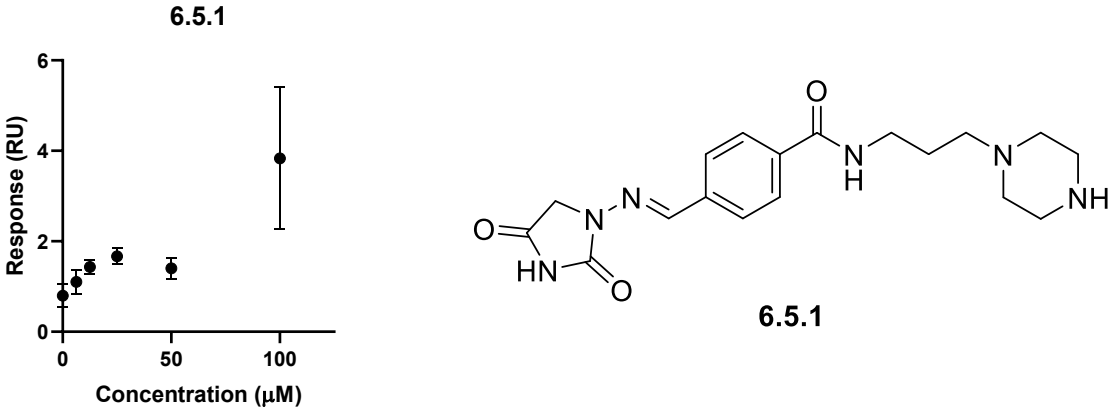


Figure 60: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by hydantoin compound **6.5.1** (structure shown) at six concentrations from 0-100 μM (n=3, mean ± SEM).

Compound **6.5.1** had been returned as the best phenyl core target in both the constrained and unconstrained docking, with a seemingly optimised chain length. In the short titrations however, there is a weak dose-dependent response from 0-25 μM, followed by a weaker response at 50 μM, then a very large response at 100 μM with significant variation between replicates. Examination of the baseline and binding reference indicate no obvious explanation to the activity. As the structure had performed well in docking and had a potentially strong binding response, it was selected for further analysis.

The compounds selected for further analysis in a longer titration were **6.5.1**, **6.5.2**, **6.5.10**, **6.6.3** and **6.7.3**. As detailed in section 5.9, the compounds were run four times, with the controls run twice, at eight concentrations, 0-400 μM. The results are shown in Figure 61 and Figure 63.

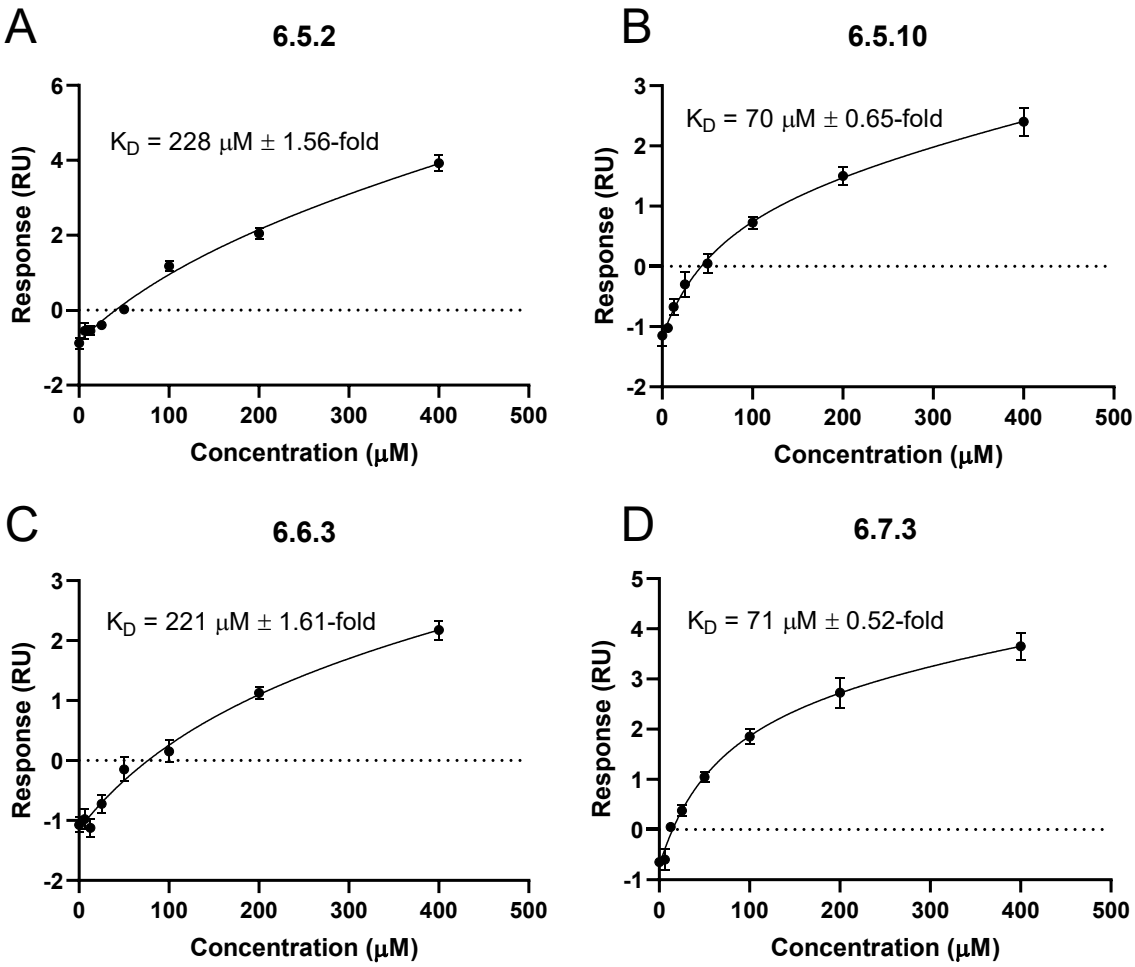


Figure 61: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by hydantoin compounds (A) **6.5.2** (B) **6.5.10** (C) **6.6.3** (D) **6.7.3**, at eight concentrations from 0-400 μM, with measured K_D values (n=4, mean ± SEM).

Matching the short titrations, compound **6.6.3** showed a reliable but weak dose-dependent response. A clear binding response starts at 25 μM, unlike the rest of the compounds which tended to show a clear response above the 0 μM concentration at 6.25-12.5 μM. The K_D was measured to be 221 μM ± 1.61-fold. The large error likely reflects the fact that the response has not come to plateau, which is reflected in the confidence interval which has a minimum of 23 μM and maximum that cannot be accurately estimated.

Compound **6.5.2** also had an initially weak response at the lower concentrations, but then increased rapidly between 50-400 μM, reaching one of the highest responses in the assay. In a similar manner to compound **6.6.3**, this meant the response had not reached plateau, giving a K_D of 228 μM ± 1.56-fold. The confidence intervals are similarly 27 μM to an incalculable maximum. While the errors between replicates are small, the trend in response does not appear to be entirely consistent with the results from the short

titrations, which showed a dose-dependent increase from 6.25 µM. No explanation can be found in the baseline or binding reference and may require further SPR repeats with a new chip to clarify the activity.

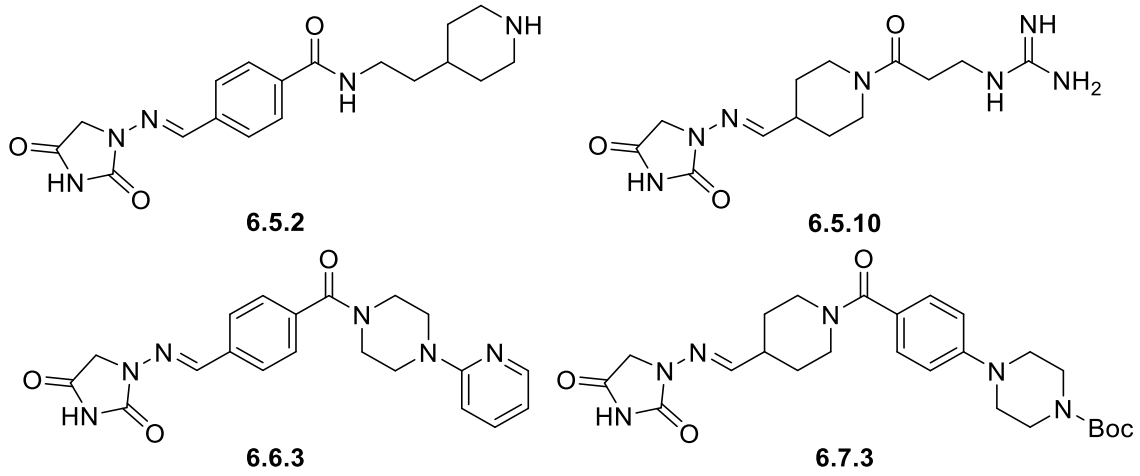


Figure 62: Structures of four compounds analysed in the 8-point SPR titration experiment, shown in Figure 61.

Compound **6.5.10** showed a good dose-dependent response from 12.5 µM, not reaching as high a response as other compounds but with good reliability, yielding a K_D of 70 µM ± 0.65-fold. However, the 95% confidence intervals remained large at 20-502 µM. Overall, this is in good agreement with the response from the short titrations and is a promising result, although the confidence intervals are likely limited concentration range of the titration.

Compound **6.7.3** also showed a reliable, steady response increasing with the concentration, with one of the highest responses at 400 µM. The K_D was calculated to be 71 µM ± 0.52-fold, with confidence intervals of 27-260 µM. This improved result and closer confidence intervals is likely due to the small errors between replicates, resulting in a well fit curve.

Both compounds **6.5.2** and **6.7.3** reached a large SPR response at the highest compound concentration, both above the theoretical R_{max}. As discussed previously for the SPR experiment, the cause of the increased response is unclear. These structures may have an increased affinity for an additional binding event at high concentrations; however, this is not evident from the titration curve, or the compound structures.

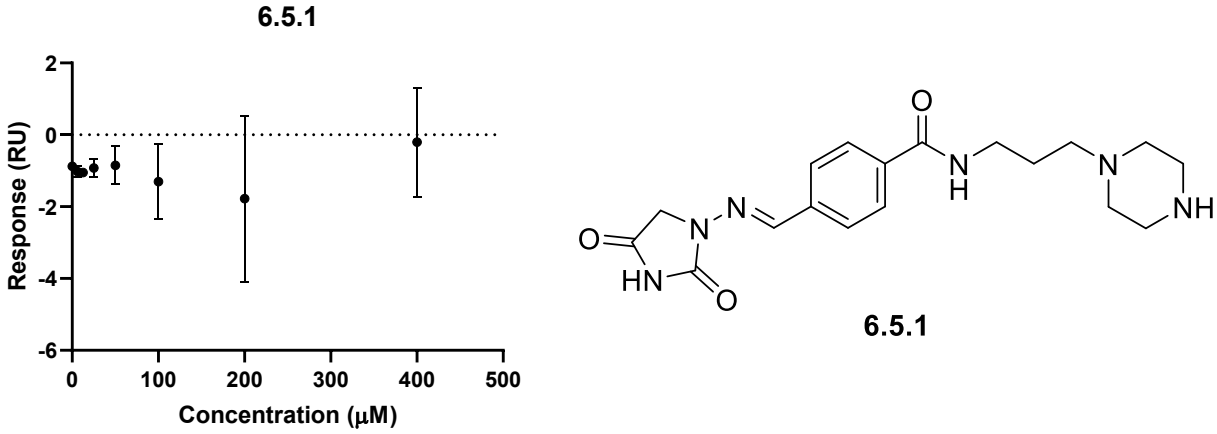


Figure 63: Binding to immobilised biotinylated $\alpha_5\beta_1$ integrin by hydantoin compound **6.5.1** (structure shown) at eight concentrations from 0-400 μM (n=4, mean ± SEM).

Finally, compound **6.5.1** showed a negative response for most concentrations, with large errors between each. While the response in the short titrations was inconsistent, this behaviour was very unexpected. Examination of the surface baseline does show a rapid degradation of the protein, specifically in the second batch of experiments, unseen for any of the other compounds. However, the first two replicates, unaffected by this behaviour, also show a poor response which indicates that the surface degradation may be unrelated. As the binding response was highly variable and dissimilar to the screening and the short titration, no K_D measurement can be made reliably.

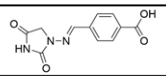
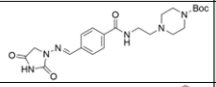
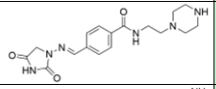
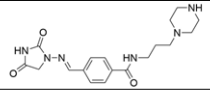
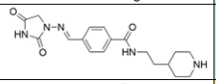
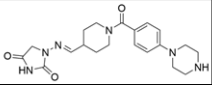
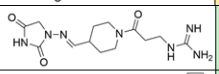
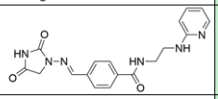
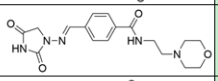
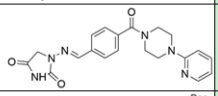
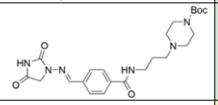
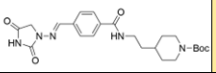
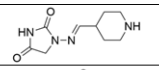
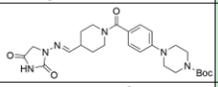
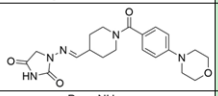
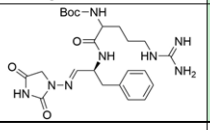
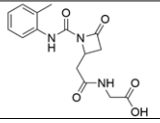
In retrospect, the initial results of compound **6.5.1** were likely no response at the low concentrations and a significant error at 100 μM, and compound **6.6.1** would have been a more valuable inclusion to the further analysis.

Analysing the structures of the compounds with K_D measurements, compounds **6.5.2** and **6.6.3** contain the phenyl core, while the much stronger binding **6.5.10** and **6.7.3** contain the piperidine core. This could indicate a higher binding potential for the piperidine core, although limited conclusions can be made from the data set. Four different basic site termini are represented with a piperidine, aminopyridine, guanidine and Boc-piperazine all represented. This may indicate that the interaction strength is largely dependent on the metal binding and linker regions, especially since compound **6.7.3**, containing the Boc-piperazine, did not show favourable docking and is not expected to form interactions with GLN 221 or ASP227.

The response of compound **6.7.3** in SPR may be indicative of potent activity for compound **6.5.5**, the deprotected form which showed non-specific binding. A priority should be to test compounds **6.7.3**, **6.5.5** and others in an orthogonal assay to validate the activity found in the SPR. The K_D values found in this project are similar to the values

found in SPR analysis of integrins with macrocyclic RGD-containing peptides found by Cheng *et al.* (50 μM), but weaker than with small molecules found by Pang *et al.* (102 nM).[120, 121] Overall the compound potency remains low and although hydantoin is a novel mechanism for binding to the MIDAS site, further work is required to increase the potency to a level close to the values found in literature.

Table 7: Number and structure of all hydantoin compounds with assay data collected, labelled with the quality of the result. Examples of poor data include unreliable results and non-specific interactions.

Name	Structure	100 µM solubility screen	500 µM solubility screen	6-point titration	8-point titration k_d / µM	Live/Dead HUVEC	Live/Dead C2C12	Scratch Wound
6.3.4								
6.3.5								
6.3.6								
6.5.1								
6.5.2					228±1.56-fold			
6.5.5								
6.5.10					70±0.65-fold			
6.6.1								
6.6.2								
6.6.3					221±1.61-fold			
6.6.4								
6.6.5						Some cell death		
6.7.2								
6.7.3					71±0.52-fold			
6.7.4								
6.8.3								
AAAAAA	Hexapeptide							
GRGDTP	Hexapeptide							
GM263								

Data quality	
	Good
	Intermediate
	Poor
	Not collected

6.11 Summary of hydantoin series

At first, simple alterations of the hydantoin containing, screening hit **3.6.6** were designed, however the amide bond in the hit compound could not be replicated. Formation of an imine bond with an aldehyde had greater precedent and was successfully completed. Compounds **6.3.5** and **6.3.6** were then produced and tested in the cell based assays, yielding an increased phosphorylation of Akt than the compounds in the zwitterion series and the literature agonist, **GM263**.

Further synthesis in the series altered the benzaldehyde linker for a piperidine and a phenylalanine, producing 15 further compounds using all three linkers, with simple variations of compound **6.3.6** and analogues with functional groups found in literature, such as guanidine and aminopyridine.

While the established synthetic route did generate a good number of analogues, some reactants for each linker were not reactive, including primary and secondary amines. This may be due to the stability or age of the chemicals as no obvious link between the structures and reactivity could be made. The reduction of the imine was attempted in several conditions, with promising reaction progression using PtO_2 and H_2 in TFA, however the purification could not be completed. With additional experiments, the amine could likely be isolated and would provide an interesting comparison to the synthesised imines, with the additional flexibility at the metal binding site the reduction would provide.

A live/dead assay of ten compounds found that only one induced cell death significantly above the DMSO control. SPR analysis of the compounds showed a dose-dependent, protein binding response for most compounds in the series, with compounds **6.5.10** and **6.7.3** measuring a K_D of 70 μM ± 0.65-fold and 71 μM ± 0.52-fold respectively. These compounds represent a lead for future synthesis.

With only limited synthesised analogues and assay data, it is difficult to perform an SAR analysis of the compounds. Changing the piperazine terminus in **6.3.6** to the piperidine in **6.5.2** was the best of the simple analogues in the SPR analysis, however the aminopyridine of compound **6.6.3** and the piperidine linked compounds **6.5.10** and **6.7.3**, with guanidine and Boc-piperazine termini respectively, showed more potent equilibrium dissociation constants. The piperidine core molecules had the best results in the SPR assay, though further synthesis and testing of compounds is required to confirm these results.

7. Conclusions and future work

Current treatments for type 2 diabetes are insufficient at providing protection against cardiovascular disease, the cause of death for 50.3% of people with T2D.[11] Recommended treatments for patients with T2D at risk of cardiovascular disease, in addition to diet and exercise changes, are GLP-1 receptor agonists and SGLT2 inhibitors.[20] These are the only classes of drug that have shown significant reduction of MACE (major adverse cardiovascular events) hazard ratios in meta-analyses, by 12% and 10% respectively.[154, 155]

The interaction between IGFBP-1 and $\alpha_5\beta_1$ via the RGD-binding domain has shown promising results for the treatment of cardiovascular disease, with upregulation of IGFBP-1 resulting in improved whole-body insulin sensitivity, improved glucose tolerance, reduced blood pressure, reduced susceptibility to atherosclerosis and enhanced vascular NO production.[43] Isolated RGD-containing hexapeptide sequences were shown to promote the same effects, but problems with metabolic stability of peptides led to a virtual shape screen of the RGD structure to find molecules that could mimic the shape and electrostatic interactions of the hexapeptides.[113] The top results were purchased and tested for their ability to enhance phosphorylation of Akt using western blotting. When tested at a concentration of 10 μM, most compounds promoted phosphorylation, equipotent to the hexapeptide positive controls.

The hit compounds were docked using the Schrodinger Suite and modifications were designed to improve interactions with the residues predicted nearby. Firstly, a series of zwitterions was designed that had similarity to integrin binders in literature. Molecular docking of compound **3.6.5** showed the carboxylic acid binding to the MIDAS cation, as seen in the endogenous ligand and literature binders. Modification to the opposite terminus in **3.6.5** from a cyclic diamine to a glycine, histidine and arginine formed improved interactions with residues GLN221 and ASP227, mimicking the endogenous ligand and returning docking gscores of -9.2 to -8.7. The series retained a piperazine-benzoic acid core from the hit compound, which was a common intermediate for synthesis.

Analogues with a flexible acid binding terminus were designed to extend into an aspartate rich pocket on the β_1 monomer of the integrin. Although the conformation of the ligand appeared strained, docking gscores ranging to -8.9 indicated this could be formed by this binding mode.

Synthesis of these targets was challenging, with late-stage intermediates becoming increasingly polar and zwitterionic. This meant that isolation of the targets was time-consuming and often unsuccessful. However, several final compounds, **5.3.6, 5.3.7** and **5.3.9** with the histidine terminus, were successfully isolated. Compound **5.2.3**, with the flexible alkyl chain, was also synthesised, enabling the alternate binding hypothesis to be tested.

Further compounds **5.4.9** and **5.6.1**, originating from a series of pyrazoles identified by a virtual high throughput screen and hit compound **3.6.4**, were synthesised based on simple changes from the hit matter to test the zwitterion design philosophy across alternate scaffolds. This series had docking gscores reaching -11.2.

These compounds, along with some intermediates, underwent preliminary evaluation in a range of assays - using C2C12 and HUVEC cells to assess cytotoxicity, western blotting to evaluate enhancement of insulin-stimulated Akt phosphorylation, scratch-wound assays with endothelial cells to determine the effect of compounds on both proliferation and migration as well as SPR to evaluate direct binding of compounds to integrin. No compound toxicity was indicated by the live/dead assay.

The Akt phosphorylation assay doses compounds and controls with insulin and during this stage, unexpected issues became apparent. The response to insulin treatment changed, with the insulin only control response increasing above the positive controls. It was not possible to identify the reason for this, but the possibilities include a different activity of a new batch of insulin or a media formulation change. Attempts to change insulin doses, measurement timepoints and cell types did not improve the baseline response to insulin for this assay. As a result, the assay was not repeated, with only two experiments run, using 100 nM and 10 nM of insulin.

The poor quality of the Akt phosphorylation data limits the conclusions that can be drawn, however without any other indications of compound activity, the general trends were examined. The zwitterion series yielded a poor, unreliable response that did not show significant responses above the baseline.

The scratch-wound assay, measuring endothelial cell migration which is impaired in T2D patients and restored by IGFBP-1 *in vitro*, was used to measure the effect of compounds compared to IGFBP-1 and RGD peptides. Overall, the compounds demonstrated a cell migration below the level observed with the IGFBP-1 control, which has been interpreted to mean that the compounds from the zwitterion series did not have an agonistic effect on integrin-mediated migration. The assay shows some compounds leading to similar cell migration to the hexapeptide GRGDSP and within the error of the DMSO control,

and some with lower cell migration. Reduced cell migration could represent antagonistic binding to an integrin or a lack of selectivity. No results showed obvious promise for the series, so synthetic focus was changed.

SPR direct binding analysis was later completed that more accurately measured the series binding activity, although issues were still present in the compound preparation, control variation and analyte response. Only one compound from the series, **5.6.1**, showed a good dose-dependent response and was measured to have an equilibrium dissociation constant (K_D) of 59 μM ± 0.64-fold. Compound **5.6.1** had the best K_D measurement in the project; however, it had not shown any promising activity in the other assay results. While these assays have not been shown to produce validated, reliable measurements of compound activity, the compound is theorised to be to be antagonistic and is therefore of lower interest to the project going forward.

Despite significant efforts in modelling and synthesis, the results of the zwitterion series did not show the desired improvement from the hit compounds and limited development opportunities were generated. The *in silico* docking of the targets had shown promising comparison to the endogenous ligand and good affinity scores, but these were not validated from assay results.

The numerous examples of zwitterionic integrin binders in literature shows that this approach can generate efficacious compounds for the activation of $\alpha_5\beta_1$ integrin, though the methods applied in this case have not yielded results. Further work on zwitterionic compounds could be undertaken with reviewed design methods.

The hydantoin modality featured on hit compound **3.6.6** stood out as a potential bioisostere to the carboxylic acid as there is literature precedence of hydantoins binding to zinc ions in proteins.[131-135] First, the binding mode was modelled and found to be reaching out to TRP157 and LYS182 residues in $\alpha_5\beta_1$, away from the endogenous binding site. Modification of the hit compound from the indane to an indanone was modelled to have a small benefit to target affinity, with the gscore improving from -7.5 to -8.4.

Recreation of the amide bond in hit compound **3.6.6** was unsuccessful despite numerous conditions attempted, so formation of an imine was completed instead, leading to the synthesis of compounds **6.3.5** and **6.3.6**. These compounds underwent preliminary evaluation in live/dead, pAkt and scratch-wound assays, yielding promising results. No toxicity was found in the live/dead assay and the compound **6.3.6** increased phosphorylation of Akt above all over compounds in both the 100 nM and 10 nM insulin experiments, with compound **6.3.5** showing a good response in the 100 nM run. In the

scratch-wound assay both compounds showed a comparable response to the GRGDSP control, with compound **6.3.6** showing high variation, with one repeat showing an increased cell migration comparable to the IGFBP-1 control.

A small library of analogues of **6.3.6** was designed and docked, with simple variations, literature motifs and simple growth vectors selected for synthesis. Piperidine and phenylalanine linkers were also designed, replacing the phenyl, with similar targets designed and docked. Piperidine, piperazine and guanidine termini generally had the best docking scores in both an open and constrained docking grid.

Synthesis of the intermediates was carried out and derivatised with the basic termini. The phenyl linker had five analogues synthesised with two intermediates also submitted for testing. Eight basic termini were not able to be coupled, where in all but one case there was little evidence of product formation. Four piperidine linked targets were synthesised, with one intermediate submitted and three syntheses failing. The targets that were not completed all were not sufficiently purified for submission, with reliable reactivity of the basic termini. Only one target, **6.8.3**, was made for the phenylalanine core, with the Boc-deprotected product not sufficiently purified for submission.

Reduction of the imine was attempted in nine conditions, finding that PtO_2 and H_2 in TFA showed significant formation of the amine. The purification was not completed due to the challenge in purifying the highly polar compounds in this project but could likely be completed with further repeats.

Live/dead analysis of ten compounds from the series showed one, **6.6.5**, that had any cell death above the controls. Compound **6.6.5** also showed significant non-specific binding in the SPR analysis and was not analysed in the titrations. Only one compound, **6.5.5**, showed any consistent non-specific response.

A six-concentration SPR titration was run for 13 of the hydantoin compounds, with six compounds, including **6.3.6**, showing a weak, non-dose-dependent response. Seven compounds had a good response, including with structures containing the phenyl and piperidine linker, and a variety of basic termini including the aminopyridine, piperidine, guanidine and a Boc-protected piperazine. Five of these compounds, **6.5.1**, **6.5.2**, **6.5.10**, **6.6.3** and **6.7.3**, with the best responses were selected for an eight-point titration to measure the K_D values. The results were varied, from compound **6.5.1** which did not show a reliable response, to the weak K_D values of **6.6.3** and **6.5.2**, 221 μM ± 1.61-fold and 228 μM ± 1.56-fold respectively, and finally to compounds **6.5.10** and **6.7.3** with K_D values of 70 μM ± 0.65-fold and 71 μM ± 0.52-fold respectively.

The SPR assay did have some issues, specifically atypical responses from the controls in the eight-point titration experiment and the overall weak response from compound binding despite the saturated surface. Therefore, it is crucial that these compounds are tested in orthogonal assays to verify the compound binding and to establish the nature of the compound activity. The cell adhesion assay (with compounds in competition with fibronectin) used by Baiula *et al.* to measure agonistic and antagonistic activity would fit these requirements and confirm the activity of the compound **GM263** that had not shown the expected results in the in-house assays.[95] The group also use a competition assay with compound incubated with integrin and applied to a fibronectin-coated plate witch would also be suitable for robust analysis of the compounds.In addition, testing of compound **6.5.5**, the deprotected form of the **6.7.3** with one of the best K_D found in the project, which may be a promising lead for future compound development.

Compound **6.5.5** had shown very good docking results, with the best score in the non-constrained and the fourth best score in the constrained model. This prediction has not been validated by the potency of compound **6.7.3** as the *in silico* interactions of **6.5.5** with GLU126 and GLN221 are not expected to be compatible with a Boc protecting group, however the potency of **6.7.3** may elute from interactions that were not modelled. Compound **6.5.10** was also scored well in docking, receiving the best score of the piperidine linked targets that were successfully synthesised in the constrained docking, with compound **6.5.5** having the second-best score.

Compounds **6.5.1** and **6.5.2** were returned with the best scores for the phenyl linked targets in the constrained model and had good responses, though compound **6.6.3** had a better K_D and had not been highlighted in the modelling. Overall, there is evidence that the constrained docking model can predict compound activity to some extent.

The piperidine linked hydantoins may present a promising lead for future work on the project, though the primary focus of future work should be to develop reliable assays to verify and understand the nature of compound activity. Molecular docking has been used to predict compound activity and has shown some success for the hydantoin series. Further investigation and comparison between target prediction and activity may elute insights to the interactions being formed in the target site and expedite compound development.

8. Experimental

8.1 General chemical synthesis methodology

Chemical reagents were purchased from Activate Scientific, Alfa Aesar, Apollo Scientific, Fisher Scientific, Fluorochem, Merck and ThermoFisher. Solvent was evaporated using a Buchi Rotavapor R-114 with a B-480 Waterbath and a CVC 3000 Vacuubrand vacuum pump or an IKA RV and water bath with a VACSTAR digital vacuum pump and a RC 2 lite recirculating chiller.

NMR data was recorded using a 2-channel Bruker 400 MHz (9.40 T) and a 2-channel Bruker AV4NEO 500 MHz (11.75 T). NMR data was analysed using MestReNova software. LCMS data was recorded using a Thermo Ultimate 3000 HPLC instrument with a UV diode array detector and a Bruker MS detector with electrospray ionisation with positive and negative switching mode and an acetonitrile/water (+ 0.1% formic acid) gradient (runtime 1.20 mins) with UV monitoring at the wavelengths of 200-340 nm.

HPLC purity analysis was completed by Dr Jeanine Williams, using an Agilent 1290 Infinity II HPLC system with a diode array detector at 254 nm. An Agilent InfinityLab Poroshell 120 EC-C18 (2.1 x 50 mm, 1.9 µm) column was used at 40°C, running a MeCN:H_2O (+0.1% TFA) gradient (5-95%) over 5 mins with a flow rate of 0.5 mL/min.

Flash silica column chromatography was conducted with a Biotage Isolera One and a Biotage Selekt, using Biotage Sfär Silica 60 µm (25 g), Biotage Sfär Silica Duo 60 µm (10 g) and Modus B irregular 40-63 µm (12 and 25 g) columns for normal-phase purification and Biotage Sfär C18 Duo 30 µm (6 and 12 g) and Modus P spherical C18 40-60 µm (4.5 and 18 g) columns for reverse-phase purification.

Automated Mass-Directed Preparative-Liquid Chromatography (MDAP) was completed using a 1290 Infinity II Preparative LC/MSD System, using a 5-95% acetonitrile/water (+0.1% formic acid) gradient.

High resolution mass spectrometry (HRMS) was acquired on a Bruker Impact II Q-TOF spectrometer equipped with a VIPHESI source using electrospray (ES) ionisation.

8.2 Computational methodology

Shape and electrostatic comparison of compounds in the eMolecules screening library was performed by OpenEye's ROCS (3.4.0.4) to the RGD sequence.

Docking studies were performed using the Schrodinger Maestro, Epik and Glide (Schrödinger Release 2021-3, Glide, Schrödinger, LLC, New York, NY, 2020). Three

protein structures of the extracellular, truncated $\alpha_5\beta_1$ integrin with an RGD ligand, 3VI4, 4WK0 and 4WK4, produced by Nagae *et al.* and Xia *et al.* were used.[59, 68]

Protein structures were downloaded from the RSCB Protein DataBank (PDB) and imported into Maestro. The structures were prepared with the default settings. No residue clashes or errors were encountered near the RGD binding site so were automatically resolved or ignored. A variety of docking grids were generated. These were centred on the RGD ligand, with a boundary of 10-15 Å from the ligand. Non-constrained docking had no additional settings. Constrained docking in this work refers to docking into a grid with interactions required with the MIDAS metal ion, and at least one interaction with either GLN221 or ASP227. This orients the docking results to align with the endogenous RGD ligand and often produced improved docking scores.

Designed compounds were prepared by LigPrep and Epik, which generated the possible stereoisomers and ionisation states of the structures. The pH range of the possible ionisation states were 7.4 ± 0.5. For the hydantoin compounds, the 'Add metal binding states' option was required as the deprotonated form of the hydantoin would not be generated otherwise.

The docking was completed with Glide using standard or extra precision mode. In general, for each of the prepared ligands, ten poses were written per ligand, with the best five poses output after post-docking minimisation.

The results were analysed by the gscores, with the highest scoring molecules assessed for pose feasibility and compared to the RGD ligand and other docking results.

8.3 Biological assay methodology

Cell culture

C2C12 cells were cultured in T75 and T25 flasks, and in 6-well plates (Sarstedt, 83.3920.300). The growth media (DMEM 32430-100; Thermo Fisher Scientific) was supplemented with 10% fetal bovine serum (FBS) and 1% antibiotic-antimyotic solution (AAS). Differentiation was initiated by rinsing fully confluent cells once with PBS and adding differentiation media (DMEM supplemented with 5% FCS).

HUVECs from pooled donors were maintained in Endothelial Cell Growth Medium 2 (ECGM2) (PromoCell) supplemented with Growth Medium 2 Supplement Mix (PromoCell) containing the following growth factors: fetal calf serum (0.02 mL mL^{-1}), Epidermal Growth Factor (recombinant human, 5 ng mL^{-1}) Basic Fibroblast Growth Factor (recombinant human, 10 ng mL^{-1}), Insulin-like Growth Factor (Long R3 IGF, 20

ng mL^{-1}), Vascular Endothelial Growth Factor 165 (recombinant human, 0.5 ng μL^{-1}), Ascorbic Acid (1 µg mL^{-1}), Heparin (22.5 µg mL^{-1}), Hydrocortisone (0.2 µg mL^{-1}). HUVECs were cultured until passage five.

Cells were passaged when confluent in the flask by washing the cells with Dulbecco's phosphate buffered saline (PBS) at pH 7.4, followed by addition of trypsin (Life Technologies, 25200072) to loosen the remaining cells from the surface. All cells were incubated at 37°C with 5% CO_2 in a humidified incubator.

In vitro treatment for the phosphorylation of Akt assay

Cells were serum starved overnight. Cells were pretreated with 100 µM RGD-containing hexapeptide (GRGDTP or RAD-containing control peptide GRADSP or small-molecule and incubated for 40 min. Cells were then exposed to 100 nmol/L recombinant insulin (I9278; Sigma Aldrich) for 10 min.

Immunoblotting

Cells were lysed using lysis buffer (FNN0011; Invitrogen) posttreatment. Lysates were clarified by centrifugation (13,000 rpm for 15 min at 4°C). Pierce BCA assay was prepared and analysed. Lysates were diluted according to the BCA assay and separated by electrophoresis through 4–12% Bis-Tris gel (NP0335; Life Technologies) and blotted onto polyvinylidene fluoride membranes. Blots were probed with protein kinase B (AKT) (9272; Cell Signaling), pAKT (Ser473) (4060; Cell Signaling) and β-actin (Sc-47778; Santa Cruz).

Imaging

The blots were imaged using GeneSys and analysed using ImageJ and GraphPad Prism 8.

LIVE-DEAD cytotoxicity assay

Assays were performed using the LIVE-DEAD Viability/Cytotoxicity Kit (Invitrogen) with minor modifications to the manufacturers protocol. HUVEC cells were seeded into transparent Poly-L-Lysine coated 96-well plates at a density of 60,000 cells per well. After 24 hrs, cells were treated with relevant compounds in ECGM-2 media (100 µM, 200 µL) (Gibco), diluted from DMSO stock (10 mM). After a further 24 hours, treated media was removed and wells were washed with DPBS (200 µL)(Sigma Aldrich). In a separate flask, Ethidium Homodimer-1 stock solution (EthD-1, component B, 2 mM, 20 µL) was added to DPBS (10 mL). To this, Calcein AM stock solution (Component A, 4mM, 5 µL) was added to give a solution of EthD-1 and Calcein AM in PBS at approximately 4 µM

and 2 μM respectively. 100 μL of this solution was added to each well and cells were incubated for 30 minutes. Wells were imaged using an Incucyte® ZOOM (Sartorius) live cell analysis system, with fluorescence measurements at 530nm and 645 nm recorded. Images were analysed using Incucyte® Zoom software and cell viability calculated with the following formula:

$$Cell\ viability = \frac{\%\ Red\ area}{\%\ Green\ area} \times 100$$

Scratch/Wound assay

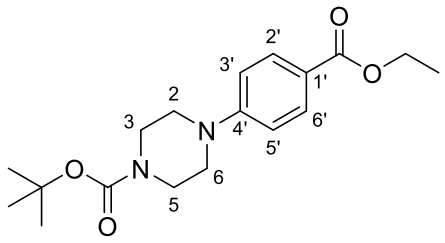
96-well IncuCyte® Imagelock plates (Sartorius) were coated using a 2% bovine gelatin solution in DPBS and placed in a 5% CO_2, humidified incubator to polymerize for 15 minutes. Prior to seeding, the excess gelatin was flushed out by two DPBS washes. HUVEC at passages 2-5 were trypsinised for 3-4 minutes in a 5% CO_2, humidified incubator and pelleted by centrifugation at 400 x g for 8 minutes. The cells were then seeded at a density of 200,000 cells/mL and pipetting 200 μL per well – or 200 μL DPBS to any empty wells. Approximately two days later, once a completely confluent monolayer of cells had formed, a wound was created using the Essen Woundmaker. The 96-well plate was slotted into the device and the scratch was created by utilizing the lever. All wells were carefully washed in 100 μL PBS and checked which wells contained a full scratch using a microscope. Compound treated—ECGM-2 aliquots were added to selected wells before being inserted into the IncuCyte® ZOOM for 24-hour image sampling. A minimum of three successful scratches per treatment condition were prepared. Images were analysed using NIH ImageJ freehand selection tool to measure the relative wound density and the resultant data were exported to Microsoft Excel for further data transformation.

8.4 Surface plasmon resonance (SPR)

Surface plasmon resonance (SPR) was completed using a Cytiva Biacore T200 with a Biacore Streptavidin (SA) chip Series S. The surface was derivatised with three injections (3 mins initially then 2 mins) of a 100 nM solution of biotinylated human $\alpha_5\beta_1$ integrin (IT1-H82WA; 2BScientific). The buffer consisted of $MgCl_2$ (1 mM), Tris (20 mM) and NaCl (150 mM), which was made up to 10x stock, diluted and filtered through a nylon 0.45 μm filter before use. Compounds were diluted to 10 mM in DMSO, before dilution in buffer to the desired concentration. DMSO was added to the running buffer to match the dilution of the samples. All solutions were run at 30 μL/min. Results were analysed using Biacore Insight software and GraphPad Prism 8 and images were produced with GraphPad Prism 8.

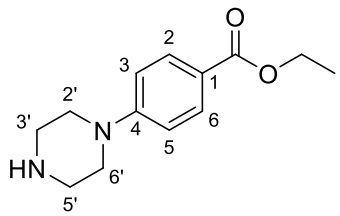
8.5 Synthesis of initial zwitterion targets

Preparation of 2-methyl-2-propanyl 4-[4-(ethoxycarbonyl)phenyl]-1-piperazinecarboxylate (**5.3.1**)[156]



2-Methyl-2-propanyl 1-piperazinecarboxylate (1.86 g, 10.0 mmol, 1.00 eq.) and potassium carbonate (2.76 g, 20 mmol, 2.00 eq.) were dissolved in DMSO (50 mL) before addition of ethyl 4-fluorobenzoate (1.46 mL, 10.0 mmol, 1.00 eq.). The reaction mixture was stirred and heated to 130°C for 24 hours. TLC showed product formation. After cooling, the reaction mixture was quenched with brine (200 mL) before the product was extracted with ethyl acetate (100 mL). The organic layer was washed with brine (3 x 100 mL), dried with $MgSO_4$, filtered and concentrated *in vacuo*. The residue was dissolved in DMSO (3 mL) for normal-phase purification (Biotage Sfär Silica, 25 g, ethyl acetate:hexane 0-25%). Fractions containing product were concentrated *in vacuo* to yield the title compound (661 mg, 1.98 mmol, 19.8%) as a colourless solid. $\boldsymbol{R_f}$ 0.72 (50% EA/hexane); $\boldsymbol{\delta_H}$ **(400 MHz, $CDCl_3$-*d₁*);** 7.94 (2H, d, *J* 9.0, H-2' & 6'), 6.88 (2H, d, *J* 9.0, H-3' & 5'), 4.33 (2H, q, *J* 7.0, CH_2), 3.59 (4H, t, *J* 5.0, H-2 & 6), 3.30 (4H, t, *J* 5.0, H-3 & 5), 1.48 (9H, s, $C(CH_3)_3$), 1.37 (3H, t, *J* 7.0, CH_3); $\boldsymbol{\delta_C}$ **(125 MHz, $CDCl_3$-*d₁*);** 166.6 (ester CO), 154.7 (carbamate CO), 153.8 (C-4), 131.2 (C-3' & 5'), 120.9 (C-1'), 114.2 (C-2' & 6'), 80.2 (CH_2), 80.1 ($C(CH_3)_3$), 60.5 (C-2 & 6), 47.8 (C-3 & 5), 28.4 ($C(CH_3)_3$), 14.4 (CH_3).

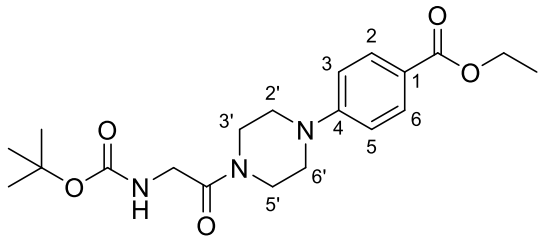
Preparation of ethyl 4-(1-piperazinyl)benzoate (**5.3.2**)[157]



2-Methyl-2-propanyl 4-[4-(ethoxycarbonyl)phenyl]-1-piperazinecarboxylate (200 mg, 0.600 mmol) was dissolved in an 4N HCl dioxane solution (6 mL). The white solid immediately became a cloudy orange solution. The reaction mixture was stirred at room temperature for 2 hours. LCMS indicated reaction completion. Reaction mixture was

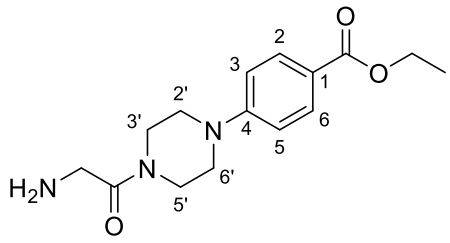
concentrated *in vacuo* to yield the title compound (199 mg, 0.600 mmol, quantitative yield) as a light brown HCl salt. **δ_H (400 MHz, DMSO-*d_6*);** 7.82 (2H, d, *J* 9.0, H-2 & 6), 7.04 (2H, d, *J* 9.0, H-3 & 5), 4.25 (2H, q, *J* 7.0, CH_2), 3.56 (4H, s, H-3' & 5'), 3.18 (4H, s, H-2' & 6'), 1.29 (3H, t, *J* 7.0, CH_3); **δ_C (100 MHz, DMSO-*d_6*);** 166.0 (ester CO), 153.5 (C-4), 131.2 (C-3 & 5), 120.2 (C-1), 114.6 (C-2 & 6), 60.5 (C-3' & 5'), 44.4 (C-2' & 6'), 42.7 (CH_2), 14.8 (CH_3); **LCMS** (Found: MH^+, 234.47. $C_{13}H_{18}N_2O_2$ requires *MH*+, 235.14).

Preparation of ethyl 4-(4-((*tert*-butoxycarbonyl)glycyl)piperazin-1-yl)benzoate (**5.3.3**)



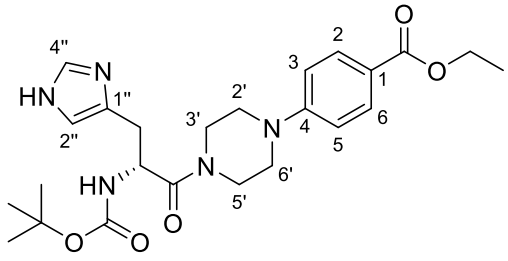
Ethyl 4-(1-piperazinyl)benzoate (70 mg, 0.26 mmol, 1.0 eq.), *N*-{[(2-methyl-2-propanyl)oxy]carbonyl}glycine (46 mg, 0.26 mmol, 1.0 eq.), *N*-[3-(dimethylamino)propyl]-N'-ethylcarbodiimide (54 mg, 0.28 mmol, 1.1 eq.) and 4-dimethylaminopyridine (48 mg, 0.39 mmol, 1.5 eq.) were dissolved in DCM (3 mL) and stirred at room temperature over the weekend. LCMS indicated product formation. The reaction mixture was diluted with DCM (12 mL) and washed with 5% acetic acid solution (15 mL), saturated $NaHCO_3$ solution (15 mL) and brine (15 mL). The organic layer was dried with $MgSO_4$, filtered and concentrated *in vacuo* to yield the title compound (82 mg, 0.21 mmol, 81%) as a brown solid. **δ_H (500 MHz, $CDCl_3$-*d_1*);** 7.93 (2H, d, *J* 8.5, H-2 & 6), 6.85 (2H, d, *J* 8.5, H-3 & 5), 5.51 (1H, br s, NH), 4.31 (2H, q, *J* 7.0, CH_2), 3.99 (2H, s, NHC*H*$_2$CO), 3.77 (2H, t, *J* 4.5, H-3'/5'), 3.55 (2H, t, *J* 4.5, H-3'/5'), 3.34-3.30 (4H, m, H-2' & 6'), 1.44 (11H, s, $C(CH_3)_3$), 1.35 (3H, t, *J* 7.0, CH_3); **δ_C (125 MHz, $CDCl_3$-*d_1*);** 167.1 (amide CO), 166.5 (ester CO), 155.8 (carbamate CO), 153.5 (C-4), 131.2 (C-3 & 5), 121.2 (C-1), 114.2 (C-2 & 6), 79.8 ($C(CH_3)_3$), 60.5 (C-3' & 5'), 47.6 (C-2' & 6'), 43.9 ($NHCH_2CO$), 42.2 (OCH_2), 28.4 ($C(CH_3)_3$), 14.4 (CH_3); **LCMS** (Found: MH^+(-Boc), 291.50. $C_{20}H_{29}N_3O_5$ requires *MH*+, 392.22).

Preparation of ethyl 4-[4-(2-aminoacetyl)piperazin-1-yl]benzoate (**5.3.4**)



Ethyl 4-(4-((*tert*-butoxycarbonyl)glycyl)piperazin-1-yl)benzoate (120 mg, 0.310 mmol) was dissolved in 4N HCl dioxane solution (3 mL) with MeOH (1 mL) to aid dissolution. LCMS after 4 hours showed a clean UV trace and mass peak. The reaction mixture was concentrated, dissolved in DMSO (1 mL) and purified by reverse-phase chromatography (Biotage Sfär C18, 12 g, 20-80% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (49 mg, 0.17 mmol, 54%) as a colourless residue with a small water impurity. **δ_H (500 MHz, MeOD-*d*$_4$);** 7.61 (2H, d, *J* 9.0, H-2 & 6), 6.70 (2H, d, *J* 9.0, H-3 & 5), 4.03 (2H, q, *J* 7.0, OCH_2), 3.48 (2H, br t, *J* 5.5, $COC\underline{H}_2NH_2$), 3.35–3.29 (4H, m, $OCNCH_2$), 3.15–3.06 (4H, m, $ArN(CH_2)_2$), 1.09 (3H, t, *J* 7.0, CH_3);); **δ_C (125 MHz, MeOD-*d*$_4$);** 170.3 (amide CO), 167.0 (ester CO), 154.1 (C-4), 130.8 (C-3 & 5), 119.9 (C-1), 113.7 (C-2 & 6), 60.2 (OCH_2), 41.4 ($OCCH_2NH_2$), 13.3 (CH_3); **LCMS** (Found: MH^+, 292.17. $C_{15}H_{21}N_3O_3$ requires *MH*, 292.17).

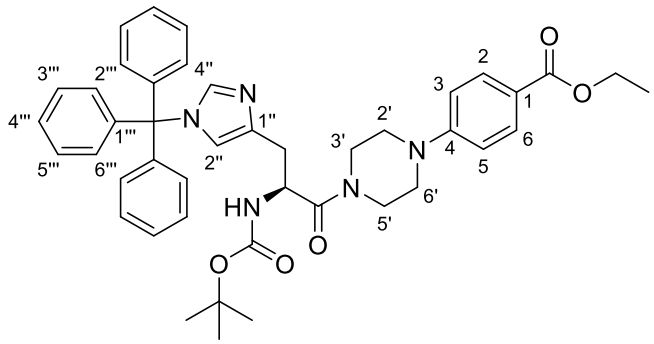
Preparation of ethyl 4-(4-((*tert*-butoxycarbonyl)-*D*-histidyl)piperazin-1-yl)benzoate (**5.3.8**)



Ethyl 4-(1-piperazinyl)benzoate (100 mg, 0.37 mmol, 1.0 eq.), *N*-{[(2-methyl-2-propanyl)oxy]carbonyl}-*D*-histidine (95 mg, 0.37 mmol, 1.0 eq.), *N*-[3-(dimethylamino)propyl]-*N*'-ethylcarbodiimide (79 mg, 0.41 mmol, 1.1 eq.) and 4-dimethylaminopyridine (67 mg, 0.55 mmol, 1.5 eq.) were dissolved in DCM (5 mL) and stirred at room temperature over the weekend. LCMS indicated product formation. The reaction mixture was diluted with DCM (15 mL) and washed with 5% acetic acid solution (20 mL), saturated $NaHCO_3$ solution (20 mL) and brine (20 mL). Significant product

remained in the combined aqueous layers, which was extracted with DCM (3 x 10 mL) and combined with the previous organic layers, then concentrated to yield the title compound (151 mg, 0.320 mmol, 87%) as a brown oil. **δ_H (500 MHz, $CDCl_3$-*d_1*);** 7.91 (2H, d, *J* 9.0, H-2 & 6), 7.53 (1H, s, H-4''), 6.82 (1 H, s, H-2''), 6.79 (2 H, d, *J* 9.0, H-3 & 5), 5.56 (1H, br s, NH-3'), 4.88 (1H, br s, *J* 7.0, NH), 4.31 (2H, q, *J* 7.0, CH_2), 3.78-3.61 (3H, m, NHC*H*CO & H-3'/5'), 3.52-3.44 (2H, m, H-3'/5'), 3.30-3.17 (3H, m, H-2' & 6'), 3.07-2.90 (3H, m, Ar''CH_2), 1.40 (9H, s, $C(CH_3)_3$), 1.35 (3H, t, *J* 7.0, CH_3); **δ_C (125 MHz, $CDCl_3$-*d_1*);** 170.6 (amide CO), 166.6 (ester CO), 155.3 (carbamate CO), 153.5 (C-4), 135.2 (C-1''), 131.2 (C-3 & 5), 120.9 (C-1) 114.0 (C-2 & 6), 106.6 (C-2''), 80.1 ($C(CH_3)_3$), 60.5 (C-3' & 5'), 50.1 (NHC*H*CO), 47.5 (C-2' & 6'), 45.0 (OCH_2), 41.7 ($ArCH_2$), 28.3 ($C(CH_3)_3$), 14.4 (CH_3), missing C-4''; **LCMS** (Found: MH^+, 471.96. $C_{24}H_{33}N_5O_5$ requires *MH*+, 472.26).

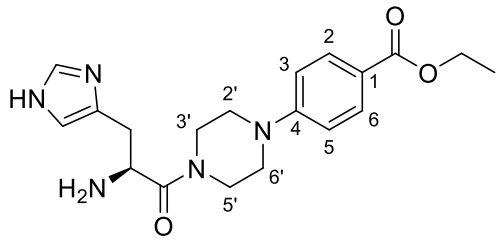
Preparation of ethyl 4-(4-(*N*-(*tert*-butoxycarbonyl)-*N*-trityl-*L*-histidyl)piperazin-1-yl)benzoate (**5.3.5**)



Ethyl 4-(1-piperazinyl)benzoate (100 mg, 0.37 mmol, 1.0 eq.), *N*-{[(2-methyl-2-propanyl)oxy]carbonyl}-1-trityl-*L*-histidine (184 mg, 0.37 mmol, 1.0 eq.), *N*-[3-(dimethylamino)propyl]-*N*'-ethylcarbodiimide (79 mg, 0.41 mmol, 1.1 eq.) and 4-dimethylaminopyridine (67 mg, 0.55 mmol, 1.5 eq.) were dissolved in DCM (5 mL) and stirred at room temperature over the weekend. LCMS indicated product formation. The reaction mixture was diluted with DCM (15 mL) and washed with 5% acetic acid solution (20 mL), saturated $NaHCO_3$ solution (20 mL) and brine (20 mL). Combined aqueous layers were further extracted with DCM (3 x 10 mL). Combined organic layers were dried with $MgSO_4$, filtered and dry-loaded onto silica for normal-phase purification. (Biotage Sfär Silica, 10 g, ethyl acetate:hexane 60-100%). Fractions containing product were concentrated *in vacuo* to yield the title compound (86 mg, 0.12 mmol, 33%) as a colourless oil. **δ_H (500 MHz, $CDCl_3$-*d_1*);** 7.94 (2H, d, *J* 9.0, H-2 & 6), 7.30 (1H, br s, H-4''), 7.32-7.27 (9H, m, H-3''' & 4''' & 5'''), 7.11-7.06 (6H, m, H-2''' & 6'''), 6.81 (2H, d, *J*

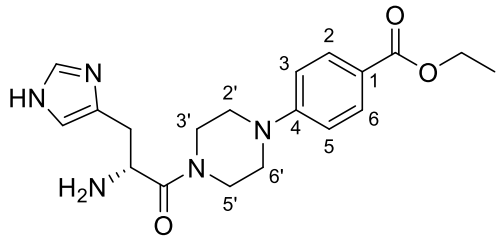
9.0, H-3 & 5), 6.62 (1H, s, H-2''), 4.96 (1H, q, *J* 7.0, NHC*H*CO), 4.34 (2H, q, *J* 7.0, OCH_2), 3.77 (2H, br s, H-3'/5'), 3.59 (2H, br s, H-3'/5'), 3.48 (1H, s, NHC*H*CO), 3.31-3.16 (4H, m, H-2' & 6'), 2.96 (1H, dd, *J* 14.5, 6.5, Ar''C*H*H), 2.84 (1H, dd, *J* 14.5, 6.5, Ar''CH*H*),1.75 (3H, br s, H-), 1.41-1.35 (12H, m, $C(CH_3)_3$ & CH_3); **δ_C (125 MHz, $CDCl_3$-*d_1*);** 170.6 (amide CO), 166.6 (ester CO), 155.2 (carbamate CO), 153.6 (C-4),142.3 (C-1), 138.6 (C-2''), 131.2 (C-3 & 5), 129.7 (C-3''' & 5'''), 128.1 (C-2''' & 6'''), 120.9 (C-4''), 119.5 (C-1''), 113.9 (C-2 & 6), 79.6 ($C(CH_3)_3$), 75.3 (C-1'''), 60.5 (C-3' & 5'), 50.0 (NHC*H*CO), 47.6 (C-2' & 6'), 45.1 (OCH_2), 41.6 ($ArCH_2$), 32.5 (C), 28.4 ($C(CH_3)_3$), 14.4 (CH_3), missing C-4'''; **LCMS** (Found: MH^+, 714.37. $C_{43}H_{47}N_5O_5$ requires *MH+*, 714.37).

Preparation of ethyl 4-{4-[(2S)-2-amino-3-(1H-imidazol-4-yl)propanoyl]piperazin-1-yl}benzoate (**5.3.6**)



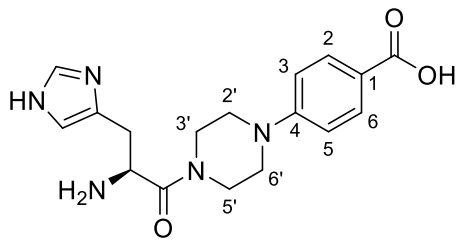
Ethyl 4-(4-(*N*-(*tert*-butoxycarbonyl)-*N*-trityl-*L*-histidyl)piperazin-1-yl)benzoate (86 mg, 0.12 mmol, 1.0 eq.) was dissolved in DCM (3 mL) before addition of TFA (60 µL, 0.78 mmol, 6.5 eq.). The reaction mixture was stirred at room temperature for 60 hours. LCMS indicated formation of significant product. The reaction mixture was concentrated and dissolved in DMSO (1 mL) for reverse-phase purification (Biotage Sfär C18, 12 g, 0-60% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (29 mg, 0.078 mmol, 65%) as a yellow oil. **HPLC** 96.8% (RT = 1.54 min); **δ_H (500 MHz, $CDCl_3$-*d_1*);** 7.94 (2 H, d, *J* 8.5, H-2 & 6), 7.56 (1 H, s, H'-2), 6.87-6.79 (3 H, m, H-3 & 5 and H'-5), 4.33 (2 H, q, *J* 7.0, OCH_2), 4.04 (1 H, br t, *J* 6.5, NH_2CH), 3.84-3.58 (4 H, m, $CON(CH_2)_2$), 3.37-3.12 (4 H, m, $ArN(CH_2)_2$), 2.92 (1 H, dd, *J* 15.0, 5.0, cis CHC*H*H), 2.76 (1 H, dd, *J* 15.0, 8.0, trans CHCH*H*), 1.37 (3 H, t, *J* 7.0, CH_3); **δ_C (125 MHz, $CDCl_3$-*d_1*);** 173.47 (amide CO), 166.67 (ester CO), 153.67 (C-4), 135.05 (C'-2), 131.37 (C-2 & 6 and C'-4), 121.21 (C-1), 114.23 (C-3 & 5), 60.66 (OCH_2), 51.64 (NH_2CH), 48.00 ($ArN(CH_2)_2$), 47.70 ($ArN(CH_2)_2$), 44.98 ($CON(CH_2)_2$), 41.82 ($CON(CH_2)_2$), 33.13 ($CHCH_2$), 14.55 (CH_3); **HRMS m/z (ES)** (Found: MH^+, 372.2036. $C_{19}H_{25}N_5O_3$ requires *MH+*, 372.2030).

Preparation of ethyl 4-{4-[(2R)-2-amino-3-(1H-imidazol-4-yl)propanoyl]piperazin-1-yl}benzoate (**5.3.9**)



Ethyl 4-(4-((*tert*-butoxycarbonyl)-*D*-histidyl)piperazin-1-yl)benzoate (381 mg, 0.810 mmol) was dissolved in 4N HCl in dioxane (10 mL) and stirred at room temperature for 60 hours. LCMS indicated complete product formation. The reaction mixture was concentrated to yield the title compound (386 mg, 0.950 mmol, quantitative yield) as a yellow HCl salt. **HPLC** >99% (RT = 1.53 min); **δ_H (500 MHz, MeOD-*d*$_4$);** 7.86 (2 H, d, *J* 9.0, H-2 & 6), 7.64 (1 H, d, *J* 1.0, H'-2), 6.92–6.88 (3 H, m, H-3 & 5 and H'-5), 4.28 (2 H, q, *J* 7.0, OCH_2), 4.15 (1 H, t, *J* 7.0, NHC*H*), 3.81-3.73 (1 H, m, $CON(CH_2)_2$), 3.66-3.57 (2 H, m, $CON(CH_2)_2$), 3.55-3.49 (1 H, m, $CON(CH_2)_2$), 3.40–3.30 (2 H, m, $ArN(CH_2)_2$), 3.18-3.09 (1 H, m, $ArN(CH_2)_2$), 2.94–2.85 (2 H, m, $ArN(CH_2)_2$ and cis CHC*H*H), 2.81 (1 H, dd, *J* 14.0, 7.0, trans CHCH*H*), 1.34 (3 H, t, *J* 7.0, CH_3); **δ_C (125 MHz, MeOD-*d*$_4$);** 174.68 (amide CO), 168.25 (ester CO), 155.29 (C-4), 136.52 (C'-2), 132.11 (C-2 & 6 and C'-4), 121.15 (C-1), 114.95 (C-3 & 5), 61.58 (OCH_2), 51.60 (NH_2CH), 48.52 ($ArN(CH_2)_2$), 48.15 ($ArN(CH_2)_2$), 45.96 ($CON(CH_2)_2$), 42.81 ($CON(CH_2)_2$), 34.58 ($CHCH_2$), 14.71 (CH_3); **HRMS m/z (ES)** (Found: MH^+, 372.2037. $C_{19}H_{25}N_5O_3$ requires *MH*+, 372.2030).

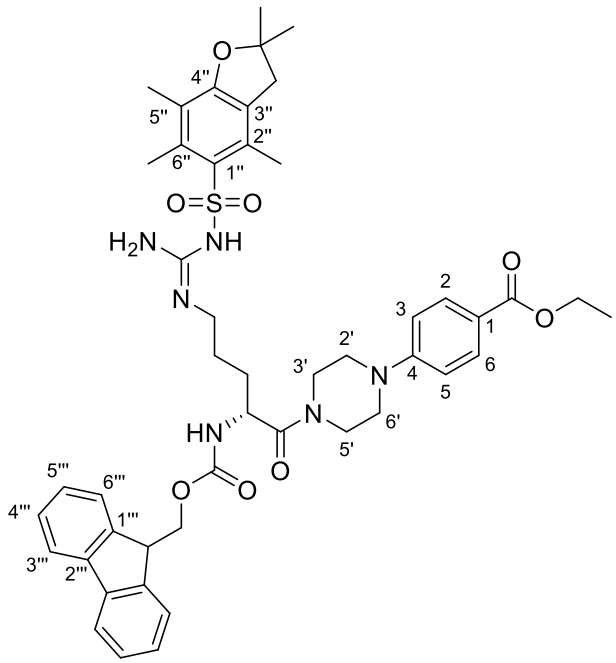
Preparation of 4-{4-[(2S)-2-amino-3-(1H-imidazol-4-yl)propanoyl]piperazin-1-yl}benzoic acid (**5.3.7**)



4-{4-[(2S)-2-amino-3-(1H-imidazol-4-yl)propanoyl]piperazin-1-yl}benzoate (29 mg, 0.080 mmol) was dissolved in 35% $NH_4OH_{(aq)}$ (4 mL) and MeOH (2 mL), then stirred at room temperature. Progress was slow and monitored by LCMS. At 230 hours LCMS

indicated clean product formation. The reaction mixture was concentrated and dissolved in DMSO (1 mL) for reverse-phase purification (Biotage Sfär C18, 4.5 g, 10-90% MeCN:H_2O (+0.1% NEt_3)). The product was co-eluted with DMSO peak during initial elution. The product was concentrated to yield the title compound (2.1 mg, 6.1 µmol, 7.6%) as a colourless residue. **HPLC** 90.5% (RT = 1.24 min); **δ_H (500 MHz, D_2O-d_2);** 7.89 (1 H, d, *J* 1.0, H'-2), 7.84 (2 H, d, *J* 9.0, H-2 & 6), 7.13 (1 H, br d, *J* 1.0, H'-5), 7.05 (2 H, d, *J* 9.0, H-3 & 5), 4.74 (1 H, dd, *J* 8.0, 6.5, NH_2CH), 3.85-3.75 (1 H, m, $CON(CH_2)_2$), 3.68–3.51 (3 H, m, $CON(CH_2)_2$), 3.49–3.39 (2 H, m, $ArN(CH_2)_2$), 3.38–3.25 (2 H, m, $CON(CH_2)_2$), 3.25–3.09 (2 H, m, $CHCH_2$); **HRMS m/z (ES)** (Found: MH^+, 344.1715. $C_{17}H_{21}N_5O_3$ requires *MH*+, 344.1717).

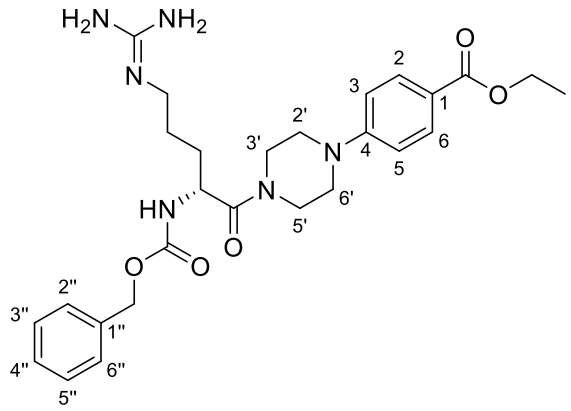
Preparation of ethyl (*R,Z*)-4-(4-(2-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-5-((amino((2,2,4,6,7-pentamethyl-2,3-dihydrobenzofuran)-5-sulfonamido)methylene)amino)pentanoyl)piperazin-1-yl)benzoate (**5.3.12**)



Ethyl 4-(1-piperazinyl)benzoate (100 mg, 0.37 mmol, 1.0 eq.), N^2-[(9*H*-fluoren-9-ylmethoxy)carbonyl]-N^5-{*N*-[(2,2,4,6,7-pentamethyl-2,3-dihydro-1-benzofuran-5-yl)sulfonyl]carbamimidoyl}-*D*-ornithine (240 mg, 0.37 mmol, 1.0 eq.), *N*-[3-(dimethylamino)propyl]-*N*'-ethylcarbodiimide (79 mg, 0.41 mmol, 1.1 eq.) and 4-dimethylaminopyridine (67.0 mg, 0.550 mmol, 1.50 eq.) were dissolved in DCM (5 mL) and stirred at room temperature over the weekend. LCMS indicated product formation. The reaction mixture was diluted with DCM (15 mL) and washed with 5% acetic acid solution (20 mL), saturated $NaHCO_3$ solution (20 mL) and brine (20 mL). Combined

aqueous layers were further extracted with DCM (3 x 10 mL). Combined organic layers were dried with $MgSO_4$, filtered and dry-loaded onto silica for normal-phase purification (Biotage Sfär Silica, 10 g, ethyl acetate:hexane 60-100%). Fractions containing product were concentrated *in vacuo* to yield the title compound (133 mg, 0.15 mmol, 42%) as a colourless solid. **δ_H (500 MHz, $CDCl_3$-*d*$_1$);** 7.93 (2H, d, *J* 9.0, H-2 & 6), 7.75 (2H, d, *J* 7.5, H-5'''), 7.57 (2H, t, *J* 6.5, H-2'''), 7.39 (2H, q, *J* 7.5, H-3'''), 7.29 (2H, t, *J* 7.5, H-4'''), 6.86 (2H, d, *J* 9.0, H-3 & 5), 6.06 (2H, br s, NH_2), 5.99 (1H, d, *J* 8.5, O_2CNH), 4.72-4.65 (1H, m, OC*H*$(CH_3)_2$), 4.47-4.39 (1H, m, N_2CNC*H*H), 4.39-4.29 (3H, m, OCH_2 & N_2CNCH*H*), 4.18 (1H, t, *J* 7.0, OCH_2C*H*), 3.84-3.58 (4H, m, H-3' & 5'), 3.49 (2H, s, NCO_2CH_3), 3.42 (2H, br s, H-2'/6'), 3.31 (2H, br s, H-2'/6'), 3.22-3.15 (1H, m, NHC*H*CO), 2.92 (2H, s, $ArCH_2$), 2.58 (3H, s, Ar6''CH_3), 2.52 (3H, s, Ar2''CH_3), 2.08 (3H, s, Ar5''CH_3), 1.79-1.53 (9H, m, NCH_2C*H*$_2C$*H*$_2CH$), 1.43 (6H, s, OCH(C*H*$_3)_2$), 1.37 (3H, t, *J* 7.0, CH_3), missing sulfonamide NH; **δ_C (125 MHz, $CDCl_3$-*d*$_1$);** 170.1 (amide CO), 166.6 (ester CO), 158.9 (guanidine CN_3), 156.8 (C-4''), 155.8 (carbamate CO), 153.4 (C-4), 143.7 (C-1'''), 143.5 (C-6'''), 141.4 (C-1''), 138.5 (C-5''), 132.4 (C-3''), 131.2 (C-3 & 5), 127.8 (C-2''), 127.1 (C-6''), 125.0 (C-Ar'''), 124.7 (C-Ar'''), 121.0 (C-1), 120.1 (C-Ar'''), 117.6 (C-Ar'''), 114.2 (C-2 & 6), 60.5 (C-3' & 5'), 50.9 (NH*C*HCO), 47.8 (C-2' & 6'), 47.4 (N_2CN*C*H_2), 47.1 (Ar''OCH), 45.2 (N_2CNCH_2*C*H_2), 43.2 (O*C*H_2CH), 41.8 (O*C*H_2CH_3), 28.6 (OCH(*C*$H_2)_2$), 24.2 (Ar3''CH_2), 19.3 (Ar6''CH_3), 18.0 (Ar2''CH_3), 14.4 (OCH_2*C*H_3), 14.2 (Ar5''CH_3), 12.5 (Ar'CH), missing $NHCHCH_2$*C*H_2; **LCMS** (Found: MH_2^{2+}, 432.85. $C_{47}H_{56}N_6O_8S$ requires *MH*$_2$*2+*, 433.20).

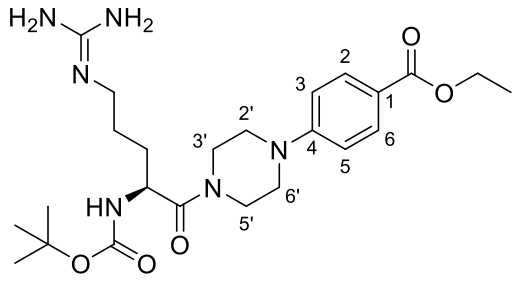
Preparation of ethyl (*R*)-4-(4-(2-(((benzyloxy)carbonyl)amino)-5-((diaminomethylene)amino)pentanoyl)piperazin-1-yl)benzoate (**5.3.10A**)



Ethyl 4-(1-piperazinyl)benzoate (100 mg, 0.37 mmol, 1.0 eq.), N^2-[(benzyloxy)carbonyl]-*L*-arginine (240 mg, 0.37 mmol, 1.0 eq.), *N*-[3-(dimethylamino)propyl]-*N*'-

ethylcarbodiimide (79 mg, 0.41 mmol, 1.1 eq.) and 4-dimethylaminopyridine (67 mg, 0.55 mmol, 1.5 eq.) were dissolved in DCM (5 mL) and stirred at room temperature over the weekend. LCMS indicated product formation. The reaction mixture was diluted with DCM (15 mL) and washed with 5% acetic acid solution (20 mL), saturated $NaHCO_3$ solution (20 mL) and brine (20 mL). The combined aqueous layers were further extracted with DCM (3 x 10 mL). The combined organic layers were dried with $MgSO_4$, filtered and concentrated *in vacuo* to yield the title compound (103 mg, 0.20 mmol, 53%) as a light brown oil. **δ_H (500 MHz, $CDCl_3$-*d*$_1$);** 7.95-7.80 (3H, m, H-2 & 6), 7.24-7.06 (5H, m, Ar''), 6.74 (2H, d, *J* 8.5, H-3 & 5), 6.50 (1H, d, *J* 5.5, NH), 5.06 (1H, d, *J* 12.5, $ArCH_2$), 4.97 (1H, d, *J* 12.5, Ar''CH_2), 4.58 (1H, br s, NHC*H*CO), 4.30 (2H, q, *J* 7.0, O*C*H_2CH_3), 3.82-3.67 (2H, m, H-3'/5'), 3.67-3.55 (1H, m, NH), 3.49-3.30 (2H, m, H-3'/5'), 3.28-3.10 (4H, m, H-2' & 6'), 3.01 (3H, s, N_2CNCH_2), 1.80-1.57 (4H, m, N_2CNCH_2C*H*$_2$C*H*$_2$), 1.34 (3H, t, *J* 7.0, CH_3); **δ_C (125 MHz, $CDCl_3$-*d*$_1$);** 170.4 (amide CO), 166.6 (ester CO), 157.4 (guanidine CN_3), 156.5 (carbamate CO), 153.4 (C-4), 136.3 (C-1''), 131.2 (C-3 & 5), 128.6 (C-2'' & 6''), 128.2 (C-4''), 128.0 (C-3'' & 5''), 120.7 (C-1), 114.0 (C-2 & 6), 67.0 (OCH_2Ar''), 60.5 (C-2' & 6'), 50.3 (NH*C*HCO), 47.3 (C-3' & 5'), 45.1 (N_2CNCH_2), 41.9 (O*C*H_2CH_3), 29.7 ($N_2CNCH_2CH_2$), 24.8 (NHCH*C*H_2), 14.4 (CH_3); **LCMS** (Found: MH^+, 525.19. $C_{27}H_{36}N_6O_5$ requires *MH*+, 525.28).

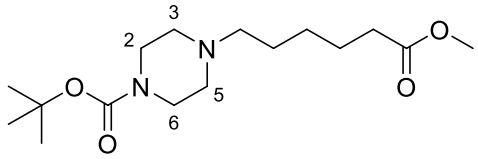
Preparation of ethyl 4-{4-[(2S)-2-{[(*tert*-butoxy)carbonyl]amino}-5-carbamimidamidopentanoyl]piperazin-1-yl}benzoate (**5.3.10B**)



Ethyl 4-(1-piperazinyl)benzoate (300 mg, 1.11 mmol, 1.00 eq.), (2S)-2-{[(*tert*-butoxy)carbonyl]amino}-5-[(diaminomethylidene)amino]pentanoic acid (304 mg, 1.11 mmol, 1.00 eq.), *N*-[3-(dimethylamino)propyl]-*N*'-ethylcarbodiimide (234 mg, 1.22 mmol, 1.10 eq.) and 4-dimethylaminopyridine (203 mg, 1.66 mmol, 1.50 eq.) were dissolved in DCM (15 mL) and stirred at room temperature for 110 hours. The reaction was monitored by LCMS, which showed unreacted starting material. The reaction mixture was diluted with DCM (15 mL), then washed with brine (2 x 30 mL), dried with $MgSO_4$ and filtered. The material was dry-loaded onto silica for normal-phase purification (Biotage Sfär Silica,

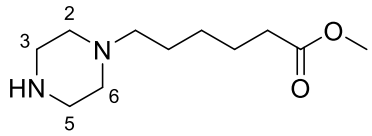
10 g, 0-15% MeOH:DCM). The fractions containing product were concentrated *in vacuo* to yield the title compound (213 mg, 0.434 mmol, 39%) as a colourless oil. **δ_H (500 MHz, MeOD-*d*$_4$);** 8.08 (2 H, br s, guanidine NH), 7.88 (2 H, d, *J* 9.0, H-2 & 6), 6.98 (2 H, d, *J* 9.0, H-3 & 5), 6.77 (3 H, br s, guanidine NH), 4.61-4.55 (1 H, m, NHC*H*), 4.30 (2 H, q, *J* 7.0, OCH_2), 3.88-3.60 (4 H, m, $CON(CH_2)_2$), 3.50-3.32 (4 H, m, $ArN(CH_2)_2$), 3.29-3.15 (2 H, m, $NH_2CNHNHCH_2$), 1.83-1.60 (4 H, m, $CHCH_2CH_2$), 1.44 (9 H, s, $(CH_3)_3$), 1.36 (3 H, t, *J* 7.0, CH_2CH_3); **δ_C (125 MHz, MeOD-*d*$_4$);** 172.49 (amide CO), 168.32 (ester CO), 158.60 (guanidino CNH), 157.81 (C-4), 157.38 (carbamate CO), 132.15 (C-2 & 6), 121.27 (C-1), 115.06 (C-3 & 5), 80.76 ($C(CH_3)_3$), 61.60 (OCH_2), 51.22 (NHCH), 46.27 ($NH_2CNHNHCH_2$), 43.06 ($CON(CH_2)_2$), 42.08 ($CON(CH_2)_2$), 39.55 ($ArN(CH_2)_2$), 30.32 ($CHCH_2$), 28.71 (($CH_3)_3$), 26.16 ($CH_2CH_2CH_2$), 14.70 (CH_2CH_3); **LCMS** (Found: MH^+, 491.11. $C_{24}H_{38}N_6O_5$ requires *MH*+, 491.30).

Preparation of *tert*-butyl 4-(6-methoxy-6-oxohexyl)piperazine-1-carboxylate (**5.3.13**)[158]



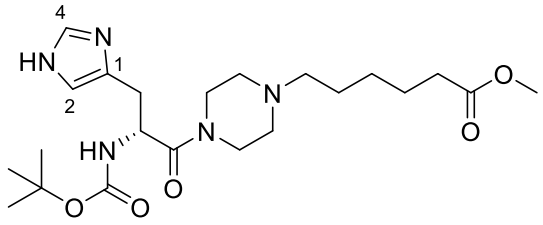
2-Methyl-2-propanyl 1-piperazinecarboxylate (279 mg, 1.50 mmol, 1.00 eq.) was dissolved in THF (1.5 mL) before addition of methyl 6-bomohexanoate (0.30 mL, 1.8 mmol, 1.2 eq.) and triethylamine (0.25 mL, 1.8 mmol, 1.2 eq.). The reaction mixture was stirred at room temperature overnight. White precipitate was formed over the course of the reaction. LCMS indicated product formation. The reaction mixture was filtered, with the solid rinsed with THF (3 x 2 mL). TLC using a $KMnO_4$ stain showed product presence. Filtrate was dry-loaded for normal-phase purification (Biotage Sfär Silica, 10 g, MeOH:DCM, 0-5%). Fractions containing product were concentrated *in vacuo* to yield the title compound (214 mg, 0.680 mmol, 45%) as a colourless oil. ***R*$_f$** 0.20 (3% MeOH/DCM); **δ_H (500 MHz, $CDCl_3$-*d*$_1$);** 3.66 (3H, d, *J* 1.0, CH_3), 3.47 (12H, d, *J* 5.5, H-2 & 6), 3.44 (4H, s, H-3 & 5), 2.43-2.33 (5H, m, NCH_2), 2.31 (2H, t, *J* 7.5, O_2CCH_2), 1.68-1.58 (2H, m, NCH_2CH_2), 1.56-46 (2H, m, $O_2CCH_2CH_2$), 1.45 (9H, d, *J* 2.0, $C(CH_3)_3$), 1.38-1.28 (5H, m, $CH_2CH_2CH_2$); **δ_C (125 MHz, $CDCl_3$-*d*$_1$);** 174.2 (ester CO), 154.8 (carbamate CO), 79.7 ($C(CH_3)_3$), 58.4 (C-2 & 6), 53.0 (C-3 & 5), 51.5 (NCH_2), 50.8 (CH_3), 34.0 (CH_2CO_2), 28.4 ($C(CH_3)_3$), 27.0 (NCH_2CH_2), 26.3 ($O_2CCH_2CH_2$), 24.8 ($NCH_2CH_2CH_2$); **LCMS** (Found: MH^+, 315.23. $C_{16}H_{30}N_2O_4$ requires *MH*+, 315.23).

Preparation of methyl 6-(piperazin-1-yl)hexanoate (**5.3.14**)[159]



Tert-butyl 4-(6-methoxy-6-oxohexyl)piperazine-1-carboxylate (203 mg, 0.650 mmol) was dissolved in 4 N HCl dioxane solution (5 mL) and stirred at room temperature for 60 hours. LCMS indicated product formation. The reaction mixture was concentrated *in vacuo* to yield the title compound (189 mg, 0.650 mmol, quantitative yield) as a colourless oil. **δ_H (500 MHz, DMSO-*d*$_6$);** 3.65 (2H, br s, NCH_2), 3.59 (3H, s, CH_3), 3.55-3.34 (H, m, H-2 & 6), 3.28-3.03 (4H, m, H-3 & 5), 2.32 (2H, t, *J* 7.5, O_2CCH_2), 1.69 (2H, br s, NCH_2CH_2), 1.60-1.50 (2H, m, $O_2CCH_2CH_2$), 1.35-1.25 (2H, m, $CH_2CH_2CH_2$); **δ_C (125 MHz, DMSO-*d*$_6$);** 173.7 (ester CO), 55.7 (C-3 & 5), 51.7 (CH_3), 48.1 (C-2 & 6), 40.4 (NCH_2), 33.4 (CH_2CO_2), 25.8 (NCH_2CH_2), 24.3 ($O_2CCH_2CH_2$), 23.1 ($NCH_2CH_2CH_2$); **LCMS** (Found: MH^+, 214.45. $C_{11}H_{22}N_2O_2$ requires *MH*+, 215.14).

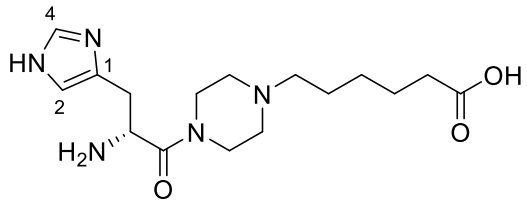
Preparation of methyl 6-{4-[(2R)-2-{[(*tert*-butoxy)carbonyl]amino}-3-(1H-imidazol-4-yl)propanoyl]piperazin-1-yl}hexanoate (**5.3.15**)



Methyl 6-(piperazin-1-yl)hexanoate (90 mg, 0.42 mmol, 1.0 eq.), N-{[(2-Methyl-2-propanyl)oxy]carbonyl}-D-histidine (107 mg, 0.42 mmol, 1.0 eq.), *N*-[3-(dimethylamino)propyl]-*N*'-ethylcarbodiimide (88 mg, 0.46 mmol, 1.1 eq.) and 4-dimethylaminopyridine (77 mg, 0.63 mmol, 1.5 eq.) were dissolved in dioxane (5 mL) and stirred at room temperature for 24 hours. LCMS showed the Boc-deprotected product mass. The reaction mixture was diluted with ethyl acetate (15 mL), washed with brine (3 x 20 mL), dried with $MgSO_4$, filtered and concentrated. Residue was dissolved in DMSO (1 mL) for reverse-phase purification (Biotage Sfär C18, 12 g, 0-80%

MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (67 mg, 0.15 mmol, 35%) as a colourless oil. **δ_H (500 MHz, $CDCl_3$-d_1);** 7.67 (1 H, s, H-2), 6.84 (1 H, s, H-5), 4.84-4.77 (1 H, m, NHC*H*CO), 3.66 (3 H, s, OCH_3), 3.53-3.30 (4 H, m, $CON(CH_2)_2$), 3.06-2.90 (2 H, m, NCH_2), 2.50-2.18 (8 H, m, $AlkN(CH_2)_2$, $ArCH_2$ and CH_2CO_2Me), 1.68-1.58 (2 H, m, $CH_2CH_2CO_2Me$), 1.51-1.44 (2 H, m, NCH_2CH_2), 1.41 (9 H, s, $(CH_3)_3$), 1.37-1.24 (2 H, m, $CH_2CH_2CH_2CH_2CH_2$); **δ_C (125 MHz, $CDCl_3$-d_1);** 174.30 (ester CO), 170.15 (amide CO), 155.48 (carbamate CO), 135.24 (C-4), 135.12 (C-2), 80.29 ($C(CH_3)_3$), 53.17 ($AlkN(CH_2)_2$), 52.76 ($AlkN(CH_2)_2$), 52.48 (CH_2CO_2Me), 51.68 (OCH_3), 50.03 (NHCH), 45.62 ($CON(CH_2)_2$), 45.40 ($CON(CH_2)_2$), 34.09 ($ArCH_2$), 31.26 (NCH_2), 28.46 ($(CH_3)_3$), 27.07 ($CH_2CH_2CH_2CH_2CH_2$), 26.41 (NCH_2CH_2), 24.90 ($CH_2CH_2CO_2Me$). Missing C-5; **LCMS** (Found: MH^+ (-Boc), 351.70. $C_{22}H_{37}N_5O_5$ requires *MH*+, 452.29).

Preparation of 6-{4-[(2R)-2-amino-3-(1H-imidazol-4-yl)propanoyl]piperazin-1-yl}hexanoic acid (**5.2.3**)

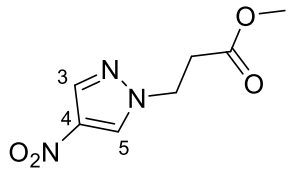


Methyl 6-{4-[(2R)-2-{[(*tert*-butoxy)carbonyl]amino}-3-(1H-imidazol-4-yl)propanoyl]piperazin-1-yl}hexanoate (67 mg, 0.15 mmol) was dissolved in 4N HCl in dioxane (3 mL) and MeOH (2 mL) and stirred at room temperature for 18 hours. LCMS indicated the Boc group was removed. The reaction mixture was concentrated and dry-loaded onto silica for reverse-phase purification (Biotage Sfär C18, 12 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). During the column chromotography the methyl ester was deprotected to the carboxylic acid. Fractions containing product were concentrated to yield the title compound (30 mg, 0.089 mmol, 59%) as a colourless oil. **δ_H (500 MHz, MeOD-d_4);** 7.70 (1 H, s, H-2), 6.96 (1 H, s, H-4), 4.38 (1 H, t, *J* 7.0, NH_2CH), 3.71-3.33 (4 H, m, $CON(CH_2)_2$), 3.02-2.88 (2 H, m, NCH_2CH_2), 2.54-2.43 (2 H, m, $AlkN(CH_2)_2$), 2.41-2.30 (3 H, m, $AlkN(CH_2)_2$ and $ArCH_2$), 2.19 (2 H, t, *J* 7.5, CH_2CO_2H), 2.10-2.01 (1 H, m, $ArCH_2$), 1.67-1.47 (4 H, m, $CH_2CH_2CH_2CH_2CH_2$), 1.40-1.28 (2 H, m, $CH_2CH_2CH_2CH_2CH_2$); **δ_C (125 MHz, MeOD-d_4);** 181.52 (acid CO), 170.99 (amide CO), 136.94 (C-2), 134.32 (C-4), 117.70 (C-5), 59.25 (NCH_2CH_2), 53.84 ($AlkN(CH_2)_2$), 53.56 ($AlkN(CH_2)_2$), 51.43 (NH_2CH), 46.10 ($CON(CH_2)_2$), 42.85 ($CON(CH_2)_2$), 38.03 ($ArCH_2$), 32.22 (CH_2CO_2H), 28.41 ($CH_2CH_2CH_2CH_2CH_2$), 27.08 ($CH_2CH_2CH_2CH_2CH_2$), 27.02

($CH_2\underline{C}H_2CH_2\underline{C}H_2CH_2$); **HRMS m/z (ES)** (Found: MH^+, 338.2183. $C_{16}H_{27}N_5O_3$ requires *MH*+, 338.2187).

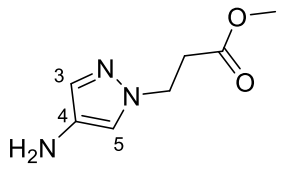
8.6 Synthesis of pyrazole zwitterion targets

Preparation of methyl 3-(4-nitro-1H-pyrazol-1-yl)propanoate (**5.5.1**)[160]



4-Nitro-1H-pyrazole (565 mg, 5.00 mmol, 1.00 eq.), methyl 3-bromopropionate (0.6 mL, 5.5 mmol, 1.2 eq.) and potassium carbonate (690 mg, 5.00 mmol, 1.00 eq.) were mixed and dissolved in DMF (10 mL). The reaction mixture was heated to 60°C and stirred for 70 hours. LCMS indicated complete product formation. The reaction mixture was diluted with water (30 mL) and the product extracted with ethyl acetate (2 x 30 mL). The combined organic layers were washed with 10% LiCl solution (50 mL). The organic layer was dried with $MgSO_4$, filtered and concentrated to yield the title compound (988 mg, 4.96 mmol, 99%) as a yellow oil. **δ_H (500 MHz, MeOD-*d_4*)**; 8.56 (1 H, d, *J* 1.0, H-5), 8.09 (1 H, d, *J* 1.0, H-3), 4.47 (2 H, t, *J* 6.5, NCH_2), 3.66 (3 H, s, CH_3), 2.97 (2 H, t, *J* 6.5, $COCH_2$); **δ_C (125 MHz, MeOD-*d_4*)**; 172.66 (ester CO), 136.71 (C-3 & 4), 131.32 (C-5), 52.42 (CH_3), 49.52 (NCH_2), 34.47 ($CO\underline{C}H_2$); **LCMS** (Found: MH+, 199.37. $C_7H_9N_3O_4$ requires *MH*+, 200.07).

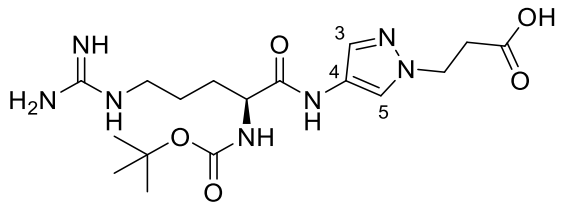
Preparation of methyl 3-(4-amino-1H-pyrazol-1-yl)propanoate (**5.5.2**)[160]



Methyl 3-(4-nitro-1H-pyrazol-1-yl)propanoate (1.86 g, 9.35 mmol, 1.00 eq.) and palladium on carbon (100 mg, 5% w/w) were mixed and dissolved in ethanol (50 mL). The mixture were degassed with N_2, then placed under an H_2 atmosphere using a balloon and stirred for 40 hours, during which the balloon was refilled twice before depletion slowed. LCMS showed reaction completion. The reaction mixture was filtered through a celite pad, washed with ethanol, and concentrated to yield the title compound (884 mg,

5.23 mmol, 90%) as a red oil. **δ_H (500 MHz, MeOD-*d*$_4$)**; 7.21 (1 H, d, *J* 1.0, H-5), 7.14 (1 H, d, *J* 1.0, H-3), 4.29 (2 H, td, *J* 6.5, 1.5, NCH_2), 3.66 (3 H, d, *J* 1.5, CH_3), 2.83 (2 H, td, *J* 6.5, 1.5, $COCH_2$); **δ_C (125 MHz, MeOD-*d*$_4$)**; 173.02 (ester CO), 132.40 (C-3), 130.22 (C-4), 120.88 (C-5), 52.28 (CH_3), 48.44 (NCH_2), 35.63 ($CO\underline{C}H_2$); **LCMS** (Found: MH+, 169.43. $C_7H_{11}N_3O_2$ requires *MH*+, 170.09).

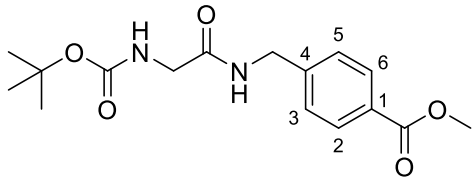
Preparation of 3-{4-[(2S)-2-{[(*tert*-butoxy)carbonyl]amino}-5-carbamimidamidopentanamido]-1H-pyrazol-1-yl}propanoic acid (**5.4.9**)



Methyl 3-(4-amino-1H-pyrazol-1-yl)propanoate (150 mg, 0.890 mmol, 1.00 eq.), N-Boc-L-arginine.HCl (304 mg, 0.980 mmol, 1.10 eq.), EDC (187 mg, 0.980 mmol, 1.10 eq.) and DMAP (22 mg, 0.18 mmol, 0.20 eq.) were mixed and dissolved in DCM (5 mL) and stirred for 18 hours. LCMS indicated the intended esterified product had formed. The reaction mixture was diluted with DCM (15 mL) and washed with sat. $NaHCO_3$ solution (20 mL). The aqueous layer was re-extracted with DCM (20 mL). The product remained in the aqueous layer so 10 M NaOH (3 mL) was added, adjusting the pH to 14. The aqueous layer was extracted with ethyl acetate (3 x 25 mL), with the de-esterified product remaining in the aqueous layer. The aqueous mixture was concentrated and dry-loaded for reverse-phase purification (Modus P 4.5 g, 0-100% MeCN:H_2O (+0.5% TFA)). An impure pink residue (648 mg) was collected which was re-purified by MDAP to yield the title compound (149 mg, 0.360 mmol, 41%) as a colourless oil. **HPLC** 98.0% (RT = 1.40 min); **δ_H (500 MHz, MeOD-*d*$_4$)**; 8.41 (1 H, br s, OH), 7.93 (1 H, s, H-5), 7.55 (1 H, s, H-3), 4.35 (2 H, t, *J* 6.5, NCH_2), 4.18-4.12 (1 H, m, NHC*H*), 3.19 (2 H, t, *J* 7.0, C*H*$_2$NHC(NH)NH_2), 2.78 (2 H, t, *J* 6.5, $COCH_2$), 1.87-1.59 (4 H, m, CH_2C*H*$_2$C*H*$_2$CH), 1.44 (9 H, s, $(CH_3)_3$); **δ_C (125 MHz, MeOD-*d*$_4$)**; 175.84 (acid CO), 171.88 (amide CO),158.68 (guanidine CN_3) 157.88 (carbamate CO), 132.00 (C-3), 123.11 (C-5), 122.24 (C-4), 80.81 (*C*$(CH_3)_3$), 55.71 NHCH), 49.85 (NCH_2), 41.90 (*C*H_2NHC(NH)NH_2), 36.85 ($CO\underline{C}H_2$), 30.66 ($CH_2\underline{C}H_2CH_2$), 28.70 ($(CH_3)_3$), 26.32 ($CH\underline{C}H_2$); **HRMS m/z (ES)** (412.2309. $C_{17}H_{29}N_7O_5$ requires *MH*+, 412.2303).

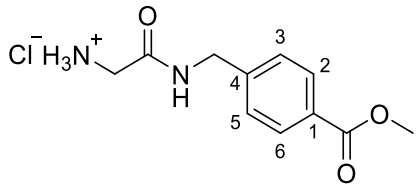
8.7 Synthesis of additional zwitterion targets

Preparation of methyl 4-{[(*N*-{[(2-methyl-2-propanyl)oxy]carbonyl}glycyl)amino]methyl}benzoate (**5.6.2**)[161]



N-{[(2-Methyl-2-propanyl)oxy]carbonyl}glycine (287 mg, 1.64 mmol, 1.10 eq.), methyl 4-(methylamino)benzoate.HCl (300 mg, 1.49 mmol, 1.00 eq.), *N*-[3-(dimethylamino)propyl]-*N*'-ethylcarbodiimide (314 mg, 1.64 mmol, 1.10 eq.) and 4-dimethylaminopyridine (269 mg, 2.20 mmol, 1.50 eq.) were dissolved in DCM (10 mL) and stirred at room temperature overnight. LCMS indicated product formation. The reaction mixture was diluted with DCM (10 mL) and washed with 5% acetic acid solution (20 mL), saturated $NaHCO_3$ solution (20 mL) and brine (20 mL). The organic layer was dried with $MgSO_4$, filtered and concentrated *in vacuo* to yield the title compound (404 mg, 1.25 mmol, 84%) as a colourless solid. **δ_H (500 MHz, $CDCl_3$-*d*$_1$);** 7.95 (2H, dt, *J* 8.5, 1.8, H-2 & 6), 7.30 (2H, d, *J* 8.5, H-3 & 5), 6.90 (1H, br s, O_2CNH), 5.35 (1H, br s, OCNH), 4.48 (2H, d, *J* 6.0, NHC*H*$_2$Ar), 3.89 (3H, s, CH_3), 3.83 (2H, s, NHC*H*$_2$CO), 1.4 (9H, s, $C(CH_3)_3$); **δ_C (125 MHz, $CDCl_3$-*d*$_1$);** 169.8 (amide CO), 166.8 (ester CO), 156.3 (carbamate CO), 143.2 (C-1), 130.0 (C-2 & 6), 129.3 (C-4), 127.4 (C-3 & 5), 80.5 (*C*$(CH_3)_3$), 52.1 (CH_3), 44.5 (NH*C*H_2CO), 43.0 ($NHCH_2$Ar), 28.3 ($C(CH_3)_3$); **LCMS** (Found: MH^+(-Boc), 223.12. $C_{16}H_{22}N_2O_5$ requires *MH*+, 323.16).

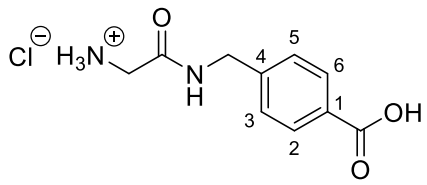
Preparation of methyl 4-[(glycylamino)methyl]benzoate.hydrochloride (**5.6.3**)[161]



Methyl 4-{[(*N*-{[(2-methyl-2-propanyl)oxy]carbonyl}glycyl)amino]methyl}benzoate (200 mg, 0.620 mmol) was dissolved in 4N HCl in dioxane (5 mL) and stirred over the weekend. The reaction mixture was concentrated *in vacuo* using a high vacuum line to yield the title compound (165 mg, 0.620 mmol, quantitative yield) as a fluffy, colourless solid. **HPLC** 100% (RT = 1.15 min); **δ_H (400 MHz, DMSO-*d*$_6$);** 9.01 (1H, br s, NH), 8.13 (2H, br s, NH_2), 7.93 (2H, d, *J* 8.5, H-2 & 6), 7.44 (2H, d, *J* 8.5, H-3 & 5), 4.43 (2H, d, *J*

6.0, NHC*H*$_2$Ar), 3.84 (3H, s, CH_3), 3.63 (2H, s, NHC*H*$_2$CO); **LCMS** (Found: MH^+, 223.11. $C_{11}H_{14}N_2O_3$ requires *MH*+, 223.11).

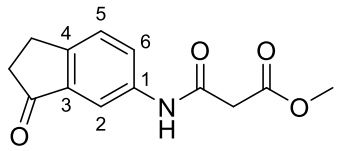
Preparation of 4-[(glycylamino)methyl]benzoic acid.HCl (**5.6.1**)[162]



Methyl 4-[(glycylamino)methyl]benzoate.HCl (55 mg, 0.21 mmol) was dissolved in ammonium hydroxide 35% solution (2 mL) and methanol (3 mL) and stirred at room temperature overnight. LCMS showed no progression so addition ammonium hydroxide (2 mL) was added and left to stir for a total of 11 days. LCMS indicated reaction completion. The reaction mixture was concentrated *in vacuo* to yield the title compound (60 mg, 0.21 mmol, quantitative yield) as a colourless solid containing salt. **HPLC** 85.7% (RT = 0.39 min); **δ_H (400 MHz, DMSO-*d*$_6$);** 8.95 (1H, t, *J* 6.0, NH), 7.95 (1H, s, COOH), 7.84 (2H, d, *J* 8.0, H-2 & 6), 7.72 (2H, br s, NH_2), 7.35 (2H, d, *J* 8.0, H-3 & 5), 4.39 (2H, d, *J* 6.0, NHC*H*$_2$Ar), 3.62 (2H, s, NHC*H*$_2$CO); **LCMS** (Found: MH^+, 208.55. $C_{10}H_{12}N_2O_3$ requires *MH*+, 209.09).

8.8 Synthesis of initial hydantoin targets

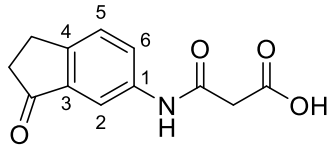
Preparation of methyl 3-oxo-3-((3-oxo-2,3-dihydro-1*H*-inden-5-yl)amino)propanoate (**6.3.2B**)



3-Methoxy-3-oxopropanoic acid (0.16 mL, 1.5 mmol, 1.0 eq.) was dissolved in ethyl acetate (5 mL) before addition of 1-propanephosphonic anhydride (T3p) (50% solution in ethyl acetate) (1.79 mL, 3.00 mmol, 2.00 eq.). The mixture was stirred for 10 minutes at room temperature before sequential addition of 6-aminoindanone (212 mg, 1.50 mmol, 1.00 eq.) and triethylamine (0.620 mL, 4.50 mmol, 3.00 eq.). The reaction mixture was stirred at room temperature overnight. LCMS showed reaction progress. The reaction mixture was diluted with ethyl acetate (15 mL) and washed with 5% acetic acid solution

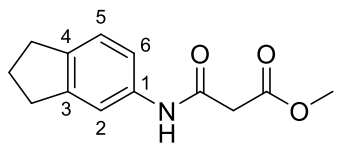
(20 mL), concentrated $NaHCO_3$ solution (20 mL) and brine (20 mL). The organic layer was dried with $MgSO_4$, filtered and dry-loaded for normal-phase purification (Biotage Sfär Silica, 10 g, ethyl acetate:hexane, 0-80%). Fractions containing product were concentrated to yield the title compound (113 mg, 0.460 mmol, 31%) as an off-white solid. **HPLC** 98.2% (RT = 1.57 min); **δ_H (400 MHz, $CDCl_3$-*d_1*);** 7.94 (1H, dd, *J* 8.0, 2.0, H-6), 7.80 (1H, d, *J* 2.0, H-2), 7.45 (1H, d, *J* 8.0, H-5), 3.82 (3H, s, CH_3), 3.51 (2H, s, $OCCH_2CO$), 3.11 (2H, t, *J* 6.0, $OCCH_2CH_2$), 2.76-2.68 (2H, m, $ArCH_2$); **δ_C (100 MHz, $CDCl_3$-*d_1*);** 127.2 (Ar), 114.6 (Ar), 52.8 (OCH_3), 41.0 ($OCCH_2CO$), 36.7 ($OCCH_2CH_2$), 25.4 ($ArCH_2$), missing low intensity peaks; **LCMS** (Found: MH^+, 247.38. $C_{13}H_{13}NO_4$ requires *MH*+, 248.09).

Preparation of 3-oxo-3-((3-oxo-2,3-dihydro-1*H*-inden-5-yl)amino)propanoic acid (**6.3.3B**)



Methyl 3-oxo-3-((3-oxo-2,3-dihydro-1*H*-inden-5-yl)amino)propanoate (112 mg, 0.450 mmol) was dissolved in MeOH (6 mL) before addition of 35% ammonium hydroxide (4 mL). The reaction mixture was stirred at room temperature for 60 hours. LCMS indicated product formation. The reaction mixture was concentrated to yield the title compound (110 mg, 0.450 mmol, quantitative yield) as a yellow oil. **δ_H (500 MHz, DMSO-*d_6*);** 8.00 (1H, d, *J* 2.0, H-2), 7.72 (1H, dd, *J* 8.5, 2.1, H-6), 7.56-7.50 (2H, m, H-5), 7.12 (1H, br s, NH), 3.23 (2H, s, $OCCH_2CO$), 3.04 (2H, t, *J* 5.5, $OCCH_2CH_2$), 2.67-2.60 (2H, m, $ArCH_2$); **δ_C (125 MHz, DMSO-*d_6*);** 206.7 (ketone CO), 169.0 (acid CO), 166.5 (amide CO), 150.5 (C-1), 138.8 (C-3), 137.6 (C-4), 127.7 (C-5), 126.4 (C-6), 112.8 (C-2), 45.1 ($OCCH_2CO$), 36.8 ($OCCH_2CH_2$), 25.4 ($ArCH_2$); **LCMS** (Found: MH^+, 233.36. $C_{12}H_{11}NO_4$ requires *MH*+, 234.08).

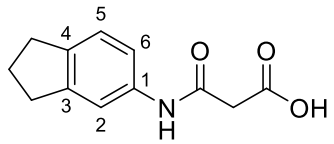
Preparation of methyl 3-((2,3-dihydro-1*H*-inden-5-yl)amino)-3-oxopropanoate (**6.3.2A**)



3-Methoxy-3-oxopropanoic acid (0.310 mL, 3.00 mmol, 1.00 eq.) was dissolved in ethyl acetate (10 mL) before addition of 1-propanephosphonic anhydride (T3p) (50% solution

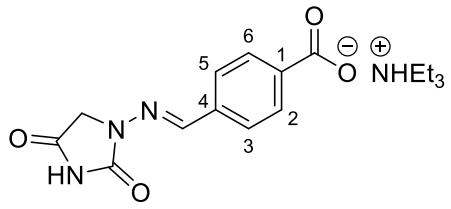
in ethyl acetate) (3.57 mL, 6.00 mmol, 2.00 eq.). The mixture was stirred for 10 minutes at room temperature before sequential addition of 5-indanamine (400 mg, 3.00 mmol, 1.00 eq.) and triethylamine (1.23 mL, 9.00 mmol, 3.00 eq.). The reaction mixture was stirred at room temperature overnight. LCMS showed reaction progress. The reaction mixture was diluted with EA (30 mL) and washed with 5% acetic acid solution (40 mL), concentrated $NaHCO_3$ solution (40 mL) and brine (40 mL). The organic layer was dried with $MgSO_4$, filtered and dry-loaded for normal-phase purification (Biotage Sfär Silica, 10 g, ethyl acetate:hexane, 10-80%). Fractions containing product were concentrated to yield the title compound (304 mg, 1.30 mmol, 43%) as an orange solid. **HPLC** 94.2% (RT = 2.43 min); **δ_H (500 MHz, $CDCl_3$-*d_1*);** 7.48 (1H, s, H-2), 7.22 (1H, dd, *J* 8.0, 2.0, H-6), 7.16 (1H, d, *J* 8.0, H-5), 3.80 (3H, s, OCH_3), 3.47 (2H, s, $OCCH_2$), 2.88 (4H, dt, *J* 15.0, 7.5, $ArCH_2$), 2.07 (2H, tt, *J* 7.5, 7.5, $CH_2CH_2CH_2$); **δ_C (125 MHz, $CDCl_3$-*d_1*);** 170.5 (ester CO), 162.6 (amide CO), 145.2 (C-1), 140.7 (C-3), 135.5 (C-4), 124.5 (C-2), 118.3 (C-6), 116.7 (C-5), 52.6 (CH_3), 41.3 ($OCCH_2CO$), 33.0 ($ArCH_2$), 32.4 ($ArCH_2$) 25.6 ($CH_2CH_2CH_2$); **LCMS** (Found: MH^+, 233.37. $C_{13}H_{15}NO_3$ requires *MH*+, 234.11).

Preparation of 3-((2,3-dihydro-1*H*-inden-5-yl)amino)-3-oxopropanoic acid (**6.3.3A**)



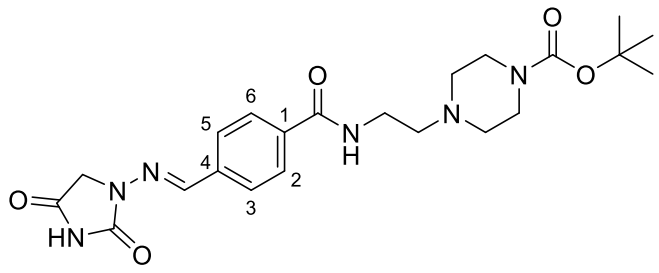
Methyl 3-((2,3-dihydro-1*H*-inden-5-yl)amino)-3-oxopropanoate (294 mg, 1.26 mmol) was dissolved in MeOH (9 mL) before addition of 35% ammonium hydroxide (6 mL). The reaction mixture was stirred at room temperature for 60 hours. LCMS indicated product formation. The reaction mixture was concentrated *in vacuo* to yield the title compound (290 mg, 1.26 mmol, quantitative yield) as an off-white fluffy solid. **HPLC** 89.0% (RT = 1.90 min); **δ_H (500 MHz, DMSO-*d_6*);** 7.50 (2H, t, *J* 8.0, H-2), 7.26 (1H, dd, *J* 8.0, 2.0, H-6), 7.13 (1H, d, *J* 8.0, H-5), 7.09 (1H, br s, NH), 3.18 (2H, s, $OCCH_2CO$), 2.81 (4H, dt, *J* 15.0, 7.5, $ArCH_2$), 1.99 (2H, tt, *J* 7.5, 7.5, $CH_2CH_2CH_2$); **δ_C (125 MHz, DMSO-*d_6*);** 169.3 (acid CO), 165.9 (amide CO), 144.6 (C-1), 138.9 (C-3), 137.7 (C-4), 124.6 (C-2), 117.7 (C-6), 115.7 (C-5), 44.9 ($OCCH_2CO$), 33.0 ($ArCH_2$) 32.2 ($ArCH_2$), 25.6 ($CH_2CH_2CH_2$); **LCMS** (Found: MH^+, 219.11. $C_{12}H_{13}NO_3$ requires *MH*+, 220.10).

Preparation of 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]benzoate triethylammonium (**6.3.4**)



1-aminohydantoin.HCl (4.56 g, 30.0 mmol, 1.00 eq.) and 4-formylbenzoic acid (4.50 g, 30.0 mmol, 1.00 eq.) were cooled to 0°C and dissolved in trimethyl orthoformate (90 mL). Dry DMF (2 mL) and triethylamine (6.30 mL, 36.0 mmol, 1.20 eq.) were then slowly added. Undissolved material dissolved after the addition of triethylamine was complete. Mixture was orange. The reactions mixture was heated to 50°C and stirred overnight. A white precipitate formed. The reaction mixture was cooled to 0°C and the precipitate filtered under vacuum. The precipitate was rinsed with ice cold methyl t-butyl ether (30 mL). The remaining solvent was removed *in vacuo*, yielding the title compound (11.1 g, 22.5 mmol, 75%) as a colourless solid. NMR indicated a 1:1.66 ratio between product and triethylamine. Collected solid was 82% product triethylamine salt. **δ_H (400 MHz, DMSO-*d*$_6$);** 7.99 (2H, d, *J* 7.5, H-2 & 6), 7.86 (1H, s, NC*H*C), 7.79 (2H, d, *J* 7.5, H-3 & 5), 4.38 (2H, s, $COCH_2N$), 2.95 (10H, q, *J* 7.0, NCH_2CH_3), 1.15 (15H, t, *J* 7.0, NCH_2CH_3); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.1 (C-2'), 168.3 (C-4'), 153.5 (acid CO), 142.3 (imine CH), 136.6 (C-1), 136.0 (C-4), 129.6 (C-2 & 6), 126.4 (C-3 & 5), 48.9 (C-5'), 45.2 NCH_2CH_3), 9.0 (NCH_2CH_3); **HRMS m/z (ES)** (Found: MH^+, 248.0677. $C_{11}H_9N_3O_4$ requires *MH*+, 248.0666).

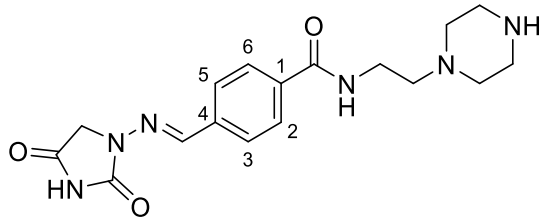
Preparation of *tert*-butyl 4-[2-({4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]phenyl}formamido)ethyl]piperazine-1-carboxylate (**6.3.5**)



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]benzoic acid (100 mg, 0.26 mmol, 1.0 eq.) was dissolved in DCM (2 mL) before addition of $SOCl_2$ (2 mL), forming a white solution. The reaction mixture was stirred and heated to reflux at 60°C for 18 hours. The reaction mixture was concentrated, removing the $SOCl_2$, and redissolved in dioxane (4 mL) forming a colourless solution. 4-*N*-(2-aminoethyl)-1-*N*-Boc-piperazine (89 mg, 0.39

mmol, 1.5 eq.) was added, immediately forming a cream-coloured suspension. The reaction mixture was stirred at room temperature for 2 hours. LCMS indicated complete product formation. The reaction mixture was concentrated and dissolved in DMSO (1 mL) for reverse-phase purification (Biotage Sfär C18, 18 g, 0-90% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (100 mg, 0.22 mmol, 85%) as a colourless solid. **HPLC** 94.0% (RT = 1.64 min); **δ_H (500 MHz, MeOD-*d*$_4$);** 7.93–7.84 (5 H, m, H-2, 3, 5 & 6 and imine CH), 4.40 (2 H, s, CO*C*H_2N), 3.56 (2 H, t, *J* 6.5, NHC*H*$_2$), 3.45 (4 H, br s, BocN(CH_2)$_2$), 2.63 (2 H, t, *J* 6.5, NC*H*$_2$$CH_2$), 2.51 (4 H, t, *J* 5.0, AlkN(CH_2)$_2$), 1.46 (9 H, s, (CH_3)$_3$); **δ_C (125 MHz, MeOD-*d*$_4$);** 170.42 (NH*C*OCH_2), 169.54 (amide CO), 156.39 (NHCON), 156.21 (carbamate CO), 144.08 (imine CH), 138.70 (C-4), 136.75 (C-1), 128.70 (C-3 & 5), 128.48 (C-2 & 6), 81.29 (*C*(CH_3)$_3$), 58.17 (CH_2*C*H_2N), 53.99 (AlkN(CH_2)$_2$), 49.72 (CO*C*H_2N), 38.01 (NHCH_2), 28.63 ((CH_3)$_3$). Missing BocN(CH_2)$_2$; **HRMS m/z (ES)** (Found: MH^+, 459.2357. $C_{22}H_{30}N_6O_5$ requires *MH+*, 459.2350).

Preparation of 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]-N-[2-(piperazin-1-yl)ethyl]benzamide (**6.3.6**)

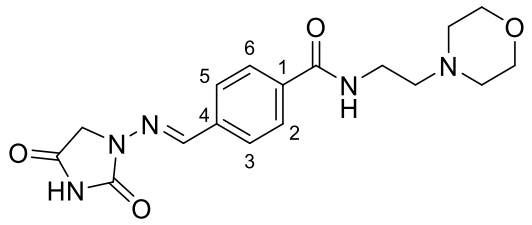


Tert-butyl 4-[2-({4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]phenyl}formamido)ethyl]piperazine-1-carboxylate (50 mg, 0.11 mmol) was suspended in 4 N HCl in dioxane (3 mL), forming a cloudy white solution. The reaction mixture was stirred at room temperature for 18 hours. LCMS indicated complete product formation. The reaction mixture was concentrated and dissolved in DMSO (1 mL) for reverse-phase purification (Modus P 4.5 g, 0-90% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were collected to yield the title compound (22 mg, 0.062 mmol, 51%) as a cream solid. **HPLC** 97.2% (RT = 0.84 min); **δ_H (500 MHz, DMSO-*d*$_6$);** 7.89 (2 H, d, *J* 8.5, H-3 & 5), 7.81 (1 H, s, imine CH), 7.76 (2 H, d, *J* 8.5, H-2 & 6), 4.32 (2 H, s, $COCH_2N$), 3.40–3.34 (2 H, m, NHC*H*$_2$), 2.71 (4 H, t, *J* 5.0, N(CH_2)$_2$), 2.44 (2 H, t, *J* 7.0, CH_2C*H*$_2$N), 2.39–2.34 (4 H, m, NH(CH_2)$_2$); **δ_C (125 MHz, DMSO-*d*$_6$);** 170.02 (NH*C*OCH_2), 165.59 (amide CO), 154.60 (NHCON), 141.43 (imine CH), 136.92 (C-4), 135.17 (C-1), 127.68 (C-3 & 5), 126.52 (C-2 & 6), 57.60 (CH_2*C*H_2N), 53.71 (N(CH_2)$_2$),

48.98 (CO<u>*C*</u>H_2N), 45.23 (NH(CH_2)$_2$), 36.73 (NHCH_2); **HRMS m/z (ES)** (Found: MH^+, 359.1827. $C_{17}H_{22}N_6O_3$ requires *MH*+, 359.1826).

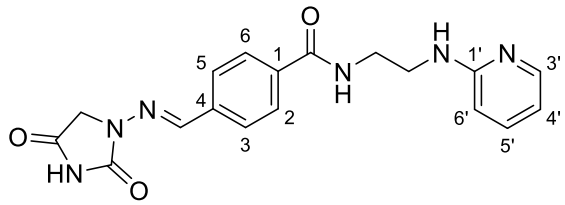
8.9 Synthesis of phenyl core hydantoin targets

Preparation of 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]-N-[2-(morpholin-4-yl)ethyl]benzamide (**6.6.2**)



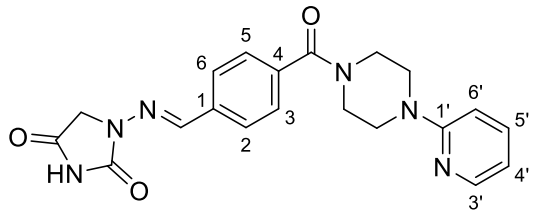
4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]benzoate triethylammonium (202 mg, 0.300 mmol, 1.00 eq.), 4-(2-aminoethyl)morpholine (43 µL, 0.33 mmol, 1.1 eq.), EDC (63 mg, 0.33 mmol, 1.1 eq.) and DMAP (7 mg, 0.06 mmol, 0.2 eq.) were mixed and dissolved in DCM (5 mL) and DMF (1 mL). The reaction mixture was stirred at room temperature for 18 hours. LCMS indicated product formation. The reaction mixture was dry-loaded for reverse phase purification (Modus P 18 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were collected to yield the title compound (62 mg, 0.17 mmol, 58%) as a colourless solid. **HPLC** 96.1% (RT = 0.99 min); **δ_H (400 MHz, DMSO-*d_6*);** 8.48 (1 H, br t, *J* 6.0, hydantoin NH), 7.89 (2 H, d, *J* 8.0, H-2 & 6), 7.83 (1 H, s, imine CH), 7.77 (2 H, d, *J* 8.0, H-3 & 5), 5.77 (1 H, br s, amide NH), 4.37 (2 H, s, NCH_2CO), 3.57 (4 H, t, *J* 4.5, N(CH_2)$_2$), 2.97 (2 H, q, *J* 6.5, NHC<u>*H*</u>$_2$), 2.47-2.35 (4 H, m, O(CH_2)$_2$), 2.12 (2 H, s, NC<u>*H*</u>$_2$$CH_2$); **$\delta_C$ (100 MHz, DMSO-*d_6*);** 169.11 (NH<u>*C*</u>OCH_2), 165.61 (amide CO), 153.57 (NHCON), 141.93 (imine CH), 136.78 (C-4), 135.27 (C-1), 127.70 (C-2 & 6), 126.60 (C-3 & 5), 66.21 (O(CH_2)$_2$), 57.36 (N<u>*C*</u>$H_2$$CH_2$), 53.31 (N($CH_2$)$_2$), 48.93 (CO<u>*C*</u>$H_2$N), 36.60 (NH$CH_2$); **HRMS m/z (ES)** (Found: MH^+, 360.1669. $C_{17}H_{21}N_5O_4$ requires *MH*+, 360.1666).

Preparation of 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]-N-{2-[(pyridin-2-yl)amino]ethyl}benzamide (**6.6.1**)



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]benzoate triethylammonium (202 mg, 0.300 mmol, 1.00 eq.), N-(2-pyridinyl)-1,2-ethanediamine.2HCl (69 mg, 0.33 mmol, 1.1 eq.), EDC (63 mg, 0.33 mmol, 1.1 eq.) and DMAP (7 mg, 0.06 mmol, 0.2 eq.) were mixed and dissolved in DCM (5 mL) and DMF (1 mL), forming an off-white suspension. The reaction mixture was stirred at room temperature for 70 hours. LCMS indicated reaction completion. The reaction mixture was dry-loaded for reverse-phase purification (Modus P 18 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Significant co-elution occurred, so the product was repurified (Modus P 4.5 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (17 mg, 0.047 mmol, 16%) as a cream solid. **HPLC** 98.6% (RT = 1.25 min); **δ_H (400 MHz, DMSO-*d*$_6$);** 11.29 (1 H, br s, hydantoin NH), 8.70 (1 H, s, amide NH), 7.98 (1 H, d, *J* 5.5, H'-3), 7.92 (2 H, d, *J* 8.5, H-2 & 6), 7.84 (1 H, s, imine CH), 7.77 (2 H, d, *J* 8.5, H-3 & 5), 7.40-34 (1 H, m, H'-4), 6.64 (1 H, s, ArNH), 6.50-6.46 (2 H, m, H'-5 & 6), 4.37 (2 H, s, $COCH_2N$), 3.43 (4 H, s, NHC$\underline{H}_2$C$\underline{H}_2$NH); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.71 (NH$\underline{C}OCH_2$), 166.31 (amide CO), 159.26 (C'-2), 154.17 (NCONH), 147.97 (C'-6), 142.34 (imine CH), 137.27 (C'-4), 137.16 (C-4), 135.71 (C-1), 128.20 (C-2 & 6), 127.03 (C-3 & 5), 112.10 (C'-5), 108.65 (C'-3), 45.60 (N$\underline{C}H_2CO$), 38.06 $ArNHCH_2$), 34.51 (CONH$\underline{C}H_2$); **HRMS m/z (ES)** (Found: MH^+, 367.1515. $C_{18}H_{18}N_6O_3$ requires *MH*+, 367.1513).

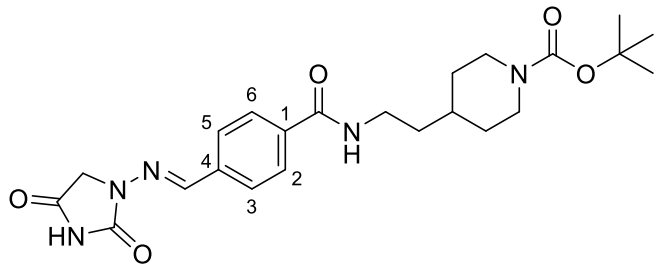
Preparation of 1-[(E)-({4-[4-(pyridin-2-yl)piperazine-1-carbonyl]phenyl}methylidene)amino]imidazolidine-2,4-dione (**6.6.3**)



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]benzoate triethylammonium (202 mg, 0.300 mmol, 1.00 eq.), 1-(2-Pyridinyl)piperazine (47 µL, 0.33 mmol, 1.1 eq.), EDC (63

mg, 0.33 mmol, 1.1 eq.) and DMAP (18 mg, 0.15 mmol, 0.50 eq.) were mixed before addition of DMF (5 mL), forming a fine white suspension. The reaction mixture was stirred for 18 hours at room temperature. LCMS indicated product formation. The reaction mixture was dry-loaded for reverse-phase purification (Modus P 18 g, 0-90% MeCN:H_2O (+0.1% NEt_3)). A colourless, gummy solid was collected (178 mg, 151%) and NMR indicated the product contained impurities. Material was redissolved in MeOH and purified by MDAP. Fractions containing product were concentrated to yield the title compound (16 mg, 0.040 mmol, 14%) as a colourless solid. **HPLC** 93.2% (RT = 1.54 min); **δ_H (500 MHz, DMSO-*d_6*);** 11.31 (1 H, br s, hydantoin NH), 8.13 (1 H, dd, *J* 5.0, 2.0, H'-6), 7.85 (1 H, s, imine CH), 7.78 (2 H, d, *J* 8.0, H-2 & 6), 7.56 (1 H, ddd, *J* 9.0, 7.0, 2.0, H'-4), 7.52 (2 H, d, *J* 8.0, H-3 & 5), 6.85 (1 H, d, *J* 8.5, H'-3), 6.67 (1 H, dd, *J* 7.0, 5.0, H'-5), 4.38 (2 H, s, $COCH_2N$), 3.82-3.18 (8 H, m, piperazine CH_2); **δ_C (125 MHz, DMSO-*d_6*);** 169.02 ($NH\underline{C}OCH_2$), 168.65 (amide CO), 158.71 (C'-2), 153.45 (NHCON), 147.59 (C'-6), 142.08 (imine CH), 137.68 (C'-4), 136.70 (C-1), 135.48 (C-4), 127.69 (C-3 & 5), 126.78 (C-2 & 6), 113.39 (C'-5), 107.33 (C-3), 48.97 ($CO\underline{C}H_2N$), 44.46 (br, piperazine CH_2); **HRMS m/z (ESI)** (Found: MH^+, 393.1674. $C_{20}H_{20}N_6O_3$ requires *MH*+, 393.1670).

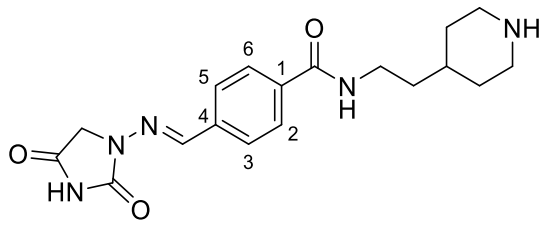
Preparation of *tert*-butyl 4-[2-({4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]phenyl}formamido)ethyl]piperidine-1-carboxylate (**6.6.5**)



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]benzoate triethylammonium (202 mg, 0.30 mmol, 1.00 eq.), 4-(aminoethyl)-1-N-Boc-piperidine (75 mg, 0.33 mmol, 1.1 eq.), EDC (63 mg, 0.33 mmol, 1.1 eq.) and DMAP (18 mg, 0.15 mmol, 0.50 eq.) were mixed and dissolved in DMF (5 mL). The reaction mixture was stirred at room temperature for 60 hours. LCMS indicated product formation. The reaction mixture was dry-loaded for reverse-phase purification (Modus P 18 g, 0-90% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (28 mg, 0.059 mmol, 20%) as a colourless solid. **HPLC** 98.0% (RT = 2.55 min); **δ_H (500 MHz, DMSO-*d_6*);** 11.30 (1 H, br s, hydantoin NH), 8.50 (1 H, t, *J* 5.5, amide NH), 7.90 (2 H, d, *J* 8.5, H-2

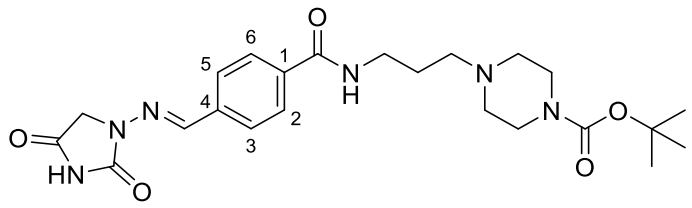
& 6), 7.83 (1 H, s, imine CH), 7.76 (2 H, d, *J* 8.5, H-3 & 5), 4.37 (2 H, s, NCH_2CO), 3.91 (2 H, d, *J* 12.5, eq. $N(CH_2)_2$), 3.29 (2 H, t, *J* 6.5, $NHCH_2$), 2.78-2.58 (2 H, m, ax. $N(CH_2)_2$), 1.68 (2 H, d, *J* 13.0, eq. $CH(CH_2)_2$), 1.53-1.43 (3 H, m, $CH(CH_2)_3$ and ax. $CH(CH_2)_2$), 1.39 (9 H, s, $C(CH_3)_3$), 1.04-0.94 (2 H, m, CH_2CH_2CH); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.02 ($NHCOCH_2$), 165.52 (amide CO), 153.85 (NHCON), 153.45 (carbamate CO), 142.00 (imine CH), 136.69 (C-4), 135.39 (C-1), 127.70 (C-2 & 6), 126.57 (C-3 & 5), 78.39 ($OC(CH_3)_3$), 48.93 (NCH_2CO), 36.69 ($NHCH_2$), 35.74 ($CH(CH_2)_2$), 32.89 ($CH(CH_2)_2$), 28.11 ($(CH_3)_3$). Missing $N(CH_2)_2$ and CH_2CH_2CH; **HRMS m/z (ES)** (Found: MNa+, 480.2219, MH^+ (-Boc), 358.1873. $C_{23}H_{31}N_5O_5$ requires *MH*+, 458.2398, *MNa*+, 480.2217).

Preparation of 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]-N-[2-(piperidin-4-yl)ethyl]benzamide (**6.5.2**)



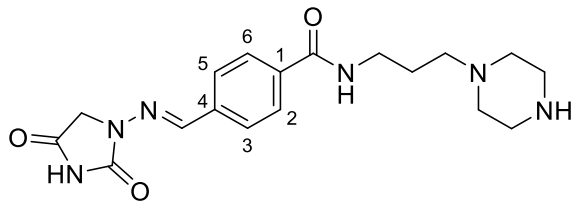
Tert-butyl 4-[2-({4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]phenyl}formamido)ethyl]piperidine-1-carboxylate (22 mg, 0.048 mmol) was dissolved in in 4 N HCl in dioxane (3 mL) and DMF (2 mL) and stirred at room temperature for 2 hours. LCMS indicated reaction completion. The reaction mixture was dry-loaded for reverse-phase purification (Modus P 4.5 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (2.6 mg, 0.0073 mmol, 15%) as a colourless solid. **HPLC** 96.85% (RT = 1.28 min); **δ_H (400 MHz, DMSO-*d*$_6$);** 8.48 (1 H, t, *J* 5.5, amide NH), 7.89 (2 H, d, *J* 8.0, H-2 & 6), 7.79 (1 H, s, imine CH), 7.75 (2 H, d, *J* 8.0, H-3 & 5), 4.28 (2 H, s, NCH_2CO), 3.29 (2 H, d, *J* 6.5, eq. $N(CH_2)_2$), 2.99 (2 H, d, *J* 12.5, ax. $N(CH_2)_2$), 2.58-2.52 (2 H, m, $NHCH_2$), 1.68 (2 H, br d, *J* 12.5, eq. $CH(CH_2)_2$), 1.49-1.39 (3 H, m, $CH(CH_2)_3$ and ax. $CH(CH_2)_2$), 1.15-1.03 (2 H, m, CH_2CH_2CH); **δ_C (100 MHz, DMSO-*d*$_6$);** 127.66 (C-2 & 6), 126.42 (C-3 & 5), 45.32 ($CH(CH_2)_2$), 36.60 ($NHCH_2$), 31.74 ($CH(CH_2)_2$), Carbon low intensity, missing quaternaries, imine CH, $NH(CH_2)_2$ and CH_2CH_2CH; **HRMS m/z (ES)** (Found: MH^+, 358.1875. $C_{18}H_{23}N_5O_3$ requires *MH*+, 358.1874).

Preparation of *tert*-butyl 4-[3-({4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]phenyl}formamido)propyl]piperazine-1-carboxylate (**6.6.4**)



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]benzoate triethylammonium (246 mg, 0.500 mmol, 1.00 eq.), *tert*-butyl 4-(3-aminopropyl)piperazine-1-carboxylate (182 mg, 0.750 mmol, 1.50 eq.), EDC (105 mg, 0.55 mmol, 1.10 eq.) and DMAP (31 mg, 0.25 mmol, 0.50 eq.) were mixed and dissolved in DMF (8 mL). The reaction mixture was heated to 40°C and stirred for 16 hours. LCMS indicated product formation. The reaction mixture was concentrated and dry-loaded for reverse-phase purification (Modus P 18 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (92 mg, 0.19 mmol, 39%) as a colourless solid. **HPLC** 86.26% (RT = 2.15 min); **δ_H (500 MHz, DMSO-*d_6*);** 8.54 (1 H, t, *J* 5.5, amide NH), 7.90 (2 H, d, *J* 8.5, H-3 & 5), 7.83 (1 H, s, imine CH), 7.76 (2 H, d, *J* 8.5, H-2 & 6), 4.37 (2 H, s, $COCH_2N$), 3.38-3.23 (6 H, m, NHC*H*$_2$ and $BocN(CH_2)_2$), 2.38-2.25 (6 H, m, C*H*$_2$N(C*H*$_2$)$_2$), 1.69 (2 H, tt, *J* 7.0, 7.0, $CH_2CH_2CH_2$), 1.39 (9 H, s, $(CH_3)_3$); **δ_C (125 MHz, DMSO-*d_6*);** 169.04 ($NHCOCH_2$), 165.59 (amide CO), 153.80 (NHCON), 153.47 (carbamate CO), 141.98 (imine CH), 136.71 (C-4), 135.39 (C-1), 127.69 (C-2 & 6), 126.57 (C-3 & 5), 78.72 ($C(CH_3)_3$), 55.58 ($CH_2N(CH_2)_2$), 52.56 ($CH_2N(CH_2)_2$), 48.93 ($COCH_2N$), 38.61 ($BocN(CH_2)_2$), 37.78 ($NHCH_2$), 28.06 ($(CH_3)_3$), 26.25 ($CH_2CH_2CH_2$); **HRMS m/z (ES)** (Found: MH^+, 473.2509. $C_{23}H_{32}N_6O_5$ requires *MH*+, 473.2507).

Preparation of 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]-N-[3-(piperazin-1-yl)propyl]benzamide (**6.5.1**)

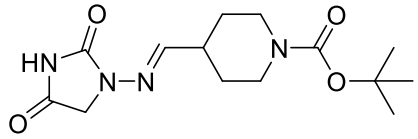


Tert-butyl 4-[3-({4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]phenyl}formamido)propyl]piperazine-1-carboxylate (80 mg, 0.17 mmol) was dissolved in 4 N HCl in dioxane (5 mL) and DMF (2 mL), forming a cloudy solution.

The reaction mixture was stirred at room temperature for 20 hours. LCMS indicated reaction completion. The reaction mixture was concentrated and dry-loaded for reverse-phase purification (Modus P 4.5 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (14 mg, 0.038 mmol, 22%) forming a trace colourless residue. **HPLC** 96.5% (RT = 1.32 min); **δ_H (500 MHz, DMSO-*d_6*);** 8.58 (1 H, t, *J* 5.5, amide NH), 7.90 (2 H, d, *J* 8.5, H-2 & 6), 7.84 (1 H, s, imine CH), 7.77 (2 H, d, *J* 8.5, H-3 & 5), 4.37 (2 H, s, NCH_2CO), 3.29 (2 H, q, *J* 7.0, $NHC\underline{H}_2$), 2.97 (4 H, br s, $NH(C\underline{H}_2)_2$), 2.50 (4 H, br s, $N(CH_2)_2$), 2.38 (2 H, t, *J* 7.0, $N\underline{C}H_2CH_2$), 1.68 (2 H, tt, *J* 7.0, 7.0, $CH_2\underline{C}H_2CH_2$); **δ_C (125 MHz, DMSO-*d_6*);** 169.52 ($NH\underline{C}OCH_2$), 166.08 (amide CO), 153.96 (NHCON), 142.46 (imine CH), 137.20 (C-4), 135.82 (C-1), 128.17 (C-2 & 6), 127.05 (C-3 & 5), 55.89 ($N\underline{C}H_2CH_2$), 50.74 ($N(CH_2)_2$), 49.41 ($CO\underline{C}H_2N$), 43.90 ($NH(CH_2)_2$), 38.13 ($NH\underline{C}HCH_2$), 26.52 ($CH_2\underline{C}H_2CH_2$); **HRMS m/z (ES)** (Found: MH^+, 373.1986. $C_{18}H_{24}N_6O_3$ requires *MH*+, 373.1983).

8.10 Synthesis of piperidine core hydantoin targets

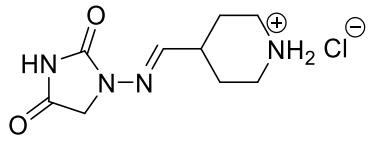
Preparation of *tert*-butyl 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidine-1-carboxylate (**6.7.1**)



1-aminohydantoin.HCl (1.52 g, 10.0 mmol, 1.00 eq.) and 1-Boc-piperidine-4-carboxaldehyde (2.13 g, 10.0 mmol, 1.00 eq.) were cooled to 0°C and dissolved in trimethyl orthoformate (30 mL), before addition of dipropylethylamine (2.10 mL, 12.0 mmol, 1.20 eq.). The flask was warmed to room temperature and stirred for 22 hours. LCMS indicated reaction completion. The reaction mixture was diluted with ethyl acetate (30 mL) and washed with brine (3 x 60 mL). Precipitate began to form during the second and third wash. The organic layer was dried with $MgSO_4$, filtered and concentrated *in vacuo* to yield the title compound (778 mg, 2.51 mmol, 25%) as a colourless solid. **δ_H (500 MHz, MeOD-*d_4*);** 7.06 (1 H, d, *J* 5.0, $NC\underline{H}CH_2$), 4.21 (2 H, s, NCH_2CO), 4.13-4.03 (2 H, m, eq. $N(CH_2)_2$), 2.98-2.83 (2 H, m, ax. $N(CH_2)_2$), 2.60-2.49 (1 H, m, $CHC\underline{H}(CH_2)_2$), 1.84 (2 H, br dd, *J* 13.5, 2.5, eq. $CH(C\underline{H}_2)_2$), 1.54-1.47 (11 H, m, ax. $CH(C\underline{H}_2)_2$ and $(CH_3)_3$); **δ_C (100 MHz, $CDCl_3$-*d_1*)**; 166.91 ($CH_2\underline{C}ONH$), 154.87 (NHCON), 153.00 (NCOO), 151.08 ($N\underline{C}HCH$), 79.86 ($O\underline{C}(CH_3)_3$), 51.42 ($N(CH_2)_2$), 48.93 ($N\underline{C}H_2CO$), 39.57

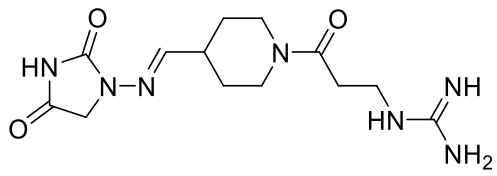
(CH<u>C</u>H(CH$_2$)$_2$), 29.37 (CH(<u>C</u>H$_2$)$_2$), 28.57 ((CH$_3$)$_3$); **LCMS** (Found: MNa$^+$, 333.15. $C_{14}H_{22}N_4O_4$ requires *MH*+, 311.17).

<u>Preparation of 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidin-1-ium chloride (**6.7.2**)</u>



Tert-butyl 4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidine-1-carboxylate (998 mg, 3.22 mmol, 1.00 eq.) was suspended in 4N HCl in dioxane (20 mL) and stirred for 168 hours at room temperature. LCMS indicated complete product formation. Solvent was removed *in vacuo* to yield the title compound (843 mg, 3.22 mmol, quantitative yield) as a pale pink solid. **δ_H (500 MHz, D_2O);** 7.16 (1 H, d, *J* 5.0, NC<u>*H*</u>CH), 4.32 (2 H, s, COCH$_2$N), 3.48 (2 H, dt, *J* 13.0, 3.5, eq. NH(C<u>*H*$_2$</u>)$_2$), 3.11 (2 H, td, *J* 12.5, 3.0, ax. NH(C<u>*H*$_2$</u>)$_2$), 2.80-2.71 (1 H, m, NCHC<u>*H*</u>) 2.13 (2 H, dd, *J* 15.0, 3.5, eq. CH(C<u>*H*$_2$</u>)$_2$), 1.87-1.75 (2 H, m, ax. CH(C<u>*H*$_2$</u>)$_2$); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.07 (CH$_2$<u>C</u>ONH), 153.39 (NHCON), 148.01 (imine CH), 66.36 (NH(CH$_2$)$_2$), 48.56 (CO<u>C</u>H$_2$N), 42.21 (NCH<u>C</u>H), 25.56 (CH(<u>C</u>H$_2$)$_2$); **HRMS m/z (ES)** (Found: MH$^+$, 211.1214. $C_9H_{14}N_4O_2$ requires *MH*+, 211.1190).

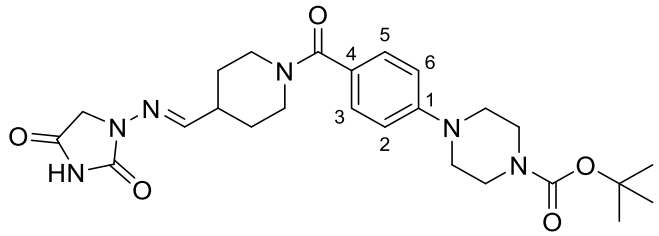
<u>Preparation of N-(3-{4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidin-1-yl}-3-oxopropyl)guanidine (**6.5.10**)</u>



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidin-1-ium chloride (68 mg, 0.28 mmol, 1.0 eq.), 3-carbamimidamidopropanoic acid (45 mg, 0.34 mmol, 1.2 eq.), EDC (59 mg, 0.31 mmol, 1.1 eq.) and DMAP (17 mg, 0.14 mmol, 0.50 eq.) were mixed and dissolved in DMF (5 mL) and stirred at room temperature for 60 hours. LCMS showed a product mass trace. The reaction mixture was dry-loaded and purified by reverse-phase chromatography (Modus P 18 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (7 mg, 0.02 mmol, 7%) as a

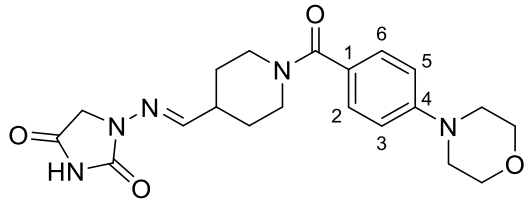
colourless solid. **HPLC** 95.48% (RT = 0.759 min); **δ_H (400 MHz, DMSO-*d*$_6$);** 7.48-7.15 (4 H, m, NHC(NH)NH_2), 6.77 (1 H, d, *J* 5.0, imine CH), 5.78 (1 H, br s, hydantoin CH), 3.82 (1 H, d, *J* 13.5, C*H*$_2$NHC(NH)NH_2), 3.70 (2 H, s, NCH_2CO), 3.07 (1 H, t, *J* 12.5, CHC*H*$(CH_2)_2$), 2.97 (2 H, q, *J* 7.5, eq. N$(CH_2)_2$), 2.70 (2 H, t, *J* 12.5, ax. N$(CH_2)_2$), 2.16 (1 H, t, *J* 7.0, $COCH_2$), 1.74 (2 H, br s, eq. CH(C*H*$_2)_2$), 1.47 (2 H, q, *J* 7.0, ax. CH(C*H*$_2)_2$); **δ_C (125 MHz, DMSO-*d*$_6$);** 168.91 (CH_2*C*ONH), 162.98 (amide CO), 157.38 (NHCON), 49.62 (CO*C*H_2N), 45.69 (NCH*C*H), 44.77 (CH_2NH), 38.10 (CO*C*H_2), 34.51 (CON(*C*$H_2)_2$), 28.52 (CH(C*H*$_2)_2$), missing imine CH, NH*C*$NHNH_2$; **HRMS m/z (ES)** (Found: MH^+, 324.1779. $C_{13}H_{21}N_7O_3$ requires *MH*+, 324.1779).

Preparation of *tert*-butyl 4-(4-{4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidine-1-carbonyl}phenyl)piperazine-1-carboxylate (**6.7.3**)



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidin-1-ium chloride (62 mg, 0.25 mmol, 1.0 eq.), 4-{4-[(*tert*-butoxy)carbonyl]piperazin-1-yl}benzoic acid (86 mg, 0.28 mmol, 1.1 eq.), EDC (53 mg, 0.28 mmol, 1.1 eq.) and DMAP (16 mg, 0.13 mmol, 0.50 eq.) were mixed and dissolved in DMF (3 mL), forming a light brown suspension. The reaction mixture was stirred at room temperature for 22 hours. LCMS showed the product mass. The reaction mixture was dry-loaded and purified by reverse-phase chromatography (Modus P 18 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (72 mg, 0.14 mmol, 58%) as a colourless solid. **HPLC** 79.38% (RT = 3.01 min); **δ_H (500 MHz, DMSO-*d*$_6$);** 8.10 (1 H, br s, hydantoin NH), 7.27 (2 H, d, *J* 9.0, H-2 & 6), 7.00 (1 H, d, *J* 5.0, imine CH), 6.97 (2 H, d, *J* 9.0, H-3 & 5), 4.14 (2 H, s, $COCH_2N$), 3.45 (4 H, t, *J* 5.0, ArN$(CH_2)_2$), 3.19 (4 H, t, *J* 5.0, BocN$(CH_2)_2$), 3.08-2.91 (4 H, m, ArCON$(CH_2)_2$), 2.60-2.53 (1 H, m, CHC*H*$(CH_2)_2$), 1.77 (2 H, br d, *J* 11.5, eq. CH(C*H*$_2)_2$), 1.50-1.36 (11 H, m, ax. CH(C*H*$_2)_2$ and $(CH_3)_3$); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.32 (NH*C*OCH_2), 169.07 (amide CO), 153.87 (NHCON), 153.39 (carbamate CO), 151.47 (C-1), 149.05 (imine CH), 128.47 (C-3 & 5), 125.85 (C-4), 114.50 (C-2 & 6), 79.04 (*C*$(CH_3)_3$), 48.50 (CO*C*H_2N), 47.42 (ArN$(CH_2)_2$), 38.32 (CH*C*H$(CH_2)_2$), 37.44 (BocN$(CH_2)_2$), 34.05 (ArCON(*C*$H_2)_2$), 29.17 (CH(*C*$H_2)_2$), 28.06 ($(CH_3)_3$); **HRMS m/z (ES)** (Found: MH^+, 499.2669. $C_{25}H_{34}N_6O_5$ requires *MH*+, 499.2663).

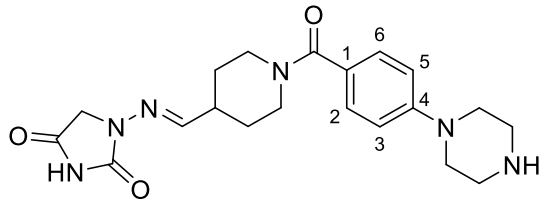
Preparation of 1-[(E)-({1-[4-(morpholin-4-yl)benzoyl]piperidin-4-yl}methylidene)amino]imidazolidine-2,4-dione (**6.7.4**)



4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidin-1-ium chloride (49 mg, 0.20 mmol, 1.0 eq.), 4-(morpholin-4-yl)benzoic acid (46 mg, 0.22 mmol, 1.1 eq.), EDC (42 mg, 0.22 mmol, 1.1 eq.) and DMAP (12 mg, 0.10 mmol, 0.50 eq.) were mixed and dissolved in DMF (3 mL) and stirred at room temperature for 16 hours. LCMS indicated product formation. The reaction mixture was dry-loaded and purified by reverse-phase chromatography (Modus P 4.5 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to yield the title compound (38 mg, 0.095 mmol, 48%) as an off-white solid.

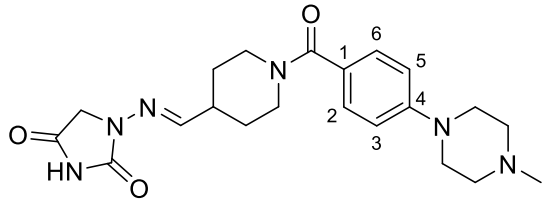
HPLC 95.8% (RT = 1.95 min); **δ_H (500 MHz, DMSO-d_6);** 7.29 (2 H, d, *J* 9.0, H-2 & 6), 7.01 (1 H, d, *J* 5.0, imine CH), 6.96 (2 H, d, *J* 9.0, H-3 & 5), 4.15 (2 H, s, $COCH_2N$), 3.73 (4 H, t, *J* 5.0, $O(CH_2)_2$), 3.17 (4 H, t, *J* 5.0, $ArN(CH_2)_2$), 3.09-2.92 (2 H, m, eq. $CON(CH_2)_2$), 2.61-2.54 (1 H, m, $CHCH(CH_2)_2$), 2.47-2.41 (2 H, m, ax. $CON(CH_2)_2$), 1.83-1.73 (2 H, m, eq. $CH(CH_2)_2$), 1.49-1.36 (2 H, m, ax. $CH(CH_2)_2$); **δ_C (125 MHz, DMSO-d_6);** 151.74 (NHCON), 149.03 (imine CH), 128.44 (C-2 & 6), 125.81 (C-4), 121.72 (C-1), 113.84 (C-3 & 5), 65.99 ($O(CH_2)_2$), 48.51 (NCH_2CO), 47.62 ($ArN(CH_2)_2$), 45.70 ($CON(CH_2)_2$), 38.07 ($CHCH(CH_2)_2$). Missing $NHCOCH_2$, amide CO, $CH(CH_2)_2$; **HRMS m/z (ES)** (Found: MH^+, 400.1979. $C_{20}H_{25}N_5O_4$ requires *MH*+, 400.1979).

Preparation of 1-[(E)-({1-[4-(piperazin-1-yl)benzoyl]piperidin-4-yl}methylidene)amino]imidazolidine-2,4-dione (**6.5.5**)



Tert-butyl 4-(4-{4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidine-1-carbonyl}phenyl)piperazine-1-carboxylate (47 mg, 0.10 mmol) was dissolved in 4 N HCl in dioxane (3 mL) and DMF (1 mL), forming a pale yellow solution. The reaction mixture was stirred at room temperature for 24 hours. LCMS indicated complete product formation. The reaction mixture was concentrated *in vacuo*, dry-loaded onto silica and purified by reverse-phase chromatography (Modus P 4.5 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions collected to yield the title compound (9 mg, 0.02 mmol, 23%) as a colourless solid. **HPLC** 94.0% (RT = 1.46 min); **δ_H (500 MHz, DMSO-*d*$_6$);** 7.25 (2 H, d, *J* 9.0, H-3 & 5), 7.01 (1 H, d, *J* 5.0, imine CH), 6.92 (2 H, d, *J* 9.0, H-2 & 6), 4.15 (2 H, s, $COCH_2N$), 3.13 (4 H, t, *J* 5.0, $ArN(CH_2)_2$), 3.06-2.92 (2 H, m, eq. $CON(CH_2)_2$), 2.84 (4 H, br t, *J* 5.0, $NH(C\underline{H}_2)_2$), 2.60-2.62-2.52 (1 H, m, $CHC\underline{H}(CH_2)_2$), 1.82-1.73 (2 H, m, eq. $CH(C\underline{H}_2)_2$), 1.47-1.36 (2 H, m, ax. $CH(C\underline{H}_2)_2$); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.90 ($CH_2\underline{C}ONH$), 169.59 (amide CO), 153.91 (NHCON), 152.62 (C-4), 149.52 (imine CH), 128.95 (C-2 & 6), 125.69 (C-1), 114.39 (C-3 & 5), 48.98 ($ArN(CH_2)_2$), 48.73 ($NH(CH_2)_2$), 45.73 ($CO\underline{C}H_2N$), 29.63 ($CH(\underline{C}H_2)_2$), 21.67 ($CH\underline{C}H(CH_2)_2$). Missing $CON(\underline{C}H_2)_2$; **HRMS m/z (ES)** (Found: MH^+, 399.2138. $C_{20}H_{26}N_6O_3$ requires *MH*+, 399.2139).

Preparation of 1-[(E)-({4-[4-(4-methylpiperazin-1-yl)benzoyl]phenyl}methylidene)amino]imidazolidine-2,4-dione

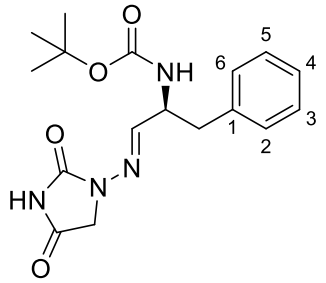


4-[(E)-[(2,4-dioxoimidazolidin-1-yl)imino]methyl]piperidin-1-ium chloride (49 mg, 0.20 mmol, 1.0 eq.), 4-(4-methylpiperazin-1-yl)benzoic acid (48 mg, 0.22 mmol, 1.1 eq.), EDC (42 mg, 0.22 mmol, 1.1 eq.) and DMAP (12 mg, 0.10 mmol, 0.50 eq.) were mixed and dissolved in DMF (3 mL) and stirred at room temperature for 18 hours. LCMS showed clean product formation. The reaction mixture was concentrated and dry-loaded for reverse-phase purification (Modus P 4.5 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were concentrated to a colourless oil (74 mg, 114%). Some unknown peaks were present in the NMR, so the product was repurified by reverse-phase chromatography (Modus P 4.5 g, 30-50% MeCN:H_2O (+0.1% NEt_3)). Fractions were collected to yield the title compound (69 mg, 0.17 mmol, 84%) as a colourless residue. **HPLC** 91.1% (RT = 1.47 min); **δ_H (500 MHz, DMSO-*d*$_6$);** 8.09 (1 H, dd, *J* 5.0, 1.5,

hydantoin NH), 7.25 (2 H, d, *J* 9.0, H-2 & 6), 7.00 (1 H, d, *J* 5.0, imine CH), 6.94 (2 H, d, *J* 9.0, H-3 & 5), 4.14 (2 H, s, $COCH_2N$), 3.23-3.17 (4 H, m, $ArN(CH_2)_2$), 3.06-2.90 (4 H, m, $CON(CH_2)_2$), 2.61-2.51 (1 H, m, NCHC*H*), 2.47-2.39 (4 H, m, $MeN(CH_2)_2$), 2.21 (3 H, s, CH_3), 1.80-1.74 (2 H, m, eq. $CH(CH_2)_2$), 1.51-1.35 (2 H, m, ax. $CH(CH_2)_2$); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.43 (CH_2CONH), 169.37 (amide CO), 153.74 (NCONH), 151.67 (C-4), 149.33 (imine CH), 128.49 (C-2 & 6), 125.31 (C-1), 113.99 (C-3 & 5), 54.48 ($MeN(CH_2)_2$), 48.54 (NCH_2CO), 47.28 ($ArN(CH_2)_2$), 45.76 (CH_3), 38.61 (NCHCH), 37.63 ($CON(CH_2)_2$), 28.02 ($CH(CH_2)_2$); **m/z (ES)** (Found: MH^+, 413.2301. $C_{21}H_{28}N_6O_3$ requires *MH*+, 413.2296).

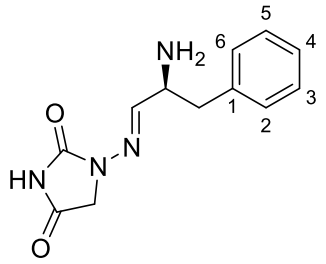
8.11 Synthesis of phenylalanine core hydantoin targets

Preparation of *tert*-butyl N-[(1E,2S)-1-[(2,4-dioxoimidazolidin-1-yl)imino]-3-phenylpropan-2-yl]carbamate (**6.8.1**)



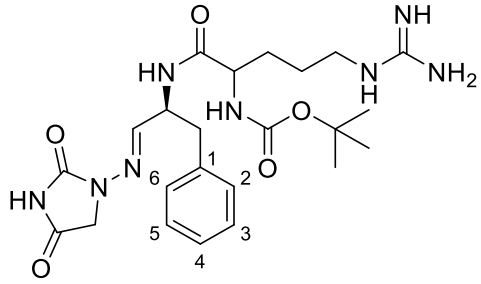
1-aminohydantoin.HCl (314 mg, 2.07 mmol, 1.00 eq.) and N-Boc-L-phenylalaninal (516 mg, 2.07 mmol, 1.00 eq.) were suspended in trimethyl orthoformate (10 mL) followed by dry DMF (2 drops) and triethylamine (0.57 mL, 4.1 mmol, 2.0 eq.) and stirred at room temperature for 90 hours. LCMS indicated reaction completion. The reaction mixture was dry-loaded and purified normal-phase (Modus B 12 g, 0-10% MeOH:DCM). Fractions were collected to yield the title compound (467 mg, 1.35 mmol, 65%) as a cream solid. **δ_H (500 MHz, DMSO-*d*$_6$);** 7.37-7.15 (5 H, m, H-2-6), 7.07 (1 H, d, *J* 8.5, NCH), 6.99 (1 H, d, *J* 4.5, CONHCO), 4.39-4.32 (1 H, m, NHC*H*), 4.19 (2 H, s, NCH_2CO), 2.97 (1 H, dd, *J* 14.0, 5.5, syn ArC*H*H), 2.80 (1 H, dd, *J* 14.0, 9.5, anti ArCH*H*), 1.37 (9 H, s, $(CH_3)_3$); **δ_C (125 MHz, DMSO-*d*$_6$);** 169.43 (CH_2CONH), 155.35 (NHCON), 153.86 (NHCOO), 146.57 (NCH), 138.44 (H-1), 129.81 (H-3 & 5), 128.55 (H-2 & 6), 126.61 (H-4), 78.42 ($C(CH_3)_3$), 53.69 (NHCH), 48.98 (NCH_2CO), 38.43 ($ArCH_2$), 28.64 ($(CH_3)_3$).

Preparation of 1-[(E)-[(2S)-2-amino-3-phenylpropylidene]amino]imidazolidine-2,4-dione (**6.8.2**)



Tert-butyl N-[(1E,2S)-1-[(2,4-dioxoimidazolidin-1-yl)imino]-3-phenylpropan-2-yl]carbamate (429 mg, 1.23 mmol) was dissolved in 4N HCl in dioxane (10 mL) forming a cream suspension. The reaction mixture was stirred at room temperature for 35 hours. The mixture was red-brown and had solidified. Addition of MeOH (3 mL) dissolved the solid into a brown solution. LCMS showed a clear product trace. The mixture was dry-loaded and purified by reverse-phase chromatography (Modus P 18 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). One major peak eluted immediately with no clear separation. A yellow oil was collected and solidified after it was concentrated *in vacuo* overnight. The solid was dissolved in MeOH and dry-loaded for reverse-phase purification (Modus P 18 g, 0-100% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were collected to yield the title compound (166 mg, 0.670 mmol, 55%) as an off-white solid. **δ_H (500 MHz, MeOD-*d*$_4$)**; 7.34-7.29 (2 H, m, H-2 & 6), 7.28-7.19 (3 H, m, H-3, 4 & 5), 7.07 (1 H, d, *J* 4.5, NCH), 4.16-4.09 (1 H, m, NCH_2CO), 3.89-3.83 (1 H, m, NH_2CH), 2.99 (1 H, dd, *J* 13.5, 7.0, ArC*H*H), 2.87 (1 H, dd, *J* 13.5, 7.0, ArCH*H*); **δ_C (125 MHz, MeOD-*d*$_4$);** 172.11 (CH_2CONH), 157.80 (NCONH), 148.77 (imine CH), 138.53 (C-1), 130.56 (C-3 & 5), 129.71 (C-2 & 6), 127.85 (C-4), 55.13 (NH_2CH), 49.51 (NCH_2CO), 42.22 ($ArCH_2$); **HRMS m/z (ES)** (Found: MH^+, 247.1186. $C_{12}H_{14}N_4O_2$ requires *MH*+, 247.1190).

Preparation of *tert*-butyl N-(4-carbamimidamido-1-{[(1E,2S)-1-[(2,4-dioxoimidazolidin-1-yl)imino]-3-phenylpropan-2-yl]carbamoyl}butyl)carbamate (**6.8.3**)



1-[(E)-[(2S)-2-amino-3-phenylpropylidene]amino]imidazolidine-2,4-dione (157 mg, 0.640 mmol, 1.00 eq.), Boc-L-arginine (192 mg, 0.700 mmol, 1.10 eq.), EDC (134 mg, 0.700 mmol, 1.10 eq.) and DMAP (39 mg, 0.32 mmol, 0.50 eq.) were dissolved in DMF (5 mL), forming a yellow solution. The reaction mixture was stirred at room temperature for 20 hours. LCMS showed a strong mass trace. The mixture was dry-loaded for reverse-phase purification (Modus P 18 g, 0-90% MeCN:H_2O (+0.1% NEt_3)). Fractions containing product were collected to yield the title compound (136 mg, 0.270 mmol, 42%) as a yellow solid. HPLC showed two, close eluting peaks, indicating that the arginine had been racemised during the coupling. All peaks in the NMR split (labelled as doublets). **HPLC** 98.7% (combined enantiomers) (RT = 1.99 and 2.06 min); **δ_H (500 MHz, DMSO-*d_6*);** 7.28-7.13 (5 H, m, H-2-6), 6.80 (1 H, d, *J* 5.0, NCH), 4.65 (1 H, br d, NHC*H*(CONH)CH_2CH_2), 3.99-3.83 (1 H, m, NHC*H*(CH_2Ar)CHN), 3.64 (2 H, d, NCH_2CO), 3.31 (2 H, s, C*H_2*NHC(NH)NH_2), 3.04-2.95 (2 H, m, $ArCH_2$), 2.83 (1 H, ddd, *J* 14.0, 12.0, 8.5), 1.57-1.43 (2 H, m, CHC*H_2*CH_2), 1.37 (9 H, d, *J* 4.0, $(CH_3)_3$), 1.32-1.21 (2 H, m, CH_2C*H_2*CH_2); **δ_C (125 MHz, DMSO-*d_6*);** 138.01 (imine CH), 129.42/129.35 (C-2 & 6), 128.09/128.01 (H-3 & 5), 126.13/126.11 (C-4), 78.04/77.92 (*C*$(CH_3)_3$), 53.90/53.60 (NH*C*H(CH_2Ar)CHN, 51.67/51.28 (NH*C*H(CONH)CH_2CH_2), 49.55/49.46 (NH*C*H_2CO), 40.23/39.98 (*C*H_2NHC(NH)NH_2), 38.19 ($ArCH_2$), 29.61/28.91 (CH*C*H_2CH_2), 28.18 ($(CH_3)_3$), 24.77/24.68 (CH_2*C*H_2CH_2). Some quaternaries were too low intensity to be assigned; **HRMS m/z (ES)** (Found: MH^+, 503.2723. $C_{23}H_{34}N_8O_5$ requires *MH*+, 503.2725).

9. References

1. UK, D. Type 2 Diabetes. https://www.diabetes.org.uk/diabetes-the-basics/types-of-diabetes/type-2 (accessed 27/06/22).
2. Grover, S. A.; Kaouache, M.; Rempel, P.; Joseph, L.; Dawes, M.; Lau, D. C.; Lowensteyn, I., Years of life lost and healthy life-years lost from diabetes and cardiovascular disease in overweight and obese people: a modelling study. *The Lancet Diabetes & Endocrinology* **2015,** *3* (2), 114-122.
3. Hex, N.; Bartlett, C.; Wright, D.; Taylor, M.; Varley, D., Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabetic Medicine* **2012,** *29* (7), 855-862.
4. Hoban, M., BUDGET 2011. Treasury, H., Ed. 2011; p 102.
5. Hofmarcher, T.; Lindgren, P.; Wilking, N.; Jönsson, B., The cost of cancer in Europe 2018. *European Journal of Cancer* **2020,** *129*, 41-49.
6. *IDF Diabetes Atlas*; International Diabetes Federation: www.diabetesatlas.org, 2021; p 136.
7. Garber, A., Obesity and type 2 diabetes: which patients are at risk? *Diabetes, Obesity and Metabolism* **2012,** *14* (5), 399-408.
8. Hackett, R. A.; Steptoe, A., Psychosocial factors in diabetes and cardiovascular risk. *Current Cardiology Reports* **2016,** *18*, 1-12.
9. Prasad, R. B.; Groop, L., Genetics of type 2 diabetes—pitfalls and possibilities. *Genes* **2015,** *6* (1), 87-123.
10. Henning, R. J., Type-2 diabetes mellitus and cardiovascular disease. *Future Cardiology* **2018,** *14* (6), 491-509.
11. Einarson, T. R.; Acs, A.; Ludwig, C.; Panton, U. H., Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovascular Diabetology* **2018,** *17* (1), 1-19.
12. Cosentino, F.; Grant, P. J.; Aboyans, V.; Bailey, C. J.; Ceriello, A.; Delgado, V.; Federici, M.; Filippatos, G.; Grobbee, D. E.; Hansen, T. B., 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD: The Task Force for diabetes, pre-diabetes, and cardiovascular diseases of the European Society of Cardiology (ESC) and the European Association for the Study of Diabetes (EASD). *European Heart Journal* **2020,** *41* (2), 255-323.
13. Lean, M. E.; Leslie, W. S.; Barnes, A. C.; Brosnahan, N.; Thom, G.; McCombie, L.; Peters, C.; Zhyzhneuskaya, S.; Al-Mrabeh, A.; Hollingsworth, K. G., Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *The Lancet* **2018,** *391* (10120), 541-551.
14. Rotenstein, L. S.; Kozak, B. M.; Shivers, J. P.; Yarchoan, M.; Close, J.; Close, K. L., The ideal diabetes therapy: what will it look like? How close are we? *Clinical Diabetes* **2012,** *30* (2), 44-53.
15. Rena, G.; Hardie, D. G.; Pearson, E. R., The mechanisms of action of metformin. *Diabetologia* **2017,** *60* (9), 1577-1585.
16. Nathan, D. M.; Buse, J. B.; Davidson, M. B.; Ferrannini, E.; Holman, R. R.; Sherwin, R.; Zinman, B., Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* **2009,** *32* (1), 193-203.
17. Group, U. P. D. S., Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *The Lancet* **1998,** *352* (9131), 854-865.
18. Griffin, S. J.; Leaver, J. K.; Irving, G. J., Impact of metformin on cardiovascular disease: a meta-analysis of randomised trials among people with type 2 diabetes. *Diabetologia* **2017,** *60* (9), 1620-1629.
19. Group, L. A. R., Effects of intensive lifestyle intervention on all-cause mortality in older adults with type 2 diabetes and overweight/obesity: results from the Look AHEAD Study. *Diabetes Care* **2022,** *45* (5), 1252-1259.
20. Marx, N.; Federici, M.; Schütt, K.; Müller-Wieland, D.; Ajjan, R. A.; Antunes, M. J.; Christodorescu, R. M.; Crawford, C.; Di Angelantonio, E.; Eliasson, B., 2023 ESC Guidelines for the management of cardiovascular disease in patients with diabetes: Developed by the task force on the management of cardiovascular disease in patients

with diabetes of the European Society of Cardiology (ESC). *European Heart Journal* **2023,** *44* (39), 4043-4140.
21. Packer, M.; Anker, S. D.; Butler, J.; Filippatos, G.; Pocock, S. J.; Carson, P.; Januzzi, J.; Verma, S.; Tsutsui, H.; Brueckmann, M., Cardiovascular and renal outcomes with empagliflozin in heart failure. *New England Journal of Medicine* **2020,** *383* (15), 1413-1424.
22. Moiz, A.; Filion, K. B.; Tsoukas, M. A.; Yu, O. H.; Peters, T. M.; Eisenberg, M. J., The expanding role of GLP-1 receptor agonists: a narrative review of current evidence and future directions. *eClinicalMedicine* **2025,** *86*.
23. Rosenstock, J.; Hsia, S.; Nevarez Ruiz, L.; Eyde, S.; Cox, D.; Wu, W.-S.; Liu, R.; Li, J.; Fernández Landó, L.; Denning, M., Orforglipron, an oral small-molecule GLP-1 receptor agonist, in early type 2 diabetes. *New England Journal of Medicine* **2025**.
24. Jastreboff, A. M.; Kaplan, L. M.; Frías, J. P.; Wu, Q.; Du, Y.; Gurbuz, S.; Coskun, T.; Haupt, A.; Milicevic, Z.; Hartman, M. L., Triple–hormone-receptor agonist retatrutide for obesity—a phase 2 trial. *New England Journal of Medicine* **2023,** *389* (6), 514-526.
25. de Jong, M.; van der Worp, H. B.; van der Graaf, Y.; Visseren, F. L.; Westerink, J., Pioglitazone and the secondary prevention of cardiovascular disease. A meta-analysis of randomized-controlled trials. *Cardiovascular Diabetology* **2017,** *16*, 1-11.
26. Chiasson, J.-L.; Josse, R. G.; Gomis, R.; Hanefeld, M.; Karasik, A.; Laakso, M., Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *The Lancet* **2002,** *359* (9323), 2072-2077.
27. Holman, R. R.; Coleman, R. L.; Chan, J. C.; Chiasson, J.-L.; Feng, H.; Ge, J.; Gerstein, H. C.; Gray, R.; Huo, Y.; Lang, Z., Effects of acarbose on cardiovascular and diabetes outcomes in patients with coronary heart disease and impaired glucose tolerance (ACE): a randomised, double-blind, placebo-controlled trial. *The Lancet Diabetes & Endocrinology* **2017,** *5* (11), 877-886.
28. Fox, C. S.; Golden, S. H.; Anderson, C.; Bray, G. A.; Burke, L. E.; De Boer, I. H.; Deedwania, P.; Eckel, R. H.; Ershow, A. G.; Fradkin, J., Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: a scientific statement from the American Heart Association and the American Diabetes Association. *Circulation* **2015,** *132* (8), 691-718.
29. Fonseca, V. A.; Handelsman, Y.; Staels, B., Colesevelam lowers glucose and lipid levels in type 2 diabetes: the clinical evidence. *Diabetes, Obesity and Metabolism* **2010,** *12* (5), 384-392.
30. Busch, C. B.; Meiring, S.; van Baar, A. C.; Holleman, F.; Nieuwdorp, M.; Bergman, J. J., Recellularization via electroporation therapy of the duodenum combined with glucagon-like peptide-1 receptor agonist to replace insulin therapy in patients with type 2 diabetes: 12-month results of a first-in-human study. *Gastrointestinal Endoscopy* **2024,** *100* (5), 896-904.
31. Thomas, D.; Elliott, E. J.; Naughton, G. A., Exercise for type 2 diabetes mellitus. *Cochrane Database of Systematic Reviews* **2006,** (3).
32. Oral & Injectable Medications for Type 2 Diabetes. https://diabetes.org/health-wellness/medication/oral-other-injectable-diabetes-medications (accessed 03/11/2025).
33. Wheatcroft, S. B.; Kearney, M. T., IGF-dependent and IGF-independent actions of IGF-binding protein-1 and-2: implications for metabolic homeostasis. *Trends in Endocrinology & Metabolism* **2009,** *20* (4), 153-162.
34. Hakuno, F.; Takahashi, S.-I., 40 years of IGF1: IGF1 receptor signaling pathways. *Journal of Molecular Endocrinology* **2018,** *61* (1), T69-T86.
35. Bae, J.-H.; Song, D.-K.; Im, S.-S., Regulation of IGFBP-1 in metabolic diseases. *Journal of Lifestyle Medicine* **2013,** *3* (2), 73-9.
36. Jones, J. I.; Clemmons, D. R., Insulin-like growth factors and their binding proteins: biological actions. *Endocrine Reviews* **1995,** *16* (1), 3-34.
37. Jones, J. I.; Busby, W.; Wright, G.; Smith, C.; Kimack, N.; Clemmons, D., Identification of the sites of phosphorylation in insulin-like growth factor binding protein-1. Regulation of its affinity by phosphorylation of serine 101. *Journal of Biological Chemistry* **1993,** *268* (2), 1125-1131.
38. Ankrapp, D.; Jones, J.; Clemmons, D., Characterization of insulin-like growth factor binding protein-1 kinases from human hepatoma cells. *Journal of Cellular Biochemistry* **1996,** *60* (3), 387-399.

39. Sala, A.; Capaldi, S.; Campagnoli, M.; Faggion, B.; Labo, S.; Perduca, M.; Romano, A.; Carrizo, M. E.; Valli, M.; Visai, L., Structure and properties of the C-terminal domain of insulin-like growth factor-binding protein-1 isolated from human amniotic fluid. *Journal of Biological Chemistry* **2005,** *280* (33), 29812-29819.
40. Forbes, B. E.; McCarthy, P.; Norton, R. S., Insulin-like growth factor binding proteins: a structural perspective. *Frontiers in Endocrinology* **2012,** *3*, 38.
41. Moreno-Layseca, P.; Streuli, C. H., Signalling pathways linking integrins with cell cycle progression. *Matrix Biology* **2014,** *34*, 144-153.
42. Kumar, C. C., Signaling by integrin receptors. *Oncogene* **1998,** *17* (11), 1365-1373.
43. Rajwani, A.; Ezzat, V.; Smith, J.; Yuldasheva, N. Y.; Duncan, E. R.; Gage, M.; Cubbon, R. M.; Kahn, M. B.; Imrie, H.; Abbas, A., Increasing circulating IGFBP1 levels improves insulin sensitivity, promotes nitric oxide production, lowers blood pressure, and protects against atherosclerosis. *Diabetes* **2012,** *61* (4), 915-924.
44. Haywood, N. J.; Slater, T. A.; Matthews, C. J.; Wheatcroft, S. B., The insulin like growth factor and binding protein family: Novel therapeutic targets in obesity & diabetes. *Molecular Metabolism* **2019,** *19*, 86-96.
45. Bhangoo, A.; Gupta, R.; Shelov, S. P.; Carey, D. E.; Accacha, S.; Fennoy, I.; Altshuler, L.; Lowell, B.; Rapaport, R.; Rosenfeld, W., Fasting Serum IGFBP-1 as a Marker of Insulin Resistance in Diverse School Age Groups. *Frontiers in Endocrinology* **2022,** *13*, 840361.
46. Brismar, K.; Hilding, A.; Ansurudeen, I.; Flyvbjerg, A.; Frystyk, J.; Östenson, C.-G., Adiponectin, IGFBP-1 and-2 are independent predictors in forecasting prediabetes and type 2 diabetes. *Frontiers in Endocrinology* **2023,** *13*, 1092307.
47. Zhang, W. B.; Aleksic, S.; Gao, T.; Weiss, E. F.; Demetriou, E.; Verghese, J.; Holtzer, R.; Barzilai, N.; Milman, S., Insulin-like growth factor-1 and IGF binding proteins predict all-cause mortality and morbidity in older adults. *Cells* **2020,** *9* (6), 1368.
48. Heald, A.; Cruickshank, J.; Riste, L.; Cade, J.; Anderson, S.; Greenhalgh, A.; Sampayo, J.; Taylor, W.; Fraser, W.; White, A., Close relation of fasting insulin-like growth factor binding protein-1 (IGFBP-1) with glucose tolerance and cardiovascular risk in two populations. *Diabetologia* **2001,** *44*, 333-339.
49. Leinonen, E. S.; Salonen, J. T.; Salonen, R. M.; Koistinen, R. A.; Leinonen, P. J.; Sarna, S. S.; Taskinen, M.-R., Reduced IGFBP-1 is associated with thickening of the carotid wall in type 2 diabetes. *Diabetes Care* **2002,** *25* (10), 1807-1812.
50. Wheatcroft, S. B.; Kearney, M. T.; Shah, A. M.; Ezzat, V. A.; Miell, J. R.; Modo, M.; Williams, S. C.; Cawthorn, W. P.; Medina-Gomez, G.; Vidal-Puig, A., IGF-binding protein-2 protects against the development of obesity and insulin resistance. *Diabetes* **2007,** *56* (2), 285-294.
51. Lay, A. C.; Hale, L. J.; Stowell-Connolly, H.; Pope, R. J.; Nair, V.; Ju, W.; Marquez, E.; Rollason, R.; Hurcombe, J. A.; Hayes, B., IGFBP-1 expression is reduced in human type 2 diabetic glomeruli and modulates β1-integrin/FAK signalling in human podocytes. *Diabetologia* **2021,** *64*, 1690-1702.
52. Gupta, M. B.; Abu Shehab, M.; Nygard, K.; Biggar, K.; Singal, S. S.; Santoro, N.; Powell, T. L.; Jansson, T., IUGR is associated with marked hyperphosphorylation of decidual and maternal plasma IGFBP-1. *The Journal of Clinical Endocrinology & Metabolism* **2019,** *104* (2), 408-422.
53. Åberg, D.; Gadd, G.; Jood, K.; Redfors, P.; Stanne, T. M.; Isgaard, J.; Blennow, K.; Zetterberg, H.; Jern, C.; Åberg, N. D., Serum IGFBP-1 Concentration as a Predictor of Outcome after Ischemic Stroke—A Prospective Observational Study. *International Journal of Molecular Sciences* **2023,** *24* (11), 9120.
54. Goodman, S. L.; Picard, M., Integrins as therapeutic targets. *Trends in Pharmacological Sciences* **2012,** *33* (7), 405-412.
55. Barczyk, M.; Carracedo, S.; Gullberg, D., Integrins. *Cell and Tissue Research* **2010,** *339* (1), 269-280.
56. Pankov, R.; Yamada, K. M., Fibronectin at a glance. *Journal of Cell Science* **2002,** *115* (20), 3861-3863.
57. A Rubtsov, M.; S Syrkina, M.; Aliev, G., RGD-based therapy: principles of selectivity. *Current Pharmaceutical Design* **2016,** *22* (7), 925-932.
58. Schumacher, S.; Dedden, D.; Nunez, R. V.; Matoba, K.; Takagi, J.; Biertümpfel, C.; Mizuno, N., Structural insights into integrin α5β1 opening by fibronectin ligand. *Science Advances* **2021,** *7* (19), eabe9716.

59. Nagae, M.; Re, S.; Mihara, E.; Nogi, T.; Sugita, Y.; Takagi, J., Crystal structure of α5β1 integrin ectodomain: Atomic details of the fibronectin receptor. *Journal of Cell Biology* **2012,** *197* (1), 131-140.
60. Takagi, J.; Strokovich, K.; Springer, T. A.; Walz, T., Structure of integrin α5β1 in complex with fibronectin. *The EMBO Journal* **2003,** *22* (18), 4607-4615.
61. Humphries, J. D.; Askari, J. A.; Zhang, X.-P.; Takada, Y.; Humphries, M. J.; Mould, A. P., Molecular basis of ligand recognition by integrin α5β1: II. Specificity of Arg-Gly-Asp binding is determined by Trp157 of the α subunit. *Journal of Biological Chemistry* **2000,** *275* (27), 20337-20345.
62. Sun, Z.; Costell, M.; Fässler, R., Integrin activation by talin, kindlin and mechanical forces. *Nature Cell Biology* **2019,** *21* (1), 25-31.
63. Springer, T. A.; Dustin, M. L., Integrin inside-out signaling and the immunological synapse. *Current Opinion in Cell Biology* **2012,** *24* (1), 107-115.
64. Su, Y.; Xia, W.; Li, J.; Walz, T.; Humphries, M. J.; Vestweber, D.; Cabañas, C.; Lu, C.; Springer, T. A., Relating conformation to function in integrin α5β1. *Proceedings of the National Academy of Sciences* **2016,** *113* (27), E3872-E3881.
65. Li, J.; Su, Y.; Xia, W.; Qin, Y.; Humphries, M. J.; Vestweber, D.; Cabañas, C.; Lu, C.; Springer, T. A., Conformational equilibria and intrinsic affinities define integrin activation. *The EMBO Journal* **2017,** *36* (5), 629-645.
66. Pang, X.; He, X.; Qiu, Z.; Zhang, H.; Xie, R.; Liu, Z.; Gu, Y.; Zhao, N.; Xiang, Q.; Cui, Y., Targeting integrin pathways: mechanisms and advances in therapy. *Signal Transduction and Targeted Therapy* **2023,** *8* (1), 1.
67. Li, J.; Yan, J.; Springer, T. A., Low-affinity integrin states have faster ligand-binding kinetics than the high-affinity state. *Elife* **2021,** *10*, e73359.
68. Xia, W.; Springer, T. A., Metal ion and ligand binding of integrin α5β1. *Proceedings of the National Academy of Sciences* **2014,** *111* (50), 17863-17868.
69. Hou, J.; Yan, D.; Liu, Y.; Huang, P.; Cui, H., The roles of integrin α5β1 in human cancer. *OncoTargets and Therapy* **2020**, 13329-13344.
70. Folkman, J., Antiangiogenesis in cancer therapy—endostatin and its mechanisms of action. *Experimental Cell Research* **2006,** *312* (5), 594-607.
71. Slack, R.; Macdonald, S.; Roper, J.; Jenkins, R.; Hatley, R., Emerging therapeutic opportunities for integrin inhibitors. *Nature Reviews Drug Discovery* **2022,** *21* (1), 60-78.
72. Fei, X.; Jung, S.; Kwon, S.; Kim, J.; Corson, T. W.; Seo, S.-Y., Challenges and opportunities of developing small-molecule therapies for age-related macular degeneration. *Archives of Pharmacal Research* **2024,** *47* (6), 538-557.
73. Boyer, D. S.; Gonzalez, V. H.; Kunimoto, D. Y.; Maturi, R. K.; Roe, R. H.; Singer, M. A.; Xavier, S.; Kornfield, J. A.; Kuppermann, B. D.; Quiroz-Mercado, H., Safety and efficacy of intravitreal risuteganib for non-exudative AMD: a multicenter, phase 2a, randomized, clinical trial. *Ophthalmic Surgery, Lasers and Imaging Retina* **2021,** *52* (6), 327-335.
74. Cirkel, G. A.; Kerklaan, B. M.; Vanhoutte, F.; Der Aa, A. V.; Lorenzon, G.; Namour, F.; Pujuguet, P.; Darquenne, S.; De Vos, F. Y.; Snijders, T. J., A dose escalating phase I study of GLPG0187, a broad spectrum integrin receptor antagonist, in adult patients with progressive high-grade glioma and other advanced solid malignancies. *Investigational new drugs* **2016,** *34*, 184-192.
75. Ludwig, B. S.; Kessler, H.; Kossatz, S.; Reuning, U., RGD-binding integrins revisited: how recently discovered functions and novel synthetic ligands (re-) shape an ever-evolving field. *Cancers* **2021,** *13* (7), 1711.
76. Kapp, T. G.; Fottner, M.; Maltsev, O. V.; Kessler, H., Small cause, great impact: modification of the guanidine group in the RGD motif controls integrin subtype selectivity. *Angewandte Chemie International Edition* **2016,** *55* (4), 1540-1543.
77. Reed, N. I.; Jo, H.; Chen, C.; Tsujino, K.; Arnold, T. D.; DeGrado, W. F.; Sheppard, D., The αvβ1 integrin plays a critical in vivo role in tissue fibrosis. *Science Translational Medicine* **2015,** *7* (288), 288ra79-288ra79.
78. Sabat, M.; Carney, D. W.; Hernandez-Torres, G.; Gibson, T. S.; Balakrishna, D.; Zou, H.; Xu, R.; Chen, C.-H.; de Jong, R.; Dougan, D. R., Design and Discovery of a Potent and Selective Inhibitor of Integrin αvβ1. *Journal of Medicinal Chemistry* **2024,** *67* (12), 10306-10320.

79. Lippa, R. A.; Barrett, J.; Pal, S.; Rowedder, J. E.; Murphy, J. A.; Barrett, T. N., Discovery of the first potent and selective αvβ5 integrin inhibitor based on an amide-containing core. *European Journal of Medicinal Chemistry* **2020,** *208*, 112719.
80. Li, J.; Fukase, Y.; Shang, Y.; Zou, W.; Muñoz-Félix, J. M.; Buitrago, L.; Van Agthoven, J.; Zhang, Y.; Hara, R.; Tanaka, Y., Novel pure αVβ3 integrin antagonists that do not induce receptor extension, prime the receptor, or enhance angiogenesis at low concentrations. *ACS Pharmacology & Translational Science* **2019,** *2* (6), 387-401.
81. Hatley, R. J.; Macdonald, S. J.; Slack, R. J.; Le, J.; Ludbrook, S. B.; Lukey, P. T., An αv-RGD integrin inhibitor toolbox: drug discovery insight, challenges and opportunities. *Angewandte Chemie International Edition* **2018,** *57* (13), 3298-3321.
82. Li, J.; Vootukuri, S.; Shang, Y.; Negri, A.; Jiang, J.-k.; Nedelman, M.; Diacovo, T. G.; Filizola, M.; Thomas, C. J.; Coller, B. S., RUC-4: a novel αIIbβ3 antagonist for prehospital therapy of myocardial infarction. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2014,** *34* (10), 2321-2329.
83. Askew, B. C.; Furuya, T.; Edwards, D. S., Ocular distribution and pharmacodynamics of SF0166, a topically administered αvβ3 integrin antagonist, for the treatment of retinal diseases. *The Journal of Pharmacology and Experimental Therapeutics* **2018,** *366* (2), 244-250.
84. Doñate, F.; Parry, G. C.; Shaked, Y.; Hensley, H.; Guan, X.; Beck, I.; Tel-Tsur, Z.; Plunkett, M. L.; Manuia, M.; Shaw, D. E., Pharmacology of the novel antiangiogenic peptide ATN-161 (Ac-PHSCN-NH2): observation of a U-shaped dose-response curve in several preclinical models of angiogenesis and tumor growth. *Clinical Cancer Research* **2008,** *14* (7), 2137-2144.
85. Zhu, J.; Choi, W.-S.; McCoy, J. G.; Negri, A.; Zhu, J.; Naini, S.; Li, J.; Shen, M.; Huang, W.; Bougie, D., Structure-guided design of a high-affinity platelet integrin αIIbβ3 receptor antagonist that disrupts Mg2+ binding to the MIDAS. *Science Translational Medicine* **2012,** *4* (125), 125ra32-125ra32.
86. Hatley, R. J.; Barrett, T. N.; Slack, R. J.; Watson, M. E.; Baillache, D. J.; Gruszka, A.; Washio, Y.; Rowedder, J. E.; Pogány, P.; Pal, S., The Design of Potent, Selective and Drug-Like RGD αvβ1 Small-Molecule Inhibitors Derived from non-RGD α4β1 Antagonists. *ChemMedChem* **2019,** *14* (14), 1315-1320.
87. Heckmann, D.; Meyer, A.; Laufer, B.; Zahn, G.; Stragies, R.; Kessler, H., Rational design of highly active and selective ligands for the α5β1 integrin receptor. *ChemBioChem* **2008,** *9* (9), 1397-1407.
88. Heckmann, D.; Laufer, B.; Marinelli, L.; Limongelli, V.; Novellino, E.; Zahn, G.; Stragies, R.; Kessler, H., Breaking the Dogma of the Metal-Coordinating Carboxylate Group in Integrin Ligands: Introducing Hydroxamic Acids to the MIDAS To Tune Potency and Selectivity. *Angewandte Chemie International Edition* **2009,** *48* (24), 4436-4440.
89. Matsuoka, K.; Watanabe, M.; Ohmori, T.; Nakajima, K.; Ishida, T.; Ishiguro, Y.; Kanke, K.; Kobayashi, K.; Hirai, F.; Watanabe, K., AJM300 (carotegrast methyl), an oral antagonist of α4-integrin, as induction therapy for patients with moderately active ulcerative colitis: a multicentre, randomised, double-blind, placebo-controlled, phase 3 study. *The Lancet Gastroenterology & Hepatology* **2022,** *7* (7), 648-657.
90. Mas-Moruno, C.; Rechenmacher, F.; Kessler, H., Cilengitide: the first anti-angiogenic small molecule drug candidate. Design, synthesis and clinical evaluation. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)* **2010,** *10* (10), 753-768.
91. Kapp, T. G.; Rechenmacher, F.; Neubauer, S.; Maltsev, O. V.; Cavalcanti-Adam, E. A.; Zarka, R.; Reuning, U.; Notni, J.; Wester, H.-J.; Mas-Moruno, C., A comprehensive evaluation of the activity and selectivity profile of ligands for RGD-binding integrins. *Scientific Reports* **2017,** *7* (1), 1-13.
92. Fraioli, R.; Rechenmacher, F.; Neubauer, S.; Manero, J. M.; Gil, J.; Kessler, H.; Mas-Moruno, C., Mimicking bone extracellular matrix: Integrin-binding peptidomimetics enhance osteoblast-like cells adhesion, proliferation, and differentiation on titanium. *Colloids and Surfaces B: Biointerfaces* **2015,** *128*, 191-200.
93. Rechenmacher, F.; Neubauer, S.; Polleux, J.; Mas-Moruno, C.; De Simone, M.; Cavalcanti-Adam, E. A.; Spatz, J. P.; Fässler, R.; Kessler, H., Functionalizing αvβ3-or α5β1-selective integrin antagonists for surface coating: a method to discriminate integrin subtypes in vitro. *Angewandte Chemie International Edition* **2013,** *52* (5).

94. Galletti, P.; Soldati, R.; Pori, M.; Durso, M.; Tolomelli, A.; Gentilucci, L.; Dattoli, S. D.; Baiula, M.; Spampinato, S.; Giacomini, D., Targeting integrins αvβ3 and α5β1 with new β-lactam derivatives. *European Journal of Medicinal Chemistry* **2014,** *83*, 284-293.
95. Baiula, M.; Galletti, P.; Martelli, G.; Soldati, R.; Belvisi, L.; Civera, M.; Dattoli, S. D.; Spampinato, S. M.; Giacomini, D., New β-lactam derivatives modulate cell adhesion and signaling mediated by RGD-binding and leukocyte integrins. *Journal of Medicinal Chemistry* **2016,** *59* (21), 9721-9742.
96. Sadybekov, A. V.; Katritch, V., Computational approaches streamlining drug discovery. *Nature* **2023,** *616* (7958), 673-685.
97. Hendrickson, J. B., Systematic characterization of structures and reactions for use in organic synthesis. *Journal of the American Chemical Society* **1971,** *93* (25), 6847-6854.
98. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977,** *267* (5612), 585-590.
99. Humblet, C.; Marshall, G. R., Three-dimensional computer modeling as an aid to drug design. *Drug Development Research* **1981,** *1* (4), 409-434.
100. Niazi, S. K.; Mariam, Z., Computer-aided drug design and drug discovery: a prospective analysis. *Pharmaceuticals* **2023,** *17* (1), 22.
101. Hassan Baig, M.; Ahmad, K.; Roy, S.; Mohammad Ashraf, J.; Adil, M.; Haris Siddiqui, M.; Khan, S.; Amjad Kamal, M.; Provazník, I.; Choi, I., Computer aided drug design: success and limitations. *Current Pharmaceutical Design* **2016,** *22* (5), 572-581.
102. Weber, I. T.; Wang, Y.-F.; Harrison, R. W., HIV protease: Historical perspective and current research. *Viruses* **2021,** *13* (5), 839.
103. Moyle, G., Drug Evaluation: Anti-infectives: Saquinavir: A review of its development, pharmacological properties and clinical use. *Expert Opinion on Investigational Drugs* **1996,** *5* (2), 155-167.
104. Kempf, D. J.; Sham, H. L.; Marsh, K. C.; Flentge, C. A.; Betebenner, D.; Green, B. E.; McDonald, E.; Vasavanonda, S.; Saldivar, A.; Wideburg, N. E., Discovery of ritonavir, a potent inhibitor of HIV protease with high oral bioavailability and clinical efficacy. *Journal of Medicinal Chemistry* **1998,** *41* (4), 602-617.
105. Hartman, G. D.; Egbertson, M. S.; Halczenko, W.; Laswell, W. L.; Duggan, M. E.; Smith, R. L.; Naylor, A. M.; Manno, P. D.; Lynch, R. J., Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *Journal of Medicinal Chemistry* **1992,** *35* (24), 4640-4642.
106. Sterling, T.; Irwin, J. J., ZINC 15–ligand discovery for everyone. *Journal of Chemical Information and Modeling* **2015,** *55* (11), 2324-2337.
107. Zdrazil, B.; Felix, E.; Hunter, F.; Manners, E. J.; Blackshaw, J.; Corbett, S.; de Veij, M.; Ioannidis, H.; Lopez, D. M.; Mosquera, J. F., The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research* **2024,** *52* (D1), D1180-D1192.
108. Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D., Molecular docking and structure-based drug design strategies. *Molecules* **2015,** *20* (7), 13384-13421.
109. Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A., A shape-based 3-D scaffold hopping method and its application to a bacterial protein− protein interaction. *Journal of medicinal chemistry* **2005,** *48* (5), 1489-1495.
110. O Ebalunode, J.; Zheng, W., Molecular shape technologies in drug discovery: methods and applications. *Current Topics in Medicinal Chemistry* **2010,** *10* (6), 669-679.
111. ROCS Overview. https://docs.eyesopen.com/applications/rocs/rocs/rocs_overview.html (accessed 20/06/2024).
112. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **2004,** *47* (7), 1739-1749.
113. Haywood, N. J.; Cordell, P. A.; Tang, K. Y.; Makova, N.; Yuldasheva, N. Y.; Imrie, H.; Viswambharan, H.; Bruns, A. F.; Cubbon, R. M.; Kearney, M. T., Insulin-like growth factor binding protein 1 could improve glucose regulation and insulin sensitivity through its RGD domain. *Diabetes* **2017,** *66* (2), 287-299.
114. Bellenie, B. R.; Cheung, K.-M. J.; Varela, A.; Pierrat, O. A.; Collie, G. W.; Box, G. M.; Bright, M. D.; Gowan, S.; Hayes, A.; Rodrigues, M. J., Achieving in vivo target depletion through the discovery and optimization of benzimidazolone BCL6 degraders. *Journal of Medicinal Chemistry* **2020,** *63* (8), 4047-4068.

115. Bickler, B. *Successful Flash Chromatography*; Biotage.com, 2018.
116. Aziz, A.; Haywood, N. J.; Cordell, P. A.; Smith, J.; Yuldasheva, N. Y.; Sengupta, A.; Ali, N.; Mercer, B. N.; Mughal, R. S.; Riches, K., Insulinlike growth factor–binding protein-1 improves vascular endothelial repair in male mice in the setting of insulin resistance. *Endocrinology* **2018,** *159* (2), 696-709.
117. Liang, C.-C.; Park, A. Y.; Guan, J.-L., In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nature Protocols* **2007,** *2* (2), 329-333.
118. Capelli, D.; Scognamiglio, V.; Montanari, R., Surface plasmon resonance technology: Recent advances, applications and experimental cases. *TrAC Trends in Analytical Chemistry* **2023,** *163*, 117079.
119. Singh, A. K.; Mittal, S.; Das, M.; Saharia, A.; Tiwari, M., Optical biosensors: A decade in review. *Alexandria Engineering Journal* **2023,** *67*, 673-691.
120. Pang, X.; Sun, X.; Gu, Y.; He, X.; Gong, K.; Song, S.; Zhang, J.; Xia, J.; Liu, Z.; Cui, Y., Discovery of C19-9 as a novel non-RGD inhibitor of αvβ3 to overcome enzalutamide resistance in castration-resistant prostate cancer. *Signal Transduction and Targeted Therapy* **2023,** *8* (1), 60.
121. Cheng, X.; Li, C.; Zhou, Z.; Wu, Z., Macrocyclic RGD-peptides with high selectivity for αvβ3 integrin in cancer imaging and therapy. *RSC Medicinal Chemistry* **2025,** *16*, 3077-3083.
122. 2BScientific Biotinylated Human Integrin alpha 5 beta 1. https://www.2bscientific.com/products/acro-biosystems/it1-h82wa/biotinylated-human-integrin-alpha-5-beta-1-itga5itgb1-heterodimer-protein-hisavitagtmtag-free (accessed 16/02/2025).
123. K, A. *Biacore T200: CMI Getting Started Guide to Surface Plasmon Resonance* Harvard University: 2025.
124. Arnett, K. Biacore T200: CMI Getting Started Guide to Surface Plasmon Resonance https://cmi.hms.harvard.edu/surface-plasmon-resonance (accessed 05/04/2025).
125. Moberg, A. L., Maria; Shuman, Cynthia; Belcher, Paul *Why sensitivity matters: reliable hit characterization for challenging targets using high-sensitivity SPR systems*; Cytiva: 2020.
126. Gawas, P. P.; Ramakrishna, B.; Veeraiah, N.; Nutalapati, V., Multifunctional hydantoins: Recent advances in optoelectronics and medicinal drugs from Academia to the chemical industry. *Journal of Materials Chemistry C* **2021,** *9* (46), 16341-16377.
127. Keppel Hesselink, J. M., Phenytoin: a step by step insight into its multiple mechanisms of action—80 years of mechanistic studies in neuropharmacology. *Journal of Neurology* **2017,** *264* (9), 2043-2047.
128. Söderholm, A. A.; Lehtovuori, P. T.; Nyrönen, T. H., Three-dimensional structure− activity relationships of nonsteroidal ligands in complex with androgen receptor ligand-binding domain. *Journal of Medicinal Chemistry* **2005,** *48* (4), 917-925.
129. Krause, T.; Gerbershagen, M.; Fiege, M.; Weisshorn, R.; Wappler, F., Dantrolene–a review of its pharmacology, therapeutic use and new developments. *Anaesthesia* **2004,** *59* (4), 364-373.
130. Dos Santos, C.; Dos Santos, L. S.; Franco, O. L., Fosfomycin and nitrofurantoin: classic antibiotics and perspectives. *The Journal of Antibiotics* **2021,** *74* (9), 547-558.
131. Yu, W.; Guo, Z.; Orth, P.; Madison, V.; Chen, L.; Dai, C.; Feltz, R. J.; Girijavallabhan, V. M.; Kim, S. H.; Kozlowski, J. A., Discovery and SAR of hydantoin TACE inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2010,** *20* (6), 1877-1880.
132. Durham, T. B.; Klimkowski, V. J.; Rito, C. J.; Marimuthu, J.; Toth, J. L.; Liu, C.; Durbin, J. D.; Stout, S. L.; Adams, L.; Swearingen, C., Identification of potent and selective hydantoin inhibitors of aggrecanase-1 and aggrecanase-2 that are efficacious in both chemical and surgical models of osteoarthritis. *Journal of Medicinal Chemistry* **2014,** *57* (24), 10476-10485.
133. De Savi, C.; Waterson, D.; Pape, A.; Lamont, S.; Hadley, E.; Mills, M.; Page, K. M.; Bowyer, J.; Maciewicz, R. A., Hydantoin based inhibitors of MMP13—discovery of AZD6605. *Bioorganic & Medicinal Chemistry Letters* **2013,** *23* (16), 4705-4712.
134. Brebion, F.; Gosmini, R.; Deprez, P.; Varin, M.; Peixoto, C.; Alvey, L.; Jary, H.; Bienvenu, N.; Triballeau, N.; Blanque, R., Discovery of GLPG1972/S201086, a potent, selective, and orally bioavailable ADAMTS-5 inhibitor for the treatment of osteoarthritis. *Journal of Medicinal Chemistry* **2021,** *64* (6), 2937-2952.

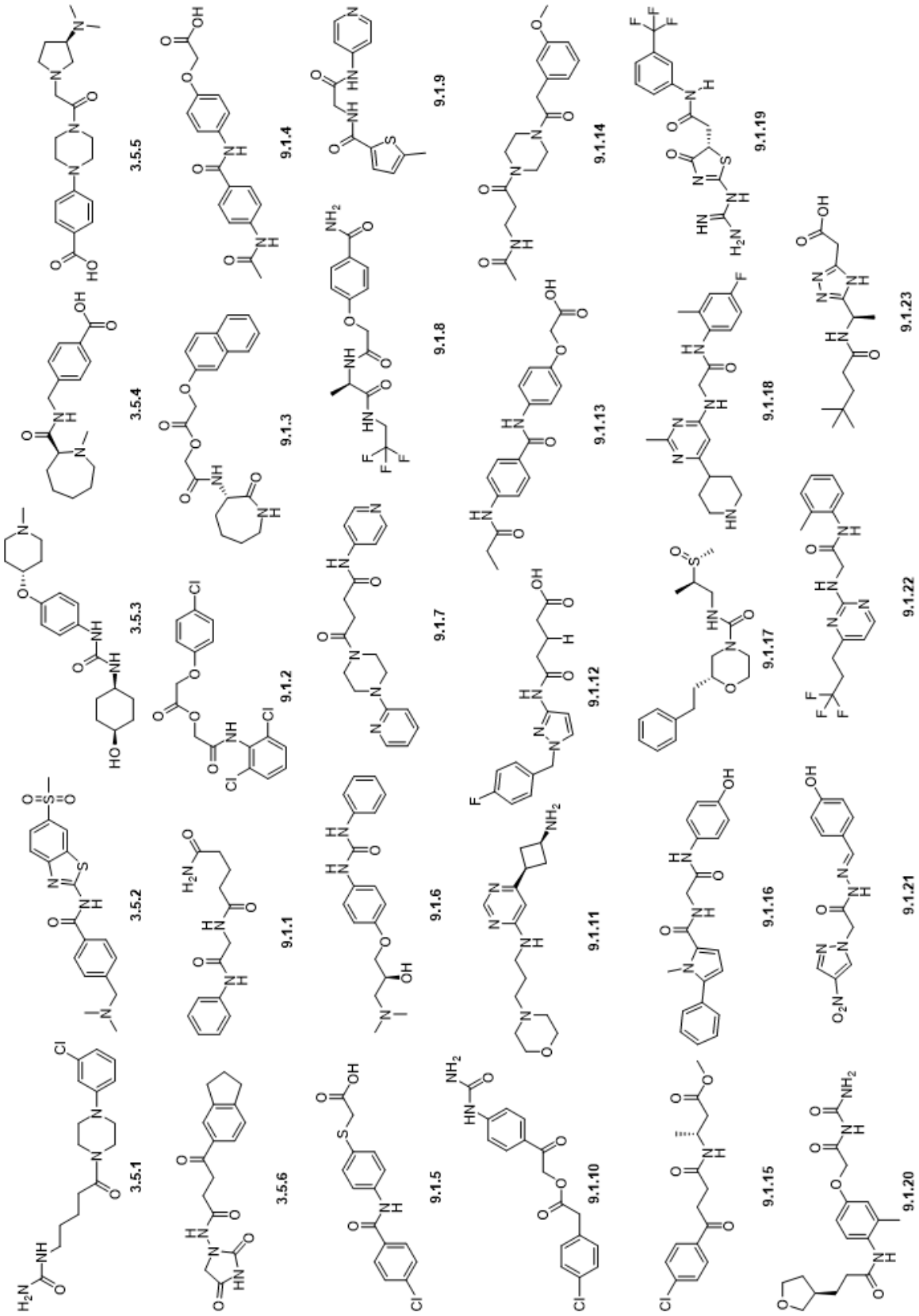
135. Meibom, D.; Wasnaire, P.; Beyer, K.; Broehl, A.; Cancho-Grande, Y.; Elowe, N.; Henninger, K.; Johannes, S.; Jungmann, N.; Krainz, T., BAY-9835: Discovery of the First Orally Bioavailable ADAMTS7 Inhibitor. *Journal of Medicinal Chemistry* **2024,** *67* (4), 2907-2940.
136. Sheppeck II, J. E.; Gilmore, J. L.; Tebben, A.; Xue, C.-B.; Liu, R.-Q.; Decicco, C. P.; Duan, J. J.-W., Hydantoins, triazolones, and imidazolones as selective non-hydroxamate inhibitors of tumor necrosis factor-α converting enzyme (TACE). *Bioorganic & Medicinal Chemistry Letters* **2007,** *17* (10), 2769-2774.
137. Corral, R. A.; Orazi, O. O., Localization of substituents in the hydantoin ring by nuclear magnetic resonance. *Spectrochimica Acta* **1965,** *21* (12), 2119-2123.
138. Safi, Z. S., Theoretical density functional study of gas-phase tautomerization and acidity of 5-methylhydantoin and its thio derivatives. *European Journal of Chemistry* **2012,** *3* (3), 348-355.
139. Bausch, M.; Selmarten, D.; Gostowski, R.; Dobrowolski, P., Potentiometric and spectroscopic investigations of the aqueous phase acid–base chemistry of urazoles and substituted urazoles. *Journal of Physical Organic Chemistry* **1991,** *4* (1), 67-69.
140. Vessières, A.; Kowalski, K.; Zakrzewski, J.; Stepien, A.; Grabowski, M.; Jaouen, G., Synthesis of CpFe (CO)(L) Complexes of Hydantoin Anions (Cp= η5-C5H5, L= CO, PPh3), and the Use of the 5, 5-Diphenylhydantoin Anion Complexes as Tracers in the Nonisotopic Immunoassay CMIA of This Antiepileptic Drug. *Bioconjugate Chemistry* **1999,** *10* (3), 379-385.
141. Dewi, M.; Syaima, H.; Wahyuningsih, S.; Lestari, W.; Rahardjo, S. In *Synthesis and characterization of complex of copper (II) with hydantoin*, Journal of Physics: Conference Series, IOP Publishing: 2023; p 012009.
142. Lippert, B., Alterations of Nucleobase p K a Values upon Metal Coordination: Origins and Consequences. *Progress in Inorganic Chemistry* **2005,** *54*, 385-447.
143. Knobloch, B.; Linert, W.; Sigel, H., Metal ion-binding properties of (N3)-deprotonated uridine, thymidine, and related pyrimidine nucleosides in aqueous solution. *Proceedings of the National Academy of Sciences* **2005,** *102* (21), 7459-7464.
144. Lippert, B., Ligand-pKa Shifts through Metals: Potential Relevance to Ribozyme Chemistry. *Chemistry & Biodiversity* **2008,** *5* (8), 1455-1474.
145. Faris, W. M.; Safi, Z. S., Theoretical investigation of tautomerism stability of hydantoin in the gas phase and in the solution. *Orient J Chem* **2014,** *30*, 1045-1054.
146. Bhikharee, D.; Elzagheid, M.; Rhyman, L.; Ramasami, P., Effect of water or ethanol on the tautomeric stability and proton transfer reaction of all possible tautomers of hydantoin: Implicit v/s explicit solvation. *Journal of Molecular Liquids* **2022,** *347*, 117942.
147. Shabanian, M.; Moghanian, H.; Hajibeygi, M.; Mohamadi, A., A DFT study of solvation effects and NBO analysis on the tautomerism of 1-substituted hydantoin. *Arabian Journal of Chemistry* **2016,** *9*, S776-S780.
148. Leggio, A.; Belsito, E.; De Luca, G.; Di Gioia, M. L.; Leotta, V.; Romio, E.; Siciliano, C.; Liguori, A., One-pot synthesis of amides from carboxylic acids activated using thionyl chloride. *Rsc Advances* **2016,** *6* (41), 34468-34475.
149. Wang, Q.; Liu, G.; Shao, R.; Huang, R., Synthesis and antivirus activity of 1, 3, 5-triazine derivatives. *Heteroatom Chemistry: An International Journal of Main Group Elements* **2003,** *14* (6), 542-545.
150. Murasawa, S.; Iuchi, K.; Sato, S.; Noguchi-Yachide, T.; Sodeoka, M.; Yokomatsu, T.; Dodo, K.; Hashimoto, Y.; Aoyama, H., Small-molecular inhibitors of Ca2+-induced mitochondrial permeability transition (MPT) derived from muscle relaxant dantrolene. *Bioorganic & Medicinal Chemistry* **2012,** *20* (21), 6384-6393.
151. Guo, X.; Shen, Z.; Yu, P.; Chu, H.; Tang, Y.; You, Q., Discovery of benzamide derivatives as potent Kv1. 5 inhibitors. *Journal of Chinese Pharmaceutical Sciences* **2012,** *21* (6), 553.
152. Buchanan, G. W.; Dawson, B. A., Aromatic imine stereochemistry as studied by 13C and 1H NMR of 15N-enriched materials. *Organic Magnetic Resonance* **1980,** *13* (4), 293-298.
153. Heinzman, S. W.; Ganem, B., Mechanism of sodium borohydride-cobaltous chloride reductions. *Journal of the American Chemical Society* **1982,** *104* (24), 6801-6802.
154. Kristensen, S. L.; Rørth, R.; Jhund, P. S.; Docherty, K. F.; Sattar, N.; Preiss, D.; Køber, L.; Petrie, M. C.; McMurray, J. J., Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-

analysis of cardiovascular outcome trials. *The Lancet Diabetes & Endocrinology* **2019,** *7* (10), 776-785.

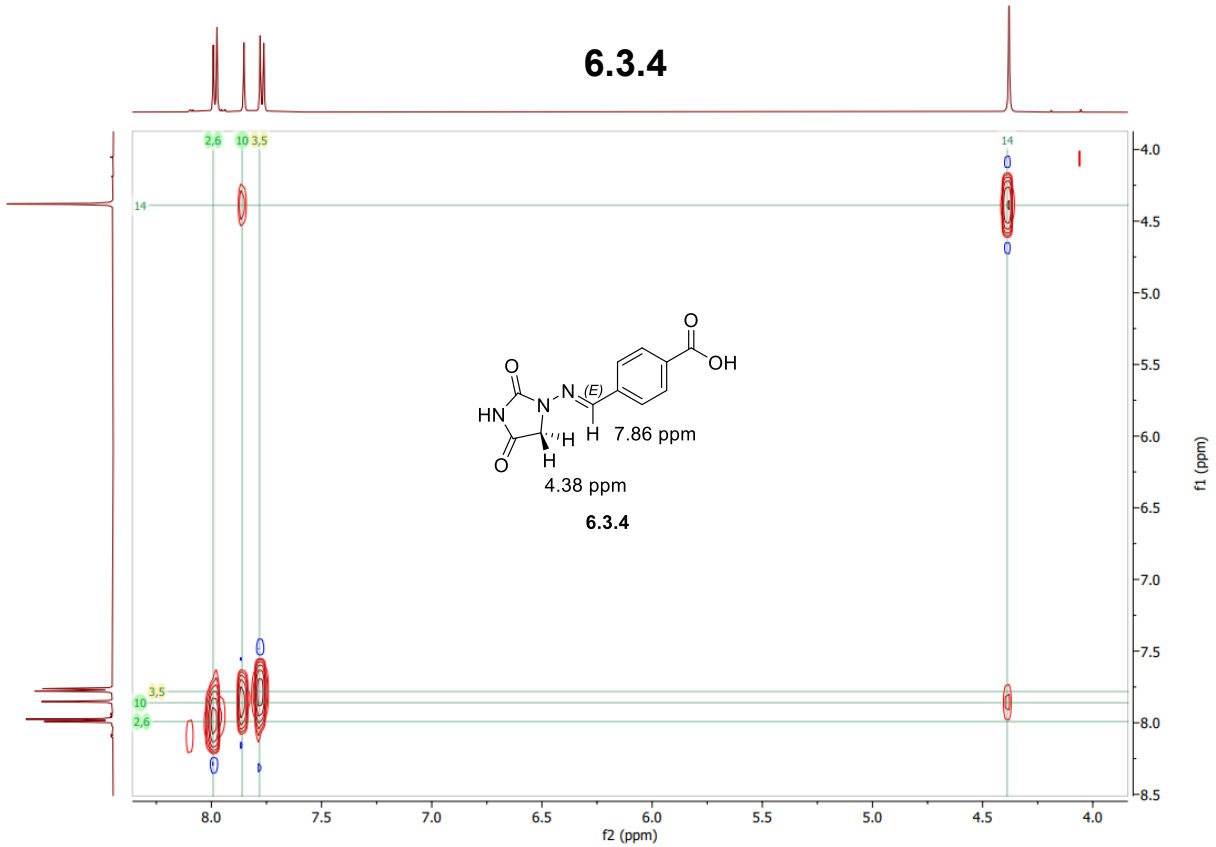
155. McGuire, D. K.; Shih, W. J.; Cosentino, F.; Charbonnel, B.; Cherney, D. Z.; Dagogo-Jack, S.; Pratley, R.; Greenberg, M.; Wang, S.; Huyck, S., Association of SGLT2 inhibitors with cardiovascular and kidney outcomes in patients with type 2 diabetes: a meta-analysis. *JAMA Cardiology* **2021,** *6* (2), 148-158.
156. Wendt, M. D.; Shen, W.; Kunzer, A.; McClellan, W. J.; Bruncko, M.; Oost, T. K.; Ding, H.; Joseph, M. K.; Zhang, H.; Nimmer, P. M., Discovery and structure− activity relationship of antagonists of B-cell lymphoma 2 family proteins with chemopotentiation activity in vitro and in vivo. *Journal of Medicinal Chemistry* **2006,** *49* (3), 1165-1181.
157. Kubota, D.; Ishikawa, M.; Yamamoto, M.; Murakami, S.; Hachisu, M.; Katano, K.; Ajito, K., Tricyclic pharmacophore-based molecules as novel integrin αvβ3 antagonists. Part 1: Design and synthesis of a lead compound exhibiting αvβ3/αIIbβ3 dual antagonistic activity. *Bioorganic & Medicinal Chemistry* **2006,** *14* (7), 2089-2108.
158. Fincham, C. I.; Bressan, A.; D'Andrea, P.; Ettorre, A.; Giuliani, S.; Mauro, S.; Meini, S.; Paris, M.; Quartara, L.; Rossi, C., Design and synthesis of novel sulfonamide-containing bradykinin hB2 receptor antagonists. Synthesis and structure-relationships of α, α-tetrahydropyranylglycine. *Bioorganic & Medicinal Chemistry* **2012,** *20* (6), 2091-2100.
159. He, S.; Dong, G.; Wu, S.; Fang, K.; Miao, Z.; Wang, W.; Sheng, C., Small molecules simultaneously inhibiting p53-murine double minute 2 (MDM2) interaction and histone deacetylases (HDACs): discovery of novel multitargeting antitumor agents. *Journal of Medicinal Chemistry* **2018,** *61* (16), 7245-7260.
160. Ding, H.-W.; Wang, S.; Qin, X.-C.; Wang, J.; Song, H.-R.; Zhao, Q.-C.; Song, S.-J., Design, synthesis, and biological evaluation of some novel 4-aminoquinazolines as Pan-PI3K inhibitors. *Bioorganic & Medicinal Chemistry* **2019,** *27* (13), 2729-2740.
161. DeMong, D.; Dai, X.; Hwa, J.; Miller, M.; Lin, S.-I.; Kang, L.; Stamford, A.; Greenlee, W.; Yu, W.; Wong, M., The Discovery of N-((2 H-Tetrazol-5-yl) methyl)-4-((R)-1-((5 r, 8 R)-8-(tert-butyl)-3-(3, 5-dichlorophenyl)-2-oxo-1, 4-diazaspiro [4.5] dec-3-en-1-yl)-4, 4-dimethylpentyl) benzamide (SCH 900822): A Potent and Selective Glucagon Receptor Antagonist. *Journal of Medicinal Chemistry* **2014,** *57* (6), 2601-2610.
162. Nair, M.; Baugh, C. M., Synthesis and biological evaluation of isofolic acid. *Journal of Medicinal Chemistry* **1974,** *17* (2), 223-226.

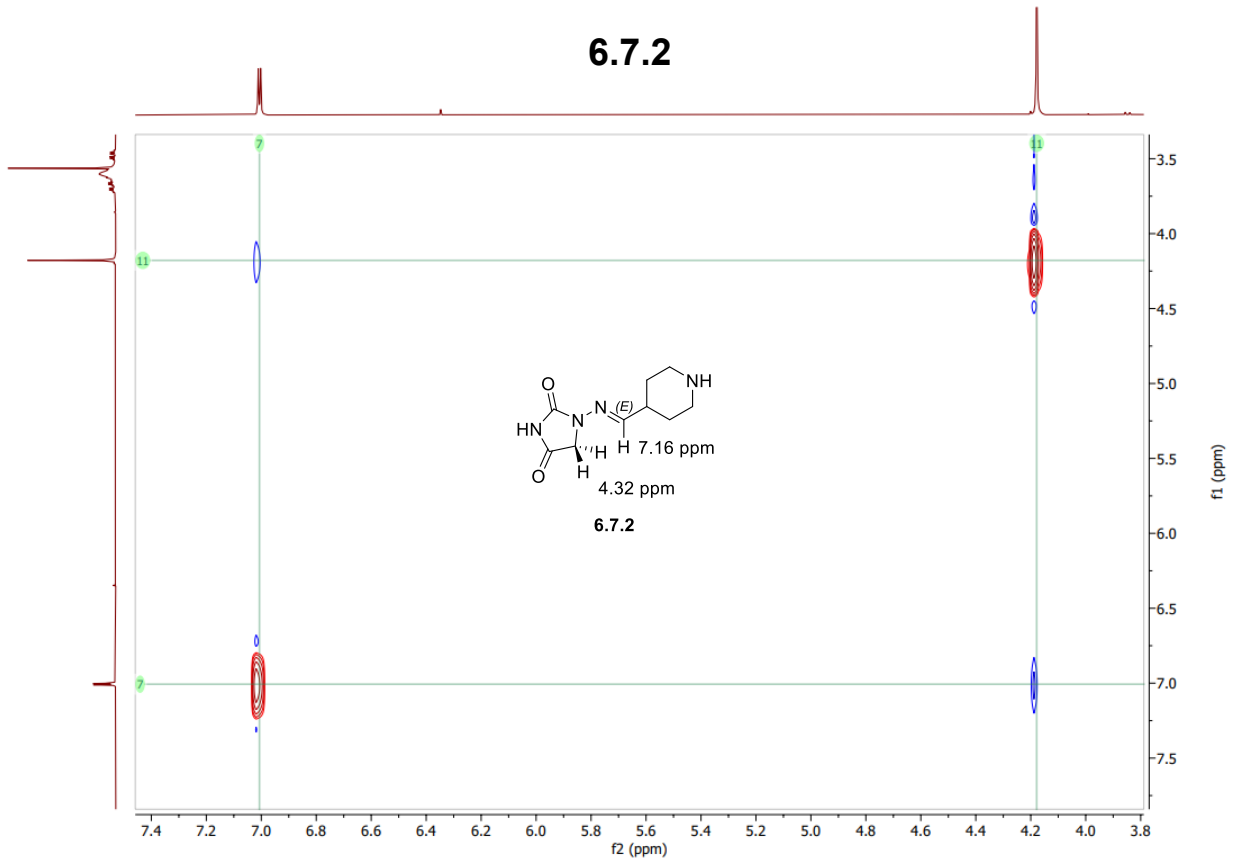
10. Appendix

10.1 ROCS hit compounds

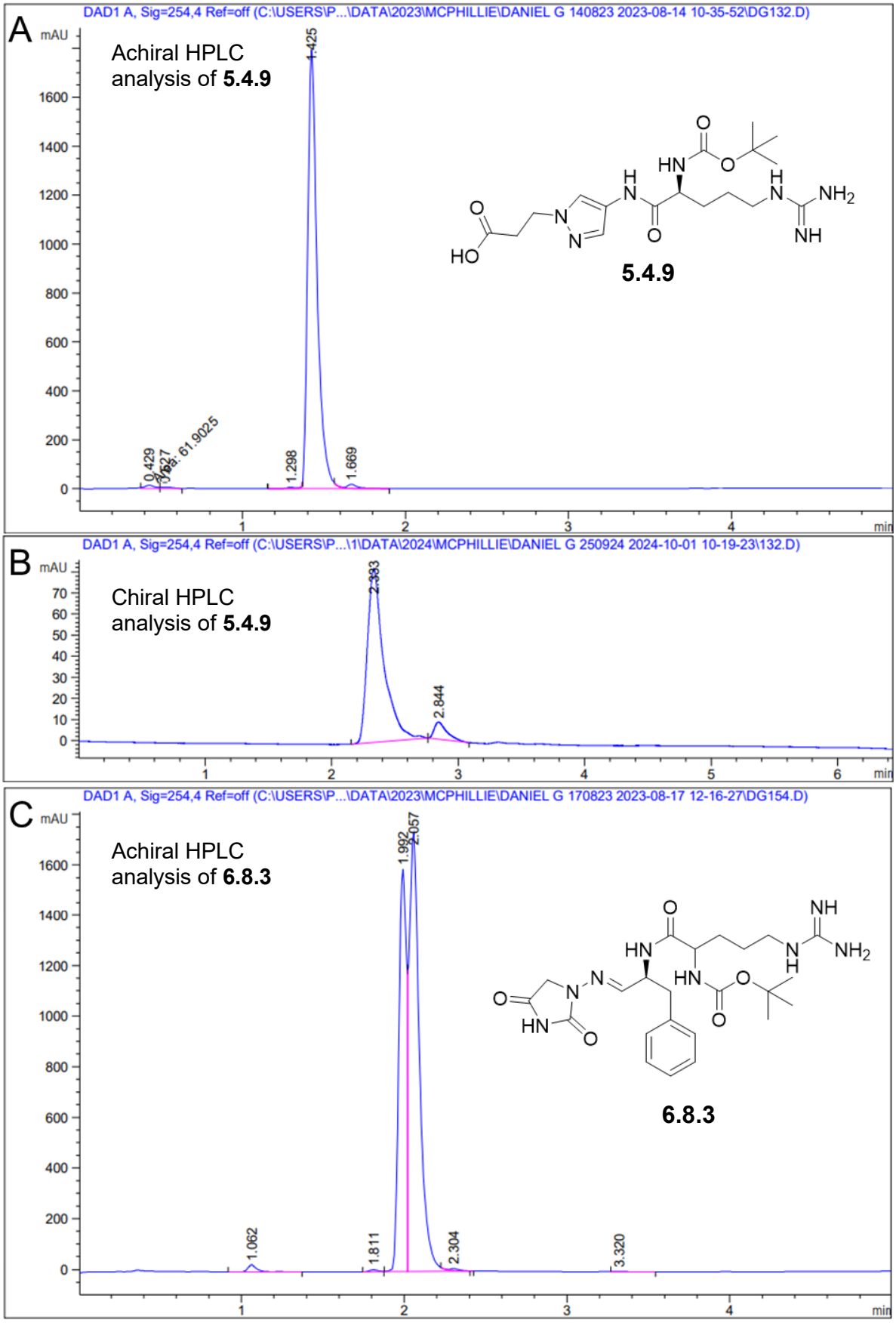


10.2 NOESY stereochemical assignment of compounds **6.3.4** and **6.7.2**





10.3 Racemisation of compounds **5.4.9** and **6.8.3**



Methodology: Achiral HLPC is detailed in section 8.1. Chiral HPLC was completed using a Chiralcel OD-RH, 150 x 4.6 mm, 5 µm column (Chiral, DAIC14724). A gradient of 0-50% MeCN:H_2O was used to elute the compounds. The elution was monitored by UV at 254 nm.