UNIVERSITY OF BIRMINGHAM

Thesis submitted to

The University of Birmingham

for the degree of

Doctor of Philosophy

The design and synthesis of fluorescent peptides and disaccharides for tumour imaging

Marcus J. Main

November 2009

SUPERVISOR: *Dr J. S. Snaith*
School of Chemistry
University of Birmingham
Edgbaston
Birmingham
B15 2TT

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

Integrin targeted fluorescent peptides for *in vivo* imaging

Integrins are cellular surface receptor proteins involved in numerous cellular signalling and adhesion processes. The $\alpha_v\beta_3$ subtype is up-regulated around sites of malignancy especially solid tumour masses. This is due to its expression on the endothelial and smooth muscle cells which are concentrated in areas of cancer, where the uncontrolled proliferation and longeitivity of cells characterises the malignant process. Sustaining such a process is achieved through the development of a labyrinth of new blood vessels to supply nutrients to, and remove waste from the accumulating tumour mass. This is the process of angiogenesis. Endothelial and smooth muscle cells bearing the $\alpha_v\beta_3$ integrin amass to form new blood vessels.

Crystallographic studies have determined that the tri-amino acid sequence RGD is crucial for the recognition and binding of extracellular proteins in the $\alpha_v\beta_3$ integrin binding site. Cyclic penta-peptides bearing this RGD motif have been found to exhibit preferential *in vivo* stability properties over their linear counterparts without affecting recognition and binding. In this study the cyclic peptide cRGDfK was synthesised using solid phase peptide chemistry and the lysine (K) residue was derivatised and proved an efficient means of attaching a fluorophore for potential diagnostic imaging applications.

Targeted fluorescent imaging is the backbone of this study. Novel lanthanide metal complexes with a variety of appended chromophores were synthesised via the Ugi four-component condensation reaction. The plan was to discern the most effective energy transfer mediated long-lifetime luminescent combinations of chromophore and metal, for subsequent attachment to the cyclic peptide. These potential long-lifetime imaging agents would allow the use of time-gated technology to eliminate troublesome background fluorescence. This chemistry was published (see page 338).

Near-IR emissive organic fluorophores were also appended to the cyclic peptide and intraveneously injected into live lung tumour bearing mice. This interesting work investigated the *in vivo* potential of these targeted probes.

Galectin targeted fluorescent disaccharides for *in vivo* imaging

Galectins are a family of intra and extra-cellular proteins which bind specifically to the β-galactoside carbohydrate residues of cellular glycoproteins. Several galectins, but most significantly galectin-3, have been found to mediate important inter- and intra-cellular processes particular to cancer cells, including signalling, proliferation and also metastasis.

Galectin-3 targeted lactosamine derivatives were synthesised with various organic fluorophores appended via an N-caproic spacer unit. These compounds were then tested against galectin expressing cell lines and the binding and subsequent emission visualised through fluorescence microscopy.

Acknowledgements

Firstly I must thank Dr. John 'The Boss' Snaith for his support, interest, patience and encouragement throughout these four short years. Thanks for being honest and keeping my feet on the ground; "That's why I'm The Boss" will always stay with me. Many thanks also for the long discussions about all things but chemistry over a lemon & ginger tea.

I thank my labmates and colleagues in the School of Chemistry for the coffees and laughs and general shenanigans. Saddles and Goose, San Andreas and Nico, Dr Russell for showing me the ropes, Clare, Tatyana, Stephen, Zara and the rest of the Preece, Cox and Davies groups for bringing up those nmrs and sharing in the fun and games on floor 7.

Profuse thanks go to the support staff and analytical department in the School of Chemistry especially Graham and Pete for your time and patience with my incessant HPLC and mass spectrometry demands. Also Neil for NMR and the chats and advice about motorbikes. For all your listening ears, help and instruction; Lianne, Nick, Louise, Pele, Terry, Stuart, Katie, Ian and Rachel, I thank you all.

I thank all the office staff past and present for being such a lovely bunch, always up for a laugh, for caring and for putting up with my teasing and boisterous banter. Helen, Pam, Nicky, Lynn, Janet, Barbara and Laura I thank you all for always making those visits down to floor 2 a bit of a giggle.

I thank our collaborators: Prof Faulkner at the Universities of Manchester & Oxford, Prof Borlak at ITEM Hannover, and Dr Carmel McConville at Cancer Research UK. I am indebted to them for their help with fluorescence spectroscopy, small animal and cellular studies respectively, and for generally being patient with an inquisitive chemist.

Many gracious thanks go to the School of Chemistry and the University of Birmingham for generously supporting me for an extra year. Also to the EPSRC for funding my project.

Thanks to Prof Preece, Prof Tuckett and Dr Parsons for visiting, being supportive and always showing an interest and asking how I am. Much appreciated.

To family and friends and those I have absentmindedly neglected to mention, I thank you all for your encouragement and pats on the back, your patience and belief in my getting this finished.

Contents

Chapter 1: An introduction to diagnostic imaging

Chapter 2: Integrin targeted fluorescent peptides for *in vivo* imaging

Chapter 2: Experimental

Chapter 3: Galectin targeted fluorescent disaccharides for *in vivo* imaging

Chapter 3: Experimental

Chapter 4: References

Chapter 5: Appendices

Chapter 6: Publication

Abbreviations

AAT	Alpha antitrypsin
Ac	Acetyl
Ac_2O	Acetic anhydride
ADP	Adenosine diphosphate
AFGP	Antifreeze glycoprotein
Arg	Arginine (R)
Asp	Aspartic Acid (D)
ATP	Adenosine triphosphate
BF_3OEt_2	Boron trifluoride diethyletherate
BOC	*tert*-Butoxycarbonyl
Bpy	2,2'-Bipyridyl
4-CCR	Four component condensation reaction
COSY-90	Correlation spectroscopy
CRD	Carbohydrate recognition domain
DCC	N,N'-Dicyclohexylcarbodiimide
DIPEA	Diisopropylethylamine
DMAP	Dimethylaminopyridine
DMF	Dimethylformamide
DNA	Deoxyribosenucleic acid
D-Phe	D-Phenylalanine (f)
ECM	Extracellular matrix
EGF	Epithelial growth factor
EI	Electron impact
Em.	Emission
ES / ESI	Electrospray ionisation
Et_2O	Diethylether
EtOH	Ethanol
Ex.	Excitation
$FADH_2$	Flavin adenine dinucleotide (reduced form)
FDG	Fluorodeoxyglucose
FGF	Fibroblast growth factor
FITC	Fluorescein isothiocyanate
FMT	Fluorecence mediated tomography
FP	Fluorescent protein
Gly	Glycine (G)
HBr	Hydrogen bromide
HCC	Hepatocellular carcinoma
HCl	Hydrochloric Acid
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HMPAO	Hexamethylpropyleneamine Oxime
HOSu	N-Hydroxysuccinimide
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum correlation spectroscopy
Hz	Hertz
IC	Internal Conversion
IR	Infra-red
ISC	Inter-System Crossing
LacNAc	N-Acetyllactosamine
LDA	
LiCl	Lithium Chloride
LiOH	Lithium hydroxide
LPA	Lysphosphatidic acid
Lys	Lysine (K)
MALDI-TOF	Matrix assisted laser desorption ionisation - time of flight

MeCN	Acetonitrile
MeI	Methyl iodide
MeOH	Methanol
MgSO$_4$	Magnesium sulfate
MMP	Matrix metalloprotienase enzyme
Na$_2$S$_2$O$_3$	Sodium thiosulfate
NaBH$_4$	Sodium borohydride
NADH	Nicotinamide adenine dinucleotide
NaHCO$_3$	Sodium hydrogen carbonate
NaOH	Sodium hydroxide
Near-IR (or NIR)	Near-infrared
NGP	Neighbouring group participation
NIS	N-Iodosuccinimide
NMA	N-Methylaniline
NMM	N-Methylmorpholine
NMR	Nuclear magnetic resonance
OsCl$_3$	Osmium chloride
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofurane-5-sulfonyl
PDGFB	Platelet-derived growth factor beta
PEG	Poly(ethylene glycol)
Pmc	2,2,5,7,8-Pentamethylchroman-6-sulfonyl
PPh$_3$	Triphenylphosphine
Py	Pyridine
PyBOP	Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosponium hexafluorophosphate
ROI	Region Of Interest
rt	Room Temperature
S$_0$	Singlet ground state
S$_1$	Singlet excited state
SOCl$_2$	Thionyl chloride
SP-C	Surfactant protein – C promoter
T$_1$	Triplet excited state
TAM	Tumour assisted macrophage
TBAI	Tetrabutylammoniumiodide
TCP	Tetrachlorophthalic anhydride
TFA	Trifluoroacetic acid
TfOH	Triflic acid (trifluoromethane sulphonic acid)
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
TosMIC	Tosylmethylisocyanide
VEGF	Vascular Endothelial Growth Factor
XRD	X-Ray Diffraction

Chapter 1: An introduction to diagnostic imaging

The overarching aim of this research was to develop efficient new synthetic routes to novel fluorescent and MRI active probes with potential future applications both in enhancing cancer diagnosis in whole body imaging, and as surgical tools to accurately highlight areas of malignancy.

1.1 Cancer & diagnostics

Modern diagnostic imaging technology is constantly improving with higher specification, more advanced and more powerful machinery becoming widespread.[1] This offers advantages in the increased speed with which imaging and diagnosis can be achieved. Earlier stage diagnosis allows more rapid therapy and improved prognosis.[2] The advance of knowledge and technology is constantly trying to keep up with and on top of disease, and diagnostic imaging is a fundamental tool in the battle against cancer.[3]

Vast changes in the incidence of, and the mortality from cancer have occurred over the past century. The statistical analysis is complex as huge advances in diagnostics and therapeutics have transpired, but currently still approximately one third of all people in England will develop cancer at some point in their lifetime. Indeed, in 2006, there were 367,000 new cases of cancer registered, and 139,000 people died from the disease. In England, £4.35 billion was spent by the NHS on cancer services in 2006, and research in the area is extensive across academia and industry.[4]
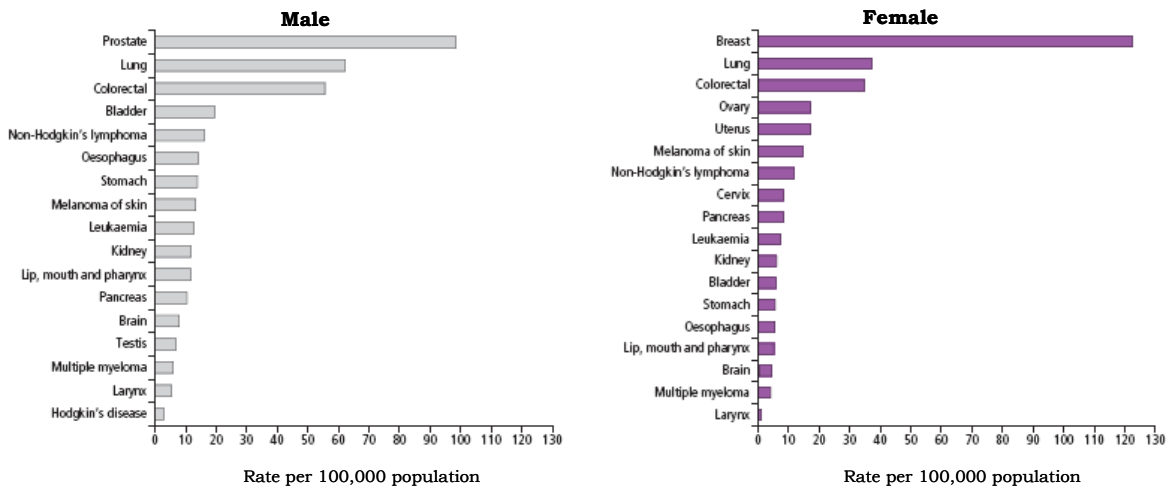
Figure 1: Age standardised major cancers in the UK 2004-2006. Adapted from Ref. 4b.

There are more than 200 types of cancer, each with different symptoms and treatment. It is a huge challenge, but roughly half of all people diagnosed with cancer now survive more than five years. More than 70% of childhood cancers are now successfully treated in comparison to 30% in the 1960s,[5] however the most prevalent cancers (breast, lung, colorectal and prostate) are most often found in the over 60 age group.[2] Overall incidence of

cancer increased by 25% in the years 1975-1995 but has since stabilised with increases in more avoidable cancers such as skin, womb and kidney, but significant decreases in stomach, cervical, bladder and lung cancer (male) cases especially **(Figure 2)**.
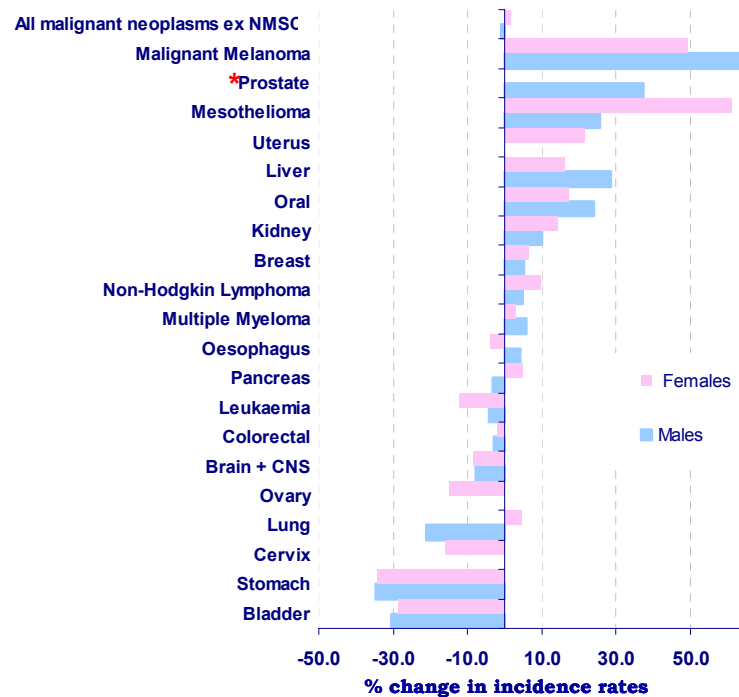
Figure 2: A representation of the changes in cancer incidence rates 1997-2006. ***** The increase in prostate cancer incidence has been enhanced by the use of PSA testing since the early 1990s which detects malignancy at an earlier and/or asymptomatic stage.[3]

The causes of cancer are diverse, but increasingly it has become apparent that lifestyle has a significant impact in many cases.[5b] Genetic inheritance factors, age, diet, sun exposure, smoke inhalation and bodyweight are all considered important factors contributing to an individual's likelihood of developing cancer.[3] As a general trend though, it is expected that cancer incidence will rise. This is due to a number of factors, including an ageing population rendering more people in the most susceptible age group.[5c] This ageing population however, is also partly due to vast improvements in general health awareness, earlier diagnosis of most diseases, and subsequent successful treatment prolonging lifetime. Therefore the overall burden upon the NHS is unlikely to change, although it will shift to different types of cancer.

Advances in diagnostic technology are contributing to the rising incidence tally, however early diagnosis is a fundamental factor in decreasing the mortality rates of many types of cancer, especially those forming solid mass type tumours, as early treatment has been proven to vastly improve prognosis.[3]

Key people involved in cancer detection, prevention and control include scientists investigating the mechanisms that cause cells to become malignant and the key components involved; those carrying out clinical trials to evaluate new treatments; the

clinicians actually conducting the diagnosis and therapy for individual patients; public health physicians implementing screening programmes and educating the public; and epidemiologists identifying high- and low-risk population sections and identifying environmental factors providing clues to carcinogenic mechanisms.[1-5]

It is within the first section of these people with which we find ourselves, the scientists; and it is the detection, monitoring and visible highlighting of cancer which is the cornerstone of this work.

The focus of this review shall centre upon diagnostic imaging and specifically the development of optical imaging agents, but to introduce this we must first look at one element of treatment in which optical imaging could have significant application: surgery.

1.1.1 Surgery

Whilst cancer diagnosis continues to rise, conversely, the chances of survival are improving as more and more therapeutics become economically viable, disease specific, and their treatment programmes more thorough.

Prior to the technological and surgical revolution in the 1970s and 80s there was an even higher death rate from cancer. However, with the advancement of imaging technology it is now possible for doctors to detect cancer at an earlier stage and surgeons are now able to excise many cancerous growths.[6]

Three-dimensional imaging techniques such as CT, PET and MRI allow members of a medical team to pinpoint the exact location of a tumour and determine its margins prior to surgery. However, to achieve successful complete surgical removal, the surgeon must ensure that all diseased cells are removed and that none of the diseased tissue remains or is damaged during surgery. This could result in the remaining cells proliferating, causing relapse and possible metastasis of the cancer.[6b] Even with the current imaging systems, during surgery the tumourous growth may not be as visibly defined as the preoperative images; the margins may be distorted and may not be visible even to the trained human eye. In order to effectively eliminate a cancerous growth it is essential to have the capacity to precisely locate and identify the nature of the diseased cells, without damaging the healthy surrounding tissue (**Figure 3**).

This has prompted the search for new methods which can highlight the tumour perimeters throughout the surgical excision process, and this is where optical imaging agents and bimodal probes have great potential.
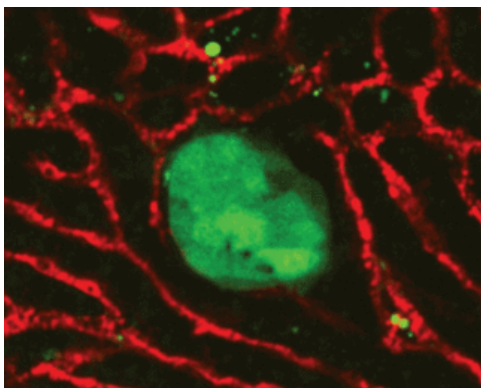
Figure 3: Highlighting tumours. Green fluorescent protein-labelled liver tumour highlighted by two-photon microscopy.[4]

This surgical application illustrates how diagnostic imaging is key to not only highlighting the tumour for diagnosis, but could also be highly applicable in the subsequent treatment. The other advantage of a probe targeted to a tumour cell-specific entity, such as the cell surface receptor integrin $\alpha_v\beta_3$ which is highly upregulated in many tumour sites, is that it offers good evidence of the biological activity and function of the malignancy. The criterion for an effective optical imaging agent offers several challenges to the synthetic chemist. The agent must be tumour specific with good binding affinity, it must be metabolically stable and offer great contrast against surrounding tissues whilst being non-toxic and have excitation and emissive properties conducive to deep tissue penetration.[7]

Another option is to incorporate a secondary function into the probe which is sensitive to an alternative diagnostic technique. This is bimodal imaging,[7b] and a tumour targeted fluorescent probe with an intramolecular paramagnetic MRI active component, for instance, could offer a versatile, non-invasive option for early accurate diagnosis, surgical removal or highly sensitive monitoring of treatment efficacy and progress, using a single molecular agent.

It is not solely the applications we are endeavouring to advance, but the chemistry and the methodology involved in the synthesis of specifically targeted fluorescence probes.

To further our understanding of where there is a place for targeted optical imaging in medicine we must briefly introduce the techniques primarily employed at present in the clinical setting.

1.2 An overview of present oncological diagnostic imaging techniques.

1.2.1 Ultrasound

Medical sonography (ultrasonography) has been widely used as a valuable, harmless diagnostic tool for over 50 years. It is a medical imaging technique most widely recognised in the field of obstetrics where it is the key modality for the study and monitoring of the unborn. It is however more widely used in the imaging of muscles, tendons, internal organs and for detecting abnormalities such as tumours.[8]

In medicine this technique uses pulses of high frequency cyclic sound energy to penetrate tissue; the reflection signature from the various media is detected and a computerised tomographical image built up. Ultrasound has limitations in that it cannot see through bone and so visualisation within the chest cavity can be somewhat restricted.[8b]

Although ultrasound has been shown to enhance inflammatory response and warm tissue through molecular friction, it does not employ the more contentious ionising radiation which is the essence of techniques such as X-ray and CT.

1.2.2 X-ray

This is a standard technique traditionally employed for observing skeletal structure and abnormalities especially in the field of orthopaedics.[9] It employs ionising X-ray radiation and a two-dimensional image is built up through the observed variation in the attenuation of the radiation according to the density of the medium in question. Hence metallic objects or bone show great contrast but soft tissue often requires further investigation to accurately offer a diagnosis.[9]

1.2.3 Computed Tomography (CT)

Since its emergence in the 1970s CT has become a gold standard diagnostic tool both alone, as a supplement to standard X-ray and ultrasonic techniques, and also in conjunction with other techniques such as Positron Emission Tomography (PET-CT).[10]

CT employs X-ray radiation to generate a large quantity of 2D X-ray images taken around a single axis of rotation. These 2D slides can then be processed using powerful geometric computer software programmes to generate a sophisticated 3D image of the inside of a patient. This 'windowing' allows the manipulation of the large volume of data collected in an increasingly short scan time to visualise the numerous physiological structures and media according to their ability to attenuate the X-ray beam.[10]

The X-ray slice is generated by rotating the X-ray source around the object with a detector at the other side of the rotating loop. Scans are taken at increments as the body is gradually passed through the loop. The scans are combined by mathematical tomography to construct a 3D image where the pixels or voxels are displayed according to the mean attenuation of the tissue on a scale called the Hounsfield scale. Windowing then creates the image display on a greyscale with over 1000 available shades of grey. The third dimension comes from the stacking of axial slices. The software can then slice through in a different, usually orthogonal plane (**Figure 4**).[10b]
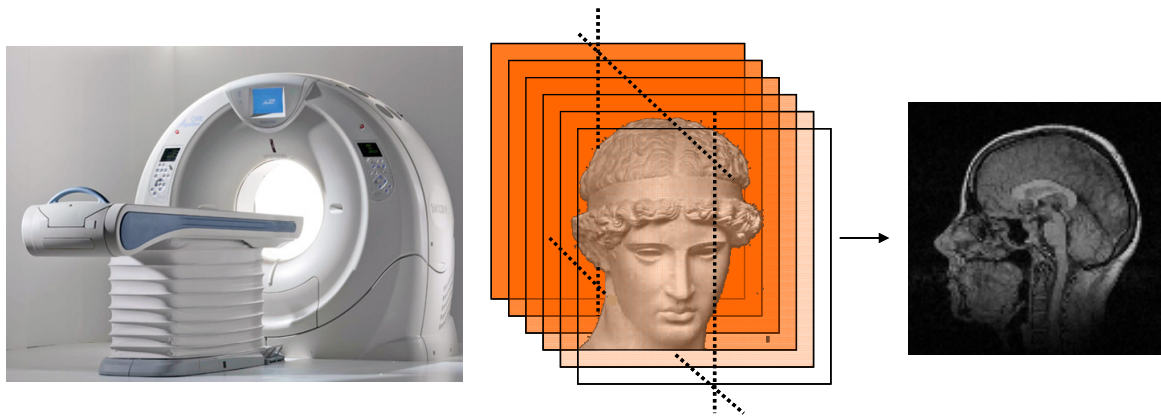
Figure 4: **Left**) A CT scanner; **Right**) The authors representation of multiplanar reconstruction.

Tomography is the method of representing on radiographic film a single 2D slice of the body through a greyscale image. By controlling the rotation of the X-ray source and detector during exposure, the target level of image remains sharp with clear tissue contrast whereas blurring occurs at all intermediate levels.[10b]

Although the maiden CT scanner was designed and built between 1967 and 1972 with financial support from the music company EMI, it was limited to taking tomographic sections of the skull and brain and employed a water tank with an ergonomic head-shaped rubber inlet in which the patient's head was enclosed for scanning. This somewhat unorthodox arrangement was designed to reduce the range of the attenuated radiation reaching the detectors; water attenuation is measured as zero Hounsfield Units (HU) and bone can reach 2000 HU, however air is -1000 HU.[10c]

The radiation dose for a particular scan is dependent of many factors including the volume to be scanned, the size of the patient, the number of previous scans, the potential requirement for future scans, and the type of scan sequence required for the desired resolution and quality of contrast.[10c] However, the number of CT scans performed nationwide continues to escalate despite the increasing understanding of the potential risks associated with the ionising radiation employed (**Figure 5**). There are several reasons for this, not least due to increasing availability and its value and reliability for an increasing number of conditions.

	Typical effective dose
Chest X-ray	0.1 mSv / 10 mrem
Chest CT	5.8 mSv / 580 mrem

Figure 5: A table showing the increased exposure to X-ray radiation during CT scanning.[10c]

CT is now employed for a whole host of disease diagnostic imaging applications through both the discovery and location of existing disease to preventative screening for those at higher risk of disease but present no immediate symptoms.

Advantages & disadvantages of CT

- High contrast resolution (fractures, ligament injuries and dislocations can be visible at 0.2mm resolution), also the power to distinguish between tissues of just 1% variation in density.

- Increasing scope of investigation can eliminate the requirement for invasive procedures (CT angiography avoids the need for the insertion of arterial catheter and guidewire).[10d]

- Vastly improved modern scanners only require an approximately 1 second exposure per slice allowing reduced exposure for scanning children without the need of sedation to ensure the child remains motionless.

- New increasingly powerful software is enabling dose reduction by filtering out more background noise.

- However, whilst improvements in radiation efficiency continue, this technique still uses moderate to high levels of ionising radiation. Some evidence exists suggesting a linear relationship between X-ray radiation dose and cancer risk. Although this risk is low with an approximate figure of 0.35% risk per scan compared with approximate 23% risk of dying from cancer naturally.[10]

- Contrast CT scans rely on intravenous administration and the agent itself can cause reactions ranging from discomfort and hotness to allergies and kidney damage to those with pre-existing renal conditions and diabetes.

Contrast agents are sometimes used to enhance the outline of tissues especially in the digestive system. Iodine and barium agents have long been employed to enhance contrast.

1.2.4 Positron Emission Tomography (PET)

This is a radiative nuclear medicine technique producing a 3D image of targeted specific functional biological processes within the body by detecting pairs of gamma rays emitted indirectly through the nuclear decay of a positron-emitting radiotracer. Images are generated according to the concentration of the radiotracer in a defined 3D space within the body and then reconstructed by computational processing software.[11]
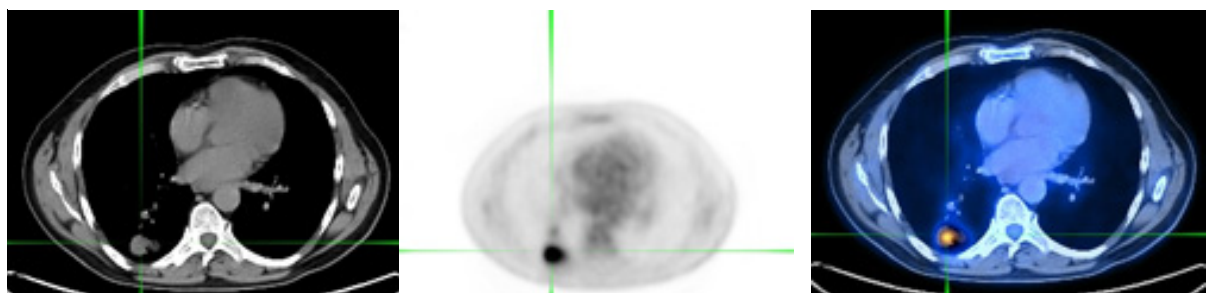
Figure 6: Transversal chest slices. **Left**) CT; **Centre**) PET; **Right**) PET-CT

Positron emission decay involves the spontaneous emission of an antiparticle of the electron, the positively charged positron. After a few millimetres the positron collides with an electron, annihilating both, and in the process producing two pairs of gamma photons which fly off at nearly 180˚ to one another, pass directly through the body and crash into the awaiting scintillator causing a flash of light which is then detected by a photomultiplier tube (**Figure 7**). This detection is dependent on the simultaneous detection (within a few nanoseconds) of the two opposite moving pairs of photons or the signal is ignored; each 'coincidence' event is listed and tens of thousands are statistically collected. The source of emission is located through a cunning software system that follows the line of response (LOR), which is the trajectory of the two gamma pairs from each decay event. A map of radioactivity is then built up by calculations of the activity of the various tissues along a LOR in relation to the quality of simultaneous coincidence events.[11]
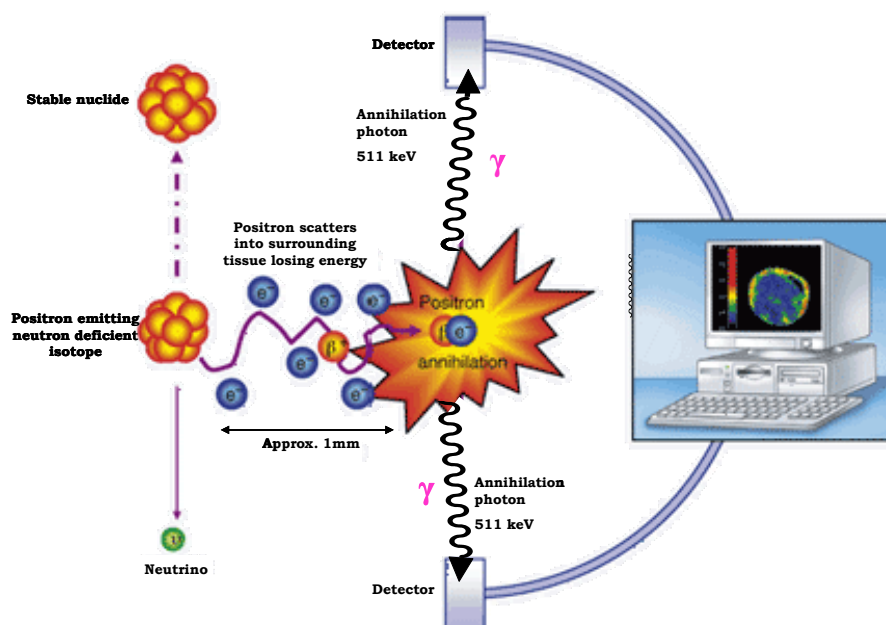
Figure 7: A cartoon of the principles of PET imaging.[11]

The radionucleotides employed are short half-life isotopes such as carbon-11 (~20 minute half life), nitrogen-13 (~10 min), oxygen-15 (~2 min) and fluorine-18 (~110 min) which are incorporated into normal biologically active compounds such as D-glucose and other analogues of this monosaccharide, without interfering with their primary function. The radiotracer bearing the isotope is obviously shortlived, and as such, PET imaging requires careful logistical planning to ensure the intravenous delivery, site specific accumulation and subsequent scan is accomplished before the tracer is rendered inert. New novel radiotracers are continually being synthesised, however the radioactive decay time constraints require a cyclotron and radiochemical laboratory in close proximity to the patient and scanner. For these reasons fluorine-18 is the preferred isotope and in ~90% of cases it is delivered to the patient intraveneously as the natural glucose derivative ^{18}F-fluorodeoxyglucose (FDG) (**Figure 8**).[11b]
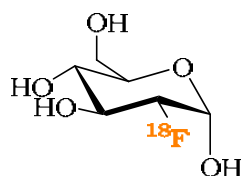
Figure 8: FDG. Post intravenous administration the patient must rest for 1 hour prior to scanning to allow an accurate measure of cellular uptake.

Of all PET scans ~90% are FDG oncological investigations and PET-CT has become a very effective technique especially in Hodgkins and non-Hodgkins lymphoma, lung malignancies and many other types of tumour, especially when looking for metastic lesions.[11c]

Just as with glucose, the FDG glucose analogue is transported into cells with the help of a family of transport proteins (GLUT-X). Inside the cell the process of glycolysis kicks into action and the FDG is phosphorylated by hexokinase. Normally glucose would now continue in the glycolytic pathway of energy production (see page 165/6), however, the glucose hydroxyl displaced by ^{18}F is required for the next step of glycolysis, and so FDG-6-phosphate is trapped intracellularly. Due to malignant, rapidly-growing and differentiating tumour cells having glycolytic rates up to 200 times higher than normal cells, the extra demand for energy is profound. Therefore the malignant cells display an increased density of the glucose transporters GLUT-1 and GLUT-3,[11d] and also exhibit higher levels of hexokinase types I and II. They also often favour a more inefficient anaerobic hypoxic pathway adding to the already enhanced demand for glucose. These combined mechanisms mediate the uptake and retention of a much higher level of FDG in tumour cells in comparison to normal tissue. This is the key to the power of PET imaging and what causes the intense emission from high glucose demand tissues such as the liver (the main organ for carbohydrate metabolism and storage), the heart, and tumours. It must be highlighted that FDG is not cancer specific. Areas of high metabolism or muscular and nervous hyperactivity, inflammation and tissue repair in the tumour vicinity can lead to ambiguity.[11b]

PET is versatile, it can assess response to therapy, especially in the case of many cancers, and the radiation dose is minimal. Limitations include cost and logistics and therefore fluorine-18 and rubidium-82 are the primary isotopes in use today. Technological advances are however increasing the emergence of on-site cyclotron and laboratory ("hot lab") facilities.[11]

MRI and CT detect organic anatomical changes whereas PET and SPECT (page 11), detect specific areas of biological activity even without noticeable anatomical discrepancy. Therefore PET is a more quantitative technique.

Recent PET research shows the versatility and power of the technique especially as studies show that molecules which target specific biological motifs can be labelled with the ^{18}F isotope (**Figure 9**). In this instance the cyclic peptide was labelled with either one, two, three or four ^{18}F atoms by direct electrophilic fluorination with [^{18}F]AcOF in TFA.[11e] The

aromatic ring of phenylalanine is not involved in receptor binding and thus the probe was effectively used to visualise $\alpha_v\beta_3$ integrin expression through the accumulation of radiative signal from adenocarcinoma tumours in mice.[11e] However, the radioactivity level was not high enough to distinguish the tumour from the uptake in other organs such as the liver.
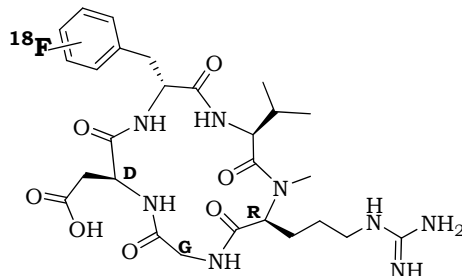
Figure 9: [18F] labelled cyclic RGD peptide targeted to the $\alpha_v\beta_3$ integrin subtype found to be upregulated in tumour sites.[11e]

1.2.5 Positron Emission Tomography – Computerised Tomography (PET-CT)

This combination of techniques has revolutionised the field of medical diagnostics. The machine is a single loop (or gantry) scanner which allows the acquisition of sequential PET and CT images in a single session with minimal inconvenience to the patient.

CT, PET and PET-CT techniques have limitations especially in the diagnosis of malignancy as low levels of disease can escape detection. Additionally, most diseases require repeat scans to analyse response to treatment and so the risk from ionising radiation escalates with X-ray, CT and PET-CT imaging.[10d]

> PET - Spatial distribution of metabolic or biochemical activity.
>
> CT - Allows the accurate alignment and correlation of PET image with acquired anatomical data.
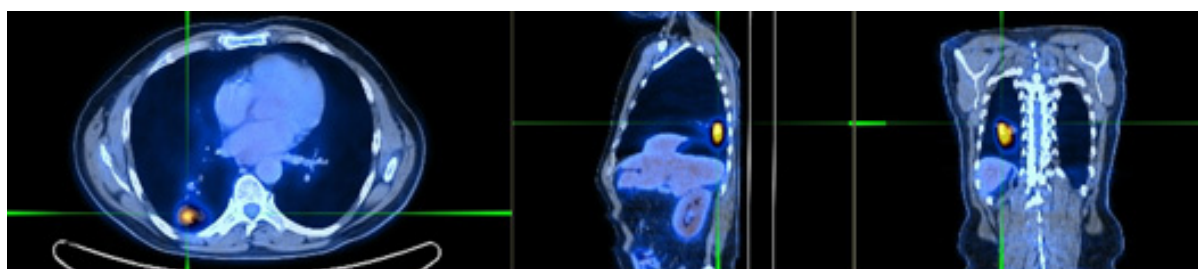
Figure 10: The accuracy offered by PET-CT in pinpointing a lung tumour.

PET-CT offers a precise 2D and 3D image construction which consequently allows more accurate surgical planning, radiotherapy, and has even begun influencing the staging of chemotherapeutic protocol.[11] An example is in stereotactic radiation therapy for cancer where marks are made on the patient's body prior to PET-CT then the slices are transferred

digitally to the linear accelerator which is then programmed to bombard the target area precisely with high energy photons for more accurate treatment.[12]

Metabolic regions are shown on a false colour coded scale of pixels (or voxels) relative to the intensity, and superimposed onto the greyscale CT.
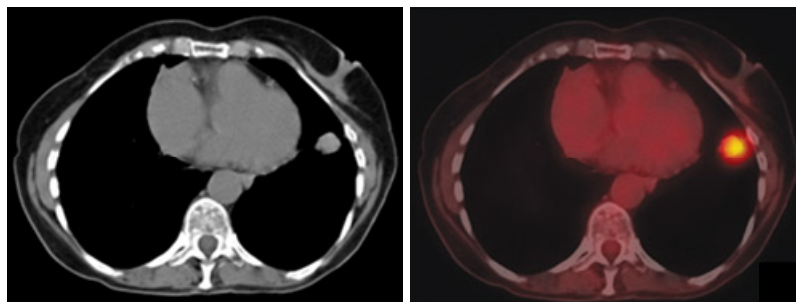
Figure 11: **Left**) A CT image highlighting a lump in the right lung; **Right**) The identical PET-CT superimposed image confirming malignancy.

1.2.6 Single Photon Emission Computed Tomography (SPECT)

This is another nuclear medicine tomographic technique dependent upon gamma radiation. SPECT is very similar to PET, however it is able to afford true three dimensional information presented typically as cross-sectional slices through the anatomy.

Just as conventional X-ray imaging is a 2D representation of 3D structure, so the SPECT image is a 2D representation of 3D distribution of a radionucleotide through the eyes of a gamma camera.

Now, SPECT is similar to PET in that the technique relies upon the *in vivo* administration of a radioactive tracer and detects gamma emission, however, the tracer used is not a positron emitter which upon electronic collision emits gamma photons in opposite directions; it directly emits gamma radiation. Therefore SPECT is far cheaper than PET for the simple reason that more stable, longer-lived radioisotopes are more easily obtainable.[13]

The acquisition of SPECT images involves a gamma camera being repeatedly rotated normally a full 360° around the body with projections (2D slices) being acquired at multiple defined angles usually every three to six degrees. Computer software then applies a tomographic reconstruction algorithm to the multitude of 2D projections to yield a 3D dataset. Advances in this technology has seen the introduction of two or three gamma camera heads simultaneously acquiring two or three projections as each head requires just 180 or 120 degree spacing about the loop.

As SPECT affords accurate 3D localisation it can provide precise information on specific internal organ function, tumour growth and many other applications such as in myocardial perfusion imaging where ischemic heart disease is detected by measuring a radiotracer in the blood flow when the heart is put under stress.[13b]

SPECT radioisotopes.

- Technetium-99m is the best known gamma emitting metastable isotope with a half life of approximately six hours. Generally it is attached to pharmaceuticals for specific functional targeting. It has a large range of diagnostic applications such as in combination with tin where the ^{99m}Tc binds to red blood cells and is used in the diagnosis of circulatory system disorders.

  [99mTc]-HMPAO is brain specific and is taken up by brain tissue in a manner proportional to blood flow to that specific locality.[13c] Thus cerebral blood vessels and metabolism can be assessed and for difficult-to-diagnose diseases such as Alzheimers and Dementia. SPECT using this radionucleotide has shown accuracy as high as 88%.

- Iodine-123 is a popular and useful tumour scanning isotope with a halflife of approximately 13 hours.

- Indium-111 ($t_{1/2}$ = 67 hours) is often used in conjuction with technetium-99m for the specific imaging of white cells.

As with PET, scanning is time consuming and the patient must lie still. Intense areas of activity such as the bladder where the probe is concentrated prior to excretion, can obscure activity in adjacent areas. Internal attenuation of the gamma radiation can lead to the underestimation of deep tissue activity in comparison to more superficial tissue and for this reason most SPECT scanners now have integrated X-ray-CT capabilities as this provides an attenuation map of tissues and thus allows for computational correction of the gamma deep tissue attenuation.

SPECT scans are of lower resolution than PET as the PET scanner detects time coincident emission of the two photons giving enhanced radiation event localisation.

1.2.7 Magnetic Resonance Imaging (MRI)

In 1971 Raymond Damadian showed that the nuclear magnetic relaxation times of tumours differed from surrounding tissue, thus the door was opened to the development of this technique as a powerful imaging weapon in the armoury of the clinician.[14a]

MRI is of interest to this study for the simple reason that an important contrast agent is based upon the paramagnetic metal gadolinium (Gd). This metal is readily available for use in synthesis as, for example, a chloride, citrate or trifluoromethansulfonate Gd(III) salt, and is reliably chelated in a variety of macrocycles which can be appended to targeting vectors.

The principles of MRI

Subatomic particles such as protons and neutrons possess the quantum mechanical property of spin. Certain nuclei have non-zero spin through possession of unpaired protons or neutrons, and therefore display a magnetic moment. The spin states of a spin ½ nucleus

such as a proton (^{1}H), are referred to as either up or down.[14b] Upon application of a strong magnetic field the spins all align around the axis in the direction of the applied field (**Figure 12**).
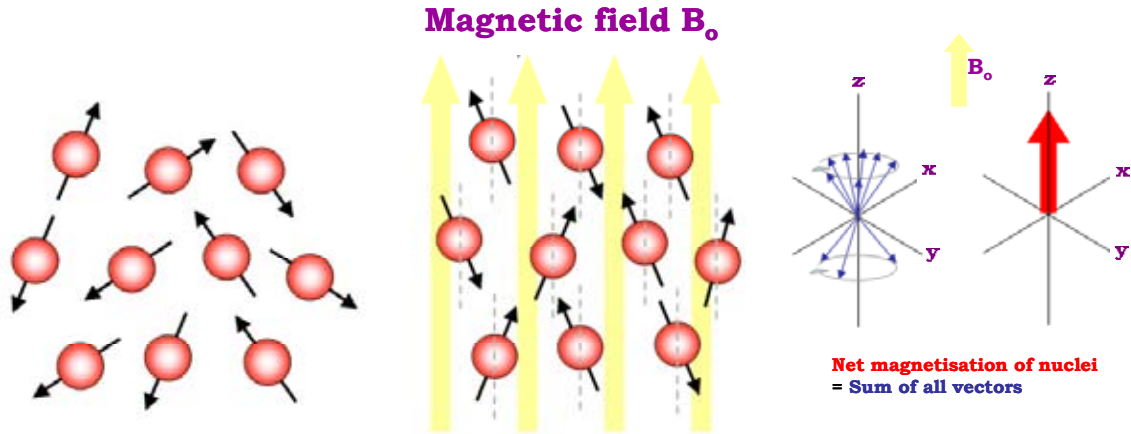
Figure 12: A diagram of subatomic particular spins aligning in a magnetic field B_0.

A radiofrequency (RF) electromagnetic pulse is then applied perpendicular to the magnetic field causing protons to absorb energy and enter an excited state (**Figure 13**). Upon deactivation of the pulse the protons relax back down to a ground state with the emission of a resonant radiofrequency pulse which is detected by the scanner. The frequency of the emitted signal is dependent upon the strength of magnetic field.
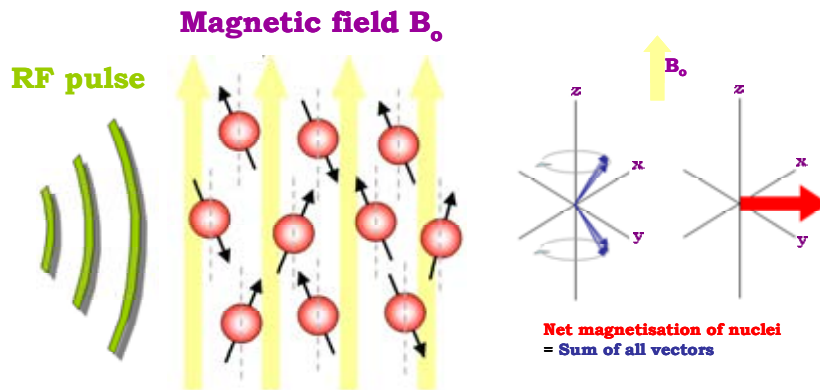
Figure 13: The effect of the radio frequency pulse.

The make up of the human body is predominantly water and fat.[14c] Both of these compounds are rich in hydrogen atoms, indeed in total, the human body consists of approximately 63% protons. Each water molecule bearing two hydrogen nuclei, or protons, of point positive charge, means that in the described powerful magnetic field these protons will align and point in the direction of the field. The location of these water protons can

then be determined by application of additional magnetic fields during the scan allowing the build up of an image.

An attractive property of MRI is the tremendous contrast between the various soft tissues of the body.[14d] Diseased tissues such as tumours can readily be detected due to the water protons within different environments relaxing to their most stable state at different rates. This relaxation of *in vivo* water protons is the backbone of MRI and can be further enhanced by the oral, direct injection or intravenous administration of a paramagnetic contrast agent.

Contrast is observed when variations in the strength of nmr signal is recorded from different physiological locations. This is dependant upon the relative density of the excited water protons, on the different relaxation times of nuclei after a pulse sequence, and often other parameters depending on the specific type of scan. Contrast can be controlled and manipulated through the application of specific pulse sequences. For example in the brain, T^1 weighted MRI causes the nerve connections of grey matter to show up white, grey matter grey and cerebrospinal fluid dark (**Figure 14**).[14e/f]

The combination of field gradients and radio frequency excitation is then used to create an image on a greyscale.
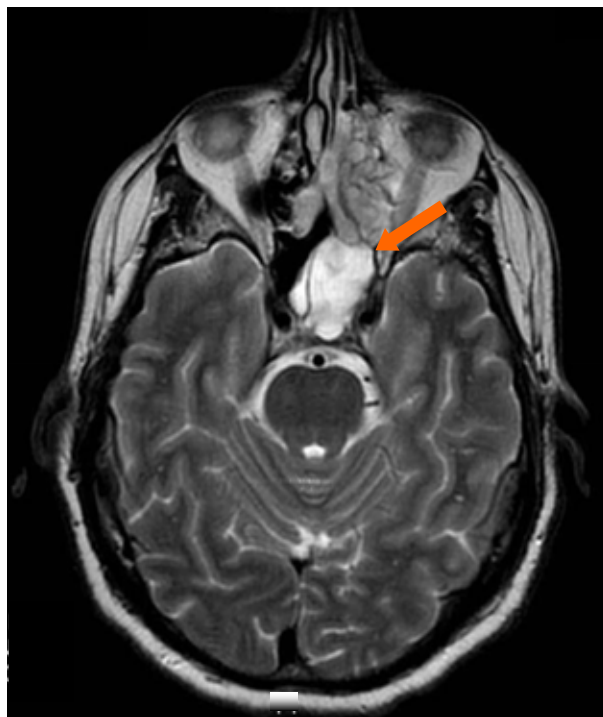
Figure 14: An mri scan revealing a site of neuroblastoma.[14f]

1.2.8 Contrast agents

Of particular interest to this study is the situation where it is not possible by adjusting scanning parameters alone, to generate enough image contrast to adequately show the area of interest. In this situation a contrast agent is required to enhance the definition of fine structure or to more accurately distinguish between areas of similar signal strength. For

the bowel and stomach this may just be water, however most contrast agents are employed to take advantage of their magnetic properties. Commonly these are gadolinium (Gd) based agents. Gadolinium exhibits a magnetic moment due to its unpaired electrons, is a paramagnetic element, and interacts with the external magnetic field.
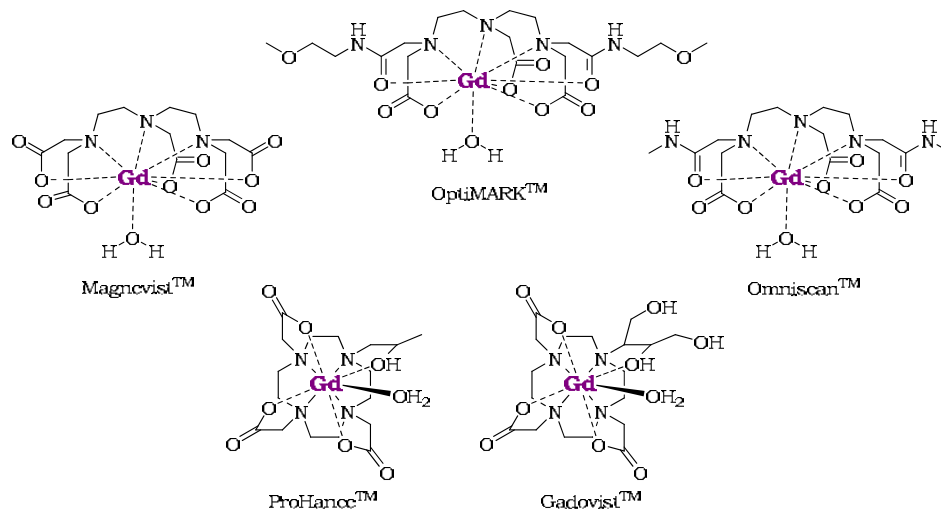
Figure 15: The structures of five approved gadolinium clinical contrast agents.[14g]

Gadolinium in the (3+) oxidation state exhibits seven unpaired f-electrons. This causes appended and local water molecules to relax more rapidly enhancing the quality of the image. Gd enhanced tissues and fluids appear extremely bright on T^1 weighted images.

Super paramagnetic nanoparticles such as iron oxide have also recently become available and these appear very dark on T^{2*} weighted images and are useful for liver imaging as abnormal areas such as tumours, appear light.

A comparison:

- CT offers good spatial resolution and the ability to discern between different structures which are of negligible distance apart.
- MRI offers comparable resolution and is excellent at distinguishing between two arbitrarily similar but not identical tissue types. Optimisation of the pulse sequence provides image contrast based on chemical sensitivity. MRI cannot, however, be used in parallel with other imaging techniques such as in PET-CT.

1.2.9 Bimodal imaging agents.

An MRI scan allows members of a medical team to pinpoint the exact location of a tumour and determine its margins prior to surgery. To achieve the successful surgical removal, the surgeon must ensure that all diseased cells are removed. This often leads to a greater than desirable amount of healthy surrounding tissue also being removed unnecessarily, causing trauma.[15] Even with the current imaging systems, during surgery the tumourous growth may not be as visibly defined as the preoperative images; the margins may be distorted and

may not be visible even to the trained human eye. In order to effectively surgically eliminate a cancerous growth it is essential to have the capacity to precisely locate and identify the nature of the diseased cells, without damaging the healthy surrounding tissue.

This has prompted the search for new methods which do not rely on the limitations of the human eye.[15b]

One method of highlighting the exact margins of a tumourous growth is through the use of a single molecule multimodal *in vivo* imaging platform.[15c]

This allows the combination of techniques, such as magnetic resonance imaging (MRI), positron emission tomography (PET), and single photon emission computed tomography (SPECT) to generate sophisticated images *in vivo.*

The most advantageous use of *in vivo* imaging is the ability to combine different imaging approaches to generate information with extremely high resolution whilst diseased cells are still functioning within the patient. It balances the positive and negative features of each imaging technique and with the correct combination of techniques, this platform has the capability to revolutionise the diagnosis, treatment and surgical management of malignant growths.

However, conventional imaging modalities provide little or no information regarding the specific biological markers on newly formed blood vessels, tumour cells and the molecular changes throughout therapy. Molecular imaging takes advantage of the traditional imaging technology previously described, but also introduces indicative markers of tumour development at different stages. Molecular imaging is non-invasive.

1.2.10 Roger Tsien & visualising cell biology

Advances in molecular biology, organic synthesis and materials science have provided the biologist with new classes of fluorescent probes for visualising cellular processes.[16a] The impressive 2008 Nobel prize winning work of Roger Tsien has centred around the development and application of visible fluorescent proteins (FPs) from jellyfish and coral.

A revolution in cellular biological imaging has been brought about by the discovery, gene cloning and application of the green fluorescent protein (GFP) from the jellyfish *Aequorea Victoria.*[16] With an emission maxima at 508 nm, the incredible ability of GFP to generate highly visible and efficient internal fluorescence is both fascinating and valuable. The chromophore in GFP is generated by spontaneous cyclisation and oxidation of three amino acids buried at the heart of a beta barrel (**Figure 16**). Crucial to the discovery and application of this fluorophore is the fact that it is independent of any jellyfish-specific enzymes and hence the gene can be applied to other organisms.[16b]
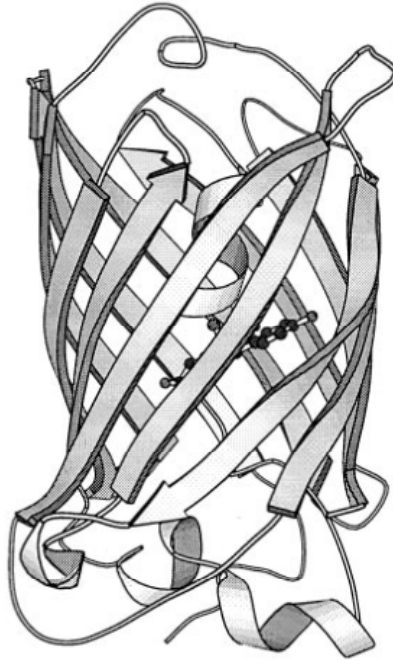
Figure 16: A 3D structure of GFP showing the beta barrel bearing the chromophore.[16c]

GFP is just one of a large family of different coloured marine fluorescent proteins with variations in structure and the environment in which they fluoresce. Laboratory mutagenesis has further enhanced brightness and quantum yield. There has been an explosion in the availability of different FPs and they promise a wide variety of imaging options to the biologist especially when used in multilabelling experiments where minimal crosstalk between discreet excitation and emission channels is achievable.[16c] However, 598 nm and 655 nm are the maximum emissive wavelengths of FPs thus far and this is still within range of biological interference.

A drawback in the application of FPs in live mammalian studies has been the poor tissue penetration of the required excitation light. Bacteria-derived near-IR fluorescent proteins have been engineered to emit via a spontaneously incorporated biliverin chromophore at 708 nm. This chromophore is generated and metabolised endogeneously in adult humans to the extent of 300-500 mgs per day.[16a]

Thus the idea of boosting and utilising pre-existing *in vivo* near-IR excitable and emissive chromophores through engineering their adhesion into the chromophore-binding domain of a bacterium, is a novel means of accessing near-IR luminescence from within. However, the localisation and specificity of near-IR autofluorescent proteins to certain cell or protein types is uncertain and the extent to which they can be derivatised to target and visualise specific biological entities remains to be seen.[16d]

1.2.11 Optical imaging

1.2.11.1 Fluorescence Imaging

The original development of fluorescent probes in the 1970's concentrated on providing a measurement of the cellular concentration of Ca^{2+} ions. Calcium is critical for life and is fundamental to many biological processes as a messenger and regulator. A means of monitoring Ca^{2+} ions selectively in the presence of the typically far more prevalent Mg^{2+} ions represented a considerable challenge, not least because changes in Ca^{2+} ion concentration occur rapidly, and hence the probe must have selectivity, and rapid binding and release kinetics.

Tsien realised that it would be preferable to have an indicator that responded to calcium binding by shifting wavelengths whilst maintaining strong fluorescence rather than the previous probes which relied upon an just an increase in fluorescence intensity upon Ca^{2+} binding.[17] The probes developed are still used today (**Figure 17**).
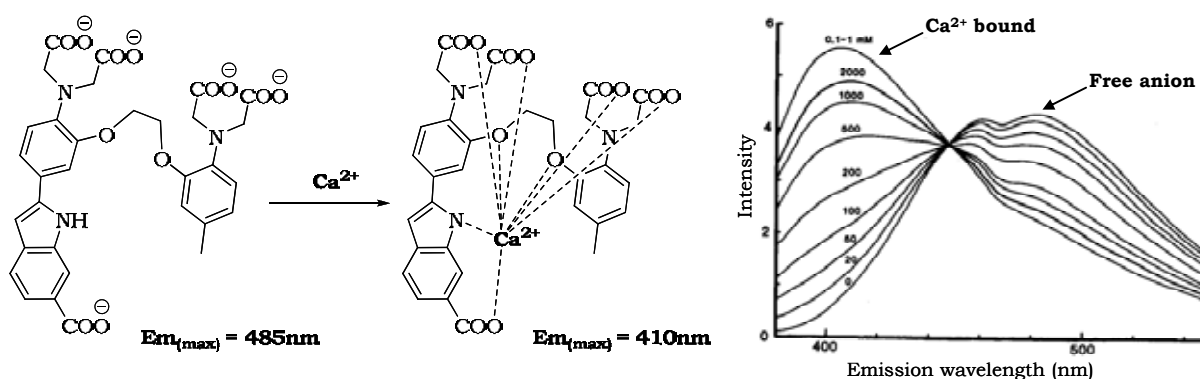
Figure 17: Tsein's INDO-1 calcium probe. Adapted from Refs 17a & b.

The INDO-1 probe involves an eight-coordinate EGTA-like binding site attached to a conjugated chromophoric system. When the ligand is free of metal, the amine lone pairs are conjugated with the chromophoric system and act as good electron donors to stabilise the excited state of the chromophore, resulting in long wavelength emission. When the Ca^{2+} is bound, the nitrogen lone pairs are involved in the metal binding reducing the ligand's ability to donate electrons to form a charge-transfer excited state and thus the emission shifts to a lower wavelength.

Whilst fluorescent imaging techniques such as the INDO-1 calcium probe (**Figure 17**) are highly sensitive there is a problem in that biological chromophores such as tyrosine and tryptophan also absorb and emit light in the same region.[19] This is known as autofluorescence and it results in background noise from the sample interfering with the signal from the luminescent probe. The separation of the two signals is very difficult. Presently the favoured strategy for avoiding this problem is to employ either a species whose emission is much longer lived than the typically very short lived background

fluorescence (~10 ns), or to employ a species whose emission is of a much longer wavelength out of the range of biological noise.

1.2.12 Timegating

Longer luminescent lifetime probes typically either involve a metal complex, or organic species whose emission is through the formally spin-disallowed route leading to long-lived luminescence (phosphorescence). Therefore the luminescent signal can be easily separated from the autofluorescence by employing a detection time-delay whereby the detector is not switched on until the undesirable signal from the sample has diminished. This is known as timegating (**Figure 18**).[17b]
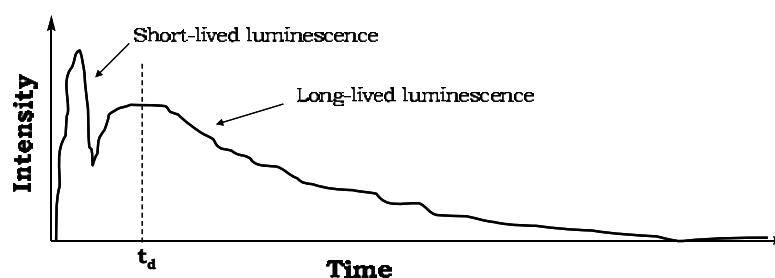
Figure 18: Timegating. A time delay (t_d) of typically 10-50ns is applied to eliminate undesired background noise.
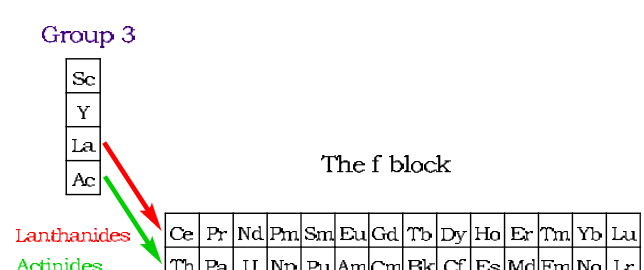
This hugely improves the signal to noise ratio, and has become commonplace in the visualisation of metal complexes in biological media, especially lanthanides complexes.

Chapter 2: Integrin targeted fluorescent peptides for *in vivo* imaging

This chapter follows on from chapter 1 with an introduction to the lanthanides as potential luminescent probes for targeted imaging. The biology surrounding the integrin target will also be covered, as will our synthetic strategy to specifically target this integrin with a cyclic peptide bearing the luminescent lanthanide complex.

2.1 Lanthanides and Luminescence.

A previously mentioned lanthanide was gadolinium (Gd).[14g] Its widespread use in clinical MRI as a macrocycle-chelated contrast agent was discussed. The reasons for this function will presently be made more apparent, however, there are other properties of certain lanthanide metals that render them attractive for potential luminescent applications.

Ln metal	Ln Metal Electronic Configuration	Ln^{3+} Electronic Configuration
Cerium (Ce)	[Xe] $4f^1 5d^1 6s^2$	[Xe] $4f^1$
Praseodymium (Pr)	[Xe] $4f^3 6s^2$	[Xe] $4f^2$
Neodymium (Pm)	[Xe] $4f^4 6s^2$	[Xe] $4f^3$
Promethium (Pm)	[Xe] $4f^5 6s^2$	[Xe] $4f^4$
Samarium (Sm)	[Xe] $4f^6 6s^2$	[Xe] $4f^5$
Europium (Eu)	[Xe] $4f^7 6s^2$	[Xe] $4f^6$
Gadolinium (Gd)	[Xe] $4f^7 5d^1 6s2$	[Xe] $4f^7$
Terbium (Tb)	[Xe] $4f^9 6s^2$	[Xe] $4f^8$
Dysprosium (Dy)	[Xe] $4f^{10} 6s^2$	[Xe] $4f^9$
Holmium (Ho)	[Xe] $4f^{11} 6s^2$	[Xe] $4f^{10}$
Erbium (Er)	[Xe] $4f^{12} 6s^2$	[Xe] $4f^{11}$
Thulium (Tm)	[Xe] $4f^{13} 6s^2$	[Xe] $4f^{12}$
Ytterbium (Yb)	[Xe] $4f^{14} 6s^2$	[Xe] $4f^{13}$
Lutetium (Lu)	[Xe] $4f^{14} 5d^1 6s^2$	[Xe] $4f^{14}$

Figure 19: The lanthanides.

Until recent years the chemistry of the 'f' block elements of the periodic table, the lanthanides and actinides, was often ignored in the training and experience of the chemist. These rare earth metals however, are finding increasing technological, catalytic and biological applications.

It was in 1794 that the Finnish chemist J. Gadolin isolated *yttria* from a mineral discovered at the village of Ytterby, near Stockholm in Sweden. Rather than consisting of the oxide of a single element, as Gadolin thought, it was subsequently quite remarkably shown that yttria contained the oxides of ten new elements: Y; Gd; Tb; Dy; Ho; Er; Tm; Yb; Lu & Th!

A short time later Klaproth, Berzelius and Hisinger isolated another oxide, *ceria*, and La, Ce, Pr, Nd, Sm and Eu were identified therein.[18]

Whilst our attentions shall be concentrated upon the lanthanides, it is worth mentioning that the actinides, other than the naturally occurring thorium (Th), protactinium (Pa) and uranium (U), are found either within nuclear explosion debris (Es and Fm), or are

synthesized as a result of the nuclear bombardment of a parent element such as U, Pu or Am. These elements are a far more recent discovery (1940 – 1971).

To brand lanthanides as 'rare earth metals' is somewhat of a red-herring as they are all actually relatively abundant. Cerium is the 26th most abundant element on earth, neodymium is more prolific than gold. The least common, thulium (Tm), is still more abundant than iodine.

Shortly after 1947 it was N. Bohr who ascertained that the fundamental characteristic this family of fourteen elements possessed is that the fourth atomic shell could contain 32 electrons, 14 more than the third. Hence these extra electrons are placed in the new 4f orbital and the first 'f' series was born. It is these 4f electrons that give the lanthanides their unique and fascinating properties.[18b]

2.1.1 Lanthanide coordination chemistry

Lanthanide element ions are large 'hard' acids and form high coordination number complexes, preferably with donor atoms such as N, O and F. The partially occupied 4f orbitals have core-like radial distribution functions, and overlap between these orbitals and donor ligands is typically very small, making covalent contributions to bonding negligible; ligand field splittings in lanthanides are of the order of a few hundred wavenumbers (cf. the splitting in d-block complexes, which is typically of the order of tens of thousands of wavenumbers). Therefore complexes are constructed largely through electrostatic interactions. Upon crossing the lanthanide series the ionic radii decrease through lanthanide contraction; this causes a concomitant increase in charge density as, in the usual (+3) oxidation state, the charge is spread over a smaller area. This leads to stronger ionic bonding and lanthanides such as ytterbium can form very stable complexes.[18c]

Lanthanides typically form stable complexes through electrostatic interactions with eight or nine ligands. It is the preparation of these stable and soluble complexes that is the key to taking advantage of the luminescence properties of lanthanides.

The different types of lanthanide-chelating ligands in use today and the reasons for our choice of the DO3A type macrocycle will be covered after some discussion of this luminescence.

2.1.2 Lanthanide luminescence

When a molecule absorbs light in the form of a photon, it can impart enough energy to promote an electron up to a higher orbital. The molecule is excited from a singlet ground state, S_0, up to an excited singlet state, S_1. Upon relaxation back down to the ground state the released energy is often dissipated into the local environment as heat. However, many molecules are able expel the absorbed energy as another photon; this is luminescence.[18]
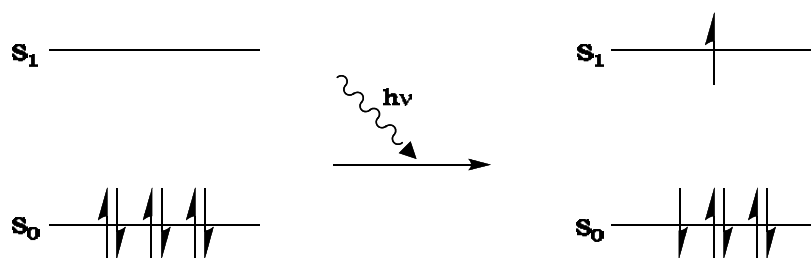
Figure 20: A representation of the excitation of an electron from ground state to excited state.

There are two types of luminescence; fluorescence (short-lived) and phosphorescence (long lived).

Fluorescence is the rapid radiative decay of a molecular excited singlet state and the luminescence lifetime is typically of nanosecond duration. Phosphorescence is a forbidden transition and as such is slow and long-lived with a lifetime on a microsecond to hour long timescale.[18b, 19]

Upon elevation to an excited state (**Figure 21**), the electron is no longer influenced by the spin of the other electron of opposite spin in the pair. Hence within the excited state there is the possibility of the two spins being opposed (singlet excited state), or parallel (triplet excited state) to one another.
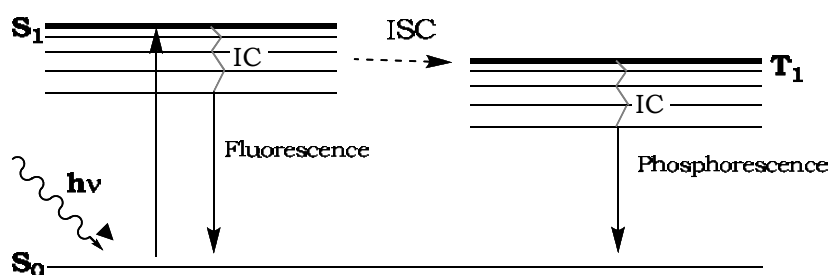
Figure 21: A simplified Jablonski diagram illustrating the fates of an excited electron. IC = Internal Conversion.

Singlet-triplet and triplet transitions are formally disallowed, but occur through a phenomenon known as spin-orbit coupling where the magnetic field generated by the electronic orbital angular momentum interacts with the spin of the electron. This can lead to the electron flipping to the opposite spin and occupying a lower energy triplet (T_1) state through the process of inter-system crossing (ISC). It is this relaxation from the triplet excited state (T_1) to the singlet ground state (S_0) that results in phosphorescence. This is a forbidden transition. It is forbidden for an electron to change orbital and flip spin simultaneously, therefore the transition has a low probability of occurring and phosphorescence can take place long after the excitation pulse has been removed.[18c,19]

Although lanthanides tend not to possess emissive triplet states they are still often described as phosphorescent due to their relatively long lived emission.

Lanthanides each have a very distinct unique spectrum due to the relaxation from the singlet excited state to a number of well defined energy levels (**Figures 22 & 23**). Each

spectral peak corresponds to a specific transition and thus an emission spectrum can aid characterisation through confirmation of the presence of a specific lanthanide metal.

Terbium
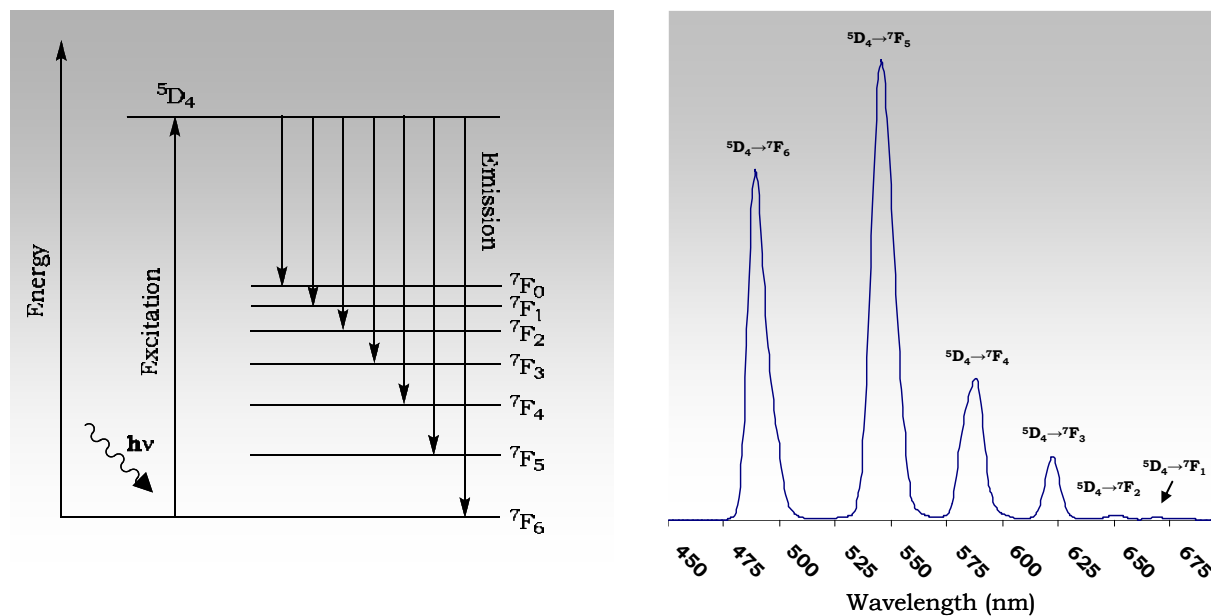
Figure 22: **Left**) An energy level diagram for terbium transitions; **Right**) A terbium emission spectrum from the author's own work illustrating the different electronic relaxation transitions.

Europium
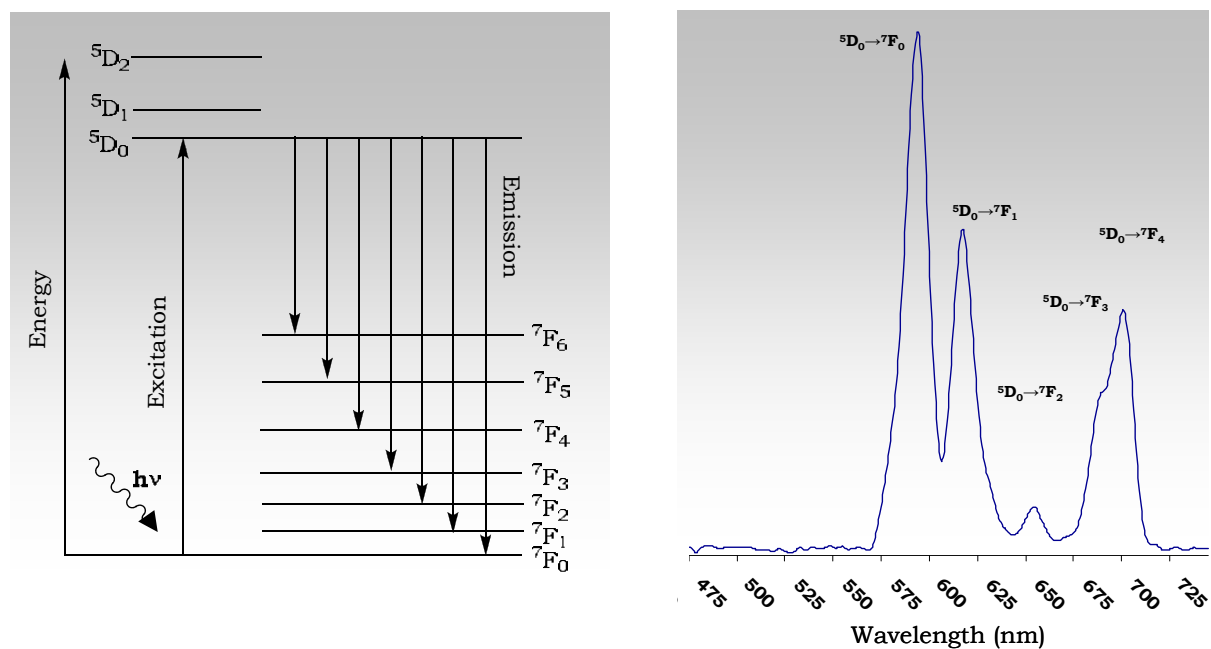
Figure 23: **Left**) An energy level diagram for europium transitions; **Right**) A europium emission spectrum from the author's own work illustrating the different electronic relaxation transitions.

Ytterbium
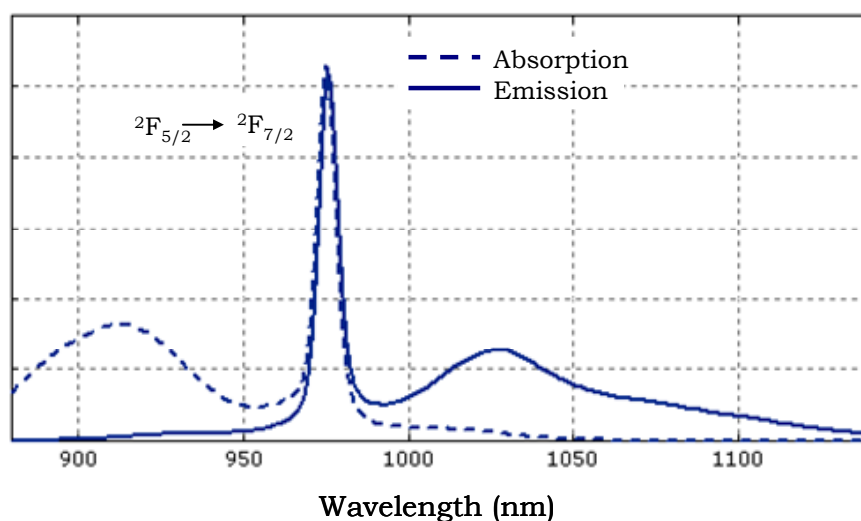
Figure 24: A ytterbium spectrum by illustrating the 980 nm emission band. Adapted from Ref. 19b.

The $^{x}X_{x}$ notation signifies the term symbol describing the specific energy level within the metal. A general form of $^{(2S+1)}L_{J}$ is applied where S is the spin multiplicity, L is the total orbital angular momentum and J is the quantum number describing the total angular momentum.[19]

With an extinction coefficient (ϵ) typically of the order of 1 $M^{-1}cm^{-1}$, the ability of lanthanide metals to absorb energy from a photon and enter an excited state is poor. This in turn leads to very weak emission in comparison to organic fluorophores which possess extinction co-efficients of the order of $1x10^{5}$ $M^{-1}cm^{-1}$. This is due to the transitions involved in the formation of a lanthanide excited state being f→f transitions which are Laporte forbidden as there is no inversion of parity. It may be argued that the d→d transitions of strongly absorbent and emissive transition metal complexes are also forbidden, and this is true, but lanthanide f-orbitals are low lying and shielded by outer $5s^{2}$ & $5p^{6}$ orbitals and as such are not disturbed by the surrounding ligand structure. Transition metal 'd' orbitals on the other hand are not so well shielded and subject to distortion from molecular vibrations which lower the symmetry and render the Laporte selection rules less applicable.[19]

However, this weak absorption characteristic of the lanthanide metals can be overcome through a phenomenon known as energy transfer (ET).

2.1.3 Energy transfer

A weakly absorbing lanthanide ion can be excited through energy transfer from an appended chromophore rather than through the disallowed and hence inefficient direct excitation of the metal itself. This is possible when an antenna, normally aromatic, chromophore absorbs light in the UV region (λ = 200-400 nm) and then directs this energy to the metal centre in an allowed transition.[19] The metal can then relax as previously described and relax through the emission of luminescence (**Figure 25**).[20]
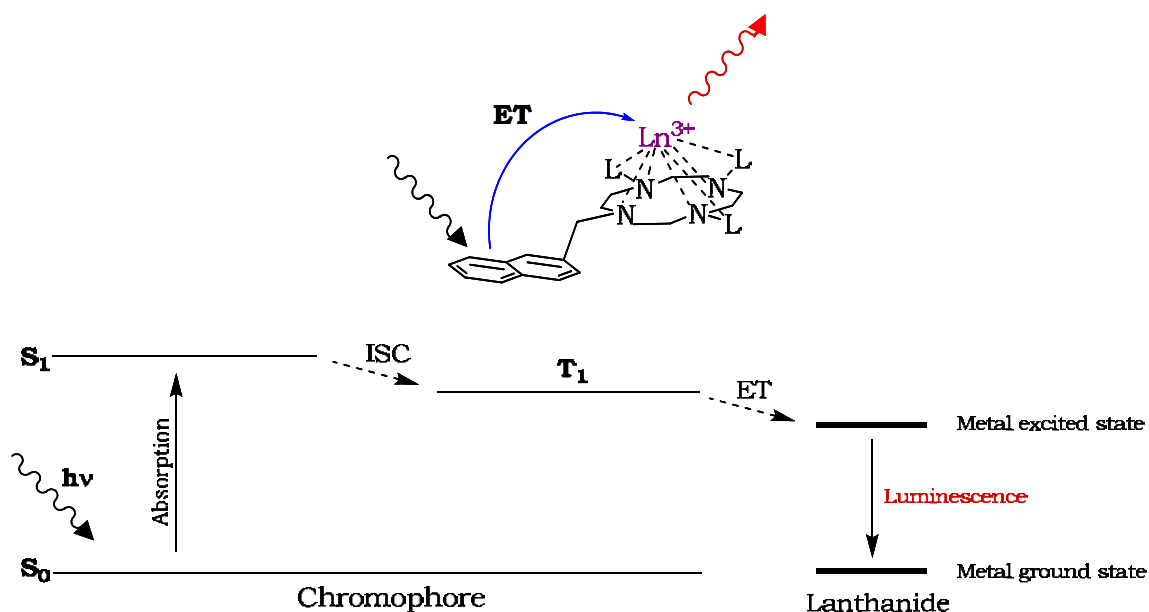
Figure 25: Energy transfer and the Jablonski diagram; adapted to include a nearby lanthanide metal.

This luminescence is known as sensitised emission and it is normally only possible if there is sufficient overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor. Another possibility is superexchange where energy transfer is actually mediated by another atom.

The mechanism for energy transfer is rationalised by one of two mechanisms, dependent upon the ligand. A conjugated system allows the overlap of π-orbitals from donor to acceptor and hence the free flow of electronic energy to excite the metal; this is known as the Dexter mechanism.[20b] However if the metal is complexed within a ligand where there is no 'through bond' energy pathway, it is thought that the energy can be transferred through space via the dipole-dipole interactions of solvent molecules. This is known as the Förster mechanism and the efficiency of the process is dependent upon the proximity of the chromophore to the metal.[20c/d]

In the case of lanthanide complexation the ligands for each of the two energy transfer mechanisms are not always obvious as both mechanisms can be occurring simultaneously. Bipyridyl (bpy) and (terpy) cryptand complexes are conjugated systems, whereas the DTPA and DOTA azamacrocycles are not and the favoured pathway of energy transfer must be via the Förster mechanism (**Figure 26**). However, in those systems where there is no obvious conjugated π-system pathway for conventional Dexter energy transfer it is still not always straightforward Förster energy transfer as the small spectral overlap between aryl donor and lanthanide acceptor state prohibits this mechanism somewhat. Indeed, all four compounds in **Figure 27** are likely to have significant Dexter superexchange contributions.[21]
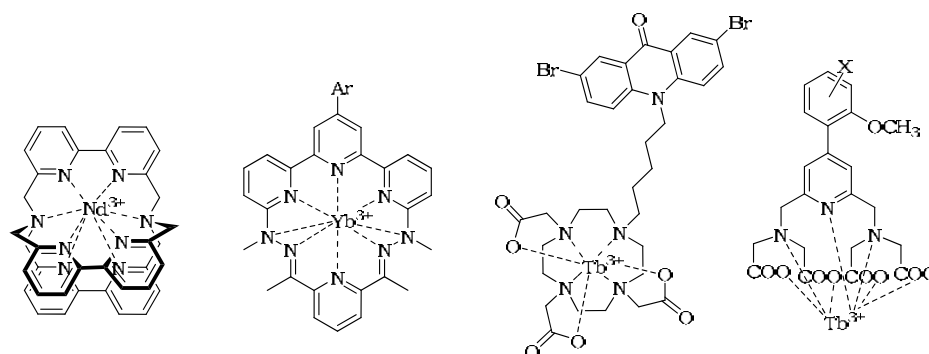
Figure 26: Examples of effective means of constructing energy-transfer lanthanide coordination complexes.[19]
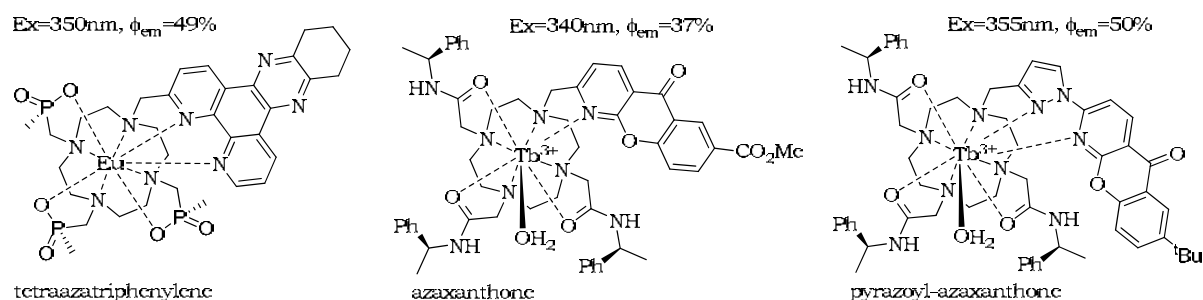
Figure 27: Examples of tetraazamacrocyclic systems where excitation of the appended chromophore leads to efficient metal centred luminescence.[21b]

2.1.4 Luminescence lifetime (τ).

The primary motivation for taking advantage of lanthanide luminescence brings us back to long-lived emission which allows time-gating to eliminate background noise. This lifetime is especially evident in Eu and Tb which possess lifetimes (τ) of up to 5 ms.[22]

The lifetime of a luminescent species is calculated by measurement of the intensity of luminescence at increasing time intervals after the original excitation and using mathematical formulae (**Figure 28**), to derive a calculated fit to the observed data.

$$\tau = 1/k_{obs}$$

$$I_t = I_0 \exp(-k_{obs}t)$$

$$I_t = I_0 \exp(-t/\tau)$$

τ = Luminescence lifetime
I = Intensity
t = time
k = rate constant

Figure 28: Calculation of the luminescence lifetime (τ).

The rate constant (k_{obs}) is derived from all the molecular processes including luminescence and non-radiative decay.

2.1.5 Quenching

Non-radiative relaxation can occur through the intersystem crossing (ISC) process, but is predominantly used to describe the transfer of vibrational energy to the molecules in the

surrounding environment. This can have an adverse effect upon lanthanide luminescence as the metal excited state has good Franck-Condon spectral overlap with the vibrational harmonics of nearby O-H oscillators. Therefore if there is local protic solvent, the preferred route of decay is through the O-H oscillations, and luminescence intensity and lifetime is quenched. Lanthanides with many electrostatically coordinated water molecules show very weak luminescence whereas just one or two appended water molecules allows strong luminescent emission and long lifetimes (**Figure 29**).[19]

The lifetime of the complex is therefore influenced by the hydration state of the metal. Moving from H_2O to D_2O, the overlap is reduced due to the lower energy of the O-D oscillation and the consequent energy match between the lanthanide excited state and higher harmonics of the O-D oscillator. This reduces the quenching and enhances the luminescent lifetime. From the differences in luminescent lifetime from the two hydration states in the deuterated and non-deuterated solvent, it is possible to quantify the number of water molecules bound to a certain complex.[22b] This however is riddled with complexities as the N-H and C-H oscillations may or may not also be involved depending on the ligand and the energy of the emissive state of the lanthanide.
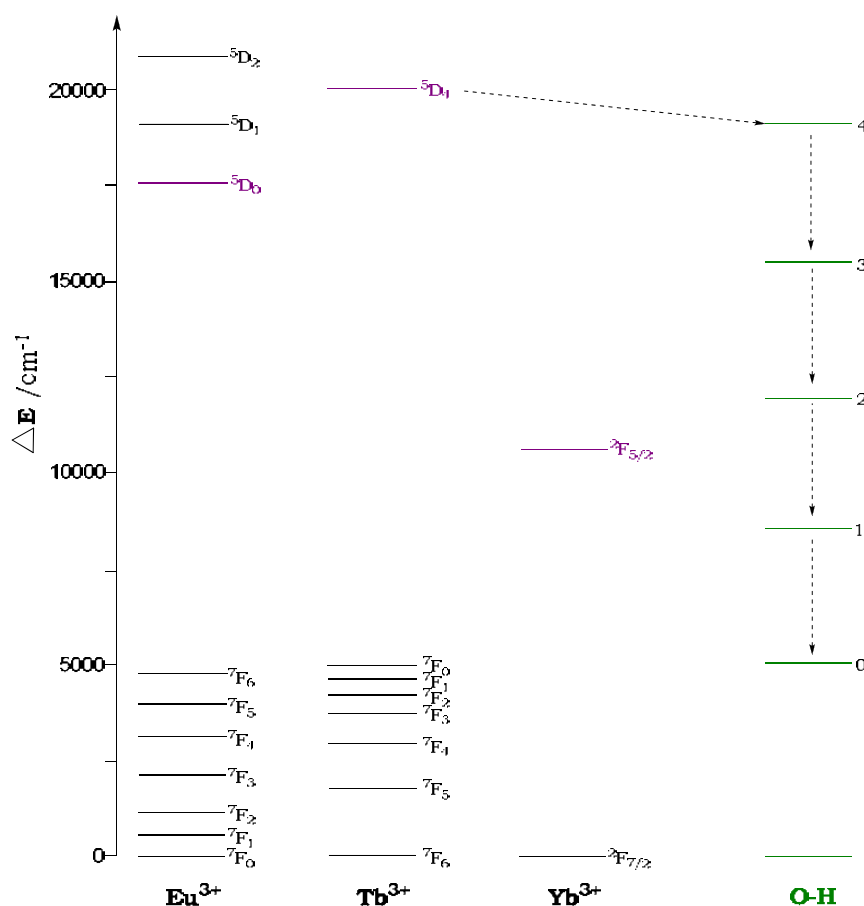
Figure 29: Energy level diagram illustrating the ground and excited states of Eu, Tb and Yb with their emissive levels shown in purple. The green scale shows quenching. The energy levels of O-H oscillations are shown with a non-radiative Tb^{3+} decay route reducing the luminescence lifetime and extinction co-efficient of the Tb excited state.[19]

However, not only will these factors need to be taken into consideration, but it also brings us back to the gadolinium species employed as MRI contrast agents.[14g] MRI images are enhanced by sharp variations in the rate of relaxation of water protons from tissue to tissue. Gadolinium has seven unpaired electrons and as such is highly paramagnetic. Stable complexes containing gadolinium with an appended water molecule which exchanges rapidly cause a reduction in the relaxation times of the ^{1}H nuclei on the nmr timescale and thus enhance the contrast. The key difference however is that T^1 relaxation in MRI is very different from the non-radiative relaxation of an excited electronic state.

2.1.6 Eu, Tb and Yb and their chromophores

The study of luminescent lanthanide complexes has predominantly centred around the use terbium and europium as the metals. This is due to their long-lived luminescence and susceptibility to energy transfer from chromophores whose emission is in the ultra violet region.[22c]

The choice of chromophore in europium and terbium complexes is limited by the large energy gap between the excited emissive state and the ground state. With terbium the energy gap is 20,300 wavenumbers and this renders energy transfer inefficient from all but the simplest chromophores such as benzyl and indolyl groups with their high energy triplet states. Larger chromophores such as naphthalene, have triplet states too close in energy to the terbium excited state, resulting in reversible and temperature sensitive transfer.

Europium has a smaller energy gap of 17,500 wavenumbers which allows a greater choice of effective chromophores with lower energy excited states. The most effective systems employ chromophores with smaller energy differences between the singlet and triplet states. For instance, the sensitising chromophore for an efficient europium transfer needs to have a triplet state of at least 2000 cm^{-1} above the europium $^5D_0$ (17200 cm^{-1}) excited state to avoid back energy transfer from the excited lanthanide ion to the triplet state of the sensitiser.[19]

Ln	Most intense emissive band
Europium	616 nm
Terbium	545 nm
Ytterbium	980 nm

Figure 30: The wavelengths of Eu, Tb & Yb emission maxima.

It is important to remember that the wavelength of europium and terbium metal-centred luminescence is still in the region where biological autofluorescence from media such as human serum (500-600 nm) can interfere with the emission signal.[22d] It is the long luminescent lifetime that is manipulated to distinguish the metal from the background.

However with ytterbium it is a different story as the emission maximum is at 980 nm in the near-IR region of the spectrum where biological interference is minimal.

The quest for near-IR emissive probes suitable for bio-imaging is becoming more and more important. It has become apparent that to be able to use long wavelength excitation is ideal as this avoids unnecessary excitation of luminescent biological species, and emission occurs at a biologically transparent wavelength.

Lanthanide complexes that emit in the near-IR include neodymium, ytterbium and erbium. The emissive states of these lanthanides are much lower in energy than the analogous states of terbium and europium, and a range of more complex and longer wavelength chromophoric ligands become available.

However there is a large disadvantage with using near-IR emissive lanthanides as luminescent lifetimes are shorter but most importantly quantum yields are lower compared to europium and terbium complexes. Near-IR lanthanide emission is also still quenched by the lower harmonics of the O-H oscillations.

Luminescent lanthanide complexes are used in biological assays.[19] Traditionally radiotracers have been employed for the detection and study of specific antibody-antigen interactions at trace concentration in biological samples. These have now been eclipsed by fluorescent assays involving lanthanides, where the metal-centred fluorescence is either directly excited through the inefficient disallowed pathway, or, when the assay shows successful antibody to analyte binding, an energy transfer excitation pathway is available to provide enhanced luminescence signifying a positive outcome (**Scheme 1**).
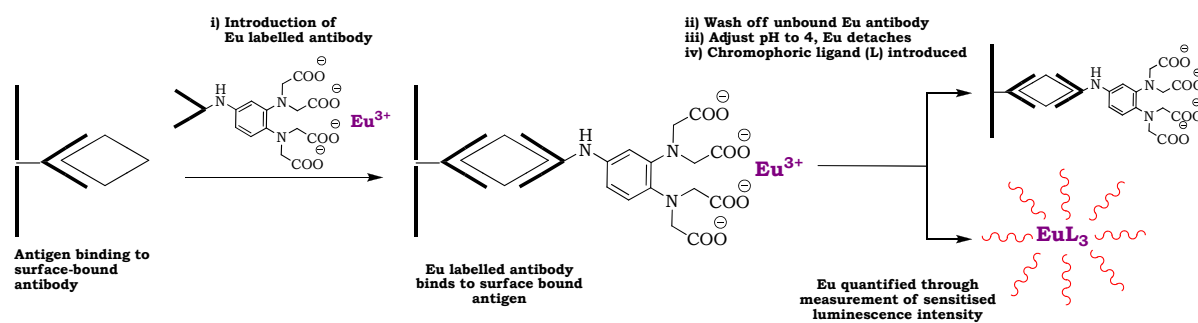
Scheme 1: Representation of a Eu based assay; the traditional DELFIA assay. Adapted from Ref. 19.

It is both the near-IR ytterbium emission and the long-lived europium and terbium emission that is of interest in this project for use in targeted luminescent probes that are not only tumour specific but also avoid the problems associated with autofluorescence through time-gated detection.

There needs to be a means of synthetically locating the chromophore at an appropriate distance from the lanthanide metal. The previously discussed Förster mechanism of energy transfer showed a relationship of r^{-6} where 'r' is the distance from chromophore to metal.[20d]

It is envisaged that a unique way of constructing the system could be through a multicomponent reaction where the metal is complexed as one component, the chromophore is second and the third would be a means of attachment to a biological entity for imaging purposes. This is where a multicomponent reaction comes into its own.

2.2 The Ugi 4-component condensation reaction (U-4CCR)

Ivar Ugi (1930-2005) was a German chemist with a penchant for one-pot multicomponent condensation reactions where more than two starting materials react to form a complex product with features of each of the reactants, in a single step.[23a] The chemistry of multicomponent reactions is now vast. The benefits have become more and more obvious with prolific yields, mild conditions and the ability to access incredibly complex structures in a single step, rather than through a lengthy and often inefficient stepwise synthetic sequence. Another attribute of multicomponent reactions is that the structure of the reaction product can be easily manipulated through the variation of one or more of the individual components.[23a/b]

The Ugi 4-component condensation reaction (U-4CCR) is analogous to the Passerini 3-component reaction. These are classic isocyanide-based multicomponent reactions where the isocyano functional group is unique in its ability to react with electrophiles and nucleophiles at the same atom, resulting in an incredible diversity of products. The components in the Passerini reaction are an aldehyde or ketone, an isocyanide and a carboxylic acid (**Scheme 2**).
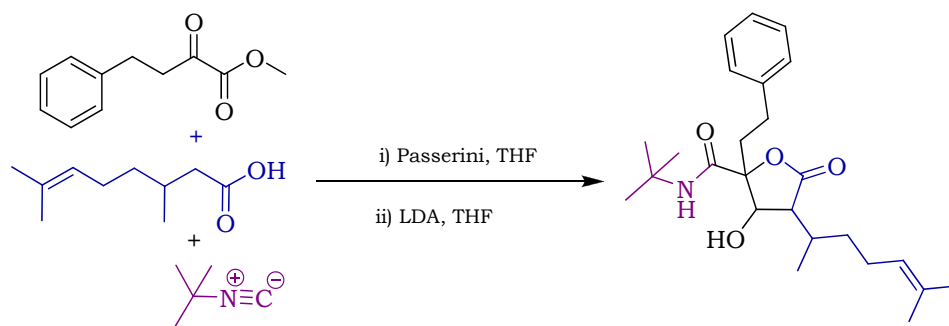
Scheme 2: An example of the Passerini reaction in the synthesis of an HIV protease inhibitor.[24]

The Ugi 4-CCR involves a fourth substituent, an amine. Mechanistically the reactions are quite different. There are a number plausible proposed mechanisms for the Ugi 4-CCR, all of which involve the pre-formation of an imine through a condensation reaction between the aldehyde and the amine. One proposed mechanism involves the formation of a heterocyclic intermediate, but the favoured route involves a Mumm rearrangement (intramolecular acylation) to form a product bearing features of each component (**Scheme 3 & Page 55**).[23b,25-27]
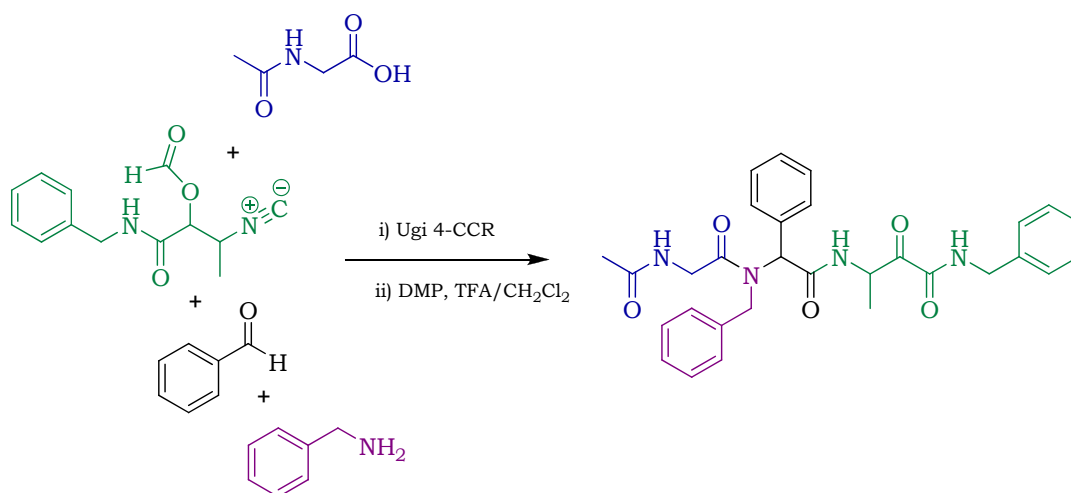
Scheme 3: An example of the Ugi 4-CC reaction in the synthesis of a HCMV protease inhibitor.[28]

A perceived synthetic drawback in this chemistry is the isocyanide component. Low molecular weight isocyanides are typically highly volatile and remarkably unpleasant compounds to handle. The stench is offensive and the scope of their reactivity renders many of them toxic. The first chemists who experimented with isocyanides worked outdoors![26,27]

Some isocyanides are now commercially available and those which are of larger molecular weight and are in solid form, are odourless. Indeed the isocyanide functional group, unusual for bearing a formally divalent carbon, has been discovered in many natural products.[29] The divalent carbon is obvious in ^{13}C-nmr where a distinctive peak is observed at ~ 100 ppm.

Isocyanides are typically synthesised from primary amines through either dehydration of the corresponding formamide, or directly from the amine precursor.[27,29b]

It was envisaged that an Ugi 4-CCR would be an effective means of constructing a sensitised luminescent lanthanide containing system which can be attached to a biological delivery motif for specific targeted imaging. The next item of importance in this discussion is the biological entity that is to be the vector in the system and as such must be specific to the types of malignancy which require highly sensitive illumination.

2.3 The vehicle and its tumour targets; RGD and Integrins

An introduction to the RGD amino acid sequence and its recognition and binding to specific cellular surface receptor proteins must begin with an overview of some of the relevant biological processes involved in malignancy and an introduction to our cancer specific target, the $\alpha_v\beta_3$ integrin.

2.3.1 Cancer and angiogenesis

There are many postulated models attempting to explain how discrete genetic variations can transform a normal healthy cell into a cancerous cell. The heterogeneous nature of cancer renders the unveiling of a general mechanism impossible, however, there are similarities between the various classes of malignant cells such that it is safe to state that the occurrence of spontaneous genetic changes is a major factor.[30] However, the purpose of these words is not to discuss carcinogenesis, but to delve beyond this unfortunate initiation, and discuss the progression of tumours and how a fundamental aspect of this, angiogenesis, relates to the research carried out during this PhD.

Angiogenesis, is the growth and development of new vasculature from existing blood vessels. The mechanism of angiogenesis and the unravelling of the numerous parties involved, with an accurate understanding of their function, is seen as an important step in the design of new therapeutic or diagnostic agents for a whole host of oncological, cardio, wound healing and other medical applications.[31]

Just as in normal healthy biological tissues, tumour survival and progression is completely dependent upon access to the cardiovascular system not only for an adequate supply of oxygen and metabolites, but also for an efficient means of waste disposal. Without access to the blood stream, most tumours would remain small and the incredible exponential growth and spread commonly perceived would not occur.

In some cases this process of new blood vessels sprouting or branching forth from pre-existing vessels is unable to keep up with the rate of tumour growth and the cancer cells furthest from the blood supply suffer from deficiencies.

However, the dependency upon vascular access is by no means consistent and varies according to the type of tumour and its stage of progression. In general however, access to the host blood supply and the generation of a special local vasculostructure is a rate-limiting factor in tumour propagation.[30]
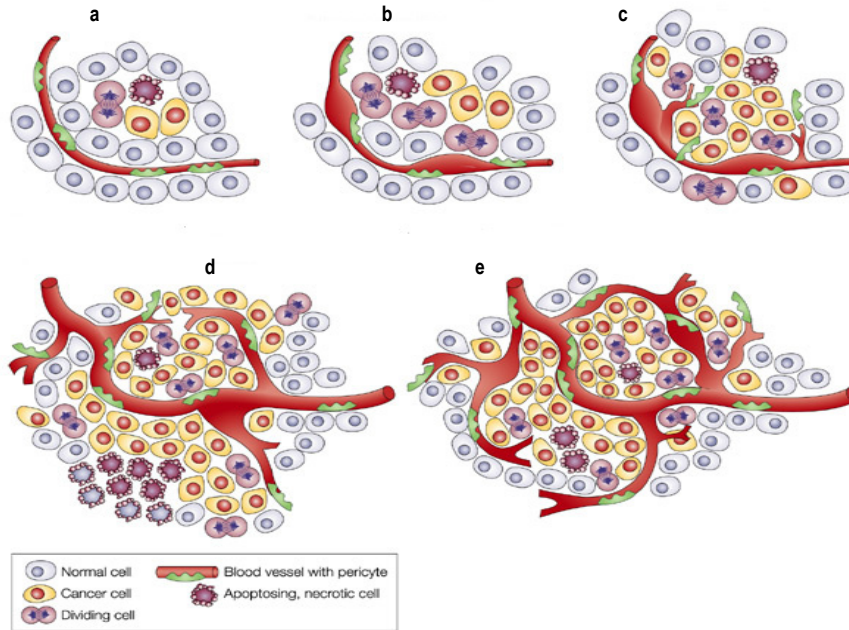
Figure 31: **a**) A dormant avascular tumour. It remains this way until an equilibrium of proliferating and apoptosing cells is achieved and angiogenesis is triggered. *Pericytes* are smooth muscle cells that help regulate blood flow. In tumours the weakness and chaotic nature of the new vessels is partly due to the inefficient recruitment of properly organised mural cells, or pericytes; **b**) The onset of angiogenesis: The dilation of the blood vessel; **c**) Sprouting; **d**) New vasculature forms and matures recruiting perivascular cells (vessel periphery 'scaffolding' cells); **e**) The blood vessels continue to sprout and grow for the lifetime of the tumour and are specifically drawn towards the pro-angiogenic factor releasing hypoxic areas to provide the essential nutrition and oxygen. Adapted from Ref. 30.

Pathologists and oncologists have long known that cancer cells located further than approximately 0.2 mm from a blood vessel have a significantly diminished survival rate with comparison to those found closer. This is due to the optimal distance for efficient oxygen diffusion being exceeded, and thus cells found within 0.2 mm can guarantee a steady oxygen supply; whereas those outside this limit suffer from a lesser to greater degree of hypoxia.[29] However, this is by no means universal, as human tumour oxygen consumption rate is variable, with some tumours, for instance prostate or germ-cell testicular tumours, having a high microvessel density in comparison with normal tissue, and others, for instance lung, mammary and colonic carcinomas having a much lower oxygen necessity.[32]

Hypoxic cells are in danger of necrosis triggered by the p53 protein, the cellular housekeeper, which orchestrates its response according to the threat posed to a cell. This rapid death and disappearance of a cell is known as apoptosis. However, the deactivation of the p53 signalling system through malignancy-directed mutation explains the survival of hypoxic cancer cells outside the effective oxygen diffusion radius. Hypoxic cells are ubiquitous in human tumours.

A deficiency of oxygen, however, is not the only quandary faced by a cell positioned at a greater than 0.2 mm distance from a blood vessel. Other factors fundamental to cell survival include the ability to acquire nutrition and discard carbon dioxide and other metabolic waste products through interaction with the vasculature. Capillary networking within normal tissue is so dense and intricate that the majority of cells are within a few cells diameter of a blood source. In especially active cells with high metabolism, cells usually have direct contact to one or more capillaries.[29] The ability of tumours to initiate and develop a blood supply through angiogenesis is fundamental to their survival, and more and more their treatment with intraveneous chemotherapeutics and steroids.

The onset of angiogenesis, or the 'angiogenic switch', can occur at any time during tumour progression depending upon the cancer type and its specific microenvironment.[30] It should also be made clear that angiogenesis is a quite normal procedure occurring in several natural biological processes including within the female reproductive organs, in tissue undergoing physiological growth and in the healing of injuries. This normal and organised regeneration and development of circulatory structure known as physiological angiogenesis, is distinctly different in character from tumour angiogenesis where the architecture is rough and chaotic. This is due to an imbalance in the variety of activators and inhibitors which coordinate angiogenesis.

There are numerous influential components involved in the process of angiogenesis and they will not all be covered in detail in this discourse, indeed the complexity of the role of many of the activators such as VEGFs are still not fully understood.
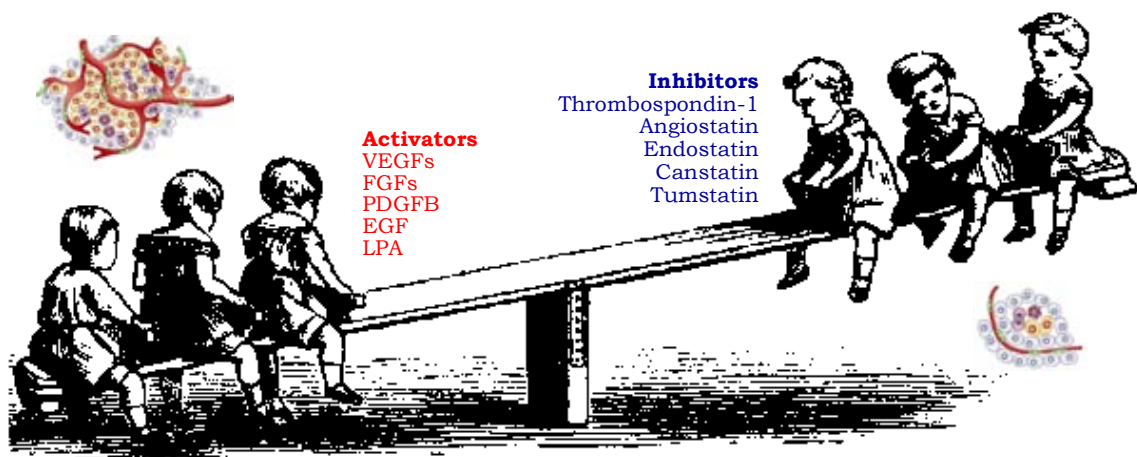
Figure 32: A cartoon illustrating factors inducing angiogenesis, activators; and those intent on disrupting this process, inhibitors.

The 'activators' are primarily growth factors released by hypoxic or pathological tissues such as tumours, and they stimulate new blood vessel creation through the recruitment, proliferation and ordering of endothelial cells.[33] The release of these signalling factors, such as the potent VEGF (Vascular Endothelial Growth Factor), is detected and the signals understood by a family of endothelial and blood vessel cellular surface receptors, known as integrins. The resulting stimulation of endothelial cells instigates migration and

accumulation at the site of signalling. Enzymes such as proteases are then released and degrade the vessel basement membrane allowing cells to escape and proliferate in the surrounding extravascular matrix initiating the formation of the labyrinth of endothelial and smooth muscle cells that is the basis of a new vessel.
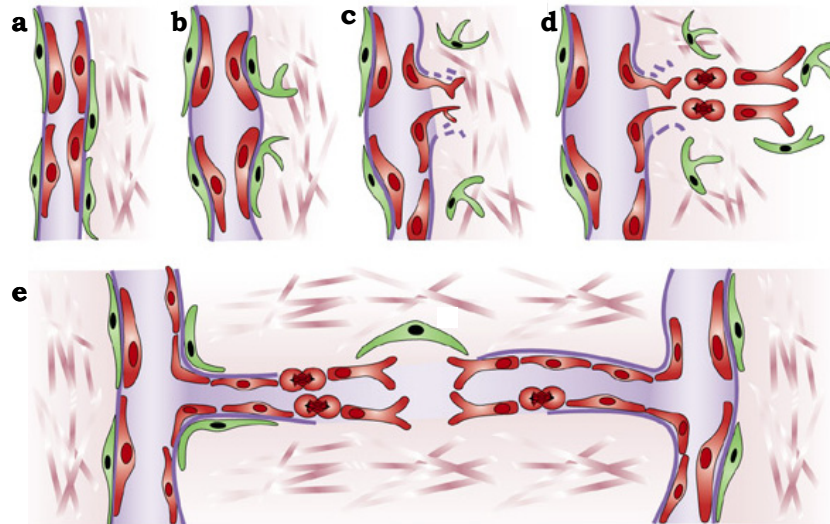
Figure 33: A diagram illustrating the basic mechanism of angiogenesis. **a**) Accumulation of endothelial (red) and smooth muscle cells (green); **b**) Smooth muscle cells detach and the blood vessel dilates as the basement membrane and ECM is degraded; **c**) Endothelial cells migrate towards angiogenic stimuli from tumour cells; **d**) Endothelial cells proliferate, presumably guided by smooth muscle cells; **e**) Endothelial cells adhere to one another alongside smooth muscle cells and a new basement membrane is formed. These sprouts will then fuse with other sprouts to create the new circulatory systems. Adapted from Ref. 30.

2.3.2 Endothelial cells

Endothelial cells are found throughout the circulatory system but originate in the bone marrow as young undifferentiated cells before being released and maturing into the endothelial cells in the blood stream.

Hypoxic regions of tumours also attract macrophages. This would, at first glance, appear to be a mechanism of cancer cell destruction as macrophages are immune cells normally deployed to scavenge and eliminate infection and abnormal cells. However, evidence suggests that some of these immune cells actually aid tumourigenesis as, being quite tolerant to hypoxia, they excrete VEGF the potent angiogenic factor which, in turn, reduces the hypoxia by attracting endothelial cells which induce capillary formation through interaction with smooth muscle cells as described above (**Figure 33**). Some tumour associated macrophages (TAMs) also secrete matrix metalloproteinase enzymes (MMPs). These enzymes enhance angiogenesis by not only liberating endothelial cells, but also through excavating space for tumour expansion through the disruption of existing tissue structure.

The mechanism of angiogenesis can occur in three ways. Sprouting, although the most commonly observed mode of blood vessel generation, is not the only type of angiogenesis.

Bridging and intussesception are two other, quite different angiogenic pathways (**Figure 34**).
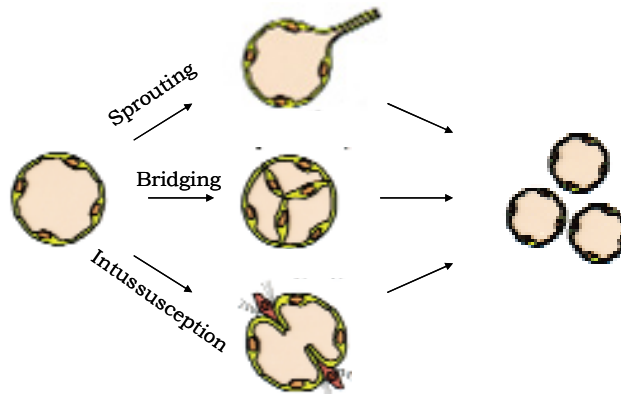
Figure 34: The pathways to new blood vessel formation. Adapted from Ref. 31.

It is upon the cellular surface of endothelial and smooth muscle cells that are found the targets of interest to this project. These are surface receptor proteins and those of particular interest are called integrins. Certain integrins have been found to be upregulated on endothelial cells which accumulate at the command of cells requiring a blood supply and hence are in great density in and around tumours.

In summary, cells lacking an adequate oxygen supply release angiogenic factors which stimulate the ingrowth and branching of capillaries. Cells involved in angiogenesis, particularly endothelial cells, bear increased levels of integrins in comparison to surrounding cells. As endothelial cells are in a high density around sites of angiogenesis it makes these integrins a potential highly specific target for the binding of luminescent probes to highlight tumour sites with great specificity.

2.3.3 Integrins

The past few years have seen the identification of significant roles for integrins in the development of cancer, especially in cell proliferation, metastasis, and angiogenesis.

Cells are dependent upon the extracellular matrix (ECM) for growth and division. The ECM consists of a series of glycoproteins, including collagens, laminins, proteoglycans and fibronectin which are oligosaccharide-appended polypeptides and are fundamental for intracellular communication (**Figure 35**). Mammalian cells possess specialised cellular surface receptors to detect and bind to elements of the ECM. One special family of these receptor proteins are the integrins.
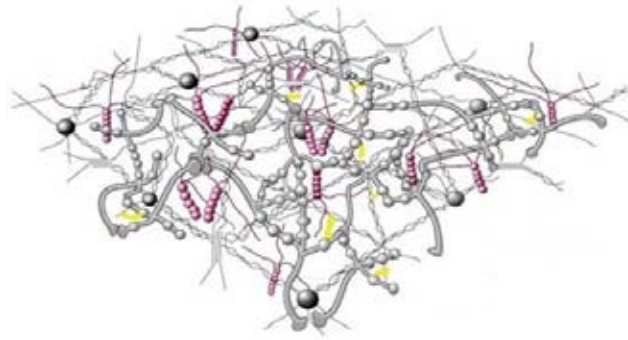
Figure 35: A representation of the complexity of the extracellular matrix (ECM). An elaborate network of proteins, glycoproteins and proteins such as collagen are entwined and interact with one another enabling intercellular signalling. Most cells require attachment to the extracellular matrix to grow, survive and proliferate. [34]

Integrins are protein heterodimers formed by the non-covalent association of an 'α' and 'β' subunit. The integrin family is composed of 18α and 8β subunits that are contained in 24 αβ heterodimeric combinations.[35] Integrins act as communication linkers allowing information to be transported between the extracellular matrix and the nucleus of the cell, and intercellularly through interactions with neighbouring cells.[36]

The structure of these membrane proteins lends them to involvement in adhesion and signalling as they possess large extracellular domains (**Figure 36**).

Integrins depend upon divalent cations (Ca^{2+}, Mg^{2+}) to bind extracellular ligands which, although being structurally diverse, are all recognised due to binding to specific amino acid residues. Binding specificity towards a particular ligand is achieved through contact points upon the integrin. High affinity binding to ligands occurs in response to so called 'inside - out' cellular signals that cause extracellular integrin structural alterations, rendering the integrin to ligand binding compatable. As a result, the ligand binding induces further structural modifications which trigger 'outside - in' communications.[37]
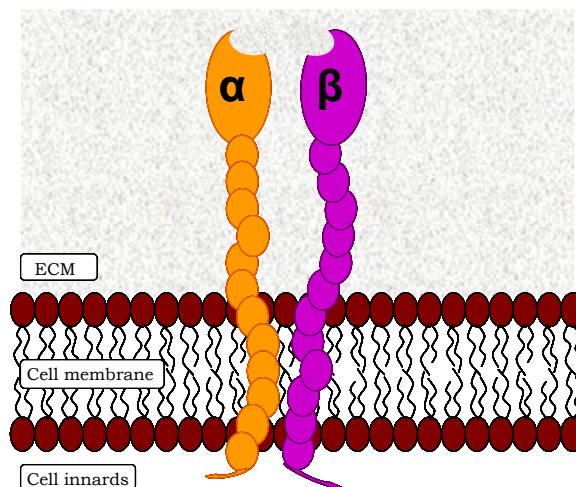
Figure 36: A cartoon of a typical αβ integrin pair.

2.3.4 Integrin $\alpha_v\beta_3$

The integrins, $\alpha_v\beta_3$ and $\alpha_5\beta_1$ have been found to play a key roles in angiogenesis, indeed it is thought that of all the integrins, the inhibition of $\alpha_v\beta_3$ signalling is one of the most important targets for the treatment of cancer.[38] Blocking of this integrin with antibodies, peptides, peptidomimetics and other antagonists has demonstrated efficient disruption to tumour growth, angiogenesis and metastasis. $\alpha_v\beta_3$ integrins are highly expressed on activated endothelial cells, new born blood vessels and on some tumour cells (melanoma, glioblastoma, ovarian and breast) making their attraction as imaging targets even more promising due to the dense congregation at tumour sites making them stand out against surrounding tissue.[38,39]

The high binding affinity of this integrin for extracellular proteins in the surrounding matrix means that they play mediation roles in the migration and accumulation of these endothelial cells and regulate their growth, survival and differentiation. It is a tri-amino acid sequence which is the crucial to the binding of extracellular entities to $\alpha_v\beta_3$ integrin known as RGD. It is this RGD sequence that will be employed in this project as the site-specific delivery motif for the luminescent lanthanide probes discussed previously.

Figure 37: Crystal structure illustrating the binging site of an $\alpha_v\beta_3$ integrin interacting with the RGD. Pdb code 1L5G.[40,41]

2.3.5 RGD

The tri-amino acid sequence arginine(R)-glycine(G)-aspartic acid(D) is shown below. In extracellular matrix proteins this motif represents only a fraction of the amino acid make-up of the whole protein but it is the small section fundamental to $\alpha_v\beta_3$ binding.
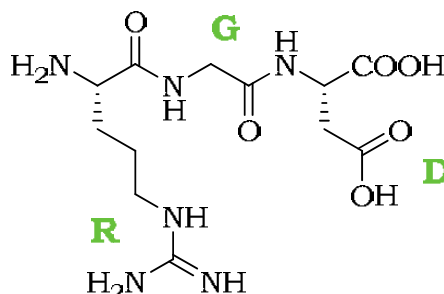
Figure 38: The tri-amino acid sequence RDG.

Proof of $\alpha_v\beta_3$ integrin binding to RGD has been shown through analysis of the crystal structure of the extracellular section of $\alpha_v\beta_3$ in complex with a cyclic RGD ligand (**Figure 39**).[41]

However the RGD sequence is also recognised as an essential binding motif for seven of the 24 integrin receptors. The roles of integrins are diverse and as such RGD may not be quite so $\alpha_v\beta_3$ integrin specific as first thought. However, it is the increased density of integrin $\alpha_v\beta_3$ receptors in tumour sites that still renders the ongoing research of great potential.

2.3.6 Previous $\alpha_v\beta_3$ integrin targeting with RGD

Ways of synthetically mimicking the RGD motif found in extracellular proteins are continuing to evolve. The field has seen the rapid development of a variety of small-molecule integrin antagonists with diverse ligand characteristics. Linear RGD peptide derivatives and those cyclised through disulphide bridges have been shown to exhibit low specificity and poor metabolic stability, being susceptible to breakdown by peptidases and proteases.[42] Through screening for the best spatial replication of the orientation of the RGD unit in the integrin binding site, high selectivity and binding affinity were recorded for cyclic pentapeptides with a D-phenylalanine amino acid unit.[38] Of particular interest is the cyclic peptide cRGDfV which has been further optimised through simple N-methylation to afford c(RGDf(NMe)V) which is the glioblastoma drug Cilengitide™ presently in phase III clinical trials (**Figure 39**).[39]
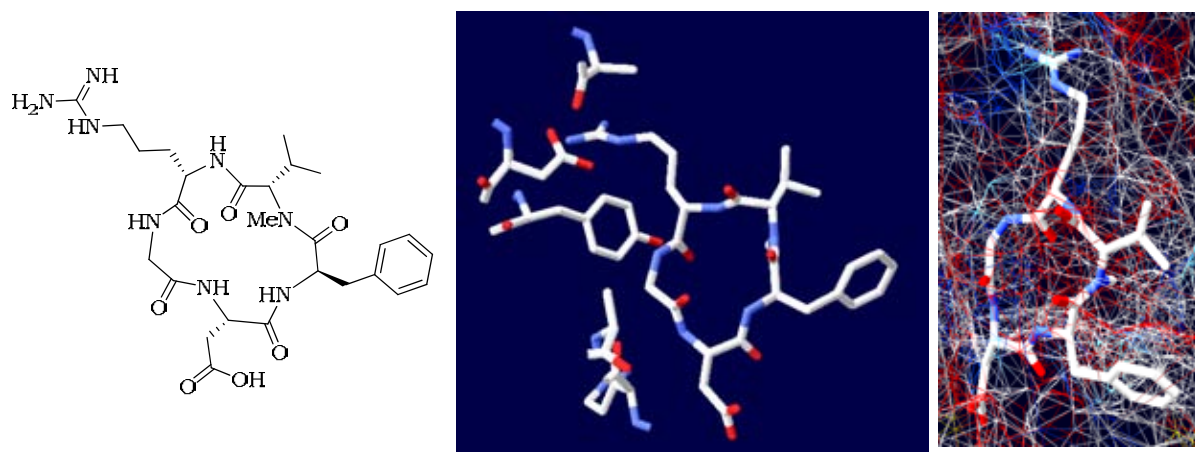
Figure 39: **Left**) The cyclic pentapeptide derivative c(RGDf(NMe)V; (Cilengitide™).[38] **Centre**) Highlighted cutout of the crystal structure showing c(RGDf(NMe)V in complex with key $\alpha_v\beta_3$ binding site residues. **Right**) A surface map illustrating the availability of the valine residue for manipulation. From PSDB (1L5G) & adapted from Ref. 41.

Although efficient as an $\alpha_v\beta_3$ antagonist the structure of these cyclic valine containing pentapeptides greatly limits their potential for further derivatisation to include alternative signalling units. In our case we need a position within the cyclic peptide which does not interfere with the integrin binding site and is available for elaboration with a luminescent probe.

The only two amino acid residues not involved in binding are D-phenylalanine and valine. As the D-phenylalanine residue is important for orientation and stability purposes, substitution of the valine residue with lysine was deemed a perfect means with which to allow conjugation with chelators, prosthetic groups and dyes at a greater distance from the active site. Reported results were very satisfactory with the lysine residue having no significant impact upon binding affinity.[39,43]

Technitium 99m radionucleotide targeting using RGD derivatives has been investigated quite thoroughly for *in vivo* SPECT imaging potential with mixed results. ^{99m}Tc-DKCK-cRGDfK (**Figure 40**), showed specific uptake in vivo, but suffered from high renal accumulation and low metabolic stability.[44]
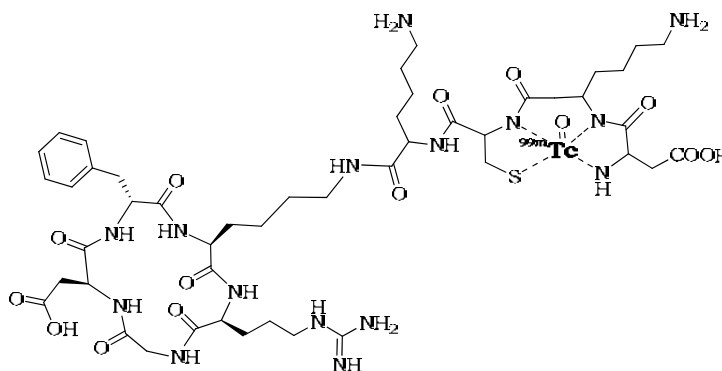
Figure 40: Technitium 99m radioactive $\alpha_v\beta_3$ integrin targeted probe ^{99m}Tc-DKCK-cRGDfK.[44]

There is ample evidence for the potential of macrocyclic chelation of other radio metals such as [111]In (SPECT); ^{68}Ga, ^{64}Cu, [86]Y (PET), and their subsequent conjugation to RDG peptides. Also reported is the effective incorporation of the excellent ^{18}F-galactose derivative (**figure 41**).[39] Other analogues of RGD rather than cyclic pentapeptides have also been successfully explored for $\alpha_v\beta_3$ targeting, including multimeric ^{18}F appended species where improved tumour accumulation and integrin binding affinity was observed upon multimerisation. Increased stability was observed using D-lysine to join the cyclic peptide units through PEG linkers.[45]
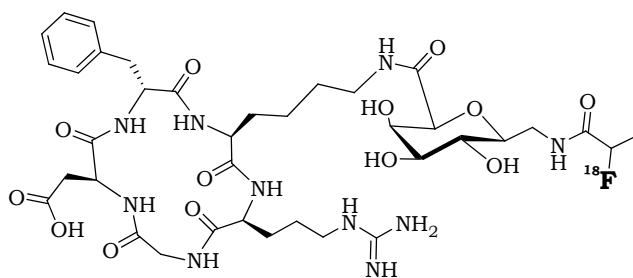
Figure 41: [^{18}F] Galacto-RGD. The 'gold standard' for clinical PET imaging of $\alpha_v\beta_3$ integrin expression. The galacto moiety is to increase hydrophilicity. In comparison to directly fluorinated cyclic RGD peptides and ^{125}I radiolabelled analogues, this galacto derivative shows superior tumour uptake and impressive PET contrast.[39]

The laboratory of Chen developed a series of radiolabelled cyclic peptide ligands to target the integrin $\alpha_v\beta_3$, to investigate its angiogenic function, and direct the development of a successful antiangiogenic therapy based on integrin antagonism.[46] Chen also investigated PET diagnostic probes using radiotracers. This involved the use of both ^{64}Cu and ^{125}I radioisotopes appended to cyclic RGDyK (y = D-tyrosine).[46-48] Of particular interest in the ^{125}I study was the fact that the introduction of a polyethylene glycol (PEG) unit onto the cyclic peptide prior to labelling with ^{125}I enhanced retention in the tumour, showed lower kidney uptake and improved the rate of blood clearance.[48] The ^{64}Cu study was similarly enhanced by a long PEG spacer unit and with a half life of 12.8 hours the probe was shown to bind and show good tumour specificity up to four hours post injection (**Figure 42**).[47]
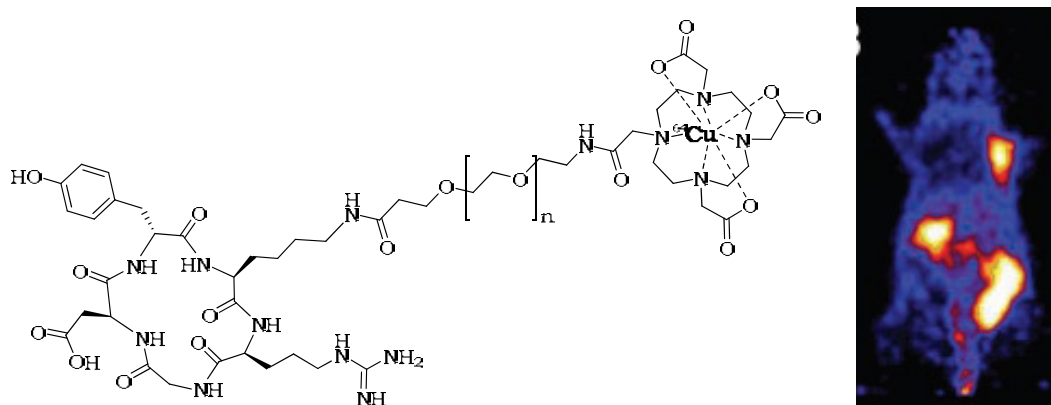
Figure 42: The c(RGDyK)-PEG-DOTA[64]Cu probe and a PET image showing tumour-specific accumulation.[49]

2.3.7 RGD & the near-IR (NIR)

Whilst the construction, binding affinity, metabolic stability and tumour specificity of the RGD section of the probe is of great importance, the primary interest and concern of this project is the use of RGD in optical imaging. Near-IR optical imaging is gaining increasing importance for powerful preclinical disease study and operative endoscopic fluorescence imaging. The beauty of near-IR fluorescence is that biological absorbance is minimal and the excitation wavelength is out of range to incur the problematic biological chromophoric interference. Therefore photon tissue penetration of 7-14 cm is theoretically adequate for a whole host of applications.[50]

Several integrin-targeted NIR probes have been developed and there are a variety of commercially available NIR dyes. An interesting and very recent development with only one reported study to date is fluorescence mediated tomography (FMT). This technique holds great promise and involves the fluorescence emission being resolved quantitiatively, and processed to give an accurate three-dimensional image analogous to PET, but without the expensive and logistically problematic radionucleotides.[51]

The cyclic(RGDfK)-Cy5.5 (**Figure 43**) probes were used with great success in the targeted imaging of mice bearing subcutaneous human xenografts, however this is regarded with a level of scepticism as xenografts can be imaged without looking at the bodywide expression of the probe due to the tumour being grown just under the skin and the scan depth being set to the depth of the protruding tumour. It is also well known that Cy 5.5 is expensive and unstable.[52] However, the clear accumulation of the probe at the tumour site and the efficacy of the tissue penetration from the relatively low wavelength near-IR emitter Cy 5.5 ($\lambda_{em.}$ = 650 nm) bears testament to the potential of RGD probes in optical imaging.[53]
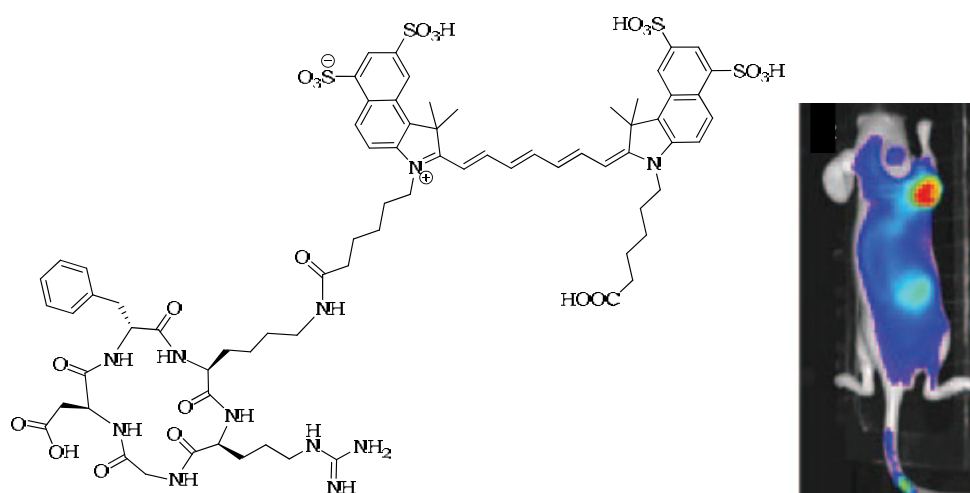
Figure 43: **Left**) c(RGDfK)-Cy5.5. **Right**) An example of the xenograft tumour image generated using a fluorescence spectrometer.[54]

RGD is not the only peptidic probe with selectivity towards $\alpha_v\beta_3$ integrin. Indeed other more complex analogues have demonstrated good binding affinity and problems associated with rapid wash out from the tumour have been circumvented through incorporation of a biotin-

streptavidin construct.[39] Indeed, slightly altered amino acid sequences such as in the sophisticated fluorescent ligand cypate-GRDSPK, have shown highly specific uptake and concentration within $\alpha_v\beta_3$ expressing tumours. Interestingly in this instance the binding was found to be mediated exclusively through interaction with the β_3 subunit.[55]

Although the RGD interaction with $\alpha_v\beta_3$ integrin has been quite thoroughly investigated for PET and SPECT imaging purposes, its potential has yet to be fully realised in fluorescence imaging and so remains a very active field of research to which we will add.

2.4 Molecular imaging & multimodality on the nanoscale

The development of emissive optical probes is becoming essential to the cellular and biochemistry fields not so much for diagnostic imaging, but more to glean information about the structure and biology of specific physiological functions. It has become apparent that the emitted photons carry information in their energy, lifetime and polarisation which can be processed and interpreted into valuable information about the local cellular environment including the concentration of anionic metabolites and cationic species such as Ca^{2+} & Zn^{2+}, and pH information, depending on the type of probe.[39]

Molecular imaging is defined as the characterisation and measurement of biological processes at the molecular level.[56] Whereas conventional imaging techniques detect anatomical and functional changes, they provide little or no information on specific molecular biomarkers such as growth factor receptors, protein kinases and cell adhesion molecules which are cancer specific.[56,57] Detection of these biomarkers allows earlier diagnosis, prompt treatment and improved prognosis. Optical imaging is a type of molecular imaging and multiple probes with different spectral character could be used for multichannel imaging. The major drawback is the poor tissue penetration of light and hence this technology may be limited to more shallow monitoring and diagnostics, endoscopy and operative visualisation. However, if the probe, for instance, has both optical and PET active character, then there are clear benefits in that pre surgical pinpointing can be achieved through the intramolecular PET active positron emitting radioisotope, and then during the subsequent operative excision of the tumour, the optical component can be employed to afford tremendous accuracy.
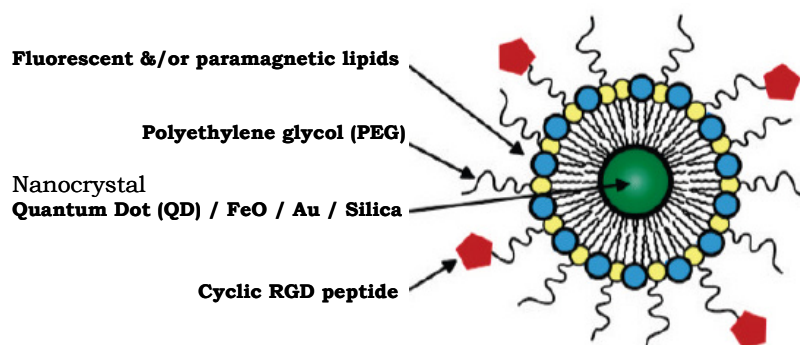
Figure 44: An example of a paramagnetic and fluorescent lisosome nanoparticle for target-specific imaging. Adapted from Refs 58 & 59.

Several types of nanoparticles have demonstrated multimodal diagnostic imaging potential, especially lipid aggregates. The main focus is on lipid-based nanoparticles with MRI and optical imaging capabilities. Liposomal nanoparticles (**Figure 44**) are flexible platforms that allow the incorporation of fluorescent molecules such as rhodamine and Cy5.5, and chelated metals such as the paramagnetic Gd^{3+}. Reported studies have shown great promise and the size of the spherical nanoparticles and quantum dots (QDs) allows multiple sites for the three components; the RGD peptide, the fluorescent marker and the paramagnetic MRI contrast metal complex, thus enhancing the signal concentration.[58]

What must be remembered is that conventional organic dyes such as rhodamine and fluorescein are prone to photobleaching which renders them unstable and limits their storage time and efficient excitation. Quantum dots are controversial due to their inherent toxicity and chemical instability;[118] thus the benefits of photo- and chemically stable near-IR dyes with their long wavelength excitation and emission, makes them an even more attractive prospect.

2.6 A summary of the cancer biology discussed.

The biology and causes of oncogenesis are bewilderingly complex, indeed if it were fully understood we wouldn't be where we are today. However, a clear picture of the many processes and entities involved continues to emerge. The big problem with most therapeutics is the lack of disease-cell specificity, but the targeting and inhibition of many specific contributory parties continues apace. However, the purpose of this research is not therapeutics and inhibition, but diagnostics, and more precisely the development of synthetic strategies towards targeted powerful optical fluorescence imaging agents.

The intention of the previous pages, whilst admittedly a chemist's eye view of biology, is to offer the reader an understanding of some of the important mechanisms in the propagation and spread of tumours; and through this to home in on some of the principal entities on a cellular level. It is through an understanding of the mechanism of action of the integrin discussed that we highlighted a tumour-specific target.

It is hoped that the attention afforded to the $\alpha_v\beta_3$ integrin was enough to convince that this cellular surface protein is prolific around tumours due to its enhanced expression on the endothelial and other cells that are fundamental to the growth and spread of tumours through angiogenesis.

The RGD tripeptide sequence will be the delivery vehicle in this study and it will be manipulated to incorporate a lanthanide metal centred luminescence system without interfering with the integrin binding site.

2.6 The original synthetic plan

The proposed synthesis was broken down retrosynthetically by simply deciding that the coupling of the fully deprotected cyclic peptide to the fluorescent probe should, if possible, be the final step of the synthesis (**Scheme 4**). This was due to totally different synthetic protocols being devised for the two moieties, and also due to the risk of complications in further synthetic steps being enhanced with the larger more functionally complex molecule.
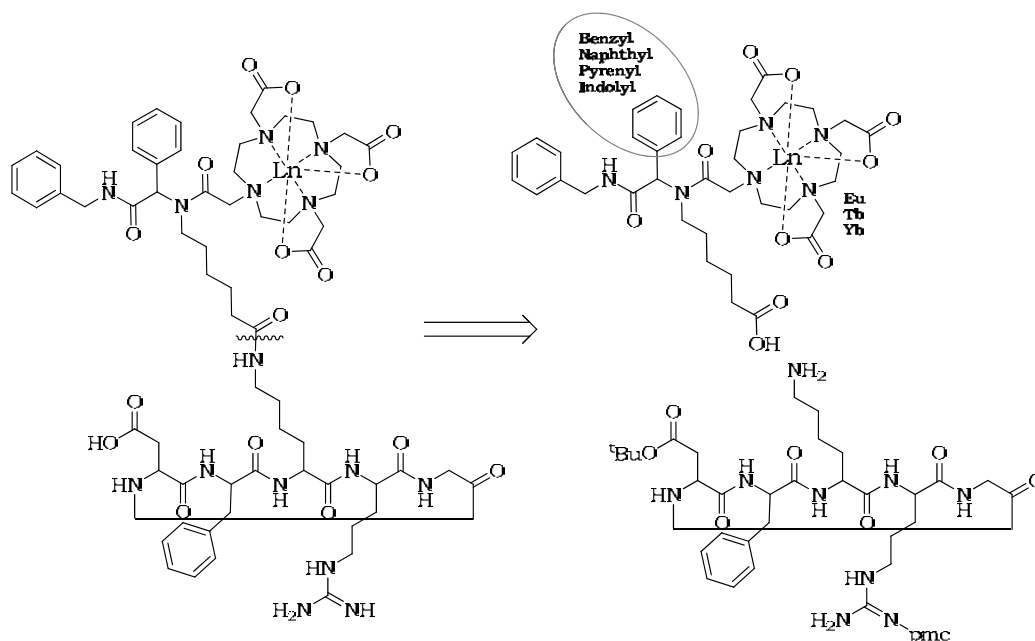
Scheme 4: **Left**) The complete luminescent lanthanide probe. **Right**) The key disconnection revealing the protected cyclic peptide cDfKRG below, and the lumophore above.

2.6.1 Synthesis of the sensitised luminescent lanthanide complex

The initial section of the synthetic strategy involved discovering the most efficient energy transfer combinations of lanthanide metal and intramolecular chromophore. The Ugi amine functionality is where, further downstream, we intend to install an attachment point to the peptide vector, therefore initially a bulky model amine will be employed to discern the optimal conditions and route to the luminescent products.

The backbone of the luminescent system will be achieved through an Ugi 4-CCR. It is unknown at this stage the exact conditions which will be required as the reaction components are relatively bulky. The reaction however will be carried out in a polar protic solvent such as EtOH as this type of solvent has been found to promote these condensation reactions.[60] The other components of the Ugi reactions are either commercially available (the aldehydes and model amine), or, in the case of the carboxylic acid macrocycle and isocyanide would be accessed through a number of steps.
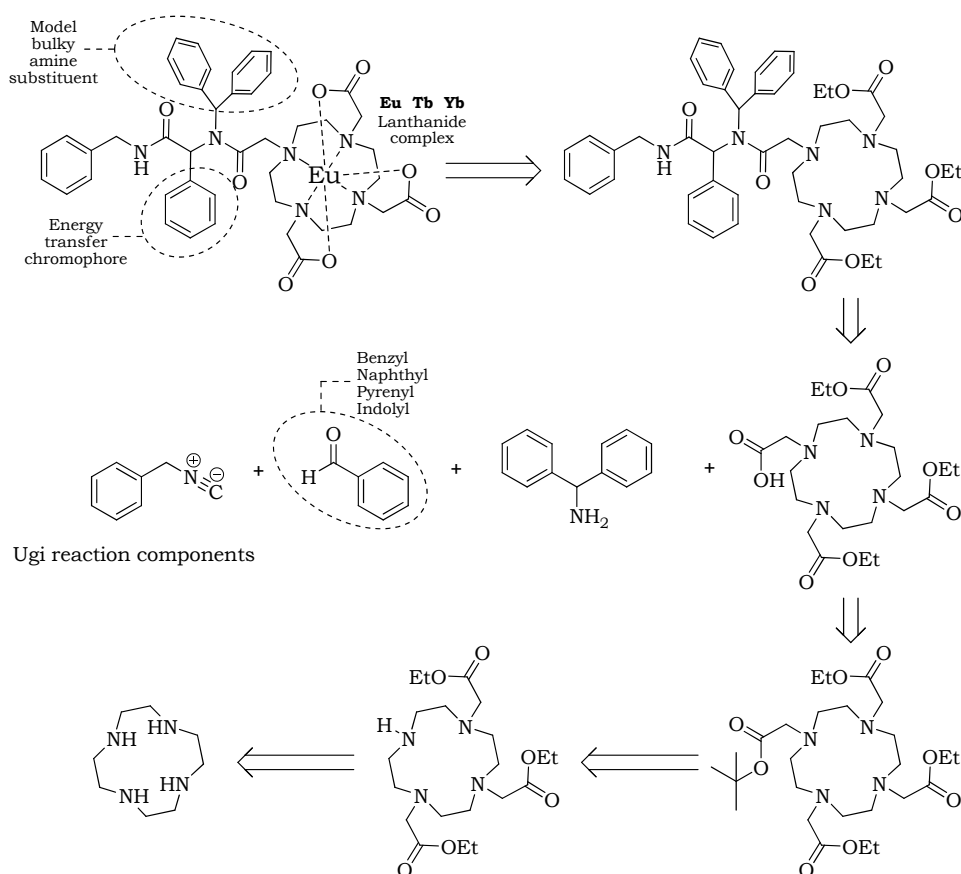
Scheme 5: Retrosynthetic analysis of the model compounds.

The carboxylic acid complex will be derived from the tetraazadodecyl macrocycle, cyclen. There is plenty of literature precedent for the selective functionalisation of cyclen through alkylations with bromoacetates under basic conditions. However there does not seem to be a consistent method of selectively forming the tris-functionalised product.

Literature conditions will be attempted and streamlined to afford a reliable and high yielding route to a selectively tris-alkylated cyclen, using ethyl bromoacetate. The remaining secondary amine will then be alkylated with tert-butyl bromoacetate prior to selective ^{t}Bu ester cleavage of the same ester to reveal the mono-carboxylic acid ready for the Ugi reaction (**Scheme 5**).

The isocyanide species will be a simple benzyl isocyanide which will be synthesised through a simple condensation between ethyl formate and benzylamine to provide the formamide which will then be dehydrated to the isocyanide.

After successful completion and characterisation of the Ugi 4-CCR products, there are just two remaining steps to complete the synthesis. Basic hydrolysis of the ethyl esters with sodium, potassium or lithium hydroxide followed by lanthanide complexation from the trifluoromethanesulfonate, citrate or chloride salt.

These final compounds will then be tested for their potential as the energy transfer luminescent probe of choice. Reactions thereafter will then employ the effective sensitised luminescent combinations in a system where an attachment arm is incorporated into the

system in place of the bulky diphenyl methyl amine to allow attachment to the cyclic peptide vector.

The spacer units of choice are ethylene glycols as it is thought likely that this will enhance the water solubility of the luminescence probe (**Scheme 6**). This may or may not be necessary however, and a simple alkyl spacer such as amino caproic acid could be swiftly made into a suitable Ugi amine through acid catalysed esterification of the carboxylic acid. The ethylene glycol amine would be created through several well-precedented functional group interconversion steps from the diol.
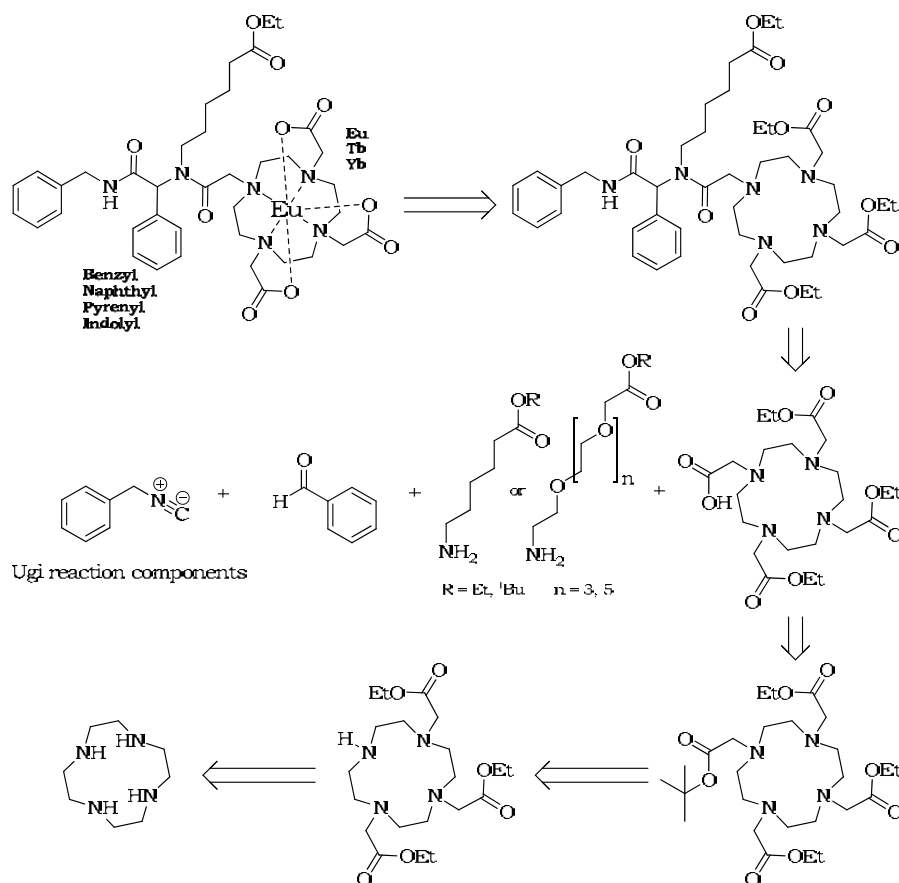
Scheme 6: Retrosynthetic analysis of the final compounds.

2.6.2 Solid phase peptide synthesis (SPPS)

The cyclic peptide would be constructed on a solid support using well known and highly efficient Fmoc SPPS. Since Merrifield first devised solid phase chemistry,[61] this technique has revolutionised peptide chemistry and the variety of different solid supports, protecting groups and coupling reagents now available is diverse and enables thorough experimentation and optimisation.

Of importance to the solid-phase construction of our pentapeptide is the choice of resin. The lysine protecting group which must be selectively cleavable. The precedent we shall endeavour to follow as much as possible is a route to the same cyclic peptide using a palladium-cleaved allylic lysine protecting group (Alloc), and 2-chlorotrityl resin as a highly acid labile solid support. However, as Fmoc-Lys(IvDde) is more readily available and as

IvDde can be selectively cleaved using hydroxylamine, this will be employed as our lysine protecting group.

Upon successful completion and characterisation of the linear penta-peptide, DfKRD, the protecting group strategy will then determine the extent to which derivatisation of the peptide through the lysine side chain can be carried out on the solid phase without compromising steps further on in the synthesis such as the fundamental cyclisation.
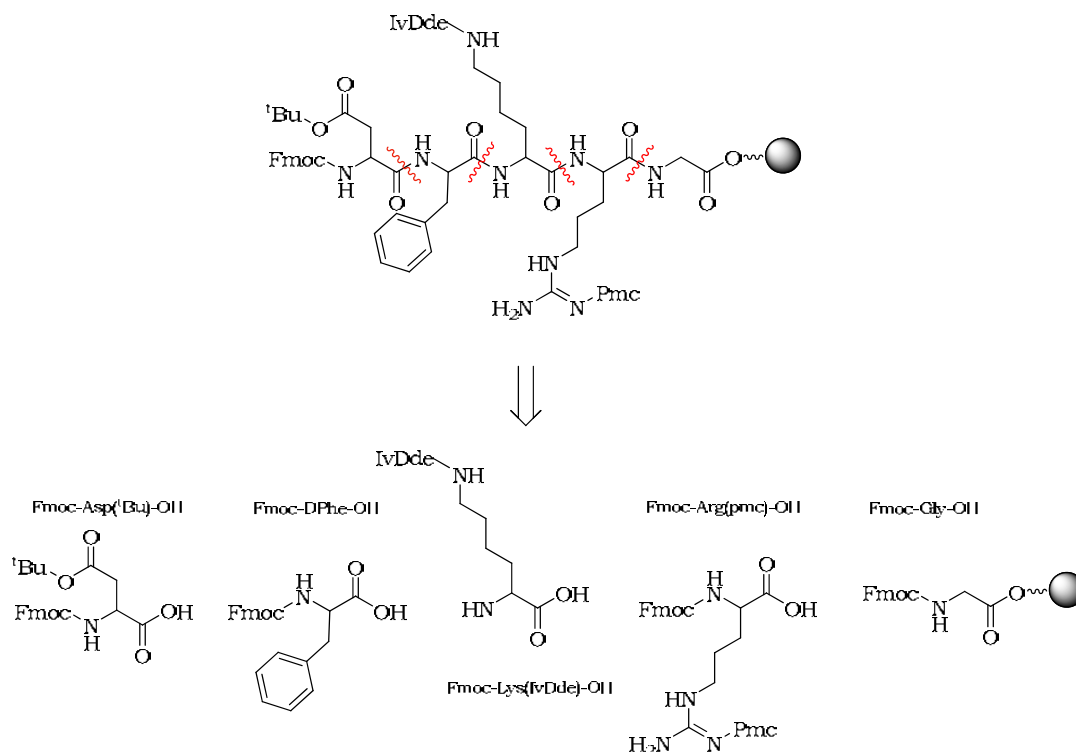
Scheme 7: Retrosynthetic analysis of the SPPS of DfKRG.

Protecting group strategy is fundamental to successful solid phase peptide synthesis. This must include consideration of the reagents required for the cleavage of the linear resin-bound peptide from the solid support.

The plan centres around the ability to selectively cleave the lysine protecting group on or indeed off the resin for derivatisation with the luminescent lanthanide probe, prior to simultaneous concentrated acid removal of the arginine and aspartic acid protecting groups (Pmc or Pbf, and ^{t}Bu), as the final step. This is not ideal as the luminescent section must be acid stable, and therefore the extent to which elaboration of the lysine free amine side chain is possible on the solid phase will be investigated. Should any subsequent steps such as the cyclisation fail, this strategy still allows the cleavage of the linear peptide from the resin and solution phase cyclisation prior to selective lysine deprotection and derivatisation.

The cleavage of the linear peptide from the resin must be selective as the cyclisation step requires that all other side chains remain protected so as to not interfere with this potentially tricky step. Thus a 2-chlorotrityl resin will be employed as this is highly acid

labile and will cleave with dilute TFA (1%) affording the linear peptide with just the carboxylic acid on the glycine terminus, and the free amine on the aspartic acid terminus, without affecting the Pmc or ᵗBu groups (**Scheme 8**).
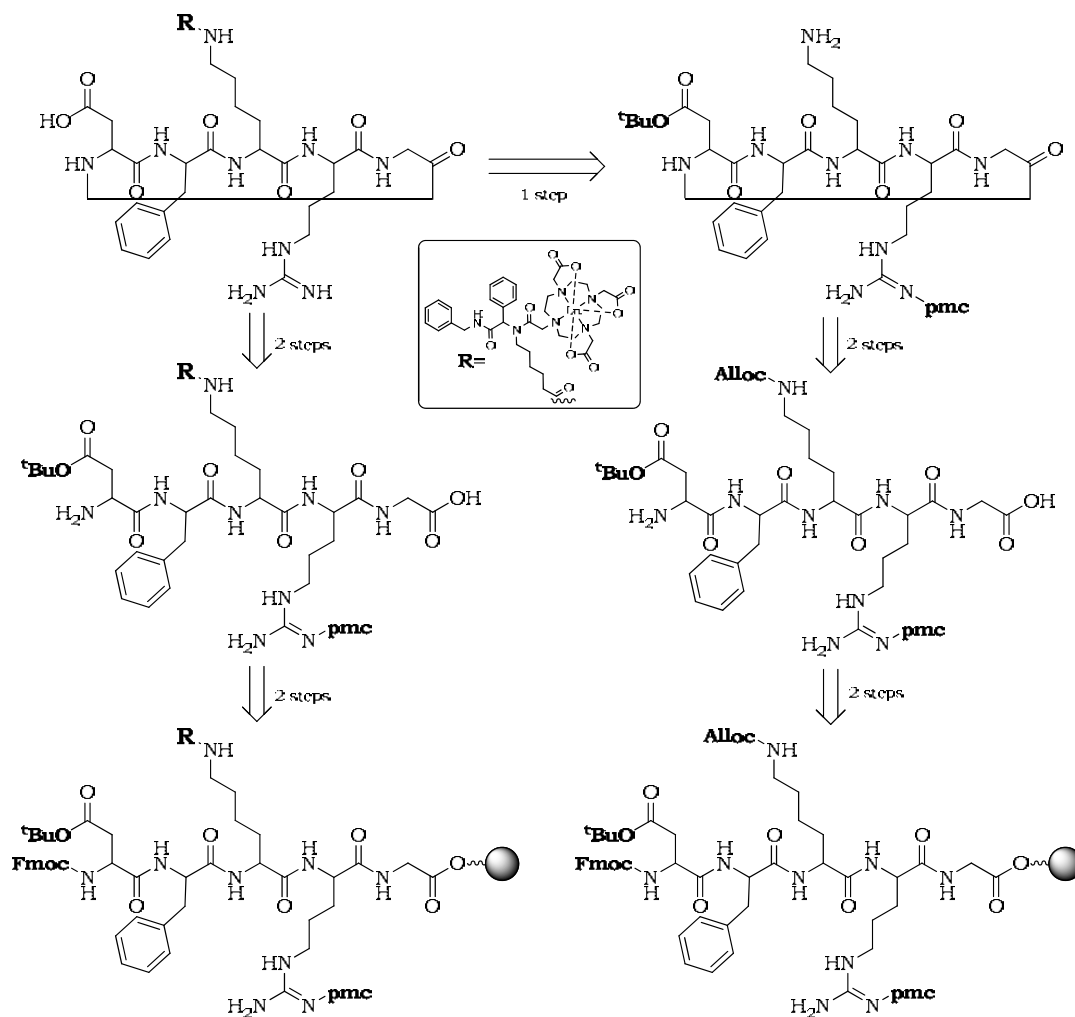
Scheme 8: The retrosynthetic routes to final labelled peptide product (**top left**).

There are many variables and deducing the most efficient reaction route will take time. Obviously with a selectively cleavable lysine side chain it would seem prudent to leave the peptide on the resin for as long as possible and attempt to couple the fluorophore to the free lysine and cleave, cyclise and deprotect as the final steps (**Scheme 8**). This would be the perfect route as just one or two final HPLC purifications would be required.

Indeed another even more efficient route to the final labelled product will be investigated and that is the possibility of coupling a spacer unit onto the lysine on the solid phase and performing the Ugi reaction, ester hydrolysis and lanthanide complexation all on the solid phase prior to resin cleavage, peptide cyclisation and final deprotection. Given the steric bulk imparted by some of the constituents this idea does seem ambitious. Solid phase Ugi reactions using very simple substituents have been reported,[62] so the extent to which this could possibly aid the synthesis of the luminescent lanthanide complex will be investigated.

2.7 Results and discussion

2.7.1 Ugi components

As described in the introduction and synthetic plan, the emphasis in the first phase of the synthesis was to establish a robust and reliable pathway to model Ugi-derived fluorophores which at a later date through the exchange of one substituent, the Ugi amine, would provide a means for attachment of the fluorophores to biological delivery motifs for targeted imaging. Further to this, these novel model compounds would also prove which combinations of lanthanide and chromophore offer the most efficient energy transfer for enhanced lanthanide fluorescence.

2.7.1.1 Macrocycle synthesis

The opening chemistry involved the construction of the only non-commercially available component for our Ugi 4-component condensation reactions, the macrocyclic carboxylic acid **03**. Beginning from cyclen, the literature precedent for this chemistry is rich in the use of temporary amino protecting groups or sterically hindered reagents allowing selective mono-alkylation before a deprotection step and final tris-alkylation gives rise to the desired tris-functionalised products in a four-step sequence (**Scheme 9**).[63]
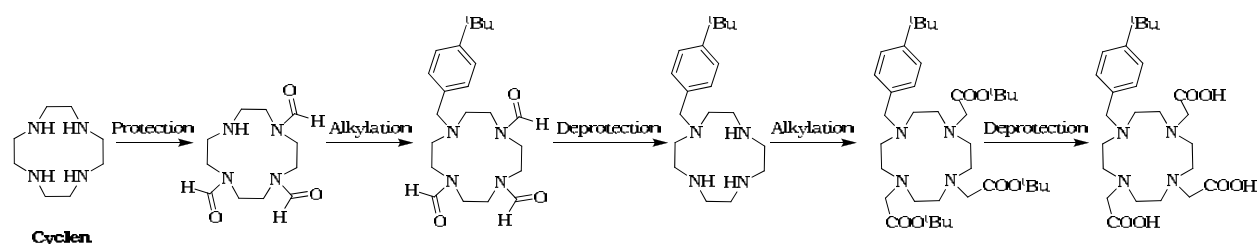
Scheme 9: An example of a lengthy route to selective functionalisation of cyclen.[64]

This seemed quite an inefficient method for accessing these tris N-alkylated cyclens, and it was envisaged that through a single step with just over three equivalents of alkylating agent we could afford the desired product. However, the precedent for this is inconsistent and unreliable.[65]

It was desirable to selectively install three ethyl esters before performing an alkylation of the final amine with tert-butyl bromoacetate, thereby allowing selective acid hydrolysis to afford the mono-carboxylic acid **03**. The ethyl esters would then be hydrolysed under basic conditions after the Ugi reaction, ready for lanthanide complexation (**Scheme 10**).
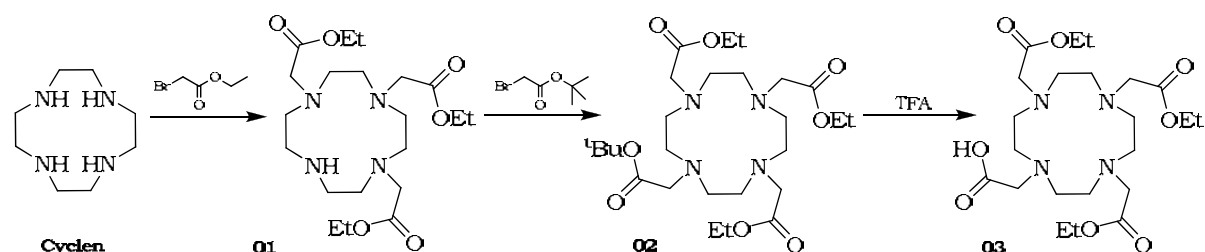
Scheme 10: The strategy for the synthesis of **03**.

Literature protocols for this first step were attempted and found to be profoundly unreliable with 20-50% yields.[63,66] The reported conditions obviously require a base and the protocols attempted combined either $NaHCO_3$, Na_2CO_3 or K_2CO_3 with acetonitrile as the solvent of choice. This yielded a combination of the bis-, tris-, and tetra-alkylated products, from which the desired thrice substituted product was apparently isolated through an unreproducible recrystallisation from hot CH_2Cl_2. The percentage of desired tris-alkylated product was difficult to judge from the crude nmr spectra, and after purification by flash column chromatography the nmr spectra ranged from poor to incomprehensible in a range of solvents. A consistent feature however, was that the resolution of the 16 central cyclen protons was not good and a broad mound covered the area from 2-4 ppm. This is a common problem in cyclen derivatives and further investigation showed that increased temperature nmr experiments increase the resolution and peak sharpness suggesting a reduced rotation issue within the molecule causing the poor resolution on the nmr timescale.[64,66]

After many fruitless attempts using different bases and purification techniques including chromatography (the products streak badly on silica TLC), it was the discovery of a pair of papers by Wong *et al* that began to shed some light on the matter.[67,68] Through studies of different solvent and base combinations it became clear that the choice of solvent is of utmost importance in this reaction. Wong reports that the reaction was attempted in MeCN without base, or with one of the following bases: pyridine, K_2CO_3 or Et_3N. The latter was found to be not only by far the highest yielding, but tris-selective when alkylation was performed using 3.5 equivalents of *tert*-butyl bromoacetate. Interestingly, with K_2CO_3 as base the outcome was unfortunately an almost equal mixture of bis- tris- and tetra alkylated products. However, weakly polar aprotic solvents such as chloroform and dichloromethane were found with different alkylating bromoacetates to perform consistently, and interestingly, to form selectively and quite specifically either the bis- or tris- products.
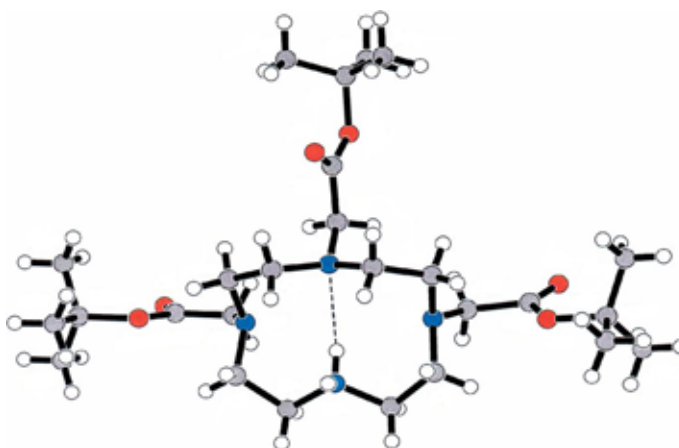
Figure 45: The X-ray crystal structure of a thrice alkylated cyclen derivative showing cross-ring H-bonding.[67]

Strongly polar aprotic solvents such as MeCN and DMF, and polar protic solvents such as MeOH were found to lead to substantial losses in yield and selectivity. Further to this, the

combination of CHCl$_3$ and Et$_3$N with quantities of electrophile from two to eight equivalents revealed that no trace of tetra-alkylated product was created even with a great excess of the electrophile. The reason for this was revealed upon examination of the crystal structure. XRD analysis showed contraction of the macrocycle from a square (3333) conformation to a 'squashed' rectangular (4242) conformation as under the reaction conditions the free 2° amine is protonated to a hydrochloride salt and a hydrogen bond is formed to the opposite amine nitrogen precluding the formation of the tetra-alkylated product (**Figure 45**).[67]

The electrophile itself was also found to be an important factor as there may be an element of steric crowding around the macrocycle if bulky electrophiles are used. In general, bromides were found to be the leaving group of choice in these alkylations. However, ethyl bromoacetate has not been reported in literature synthetic protocol, and this being our elecrophile of choice we had to rely on similar protocols for guidance.

It was reasoned that an inorganic base would be advantageous from a practical perspective as Et$_3$N leads to TLC complications (it is visible with the KMnO$_4$ stain and streaks up the plate concealing other reaction constituents). Therefore as previous literature, albeit unreliable, employed NaHCO$_3$, it was decided to first attempt a combination involving this base with CHCl$_3$ in place of MeCN.

This led to consistent 70-80% yields of the desired product. The product appears as a creamy white precipitate which is collected by filtration. We were also able to develop a gradient eluent system to very effectively purify the compound on a short flash silica column rather than relying upon the various unreliable reported recrystallisations from hot CH$_2$Cl$_2$ or 9:1 acetone/H$_2$O.[64]

The alkylation of the fourth amine was achieved quite readily following reported conditions.[67] By switching the solvent to THF and using Et$_3$N as base, the product **02** was afforded in excellent yield. The selective hydrolysis of the tert-butyl ester was performed readily under the standard concentrated CH$_2$Cl$_2$/TFA conditions, but at this stage flash column chromatography became impossible as the product was stuck fast on the TLC baseline. It was hoped that repeated co-concentrations with CH$_2$Cl$_2$ and triturations with Et$_2$O would remove all residual TFA to provide us with an adequately pure product, however the TFA proved highly tenacious leaving the product as a viscous light-brown oil. To fully characterise the product and obtain a satisfactorily well-resolved nmr spectrum an HPLC method was employed using an MeCN gradient to afford the pure carboxylic acid product **03** as a fine white powder after lyophilisation, in excellent yield (**Scheme 11**).
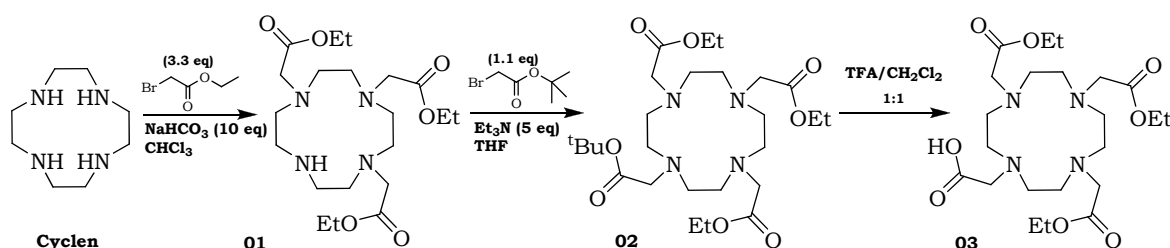
Scheme 11: The successful reagents and conditions for preparation of the Ugi carboxylic acid **03**.

2.7.1.2 Solid phase macrocycle synthesis.

Later in the project, after the Ugi phase, when exploring the possibility of attaching a gadolinium species to the peptide for MRI studies, attempts were made to construct DO3A macrocycles using solid phase methodology as we were having problems with gadolinium complexation. There is precedent for solid phase organic synthesis in the construction of DO3A type macrocycles.[69] The advantages are significant in that solid phase synthesis eliminates lengthy purification steps and can offer remarkable selectivity as one of the cyclen amines is protected through the attachment to the resin. The research group of Oliver has illustrated the versatility of solid phase organic synthesis by constructing sophisticated DOTA conjugates. [69,70]

Successful attempts were made to selectively install three different linker units on the solid phase using bromoacetic acid, para-bromomethyl benzoic acid and fmoc-aminocaproic acid (**Scheme 12**). The methodology was consistently rigorous and the products were characterised at each stage through taking a few resin beads and employing a mini-cleavage protocol to confirm the successful transformation by mass spectrometry, and measure purity by HPLC. A minor problem with solid phase synthesis is that at each step it is difficult to measure the yield as HPLC is non-quantitative, and to get an accurate measure of the yield the whole product must be cleaved from the resin thus terminating the synthesis. However, HPLC does give a good measure of purity.
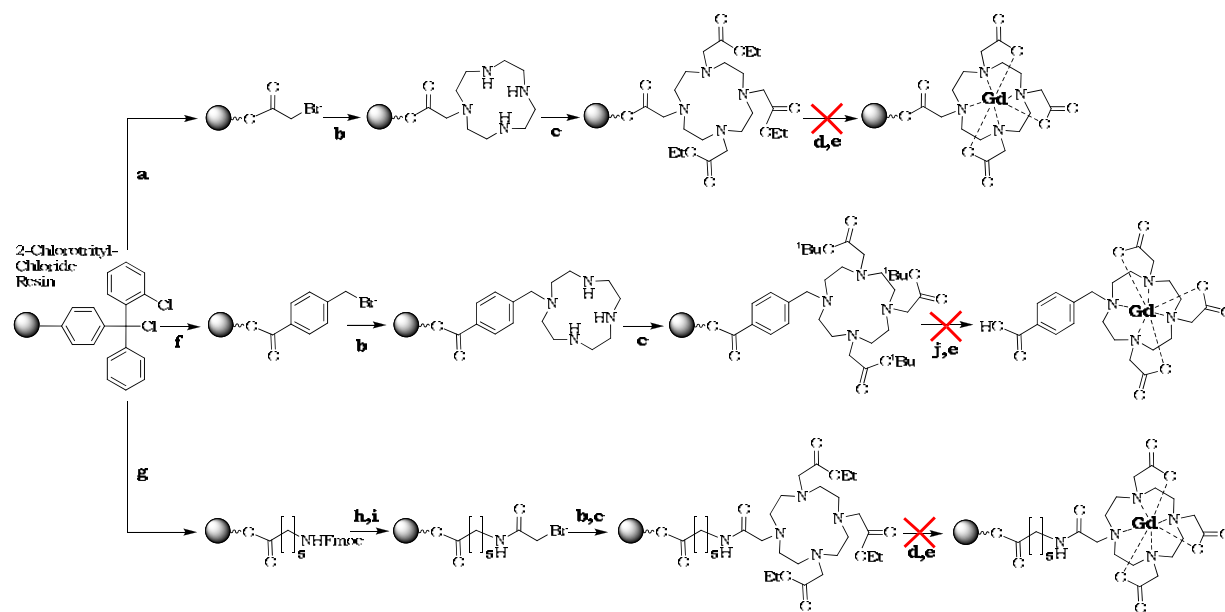
Scheme 12: The solid phase attempts at cyclen functionalisation. **a)** Bromoacetic acid (10 eq.), DIPEA, CH₂Cl₂; **b)** Cyclen (5 eq.), DIPEA, CH₂Cl₂; **c)** Ethyl or ᵗButyl bromoacetate (15 eq.), Et₃N, CH₂Cl₂; **d)** KOTMS (15 eq.), CH₂Cl₂; **e)** Gd(OTf)₃ (3 eq.), MeOH/CH₂Cl₂; **f)** p-Bromomethyl benzoic acid (10 eq.), DIPEA, CH₂Cl₂; **g)** Fmoc aminocaproic acid (5 eq.), DIPEA, CH₂Cl₂; **h)** 20% piperidine / DMF; **i)** Bromoacetyl bromide (1.2 eq.), DIPEA, DMF, -40°C; **j)** TFA / CH₂Cl₂.

Without a chromophore a molecule is often invisible to the HPLC UV detector; so a benzylic group (para-bromomethyl benzoic acid) was employed to compensate for this and also to offer a more bulky and rigid spacer unit between the macrocycle and the terminal acid which is appended to the resin.

The selective functionalisation of cyclen was remarkably efficient and in future this is the preferred route to the tris-ethylester mono-carboxylic acid product **03**. Although often a great excess of reagents are employed in solid phase synthesis, the more valuable, such as cyclen, can easily be recycled. However, the gadolinium complexation step could not be accomplished whilst the macrocycle was on the solid phase, neither, in the instance of the benzylic system where the TFA cleavage of the tert-butyl protecting group also cleaved the product from the resin, could the metal be complexed in solution.

As plenty of **01, 02** & **03** had been synthesised in the earlier solution phase synthesis, it was decided that there was no further requirement to continue this solid phase synthesis as a means of introducing gadolinium to the macrocycle. We had, however, proved its efficacy and versatility as an alternative means of constructing mono- and tri-functionalised DO3A derivatives.

2.7.2 The Ugi 4-component condensation reactions

As mentioned in the introduction and synthetic plan, the strategy behind using this multi-component reaction was to bring together the functionality required for lanthanide complexation. An intramolecular antenna chromophore to mediate energy transfer, and a means of tagging the final fluorophore to the biological delivery motif were to be united in a single step.
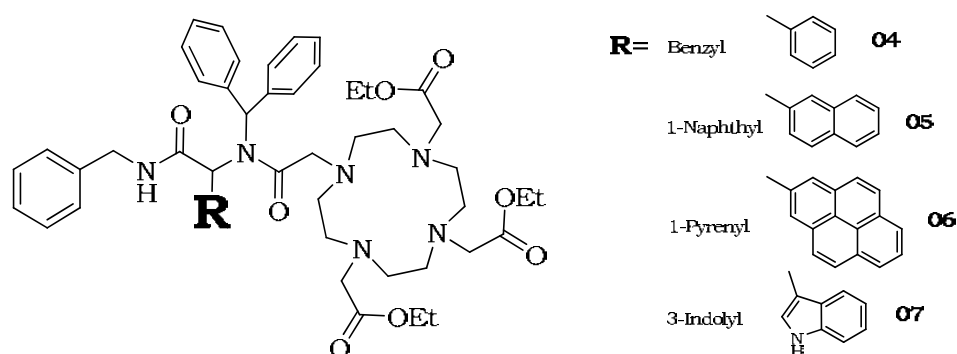
Figure 46: The Ugi products.

This Ugi multi-component reaction is versatile and is dependent upon a reactive isocyanide component for the successful outcome. However, although crucial for the reaction, the isocyanide functionality is the only section of the final compound which is redundant when it comes to the excitation and fluorescence.

The synthetic plan proposed the exploration of different combinations of chromophore and lanthanide metal to establish the systems where energy transfer was most efficient.

Therefore in this Ugi multicomponent condensation reaction the four different chromophores were introduced as the aldehyde component. The lanthanides were to be introduced and complexed in the macrocycle as the final step of the sequence.

The amine is a model bulky amine, aminodiphenylmethane, and would be exchanged for a linker unit upon the discovery of an efficient chromophore – lanthanide combination. For the purpose of increasing UV visibility in the HPLC purification protocol and latterly due to commercial availability, benzyl isocyanide was used as the isocyanide component. This was initially synthesised through dehydration of the formamide. However not only was this process decidedly unpleasant due to the high temperature vacuum distillation and the inherent volatility of the compound, but the reaction was unreliable and low yielding (30-46 %). For these reasons a commercial source was sought and it was thereafter purchased from Aldrich.
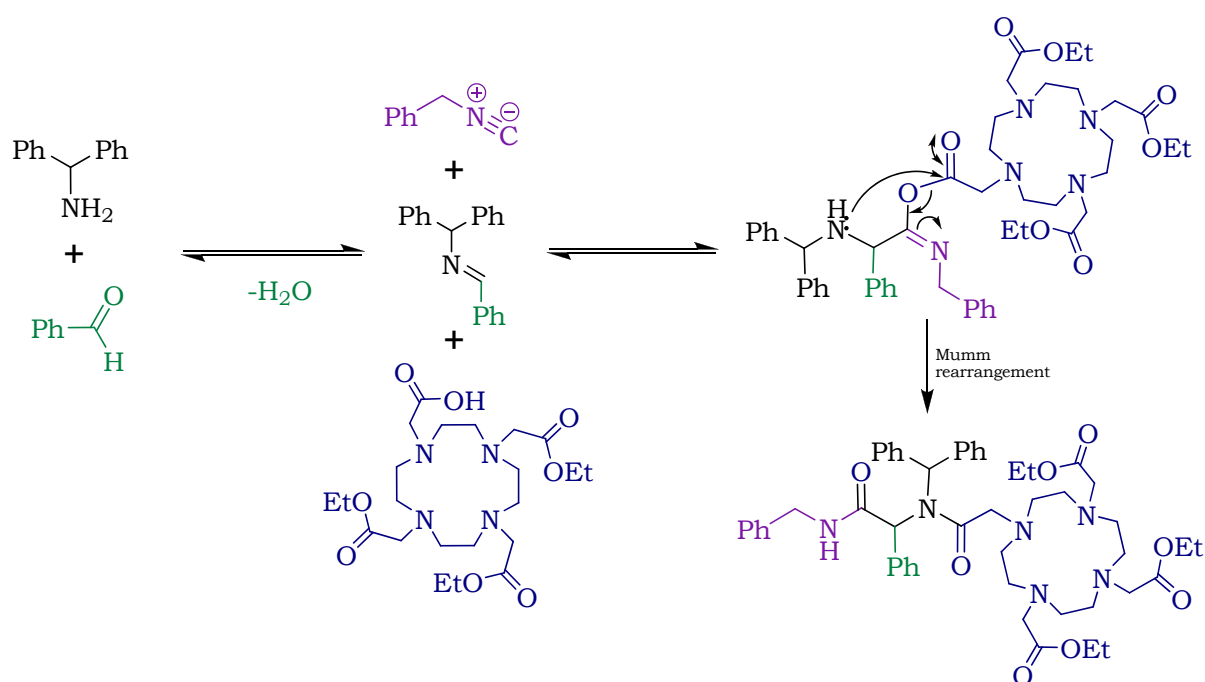
Scheme 13: A proposed mechanism of the Ugi 4-CCR reaction showing the Mumm rearrangement as the final step.

The Ugi reactions were performed using equimolar quantities of each of the four components and as polar protic solvents are typically employed in Ugi reactions we chose EtOH. Due to probable steric demands around the heterocyclic intermediate (**Scheme 13**), an elevated temperature of 65 °C was required. The reactions were followed clearly by TLC where our previous eluent system of $CH_2Cl_2/MeOH/NH_{3(aq)}$ 90 : 9.95 : 0.05 was effective. Both UV and $KMnO_4$ were excellent visualisation methods and over three days the four starting material spots conglomerated into a single more powerful spot. Another excellent means of assessing reaction completion was when the TLC sample no longer gave off the offensive isocyanide odour!

Purification was achieved by flash column chromatography and the resulting nmr spectra were again poorly resolved and gave no indication of purity. In an attempt to achieve improved nmr spectra, the products were further purified by reversed phase HPLC. The resultant nmr were clean but the resolution was still inadequate for conclusive characterisation.

A variable temperature nmr experiment was attempted and although improving the resolution and sharpening some of the peaks especially the aromatics and backbone protons, the outcome was not of high enough quality for unequivocal assignment (**figure 47)**. Therefore with accurate mass spectra and clean HPLC traces the pure products were taken on to the ester hydrolysis step.
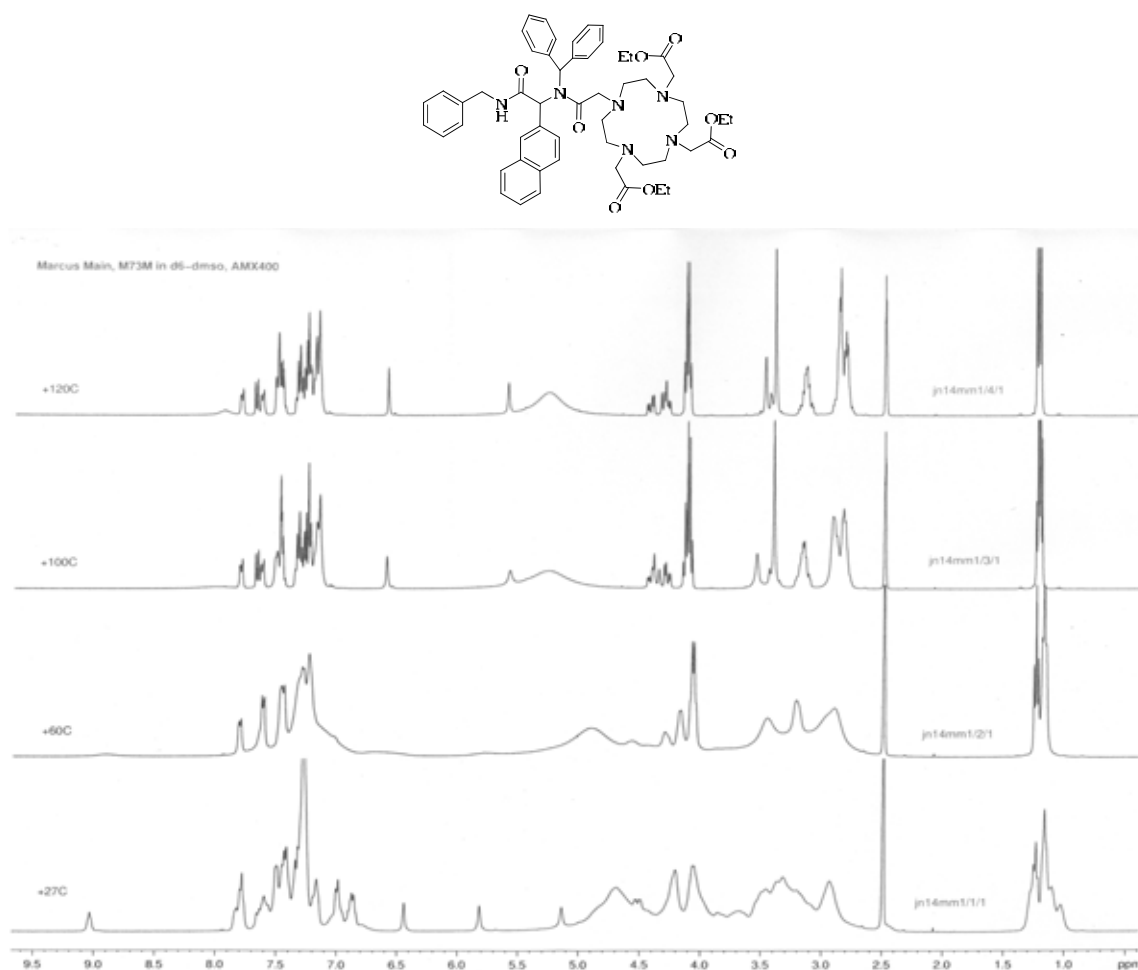
Figure 47: Variable temperature (VT) nmr data obtained for pure Ugi product **05**.

An initially somewhat confusing crystal structure of compound **05** was obtained through slow solvent evaporation (**Figure 48**). However, the confusion was found to be simply a sodium ion chelated into the macrocycle which is not an uncommon occurance in macrocycles of this type with the sodium ion thought to originate from the solvent or XRD protocol. This illustrates the susceptibility of the macrocycle to metal chelation. Thus proof of reaction was unambiguous.
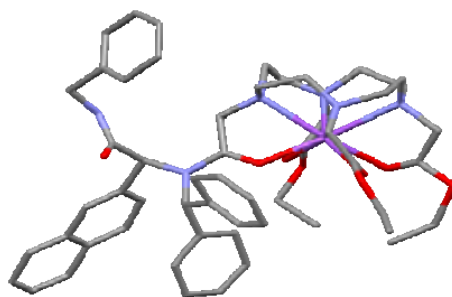
Figure 48: X-ray crystal structure of **05**. Complexed sodium ion illustrates the chelation-prone macrocycle.

2.7.2.1 Ethyl ester hydrolysis.

This step was straightforward as the products from the Ugi reaction stage were stable, clean, and afforded no solubility issues.
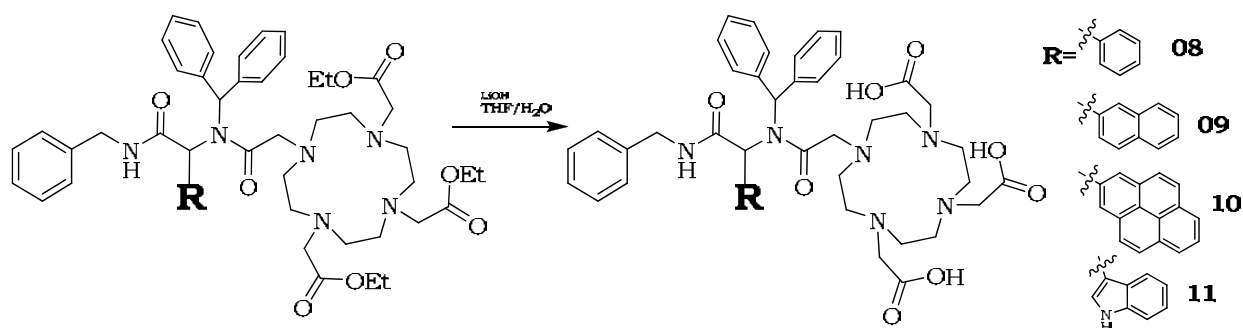
Scheme 14: Ethyl ester hydrolysis.

The conditions of choice involved use of the inorganic base lithium hydroxide in a THF/H_2O solvent mixture. Using a small excess of the base this step proved quantitative and fast. When the construction of the cyclen tris-ethyl ester mono acid was attempted on the solid phase by means of a 2-chlorotrityl resin it was found that a more organic soluble basic source was preferable and KOTMS worked perfectly in hydrolysing the ethyl esters in readiness for solid phase lanthanide complexation, in that instance with gadolinium. This complexation however either failed or we experienced the characterisation issues which was a regular drawback in our work with gadolinium.

However in this instance after a short reaction time the crude product was acidified to protonate the amines and form the zwitterions. In some instances this caused the product to precipitate out of the water (the THF was removed prior to lowering the pH), alternatively if the zwitterions did not precipitate out, the product was then purified by HPLC where now streamlined methodology was shortening run-times and affording very good yields of products as a single well defined peak.

Again the charcterisation of the free acids was impossible by nmr spectroscopy and it was left to high resolution mass spectrometry with clean HPLC traces to provide satisfactory evidence for the small quantities of products to be taken on to the final lanthanide complexation step.

2.7.3 Lanthanide complexation.
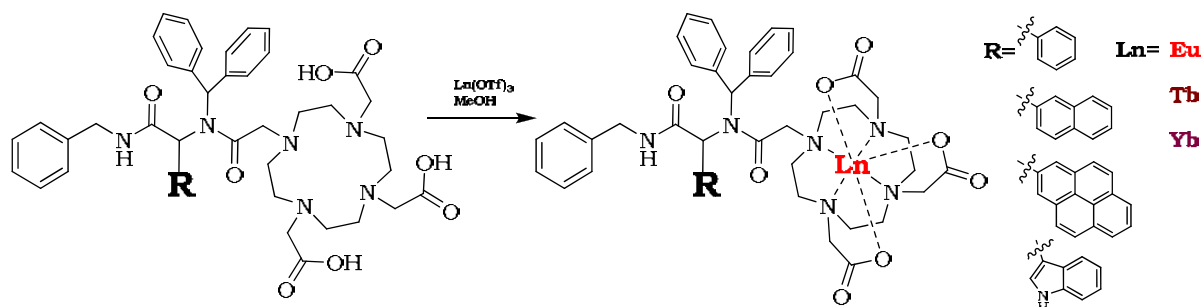
Scheme 15: The heptacoordinate Ln³⁺ complexes **12-23**. Appended water molecules have been omitted.

The three lanthanide metals chosen were rationalised in the introduction. They were Europium (Eu), Terbium (Tb), and Ytterbium (Yb) which emit long-lived phosphorescence with a luminescence lifetime of millisecond magnitude allowing timegating techniques to eradicate the interference from sample autofluorescence.

Lanthanide complexation is commonly carried out from the corresponding chloride, citrate or trifluoromethanesulfonate (triflate) salts where the metal is readily released and shows high affinity for robust macrocyclic chelation due to the multiple oxidation states and high co-ordination numbers characteristic of these extraordinary elements.

Complexation was attempted using a slight excess of the corresponding trifluoromethanesufonic acid salt in anhydrous methanol: a polar protic solvent. The reaction could be followed by analytical HPLC using a methanolic gradient method and interestingly the benzyl, naphthyl and indolyl Ugi substrates showed a pair of product peaks both giving the correct mass. The pyrenyl derivative gave just a single product peak. This will be discussed in due course.

Lanthanides are employed as nmr shift reagents. This is a phenomenon explained by the interaction of a paramagnetic metal atom with a magnetic field and the resultant spectrum is spread over several hundred ppm as opposed to the standard 12 ppm of a regular proton nmr. Nmr spectra were obtained but accurate assignments were impossible (**Figure 49**).
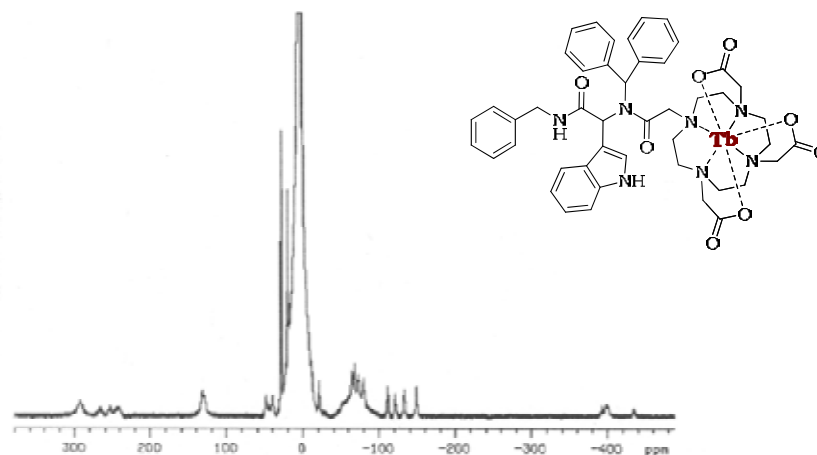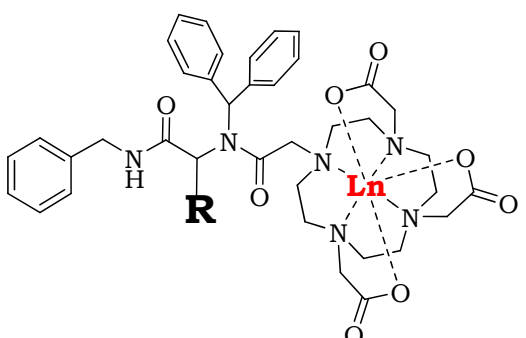
Figure 49: An example of an nmr spectrum of the Indolyl-Tb Ugi product **19**.

Once again HRMS and clean HPLC were our prime characterisation techniques but now this could be backed up by fluorescence spectra, indeed it would be this fluorescence data that would direct the synthesis of future systems where those which exhibit the greatest potential as energy transfer NIR emitting imaging probes will be taken forward.

The other nice feature which aids accurate characterisation is the beautiful isotope patterns in the mass spectra for the europium and ytterbium complexes (**Figure 51**).

This work was published in *Chemical Communications*, a copy is found at the back of this thesis.

2.7.4 The results

R		Ln	Cmpd No.	Yield %	M/z*
Benzyl		**Eu**	**12**	**86**	965
		Tb	**16**	**92**	971
		Yb	**20**	**83**	986
1-Naphthyl		**Eu**	**13**	**84**	1015
		Tb	**17**	**88**	1021
		Yb	**21**	**84**	1036
1-Pyrenyl		**Eu**	**14**	**90**	1089
		Tb	**18**	**73**	1095
		Yb	**22**	**92**	1110
3-Indolyl		**Eu**	**15**	**96**	1002
		Tb	**19**	**91**	1010
		Yb	**23**	**78**	1025

Figure 50: Table of results.* MALDI-TOF [M+Na]$^{+}$ average isotopic mass.

As explained in the publication,[71] and predicted in the introduction, it was only in certain systems that the energy transfer was actually effective in exciting the metal to enhance the emission in comparison to direct excitation of the metal itself. Energy transfer is only effective if the donor T^{1} state is of significantly higher energy than the lanthanide emissive state. Indeed it could be predicted that certain systems such as the benzyl-Yb combination would show zero energy transfer due to the disparity in energy levels. However for the completeness of the study all combinations of lanthanide and antenna chromophore were synthesised.
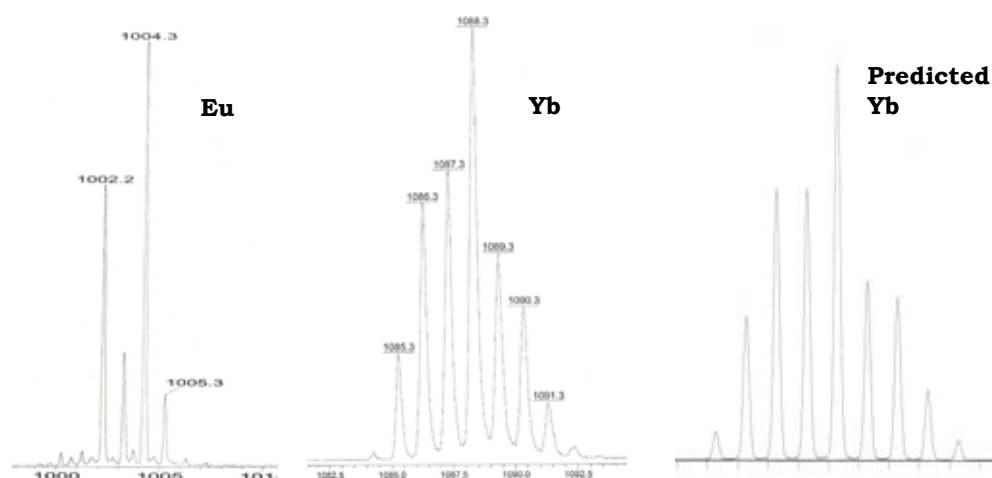
Figure 51: Examples of mass spectra illustrating the isotope patterns of Eu (**15**), Yb (**22**) and a sofware predicted Yb isotope pattern showing close correlation with the Yb pattern obtained experimentally.
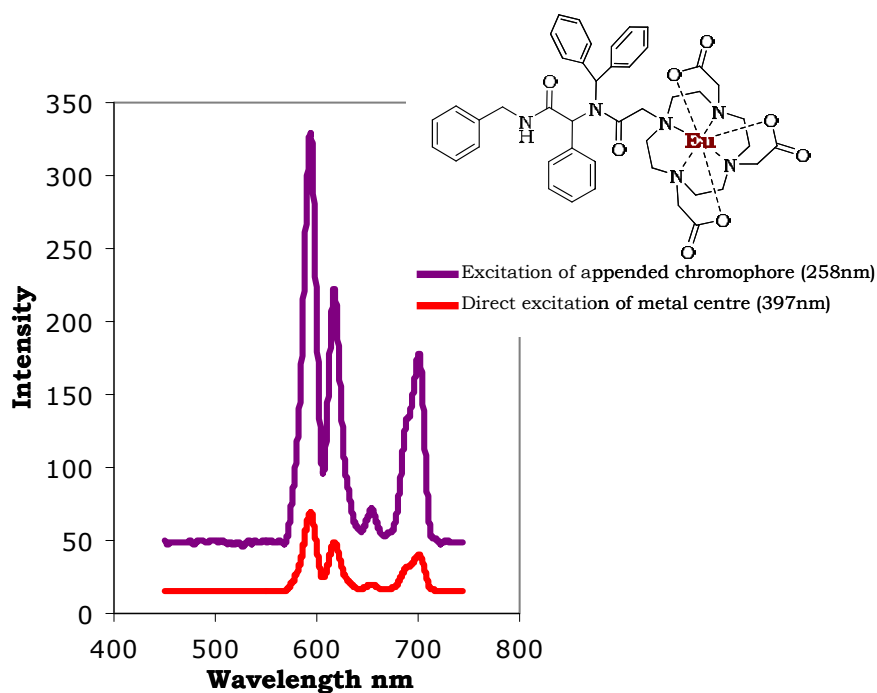
Figure 52: Fluorescence emission spectra of the Eu-benzyl (**12**) energy transfer system. See appendix 3 for fluorescence spectra.

Fluorescence spectra of Eu and Tb complexes were obtained on a standard Perkin-Elmer LS55 fluorimeter and Yb emission spectra was obtained using a pulsed Nd:YAG laser (266 nm).

Due to timegating technology eliminating background fluorescence it may initially be considered that the wavelength of the emission is irrelevant. However, it is actually of great importance as *in vivo* a further complication is biological interference. Therefore to avoid this the emission needs to be of a long enough wavelength (in the near-IR region of the spectrum) to afford the penetration required for effective detection outside of the body.

The fact that a large emissive band occurs at above 700 nm in the Eu systems is very promising as this is at the lower end of the near-IR region which we wish to expoit to achieve enhanced tissue penetration and avoid background fluorescence. Of the twelve Ugi products, seven showed positive energy transfer.

The systems showing enhanced emission, and therefore effective energy transfer were the combinations shown below (**Figure 53**).
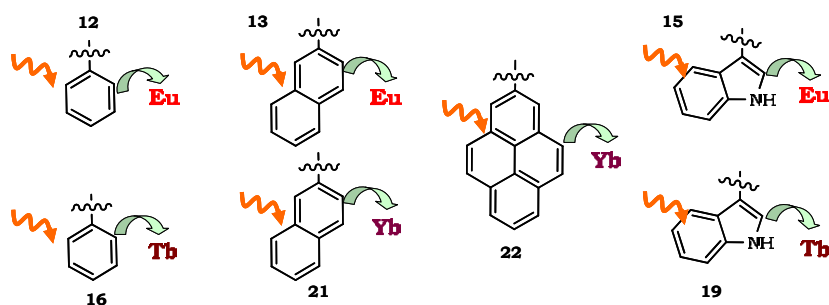
Figure 53: The effective chromophore – metal energy transfer combinations.

The rationale behind the efficacy of certain metal to chromophore combinations was discussed in detail and indeed predicted in the introduction (pages 26-27). The basic rationale being that the triplet state of the chromophore must have sufficient overlap with the singlet state of the metal but the relative energy levels must not be too close as this opens the door to back energy transfer significantly reducing the emission. This is indeed believed to be what happens with the Tb-naphthyl combination. This is all discussed further in the publication.

2.7.4.1 Characterisation

The discovery of two well separated HPLC peaks both corresponding to the correct product was interesting, and in each compound (except the pyrenyl compounds where probable steric factors prohibited the formation of isomers and the product was observed as a single peak), the two peaks were collected separately. The fluorescence spectra for the individual peaks, however, were identical and upon checking the purity of one of the collected peaks it was somewhat surprising to find that the single peak had re-equilibrated to a pair of peaks of the original ratio. The question then arose as to the rate at which re-equilibration occurs. If this is a very slow process, then the fluorescence spectra obtained a few days after purification and isolation of the separate peaks could still be the spectrum of predominantly one isomer. Whereas if the re-equilibration is rapid then the emission spectrum will be identical as it will be the equilibrated mixture of the two isomers.

To elucidate what was occuring an elegant piece of HPLC work was put into practise. We developed a rapid six minute analytical HPLC method which was programmed to record a chromatograph of a freshly collected single peak every six minutes over 72 hours with the aid of an autosampler. A semi-preparative isolation of a single peak of product was obtained and immediately the collected peak was inserted into the autosampler and the six minute cycle begun. The results are shown in **Figure 54**.
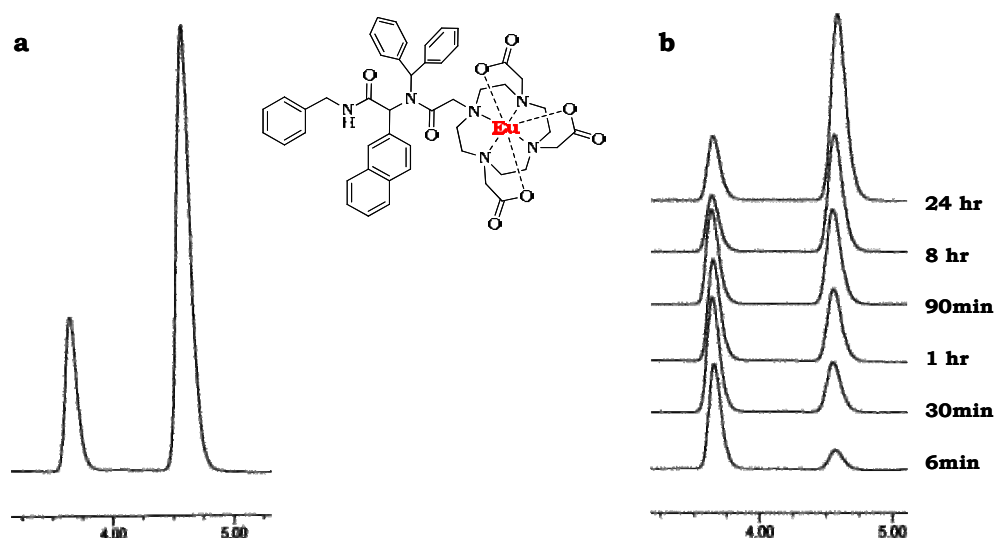
Figure 54: HPLC monitoring of the re-equilibration of the two lanthanide containing product peaks. **a**) The pure product prior to separation; **b**) the collected left hand peak re-equilibrating over 24 hours.

It became abundantly clear from this that there was absolutely no point in separating the two peaks as re-equilibration in solution begins immediately as revealed in the first trace at six minutes where the second peak has already begun to appear. Collection of the larger, longer retained peak and an identical six minute run experiment carried out revealed an identical rate of re-equilibration.

To explain the observed phenomenon we searched the literature and found that Parker had examined a similar phenomenon with a good number of quite different DOTA and DO3A ligands.[72] In lanthanide complexes of DOTA the four ethylene diamine groups adopt gauche conformations giving rise to two distinct ring conformations, λλλλ and δδδδ. In DOTA itself, bearing great symmetry and no real steric constraints, the acetate arms can also arrange in two possible ways (Λ and Δ), resulting in the appearance of a total of four stereoisomers. These adopt either a square antiprismatic (SAP), or a twisted square antiprismatic (TSAP) geometry. The key point is that these isomers may interconvert in solution through ring inversion (λλλλ ↔ δδδδ) or arm rotation (Λ ↔ Δ). This relates to our DO3A Ugi derived complex because ligand structure affects the isomeric ratio and to explain the fact that only two diastereoisomers are observed it is rationalised that the one extensively derivatised nitrogen not only renders the isomers diastereomeric, but sterically restricts the available conformations.[72] The absolute configuration of the stereogenic carbon centre will determine the helicity (Λ or Δ) of the complex and therefore the two diastereomers are most likely the ring inversion products λλλλ (SAP) and δδδδ (TSAP) (**Figure 55**).
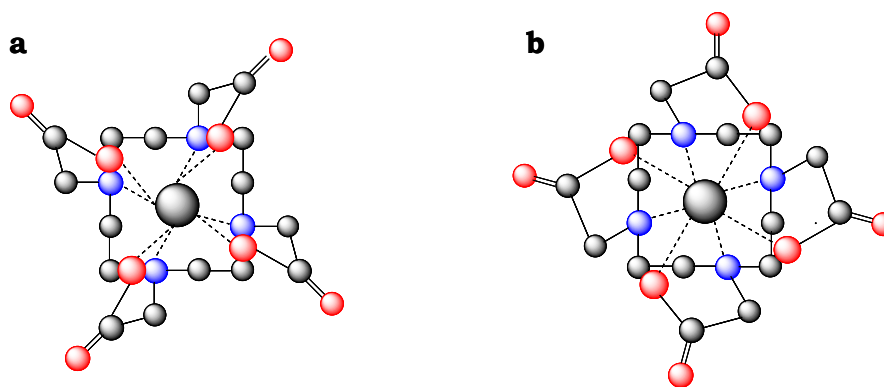
Figure 55: The two conformational diastereoisomeric products. **a**) A twisted square antiprism (TSAP); **b**) A square antiprism (SAP). Adapted from Ref. 72.

With respect to why it is not an equal mix of the two diastereoisomers (**Figure 56**), the literature suggests that the isomer ratio is dependent upon the lanthanide ion, the temperature, pressure, and the concentration of added salts. However as a general trend for larger lanthanides (La^{3+} - Nd^{3+}), the TSAP isomer dominates, whereas for smaller lanthanides (Sm^{3+} - Er^{3+}) the SAP conformation is favoured.[72,73]

This would suggest that the larger, more abundant right hand HPLC peak is the square antiprismic isomer.

Complex	Retention times T_r (min)*		Peak Area	
	A	B	A (%)	B (%)
Benzyl Eu	37.5	40.7	32	68
Benzyl Tb	37.2	40.5	39	61
Benzyl Yb	36.6	40.3	25	75
Naphthyl Eu	35.0	37.8	43	57
Naphthyl Tb	38.7	41.6	23	77
Naphthyl Yb	35.5	39.6	18	82
Pyrenyl Eu	-	41.6	0	100
Pyrenyl Tb	-	41.7	0	100
Pyrenyl Yb	-	41.7	0	100
Indolyl Eu	38.0	41.0	20	80
Indolyl Tb	38.2	41.1	27	73
Indolyl Yb	38.1	40.8	26	74

*HPLC method D
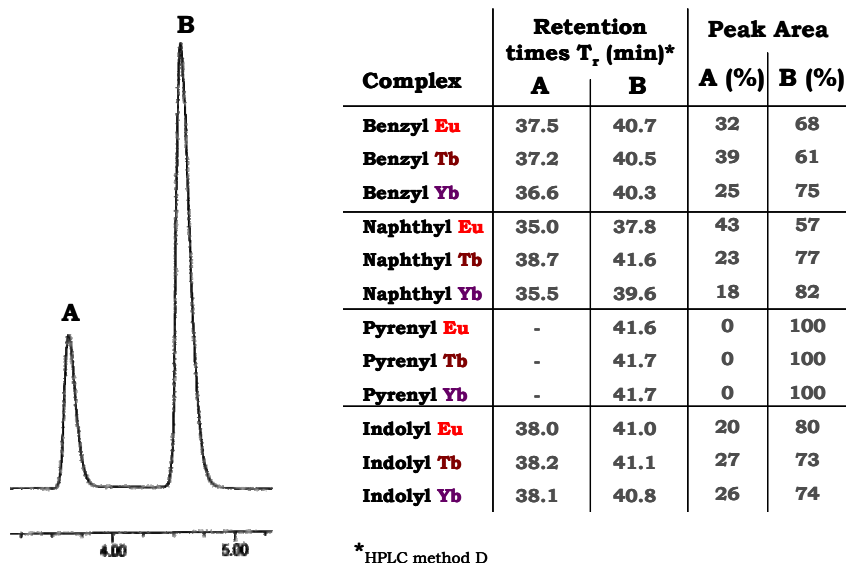
Figure 56: The HPLC ratios of peaks A:B at equilibrium for each of the final products.

Crystal structures

The final item of discussion is the crystal structural evidence obtained to unequivocally prove the success of the synthesis. Unfortunately crystals were quite challenging to grow and the only set of sufficient quality and size were the three benzyl products (**Figure 57**).
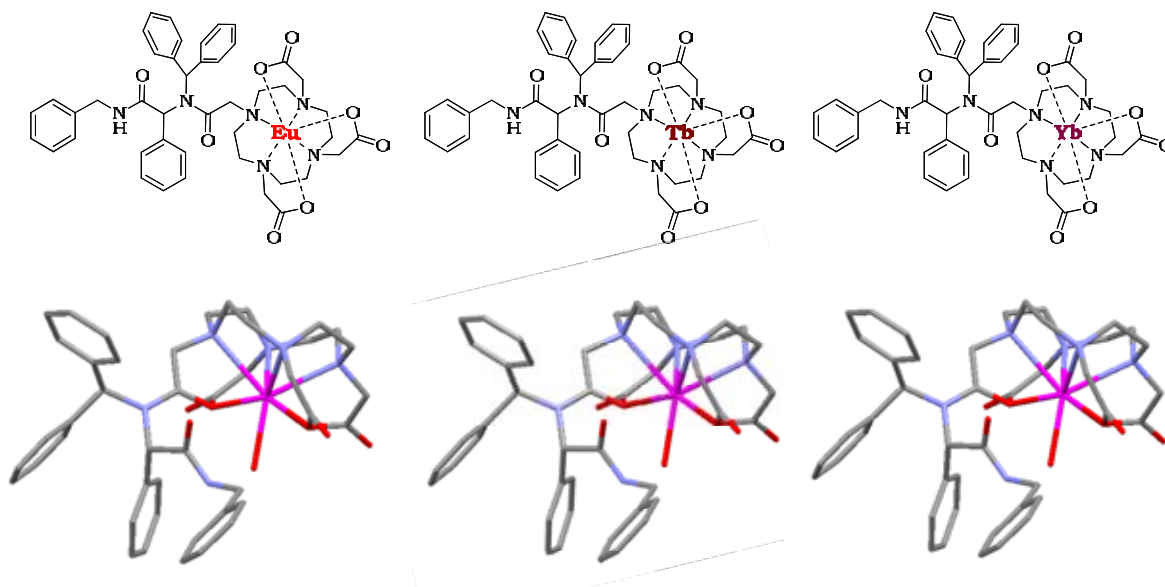
Figure 57: The three benzyl crystal structures.

From these crystals the two diastereoisomers are not isolated separately, the crystal structures are thought to be of the square antiprismatic conformations. Neither were we able to compare the different steric impacts of the four chromophores as we were only able to grow crystals of the benzylic derivatives. However the three structures, although seemingly identical, do show some subtle variations which are of interest. This arises purely because the lanthanide atoms are of different diameters with the Tb^{3+} having a smaller ionic radius than Eu^{3+}, with Yb^{3+} being smaller still due to lanthanide contraction. The crystal

63

structures show a decrease in the length of all ligand to metal bonds upon moving from Eu to Tb to Yb. These slight changes in bondlength (approx. 0.02-0.03 Å) reveal an increased tightness in the chelation of the metal upon reduction in metal radii.

2.7.5 Model study summary

To summarise this first area of study we have finetuned a reliable and high yielding synthetic route to complex luminescent lanthanide systems which exhibit intramolecular energy transfer from an antenna chromophore leading to long-lived, long wavelength emission. The ideal chromophore-metal combination was discussed and it was decided to proceed and develop this chemistry to incorporate a spacer unit with which to attach the terbium to benzyl or indolyl systems. These two systems were decided upon as they exhibited the greatest energy transfer and hence emission intensity.

2.7.6 Installation of a spacer unit
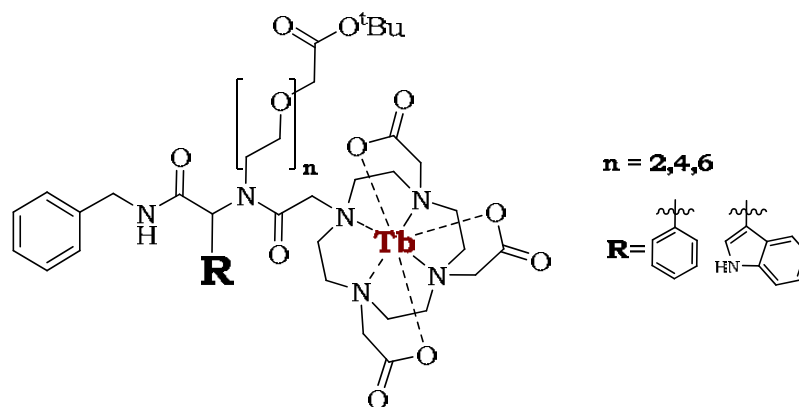
Figure 58: The initial target compounds.

Essentially the chemistry required to modify the Ugi product in a fashion which allows attachment of the luminescent system to a biological motif such as the cyclic peptides and carbohydrates (to be discussed forthwith), is identical to the chemistry developed in the synthesis of the previous model systems where the bulky diphenylmethyl amine was the amine substituent.

2.7.6.1 The choice and synthesis of linkers

Various trial Ugi 4-component condensation reactions were attempted with benzyl isocyanide, aldehyde and carboxylic acid components with a selection of spacers. It was decided to attempt the Ugi reaction initially with an oligoethylene glycol spacer unit to enhance the aqueous solubility of the probe. Thus oligoethylene glycols of two lengths were created through a number of functional group transformations to afford a free amine on one terminus and a protected acid on the other (**Scheme 16**).
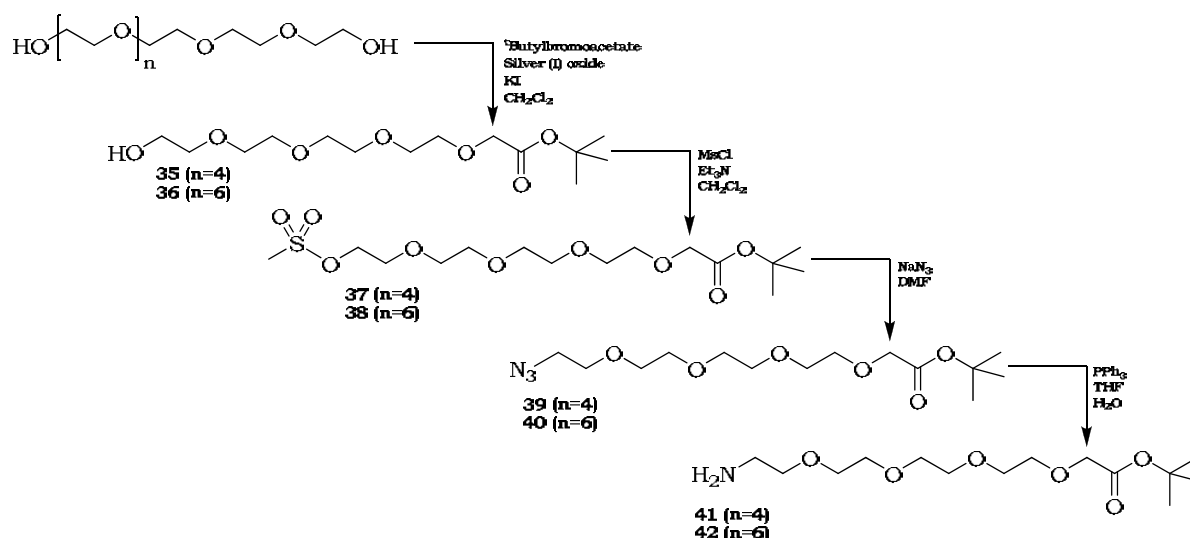
Scheme 16: The functional group transformations involved in the synthesis of ethylene glycol spacer units.

The mono-alkylation of one end of the ethylene glycol to form *tert*-butyl esters **35** and **36** was expected to be a less than 50 % yielding step as it is envisaged that the reaction must result in the formation of disubstituted product. However this was not the case and an excellent yield of monoprotected glycol was obtained through the use of silver oxide. It is thought that the mechanism involves one end of the ethylene glycol coordinating to the surface of the silver atom allowing regioselective alkylation at the other terminus.[74] The free terminal hydroxyl could then be manipulated using well known chemistry. Mesylation was achieved using mesylchloride with triethylamine as base to form a good leaving group which in the next step was substituted for an azide by simply heating the mesylate with sodium azide in DMF. To finally afford the free primary amine for Ugi chemistry, the azide was reduced to the free amine using the Staudinger reaction (triphenylphosphine in dry THF followed by quenching with H_2O). The product was then purified by careful flash column chromatography using a $CH_2Cl_2/MeOH$ gradient to ensure all of the tenacious triphenylphosphine oxide was removed prior to recovery of product with mild base added to the eluent to mobilise the product. Test reactions proved successful with the formation of **91** in reasonable yield (**Figure 59**).
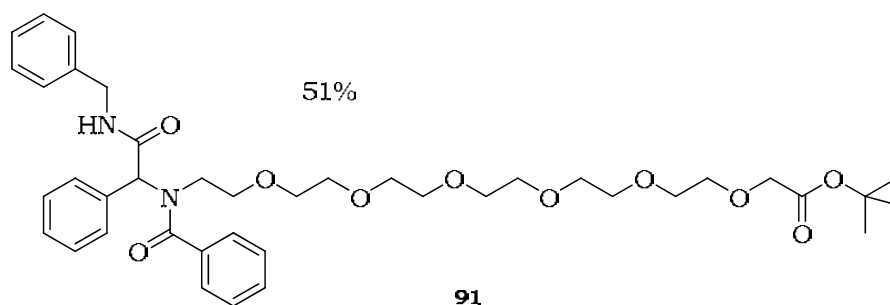
Figure 59: A successful 4-CCR test reaction product **91**.

Unfortunately however this proved completely unproductive as repeated attempts at the Ugi reaction using the components required for our energy transfer system failed. TLC analysis initially looked promising with the formation of the imine, but despite adjusting the conditions the only product consistently isolated was the Passerini reaction product (**Scheme 17**). The Passerini reaction is a three-component reaction which does not involve an amine component. We are still unsure as to what a plausible explanation for this failure may be, especially as the chemistry proved successful in the test reactions.
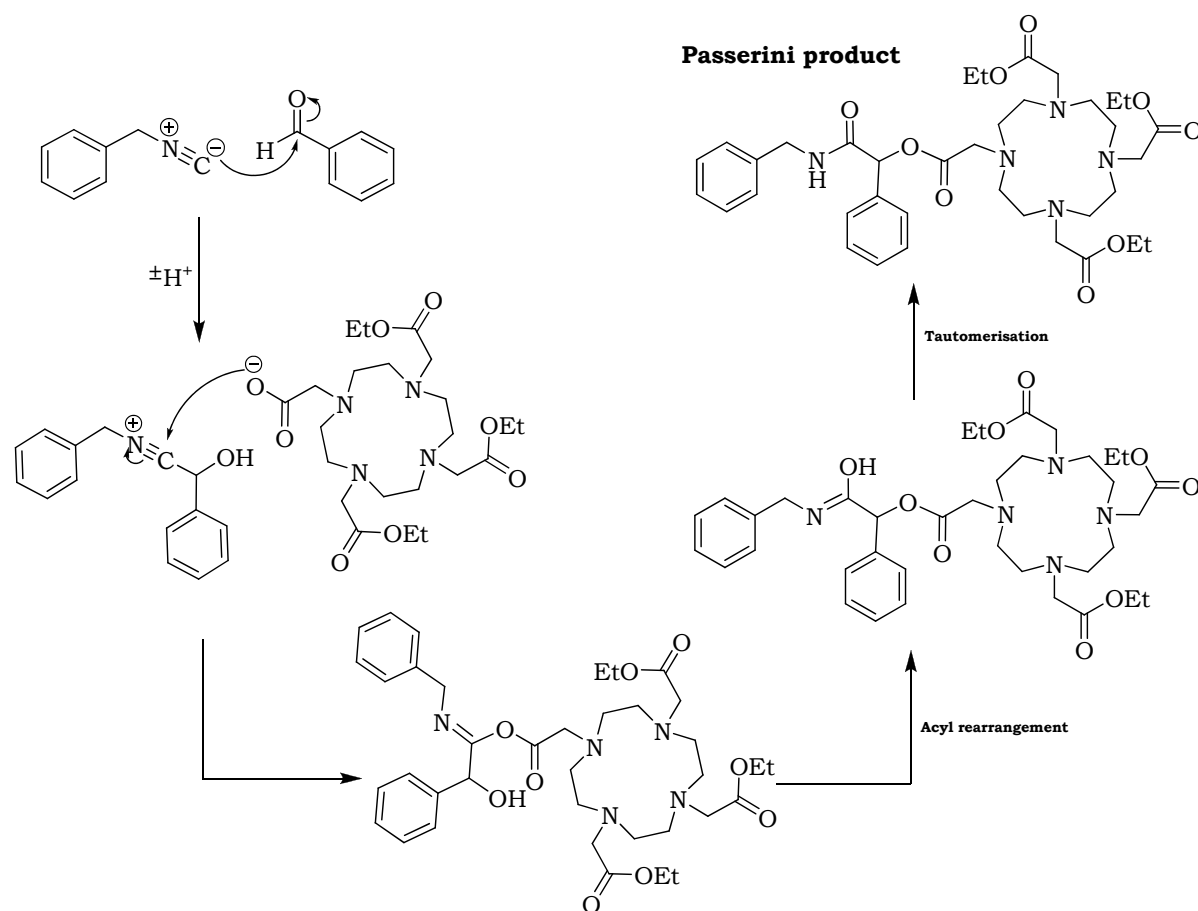
Scheme 17: The undesirable formation of the Passerini product.

This was unexpected and although disappointing, it meant that our attentions turned to the more easily accessible aminocaproic spacers (**Scheme 17**). This involved the protection of the acid as the *tert*-butyl or ethyl ester and the Ugi reaction was performed using the terminal amine. It was envisaged that although potentially reducing the water solubility of the product, when appended to a cyclic peptide this factor should be rendered inconsequential as the whole system would have sufficient aqueous solubility.

2.7.7 Solid phase Ugi reactions.

At this stage it was worth considering the possibility of achieving the Ugi 4-CCR on the solid phase as this would negate the need to assemble the cyclic peptide and the lanthanide fluorophore separately. However, trial solid-phase Ugi reactions were not as

successful as it was hoped from the literature precedent.[75] The group of Kim showed successful solid phase chemistry with small simple constituents to rapidly build a library of compounds with therapeutic potential.[62] However under similar conditions with our relatively bulky components there was no Ugi product outcome. The fact that SPPS is typically performed under mild conditions and in a polar aprotic solvent could explain the negative outcome as in a polar protic solvent such as the EtOH or MeOH favoured in Ugi chemistry, the resin shrinks and solvent penetration to mediate the reaction becomes poor. Therefore it was decided to continue building the two sections of the probe separately as this would offer greater flexibility and fewer potential problems associated with the rigidity of the resin-bound peptide and the steric demands of both the peptide protecting groups and the Ugi reaction substituents impeding the synthesis.
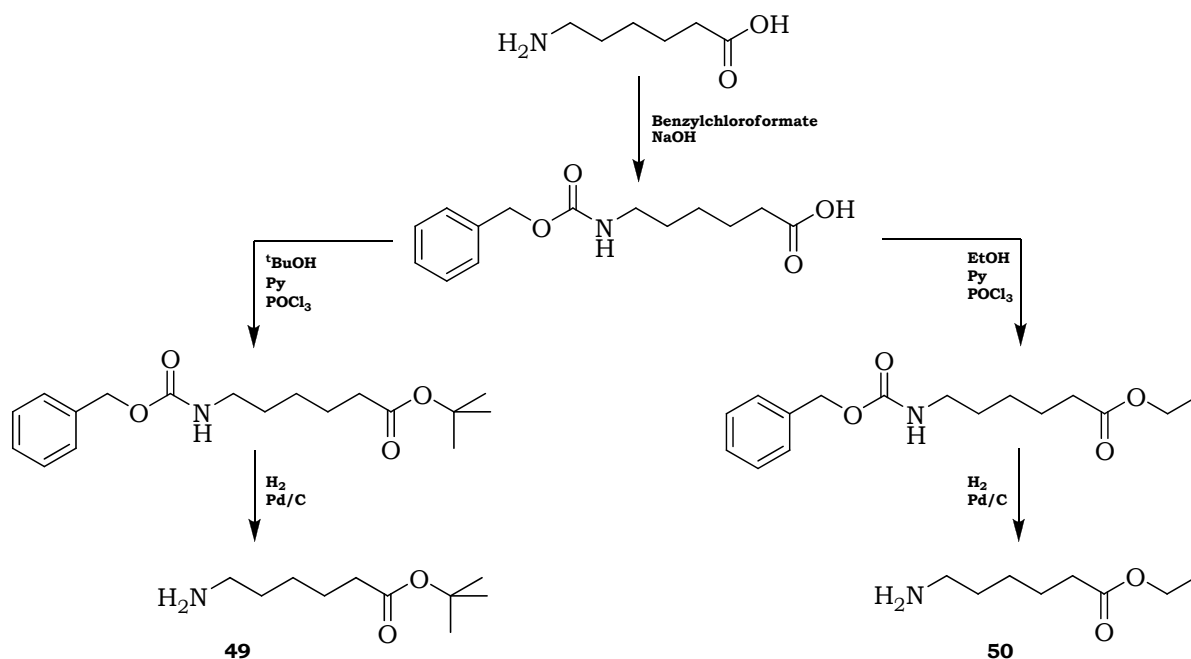
Scheme 18: Selectively hydrolysable aminocaproic acid derived spacers **49** & **50**.

The test Ugi reaction using the amino caproic spacers (**Scheme 18**) proved reliable and reasonably high yielding suggesting that somehow the PEG was prohibiting the success of the previously attempted Ugi reactions with the macrocycle. The Ugi reactions were performed with the DO3A derivative **03** as the carboxylic acid, indole & benzyl aldehydes, and benzyl isocyanide as the other substituents. The reaction proceeded straightforwardly following the protocol of our previous Ugi reactions. The consideration now was the order in which to perform the complexation and deprotection as this would influence the choice of aminocaproic ester used in the Ugi reaction. It was thought that as the caproic arm is a good distance from the macrocycle then an ethyl ester could be employed to save a step as

this would be hydrolysed simultaneously alongside the ethyl esters on the macrocycle, hopefully without also co-ordinating the metal in the next step (**Scheme 19**).

Thus the synthetic route using the caproic ethyl ester **50** in the Ugi reaction was attempted first and we were content with the positive outcome. The basic ester hydrolysis then afforded the tetra acid compounds **26** & **27** as expected, but complexation with the terbium metal failed and despite repeated attempts with alternative conditions the isolation of the desired product eluded us (**Scheme 19**). The reasons for this could be due to the caproic arm looping around and interfering with terbium chelation, or in some way influencing the geometry of the compound. Another theory is that the terbium chelates as expected but the free caproic acid of neighbouring compounds agglomerate onto the metal atom blocking our means of attachment, this however is unproven.
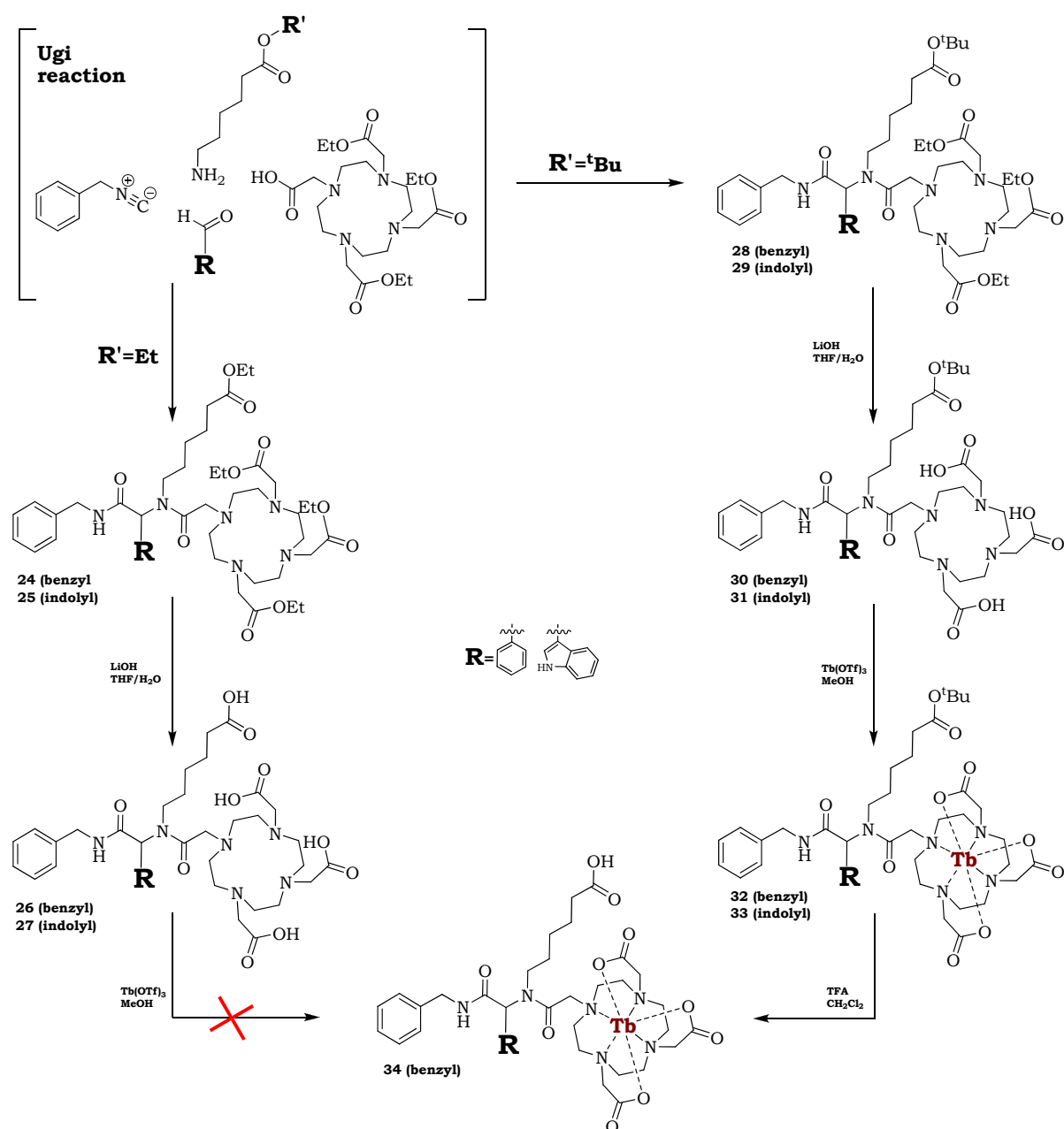
Scheme 19: The Ugi products and the route to the final compound prior to coupling to the cyclic peptide.

Thus the alternative route using the base stable *tert*-butyl ester caproic amine **49**, was employed and all steps proceeded to our satisfaction. Ethyl esters were selectively hydrolysed and a terbium ion installed in the macrocycle prior to a simple and clean acid cleavage of the *tert*-butyl ester using TFA to afford the pure product.

Thus the synthesis of our lanthanide based agents was complete but for the biological vector which, as described in the introduction, will convey the fluorophore to the specific biological binding site. The attachment of terbium complexes to the free lysine amine will proceed via a coupling reaction. However, if steric issues arise due to the bulky PyBOP coupling reagent prohibiting this attachment, there is still the option of simply pre-forming the labile succinic ester on the free caproic acid arm thus inviting nucleophilic attack from the lysine free amine in basic buffer conditions.

Thus the way became clear for the synthesis of the RGD integrin-binding cyclic peptide vector, cDfKRG.

2.7.8 Peptide synthesis
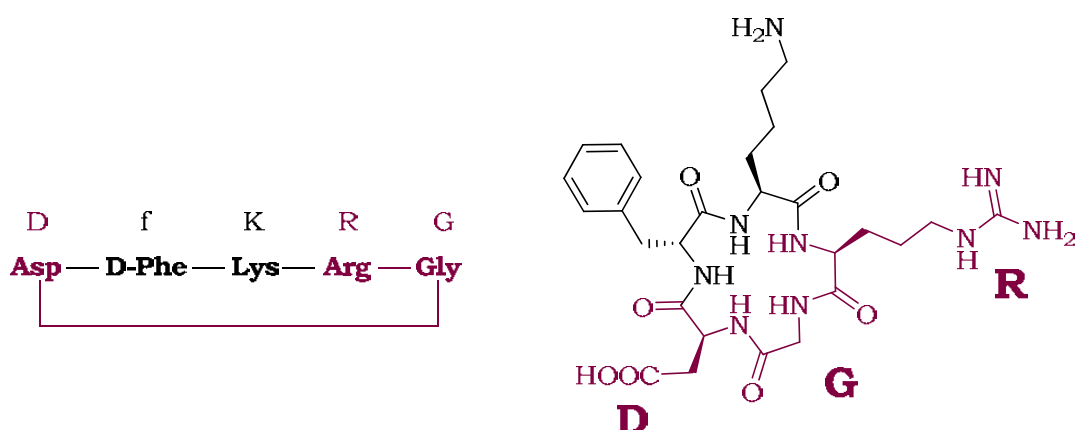
Figure 60: The integrin $\alpha_v\beta_3$ binding cyclic peptide cDfKRG.

As discussed in the introduction and synthetic plan, the linear DfKRG peptide was constructed using standard solid-phase protocol. However, the options with regard to protecting group chemistry for the selective deprotection of the lysine residue, the order in which to achieve the coupling to the fluorophore, and the stage at which to cleave the linear peptide from the resin and cyclise, were all variables that were addressed and shall be discussed in due course.

The general synthetic protocol was adapted from Dumy's solid-phase synthesis of an RGD containing cyclic pentapeptide.[76] Further derivatisation of the peptide through simple aliphatic extension units was devised to compare the efficiency of different routes to the target. The favoured approach, in accordance with our synthetic plan, was to take the synthesis as far as possible on the solid-phase with the peptide still appended to the resin.

The advantages of this are clear when it comes to the overall yield and also when it comes to the purification of advanced intermediate products by HPLC; the fewer steps in solution-phase the better when considering the economics of the synthesis.

The resin chosen was the highly acid-labile 2-chlorotritylchloride (**Scheme 20**). The reason for this was that the peptide would be selectively and rapidly cleaved from the resin in the presence of other, less acid-labile protecting groups such as the ^{t}Bu, Boc, Pmc and Pbf which are typically cleaved with concentrated TFA. The other reason was that our design never exposes the resin to acid conditions until cleavage is required.

Fmoc-glycine-2-chlorotrityl resin was commercially available and so taking this as our C-terminal residue and using the Fmoc solid-phase coupling protocol described in the experimental section, the penta-peptide was constructed (**Scheme 20**).
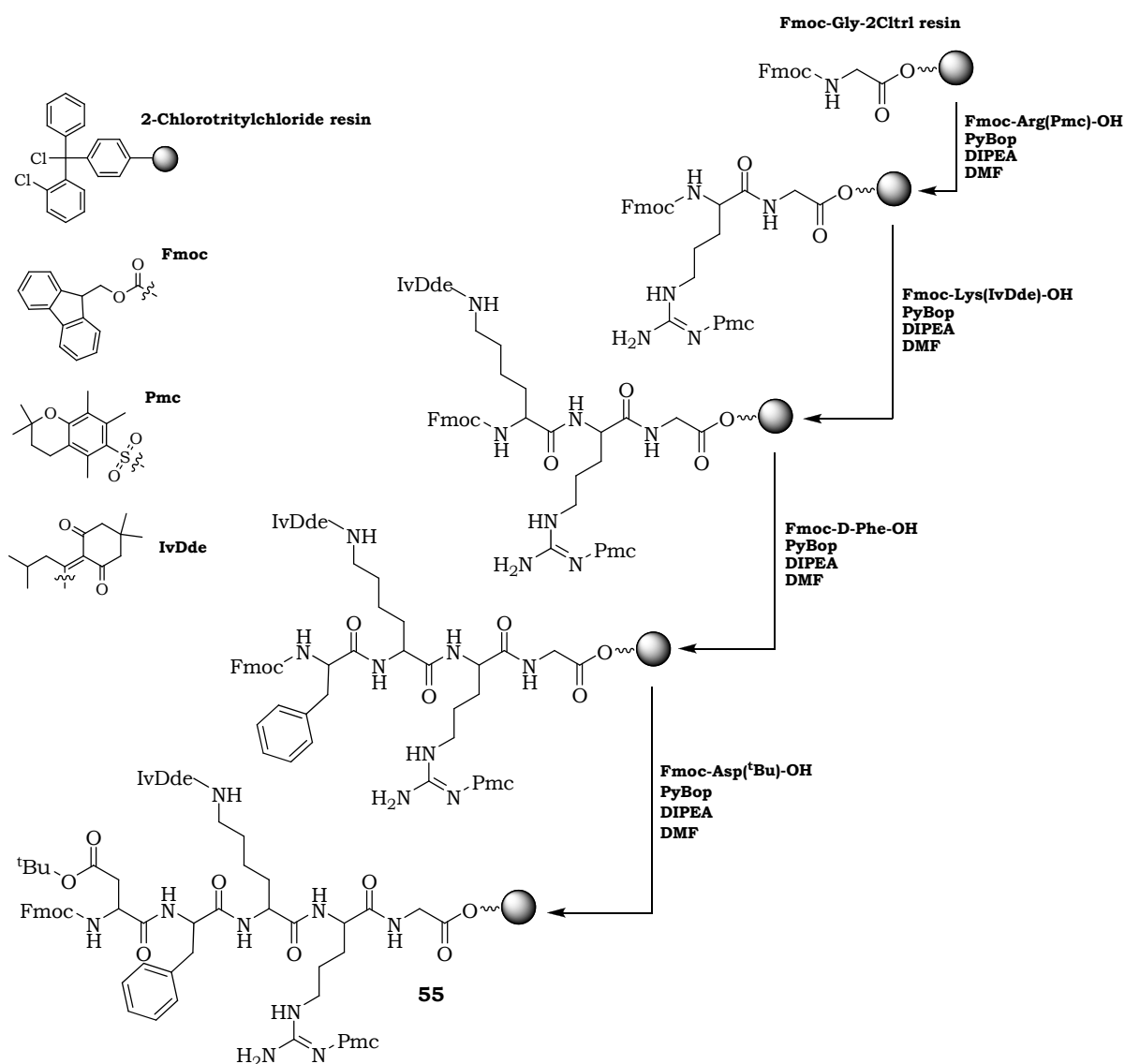
Scheme 20: The solid phase peptide synthesis (SPPS).

Our protecting group strategy dictated that as the arginine and aspartic acid residues are purely involved in integrin binding, they do not need to be manipulated during the synthesis and should be protected with groups that are cleaved under identical conditions in a single step. Commercially available is Fmoc-Arg(Pmc)-OH, where Pmc is an appropriate, strong acid labile, bulky orthogonal protecting group (**Scheme 20**). Likewise, Fmoc-Asp(tBu)-OH was readily available and this would provide a convenient, sterically less demanding acid-labile protecting group for the carboxylic acid side chain of aspartic acid. The initial plan was to employ the cheaper and more available Fmoc-Lys(IvDde)-OH as our regioselectively (hydroxylamine) cleaved lysine ε-NH$_2$ protector, but, as shall be discussed in due course, unforeseen problems arose and this had to be revised.

The linear synthesis went smoothly and progress was monitored with bead stain tests and mini-cleaves at each stage. A double coupling protocol was rigorously employed to ensure complete coupling and an HPLC trace of the final cleaved fully protected penta-peptide demonstrated impressive purity (**Figure 61**).
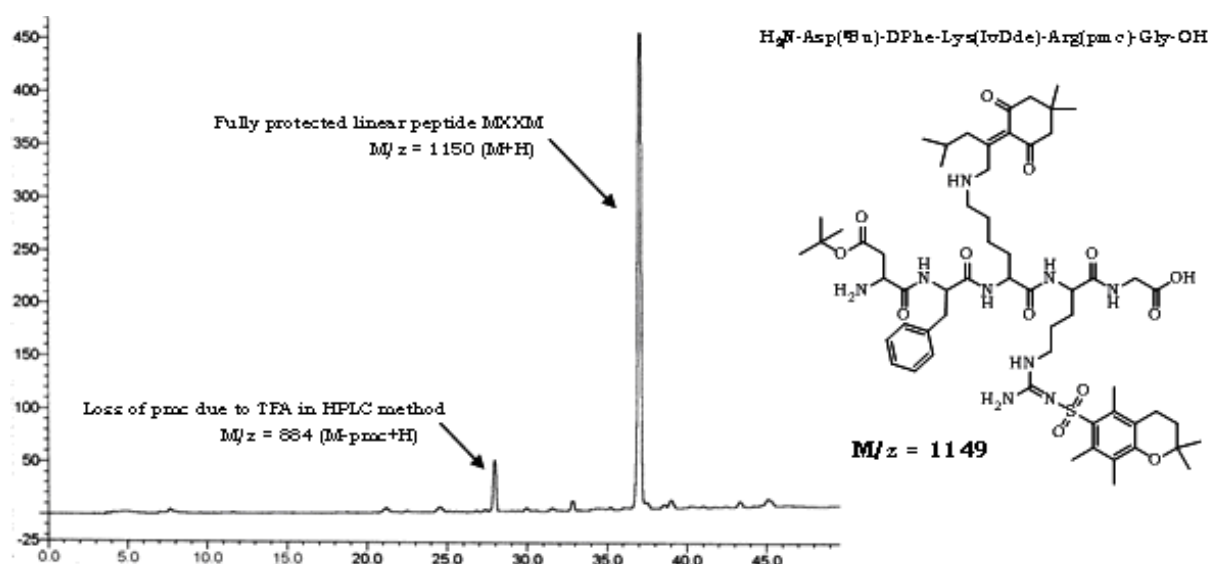
Figure 61: An analytical HPLC trace of the crude linear peptide **55** showing the incredible cleanliness of the SPPS.

We had now reached the stage where selective deprotection of the IvDde with hydroxylamine would afford the free lysine amine for solid phase functionalisation.[77] It was deemed wise to perform a test reaction to check that when a bulky fluorophore is attached that it is still possible to efficiently cleave the peptide from the resin, cyclise and deprotect the Pmc and tBu groups. To this end the peptide was subject to hydroxylamine cleavage conditions. However, this failed repeatedly with a mini cleave just confirming the remaining starting material.

The only other reported conditions for IvDde cleavage involve the use of hydrazine which could also result in Fmoc cleavage. However, a small quantity of peptide-bearing resin was

taken and hydrazine was carefully applied to the luer flask and the reaction time shortened. The mini-cleavage revealed the desired loss of IvDde and so we proceeded to couple the well known fluorescent label, fluorescein isothiocyanate (FITC) to the lysine amine using a double coupling protocol.

The crude mass spectrum of the cleaved product prior to attempted cyclisation revealed a problem, and after deliberation and consulting the literature it was confirmed that two fluorescein molecules were bonded to the peptide as the hydrazine hydrate had not only cleaved the IvDde group, but also the terminal Fmoc.[78] This means that the appended fluorescein in the terminal position prohibits the crucial cyclisation step to form the RGD unit. This drawback was confirmed when a simple solid phase coupling was attempted on the hydrazine washed peptide using benzoic acid/PyBop/DIPEA, and the bis-amide was isolated as the sole product. The IvDde group could not be cleaved selectively in the presence of Fmoc.

This did not render the resin bound IvDde penta-peptide obsolete as the way forward with this was now to perform the terminal Fmoc cleave (IvDde is piperidine stable), cleave the C-terminus from the resin and cyclise the peptide before performing solution phase IvDde cleavage with hydrazine, and coupling selectively on the lysine.

However, the desirable synthetic pathway was to perform coupling of the fluorophore on the solid-phase to eliminate unnecessary and time-consuming purification steps and so the strategy needed revising.

We acquired a quantity of Fmoc-Lys(Alloc)-OH and began the SPPS protocol all over again. Alloc is a protecting group selectively cleaved by a Pd^0 source such as $Pd(PPh_3)_4$. The synthesis proceeded in a now well practised and understood fashion with an excellent purity of crude fully protected linear peptide being revealed in the final mini-cleave.

Solid-phase Alloc removal is well established and a Tsuji-Trost type reaction with an anhydrous system of $Pd(PPh_3)_4/PhSiH_3/CH_2Cl_2$ was employed where $PhSiH_3$, is acting as an allyl scavenger. The reaction worked perfectly with no trace of Alloc in the mini-cleavage mass spectrum. A common problem that has been reported in this step is the formation of an undesirable allylamine byproduct.[79] Thankfully there was no evidence of this and the subsequent coupling of fluorescein isothiocyanate using conditions of $Py/DMF/CH_2Cl_2$, proceeded as desired with mini-cleavage revealing the product. Terminal Fmoc deprotection with piperidine and resin cleavage with dilute TFA afforded the labelled linear peptide ready for cyclisation and final Pmc and ^{t}Bu protecting group cleavage. However, the penultimate and crucial cyclisation step failed (**Scheme 21**).

54
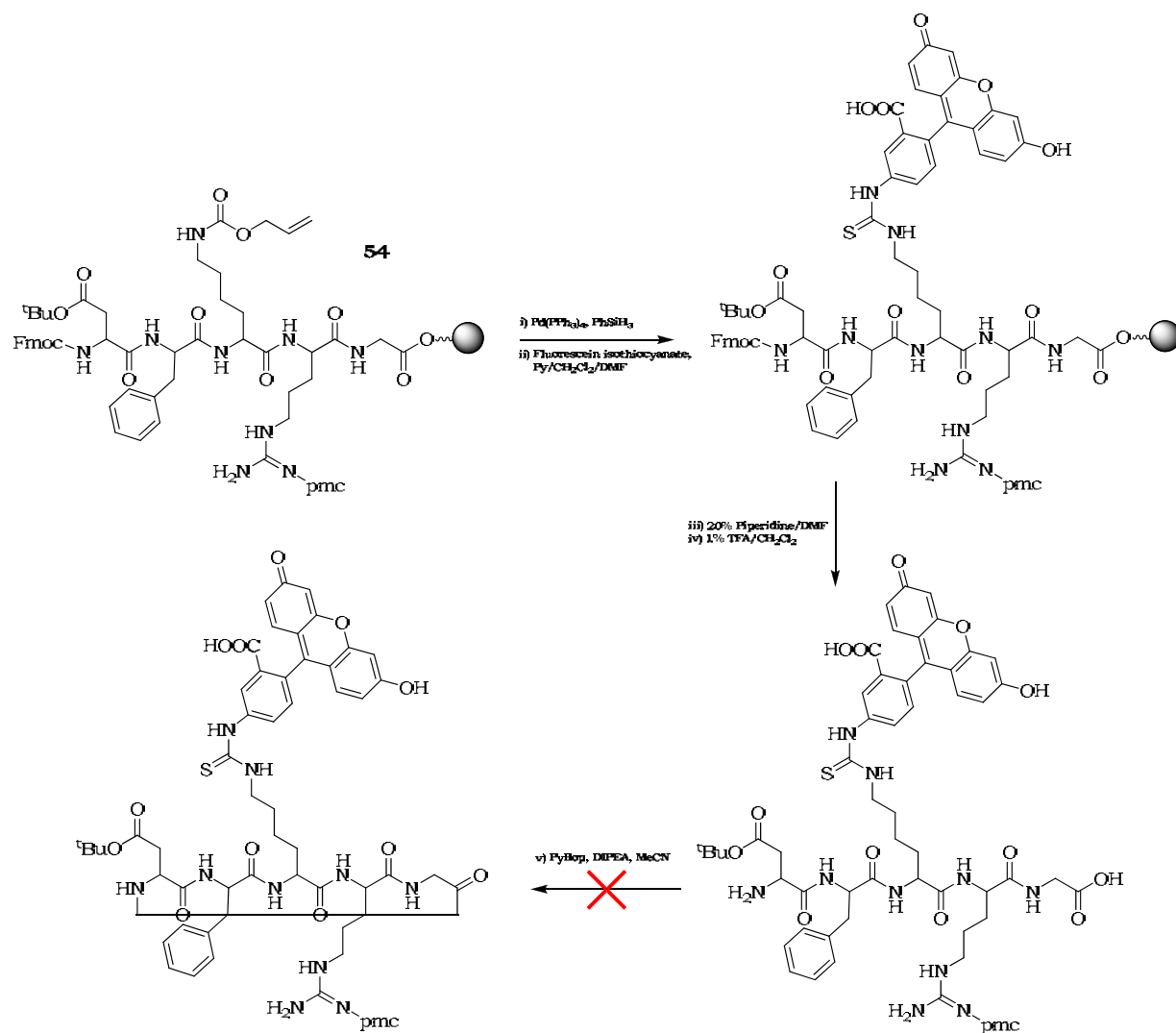
Scheme 21: The fluorescein label prohibiting cyclisation.

Increasing the temperature and reaction time offered no solution and even though a trace of product was identified it was a pitiful quantity (4 %) in comparison to the recovered starting material. The fluorescein was clearly prohibiting cyclisation, probably on steric grounds and once more the strategy required revision.

It was then realised that for the successful attachment of bulky fluorophores we could not follow the route of Dumy,[76] but must cleave from the resin and cyclise prior to selective deallyllation, fluorophore attachment and final protecting group cleavage (**Scheme 22**).
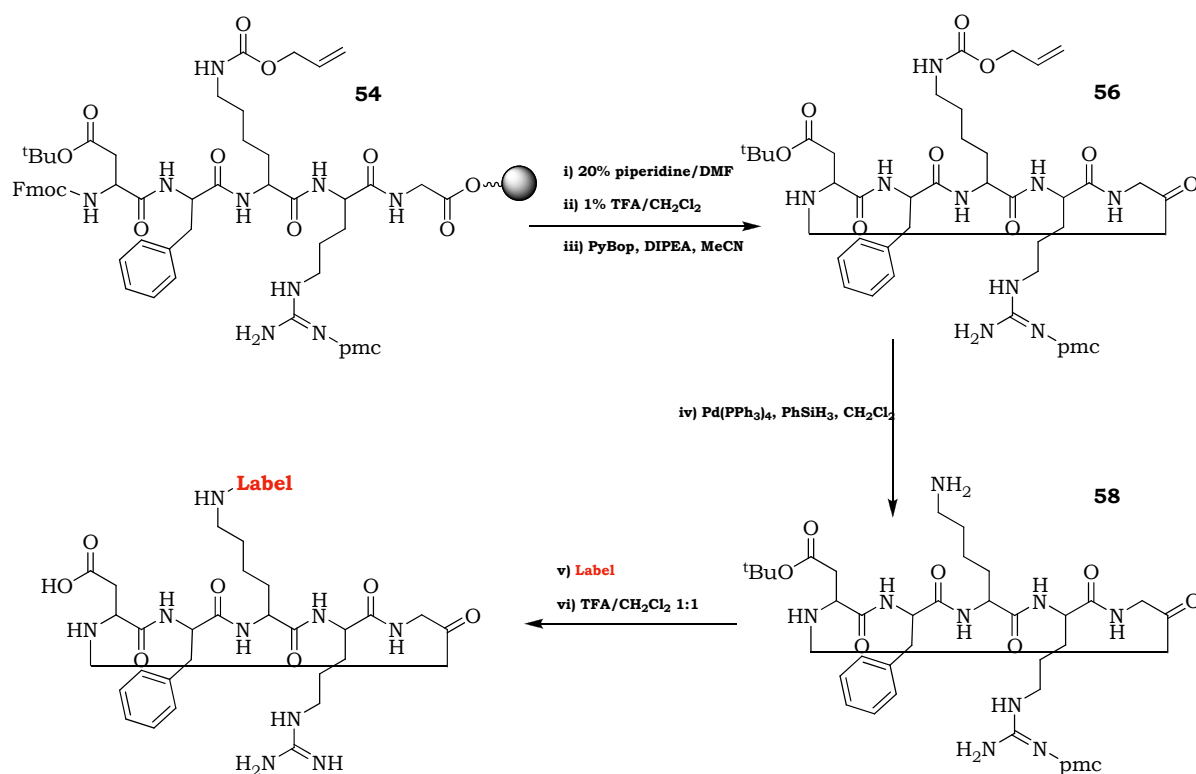
Scheme 22: The revised synthetic route to the final labelled cyclic peptide products.

Resin cleavage proceeded straightforwardly with an immediate colour change as the resin turned dark purple in the 1 % TFA/CH$_2$Cl$_2$ cleavage conditions. It was found that for optimal yield not only was it necessary to prolong the reaction for two minutes, but also to repeat until the resin stayed a deep purple colour as the resin was prone to fade upon washing with excess solvent indicating an incomplete cleavage. This process results in the use of a greater quantity of TFA than is ideal, and it is thought that although the predominant product is still fully protected, the reduced-pressure removal of the TFA must be carried out at low temperature with repeated co-concentrations to prohibit the simultaneous deprotection of the Pmc and ^{t}Bu groups as the concentration of TFA, being less volatile and more tenacious than CH$_2$Cl$_2$ will increase in the rotary evaporator. **Figure 61** shows HPLC evidence of this with a small peak showing loss of Pmc. This is immeasurable by mass spectrometry as in all the mass spectra obtained there is a clear natural fragmentation occurring within the spectrometer giving rise to [M-protecting group]$^+$ peaks (**Figure 62**). MALDI-TOF in the preferred matrix of gentisic acid containing 0.1 % TFA proved reliable enabling some excellent clean spectra showing the product peak clearly alongside the fragmentation peaks. HPLC had to be relied upon as an accurate measure of purity.
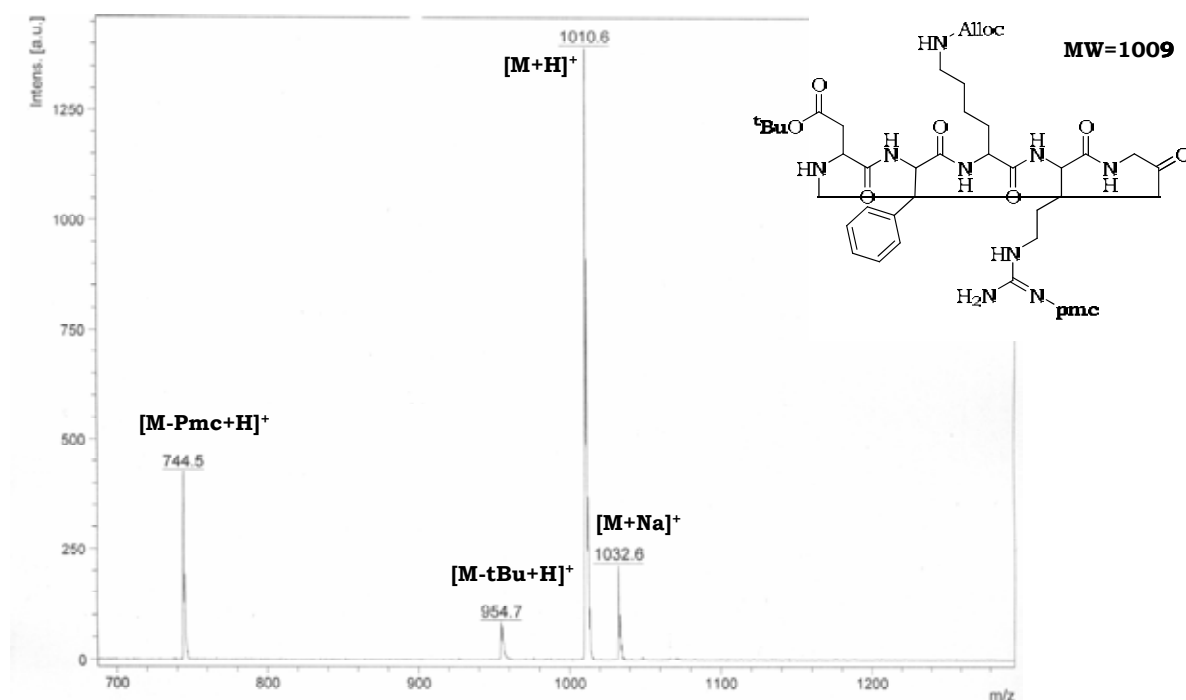
Figure 62: MALDI-TOF mass spectrum showing the fragmentation pattern of the HPLC purified cyclised Alloc protected peptide **56**.

Resin cleavage proceeded as desired and gave the linear peptide as a single peak by HPLC which was taken on crude to the cyclisation step.

2.7.8.1 Cyclisation

Cyclisation was a key step in this protocol as it is easy to see how polymerisation of the peptide may occur. However the design of the peptide favours cyclisation due to the D-phenylalanine residue reducing the conformational flexibility thereby favouring a conformation in which the termini are in close proximity. Thus under dilute coupling conditions at ambient temperature and given a three day reaction time with simple monitoring by analytical HPLC, the linear peptide peak disappeared and the cyclic peptide was recovered in excellent yield after preparative reversed-phase HPLC.

2.7.8.2 Solution-phase Alloc cleavage

Literature solid-phase Alloc cleavage conditions proved highly unreliable in solution phase. Brand new Pd(PPh₃)₄ (this compound is particularly air sensitive), did not alter the outcome. The literature cleavage methods attempted each used a different scavenger or solvent and it became obvious that the choice of nucleophile is crucial to this deallylation mechanism (**Scheme 23**).
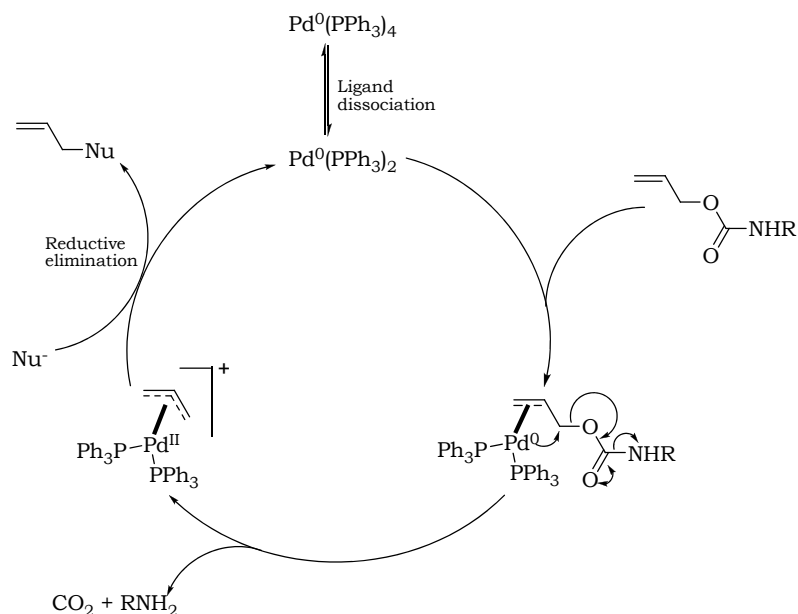
Scheme 23: The catalytic cycle for alloc deprotection, adapted from Ref. 80.

There is good precedent for solid-phase deallylations with many reported solvent and scavenger combinations which were considered for our solution phase attempt with $Pd(PPh_3)_4$ including: Me_3SiN_3/TBAF,[80] Bu_3SnH/AcOH,[80] THF/DMSO/HCl/NMM,[81] NMA/AcOH,[80] NMP/$PhSiH_3$,[82] (NMM)/AcOH,[80] and CH_2Cl_2/$PhSiH_3$.[79-84]

All solid-phase nucleophilic allyl scavengers, due to the solvents used in solid-phase chemistry, must be organic soluble and so our solution phase Alloc cleavage attempts initially also revolved around the $PhSiH_3$ protocols. These all failed with no sign of product. It is suggested that the $PhSiH_3$ is too harsh on the free cyclic peptide leading to breakdown of product. The NMM/AcOH conditions were then also attempted but with the same negative outcome. It was then realised through the discovery of a paper by Thayumanavan,[85] that allyl ethers can be cleaved under the mildly basic conditions $Pd(PPh)_3$/K_2CO_3/MeOH and it is mentioned in the same paper that allyloxycarbonyl (Alloc) groups are also cleaved under these conditions. [85]

Thus it was finally realised that in solution phase an organic base is not essential. Potassium carbonate was found to not only perform excellently but was consistent and high-yielding. The reaction was followed by hourly analytical HPLC analysis and it became clear that after three hours the cleavage was complete and that any increase in reaction time led to degradation of the product as there was a slow reduction in the size of the product peak and the appearance of many more signals. Thus an optimal time for reaction was set to three hours before prompt filtration, neutralisation and preparative HPLC purification.

This was a breakthrough and our strategy was back on track especially as the final products were now just two steps away.

2.7.9 Fluorophore attachment and revised choice of fluorophore

The choice of fluorophore was now an issue for discussion as a collaboration had been negotiated between ourselves and the Borlak group in Hannover at the Institute for Toxicology and Experimental Medicine (ITEM). This had opened an unexpected door as it would allow us to test our labelled peptide in live tumour-bearing (transgenic) mice through the use of a small animal fluorescence spectrometer with long wavelength (near-IR) capabilities. These facilities are hard to find and in order to exploit this excellent opportunity we were forced to reconsider the optimal excitation and emission wavelengths of the fluorophores. This is of particular importance in live animal models as the excitation wavelength must also be in the near-IR region not only for optimal tissue penetration and minimal interference, but also so as to not harm the mice.

The fundamental foreseeable problem with our lanthanide based probes for full body imaging application is that although the emission is of the desirable wavelength, the excitation is at a much lower wavelength in the UV blue region. This is due to the point of excitation being the local organic chromophore which then transfers energy to the metal centre. Excitation at this wavelength is not conducive to either good tissue penetration or the health of the mouse.

Therefore it was decided that although it will be good to show project completion through the formation of our final products, they would not be used in the impending mice studies, but rather saved for cellular work. Therefore it was decided to purchase a known organic fluorophore in which both excitation and emission are of near-IR wavelength to test the *in vivo* targeting potential through illumination of $\alpha_v\beta_3$ integrin binding.
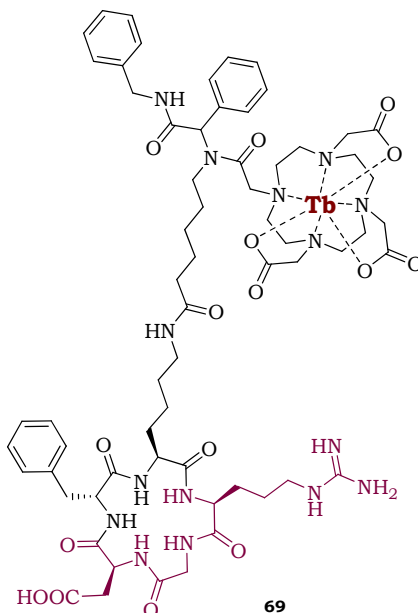
Figure 63: Final lanthanide based luminescent labelled cyclic peptide cDfK(εLn)RG, **69**.

The final product (**Figure 63**) was afforded through standard coupling conditions (PyBop/DIPEA/MeCN) and the Pmc and tBu protecting groups removed with TFA prior to isolation by HPLC.

2.7.9.1 Cyanine dyes – NIR797

Cy 7.0 and Cy 5.5 are commercially available and well known cyanine bioanalytical dyes (**Figure 64**).[86] This is a growing group of large organic fluorophores being developed for not only their photophysical properties but also for their solubility and stability characteristics. These dyes are an ideal alternative to visible fluorophores as they absorb and emit in the long wavelength nearIR region from 700-1200 nm where there is characteristically little biomolecular interference. They also exhibit large molar extinction coefficients, moderate to high fluorescence quantum yields and broad wavelength tunability.[87]
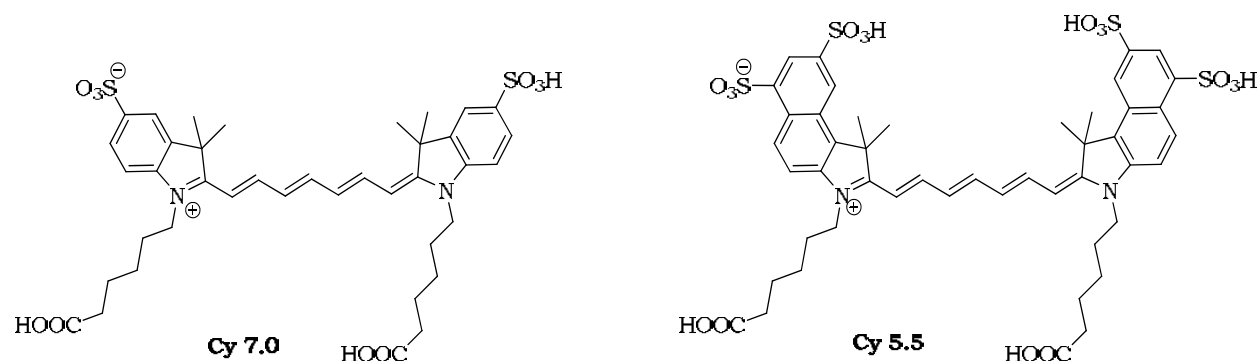
Figure 64: GE Healthcare near-IR bioconjugates Cy 7.0 (Ex 743 nm; Em 776 nm), and Cy 5.5 (Ex 485 nm, Em 667 nm).

Cy 7.0, however, is a heptamethylene cyanine dye restricted by not only poor stability but also a difficult, low yielding and unreliable synthesis. It is also prone to form non-fluorescent aggregates. Unfortunately these are common drawbacks for many of this family of dyes and the other similar derivatives reported. The complicated synthesis has led to those that are commercially available being very expensive.[87-90] However, improvements in stability and purification technique have been afforded through the incorporation of a rigid cyclohexenyl ring in the polymethine chain.
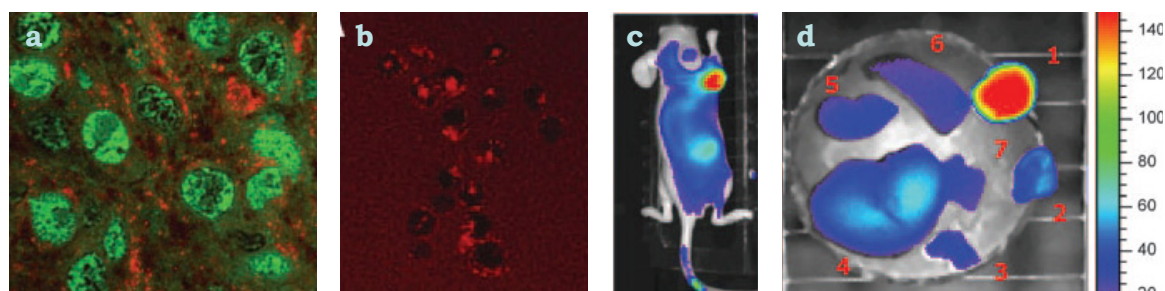
Figure 65: An example of a cRGD-Cy 5.5 conjugate used in *in vivo* fluorescence imaging. **a)** Laser scanning confocal microscopy of rat brain tumour cells; **b)** Binding and endocytosis, U87MG human glioblastoma cells; **c)** Fluorescence imaging of U87MG brain tumour xenograph; **d)** The removed tumour (1) vs other organs. [52,88]

This has led to manipulation of this central unit and, through nucleophilic substitution reactions with nucleophiles bearing an additional functional group, to allow the covalent

labelling of biomolecules.[89] NIR797 (**Figure 66**), is of the same family of cyanine dyes used in all manner of biological fluorescence assays and has recently become commercially available through Fluka, albeit in poor (~50%) purity.[90]
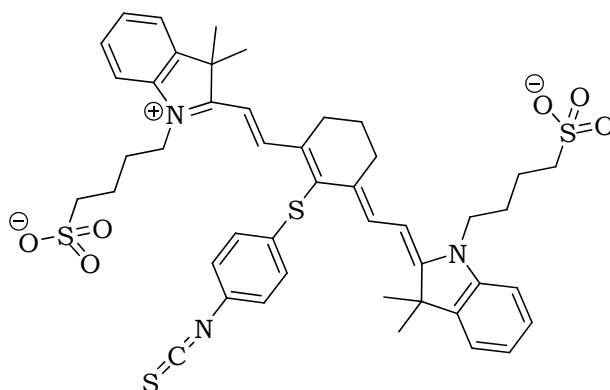
Figure 66: NIR797 (Fluka).

This dye is water soluble and has been effectively used in DNA sequencing, genetic analysis, proteomics and *in vivo* imaging.[90] The terminal isothiocyanate is not only an ideal electrophile for rapid nucleophilic attack by the lysine εNH_2 of the peptide, but also the resulting thiourea group has been found to modulate and enhance fluorescence emission decay in fluorescein labelled molecules.[91] Although expensive, NIR797 would be an excellent starting point for these investigations due to the excitation and emission maxima being in an ideal region of the spectrum in the mid-range of the near-IR (Ex. = 797; Em. = 815 nm). This dye would also offer an excellent and powerful probe for cellular and tissue slice fluorescence imaging. It has an extinction coefficient of 94100 $M^{-1}cm^{-1}$ in H_2O.[86,87]

2.7.9.2 Introduction of a spacer unit.

As mentioned in the introduction and synthetic plan, it was also deemed important to investigate the proximity of the fluorophore to the binding site through the introduction of differing length spacer units. This chemistry was perfected initially for incorporation into the Ugi reaction as the amine substituent. However, in this peptide chemistry the spacer must be an extension upon the lysine terminal primary amine and therefore the terminal group must be subject to nucleophilic attack of the amine. Therefore the ethylene glycol and caproic spacers (**Scheme 24**) were functionalised as carboxylic acids ready for coupling to the lysine amine prior to deprotection to reveal a distant free amine for fluorophore coupling.[92]
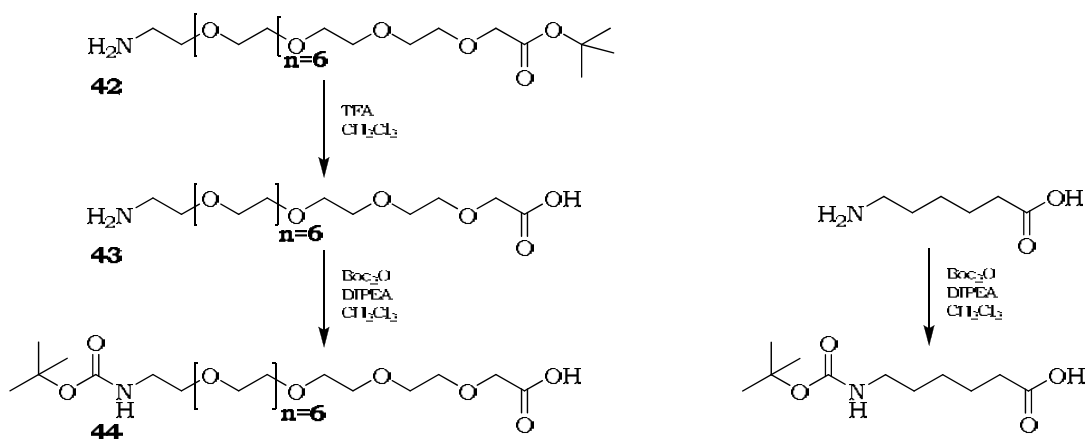
Scheme 24: The spacer units synthesised as terminal carboxylic acids for coupling to the lysine residue.

Thus the cyclised, selectively Alloc-deprotected peptide was subjected to coupling conditions with the Boc-aminocaproic acid and ethylene glycol spacer **44** to afford an elongated lysine arm in preparation for more distant fluorophore coupling.

At this point it was decided that due to the expense of NIR797 it would be wise to couple the NIR797 directly to the lysine residue and test this compound prior to investigating the efficacy of spacer units in potentially increasing binding affinity through distancing the steric bulk of the fluorophore from the active site.

After HPLC purification and MALDI-TOF mass spectrometry the product **61** (**Figure 67**) was confirmed in high purity and was ready for the mice studies in Hannover with naked cDfKRG **59** as a control or blocker.
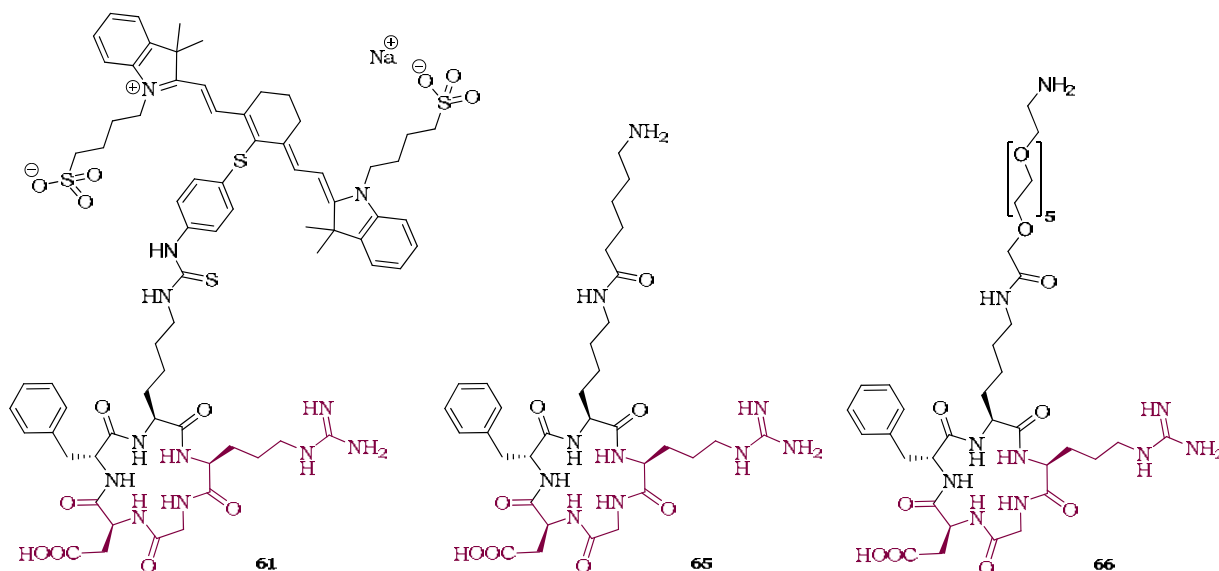
Figure 67: The final labelled NIR797 peptide **61**, and the two spacer appended peptides **65** & **66**.

2.7.10 Synthesis of Osmium 2,2'-bipyridine dyes

Staying with the theme of long-lived metal-centred fluorescence, it was decided through discussions with our collaborators that it would be prudent to also investigate osmium bipyridyl complexes.

This was our first venture into organometallic chemistry. Although identical routes to our devised luminescent product involving other complexed transition metals such as ruthenium[93-95] and rhenium[96] were located in the literature, the search yielded limited results for complexation with osmium as the metal in 2,2'-bipyridyl complexes.

The chemistry required a knowledge of the oxidation and co-ordination status of the metal and also an understanding of how solubility difficulties could be overcome through the manipulation of the counter-ions.
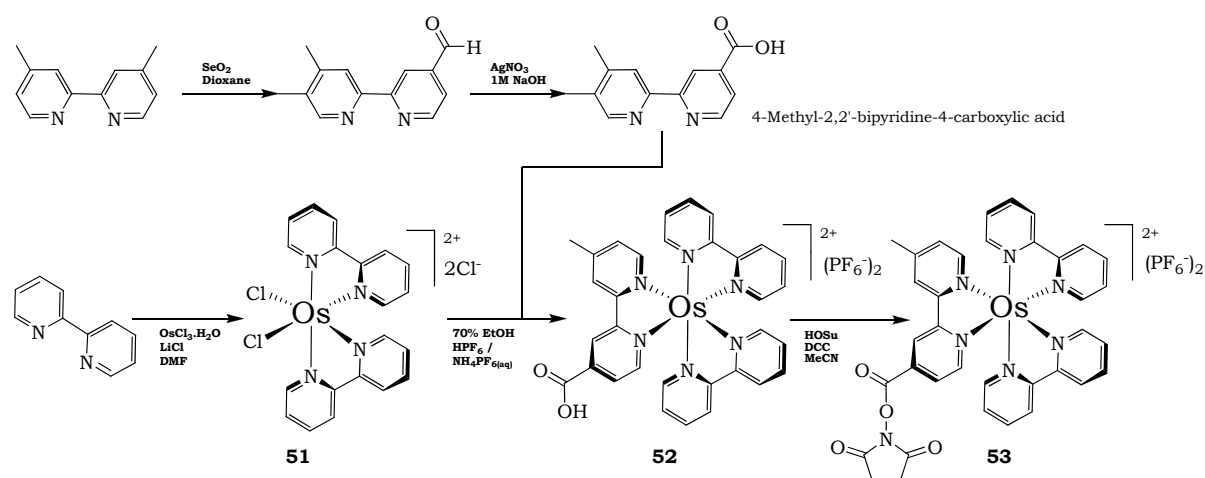
Scheme 25: The synthesis of the Osmium complex.

We were able to acquire a small quantity of the carboxylic acid intermediate 4'-methyl-2,2'-bipyridine-4-carboxylic acid in order to test the synthesis and quite rapidly access the product **53** (**Scheme 25**). This compound is prepared through selective oxidation of a single 4,4'-dimethyl-2,2'-bipyridine methyl group in a relatively poor 45 % yield.[93]

2.7.10.1 Synthesis of Cl$_2$Os(bpy)$_2$, **51**

In this step it is interesting to note how that DMF not only serves as an ideal polar-protic solvent for the dissolution of OsCl$_3$, but also reduces the Os^{3+} to Os^{2+}. The order of synthetic protocol is also important as pre-addition of the LiCl precludes the formation of the undesired 'tris' complex (bpy)$_3$Os. In comparison to the identical synthesis with ruthenium,[93] a shorter reaction duration was required as osmium is able to facilitate the binding of the relatively bulky 2,2'-bipyridyl units more rapidly.[97-99] Dissolution in acetone followed by 72 hours at -8 °C caused complete precipitation of the product which was then simply isolated by Büchner filtration as a solid of a deep brown colour.

2.7.10.2 Synthesis of (4'-Me-bpy-4-COOH)Os(bpy)₂, **52**

The inclusion of the third acid-functionalised bipyridine unit renders the solubility of the product very poor unless a change of counter ion is enacted. Acidification with hexafluorophosporic acid, followed by the careful addition of aqueous ammonium hexafluorophosphate furnishes the product with a pair of hexafluorophosphate counter anions and causes it to precipitate out of the reaction solution as a solid not dissimilar in colour to the traditional British racing green.

2.7.10.3 Synthesis of succinic ester, **53**

This final step is not fundamental as standard coupling conditions could be employed between the carboxylic acid and peptide amine to join the two parties together. However it is desirable to limit any possible side reactions with other functionality as shall be discussed forthwith when it is discovered that the ᵗBu and Pmc groups are redundant and, if left deprotected, the carboxylic acid of the aspartic acid residue would become vulnerable to coupling reagents. So the installation of a good leaving group to enable a simple nucleophilic substitution reaction under mildly basic or neutral conditions with the lysine free amine was desirable.

It was chosen to attempt to install a succinic ester due to the ease and mildness of reaction conditions in comparison to the thionyl chloride conditions required to form an acid chloride. This reaction was also precedented and N-hydroxysuccinimide afforded the product in excellent yield via a DCC coupling. However it was not anticipated how reactive the succinic ester would be and this led to problems with purification and characterisation. Ready hydrolysis of the succinnic ester in aqueous and polar protic solvents such as MeOH was apparent when standard HPLC protocols using water/organic solvent gradients revealed a combination of ester product and the carboxylic acid starting material peaks. Initially this was rationalised as an incomplete reaction, but repeated HPLC runs revealed only an increase in carboxylic acid. MALDI-TOF and electrospray mass spectrometry on the crude and HPLC pure peaks revealed again predominantly the carboxylic acid, and, when in MeOH, the MeOH adduct. Interestingly a loss of CO_2 was also always visible on the mass spectrum. Nmr of the filtered dry crude product did reveal the succinic ester CH_2 protons, suggesting a successful reaction, but the standard of the nmr was not good enough for unequivocal proof of product through measurement of peak integration. Thus it was decided that the best proof of succinic ester formation was to take a small quantity of the crude and couple to the peptide.

The result was very satisfying with the reaction followed clearly by analytical HPLC and the mass spectrum revealing the attachment of osmium complex in excellent yield with only a slight excess of the fluorophore added to the peptide in dry MeCN (**Figure 68**).
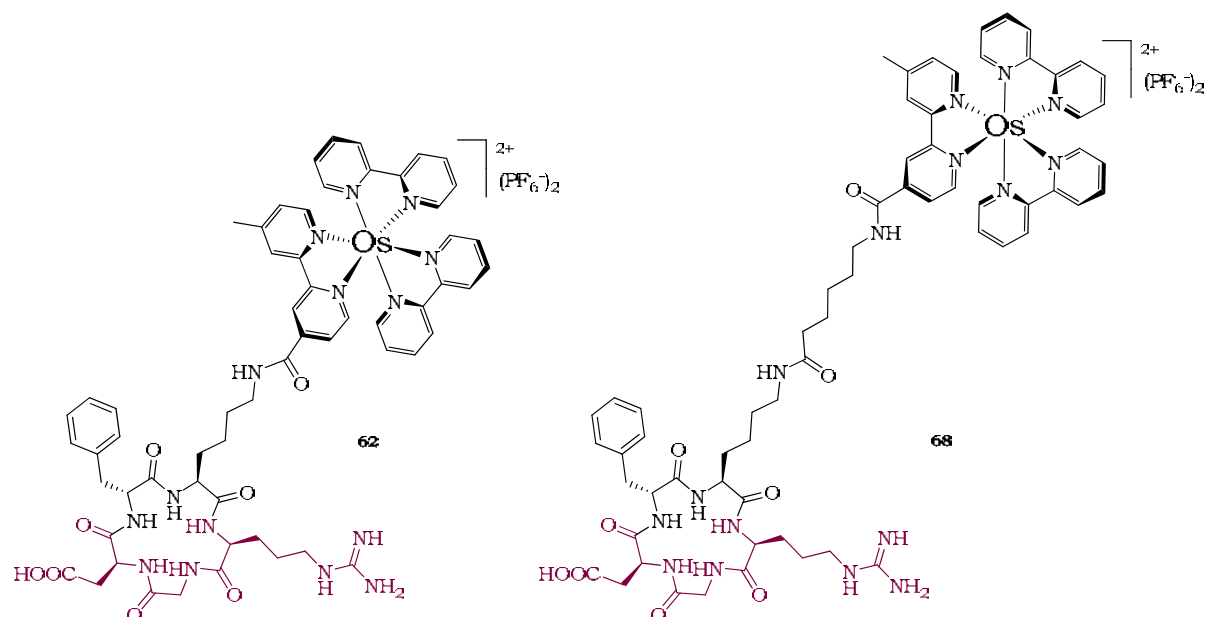
Figure 68: The fully deprotected Osmium-cDfKRG conjugates prepared for biological testing.

The final deprotection was achieved very cleanly with TFA/CH$_2$Cl$_2$ and although initially it was feared that this might replace the PF$_6^-$ counterions, we need not have worried as these proved robust. The ^{t}Bu and Pmc protecting groups were removed cleanly.
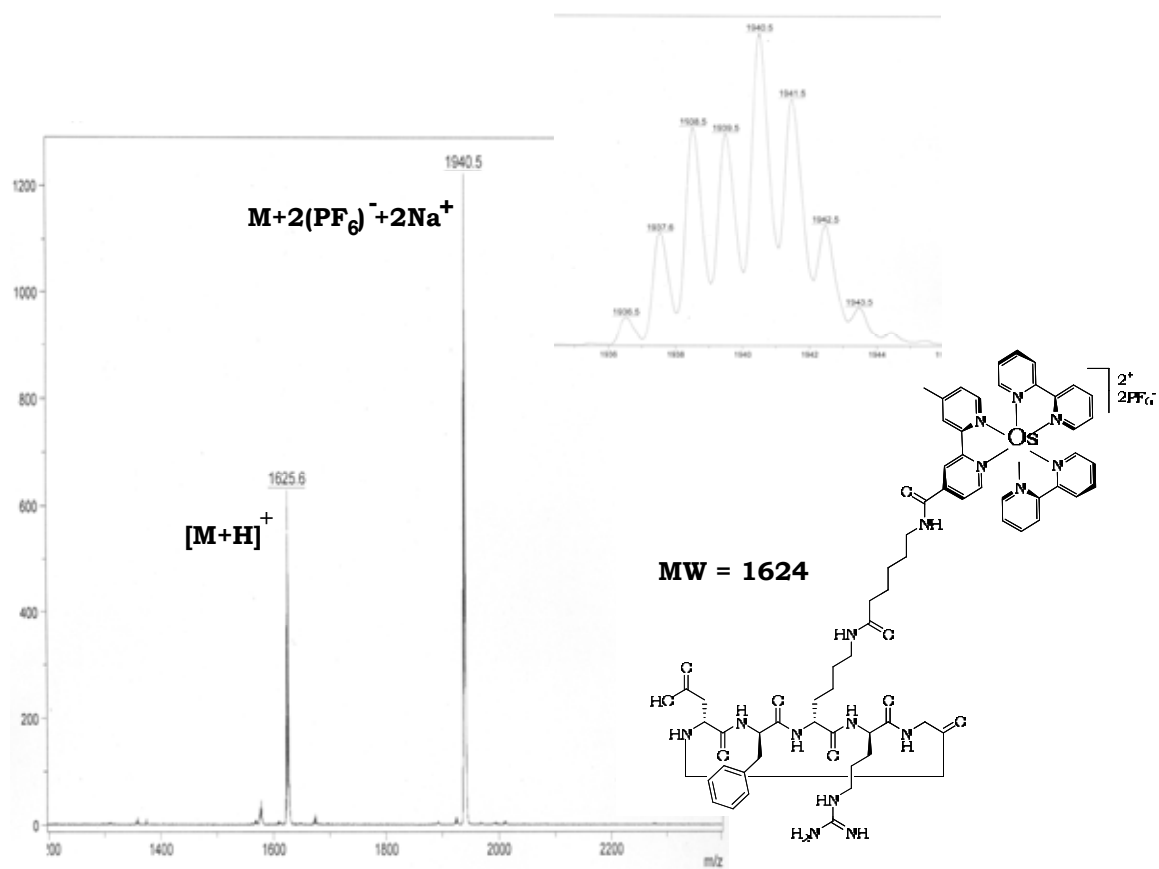
Figure 69: The MALDI-TOF of the product **68**.

Interestingly, mass spectrometry became invaluable as the PF_6^- counter ions showed on the spectrum with the $[M+2(PF_6)^-+Na]^+$ ion visible. This was very indicative of the product (**Figure 69**). The main point of interest now was to discover the photochemistry of the peptide probe. UV spectroscopy followed by fluorescence emission spectroscopy revealed that the optimum excitation and emission wavelengths were indeed in the desired near-IR region of the spectrum (**Figure 70**).
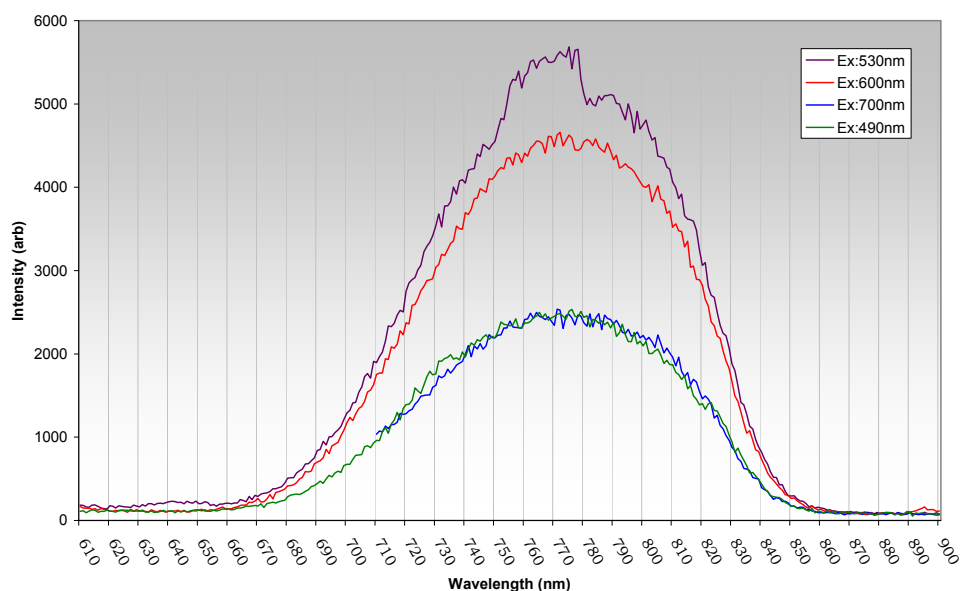
Figure 70: The fluorescence emission spectrum of Osmium conjugate **68**.

This was great news and we took to Hannover the cyclic peptide bearing both NIR797 and Osmium fluorophores for testing in transgenic mice using optical imaging technology (see the imminent biological results and discussion section and appendices 4 & 5).

2.7.11 Streamlining the synthetic route.

It became apparent due to the failure of the peptide cyclisation step after solid phase attachment of FITC, that the selective removal of the lysine protecting group was no longer necessarily fundamental to the successful outcome of the synthesis, indeed, it now appeared to be adding unnecessary steps, reducing the overall yield and efficiency of the protocol.

The rationale for this is that the free arginine side chain, being a stabilised π-system is relatively non-nucleophilic in comparison to the lysine ε-primary amine; similarly the carboxylic acid of the aspartic acid is far less nucleophilic than the lysine $ε$-NH_2. Therefore, as the peptide is now cyclic, there are no other functionalities to consider when appending a fluorophore selectively to the lysine side chain. This means that the acid cleavage of ^{t}Bu and Pmc can be accomplished prior to a final labelling step. Indeed, taking this into account, the peptide could now be built with an acid labile protecting group also on the lysine amine. This would eradicate the tricky palladium-catalysed selective deallylation

step as the linear peptide could be Fmoc deprotected, cleaved from the resin, cyclised, and totally deprotected before a final fluorophore attachment step (**Scheme 26**). This was an excellent realisation and offered a far more rapid and higher yielding route to the products with two fewer HPLC purifications.
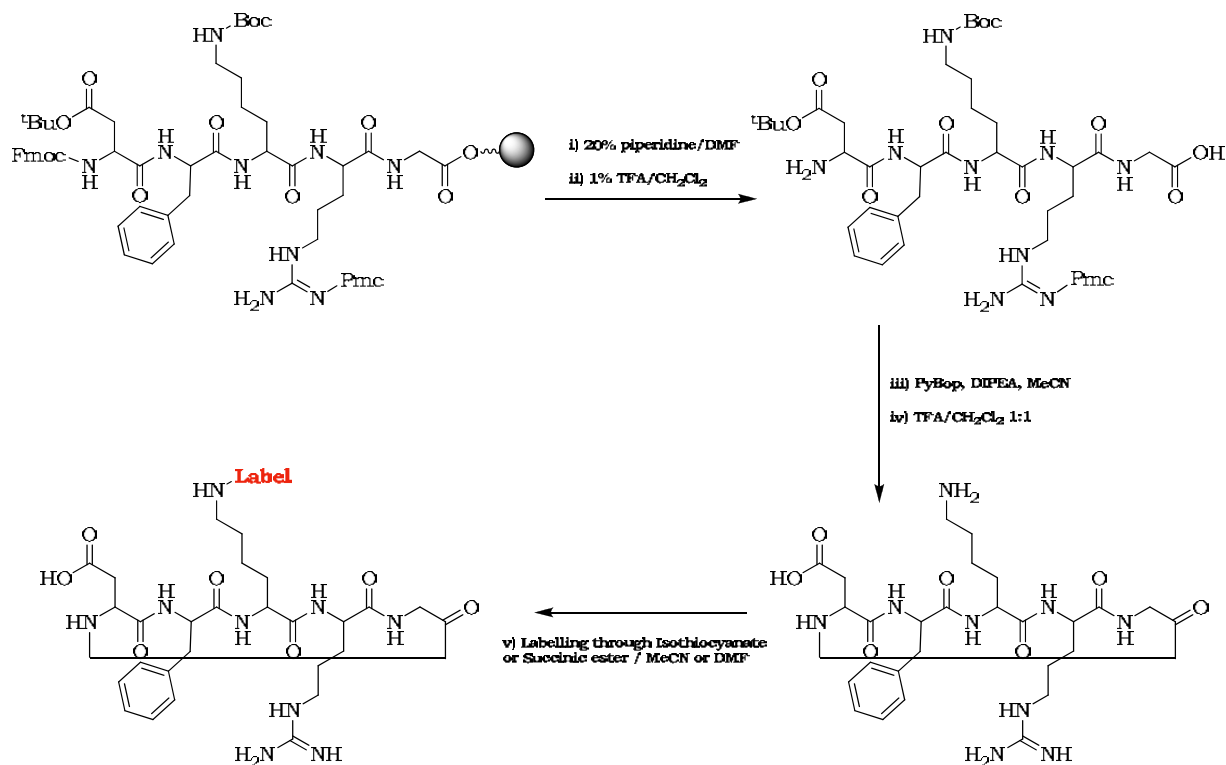
Scheme 26: Possible streamlining of the peptide synthesis and fluorophoric labelling.

As a quantity of resin bound Lys(IvDde) and Lys(Alloc) penta peptides had been synthesised by SPPS we elected to make use of these and promptly cleaved, cyclised and performed the deprotections affording a good quantity of naked cDfKRG peptide for appending the desired fluorophores.

However this strategy does not take into account the insertion of spacer units which have required coupling conditions which could lead to derivatisation of the aspartic acid and arginine side chains. It would be wise to investigate the selective elongation of the lysine arm with the spacers whilst the peptide is still on the solid phase and deduce whether they offer an improvement before redesigning the whole strategy.

2.8 Summary & conclusions

Although time-consuming, the selective manipulation of tetraazadodecane (cyclen) side-chains to efficiently afford the Ugi carboxylic acid **03** was ultimately very satisfactory. The ensuing Ugi 4-component condensation reactions were again a great success although relatively laborious in their purification. Lanthanide complexation afforded a small library of potential energy transfer mediated long-lifetime fluorophores. These were downselected according to the efficacy of the energy transfer mechanism. This work was published.

The next step was to incorporate into the fluorophores a means of attachment onto the integrin $\alpha_v\beta_3$ specific cyclic peptide vector cDfKRG, via the lysine εNH_2. This was attempted unsuccessfully using varying length ethylene glycols as the Ugi amine component which had been selectively derivatised at one end and protected at the other. However, success came in the form of an aminocaproic spacer.

The solid phase synthesis of the linear penta peptides was remarkably clean and efficient with the coupling reagent combination of PyBop / DIPEA in anhydrous DMF proving reliable and fast. The failure of the selective cleavage of the IvDde lysine protecting group directed a change of tactics with the Pd^0 cleaved Alloc protecting group then being employed to good effect. It was then discovered that the attachment of a fluorescein tag to the exposed lysine εNH_2 whilst the peptide was still resin bound, prohibited the subsequent fundamental cyclisation step, and therefore the strategy was again amended.

Cyclisation of the peptide, lysine deprotection and fluorophore coupling, followed by a final deprotection step to remove all other amino acid protecting groups afforded an effective route to the final luminescent labelled, integrin targeted cyclic peptides. A more streamlined route was also realised and future work will now be both simple and reliable. In addition, the system was further derivatised to distance the bulky fluorophores further from the peptide to reduce the possibility of binding site interference.

Due to the weak emission from lanthanide metal-centred luminescence in comparison to some commercially available fluorophores such as fluorescein and the near-infrared emitter NIR797, it was initially decided to test the specificity of the cyclic RGD peptides to their integrin targets using these dyes, prior to incorporating the originally developed Ugi derived lanthanide fluorphores. The other drawback was that timegated technology was rendered unavailable to us to capture the long lived luminescence from the lanthanide probes.

A door opened to us to test the cyclic peptide probes in lung and liver tumour-bearing mice and we took advantage of this offer to image tumour accumulation and specificity of the near-IR emissive systems *in vivo*. This work was perhaps premature but did produce some results of great promise. The tissue penetration of the emission was excellent although the anatomy of the mouse and the limitations of the imager complicated things. At the completion of this thesis, the cyclic peptide probes were being tested against tumour cells.

2.9 *In vivo* studies: small animal testing

(COST0602 collaboration. Host: Prof. J. Borlak; ITEM Fraunhofer, Hannover, Germany)
Jan & April 2009

2.9.1 Optical *in vivo* imaging
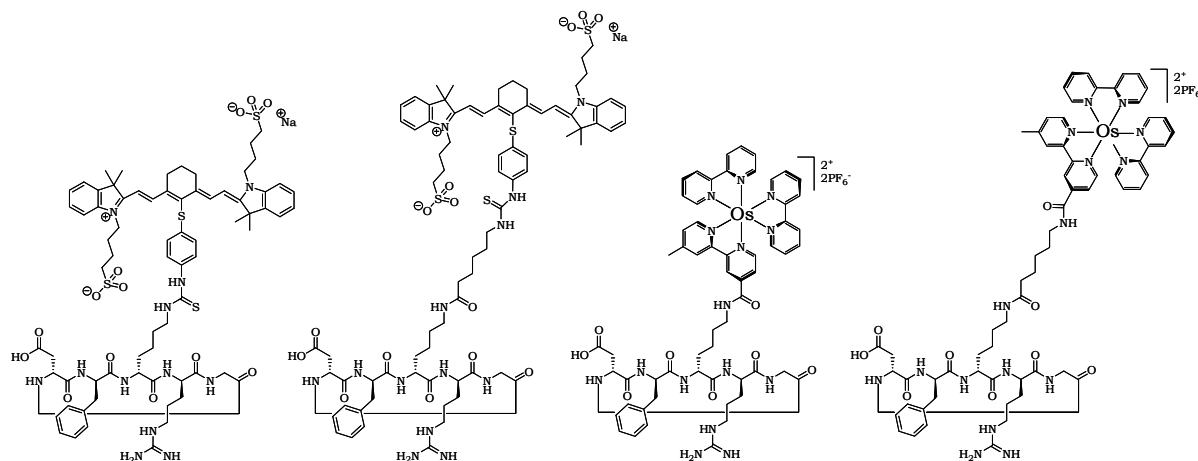
Figure 71: Cyclic DfKRG peptides with NIR797 and Osmium complex fluorophores attached either through the lysine side chain or via a caproic spacer.

2.9.2 The probes

Four compounds (**Figure 71**) were taken to Germany for whole body fluorescence imaging in mice. The purpose of the journey was to ascertain what potential the near-IR wavelength emission from these targeted imaging agents possessed for *in vivo* application.

There were several interesting and important questions that required addressing in this study that would impact upon the development of further synthetic protocol and ultimately the direction of the project:

- Is NIR797 a viable *in vivo* fluorophore?
- Is there any sign of instability and *in vivo* breakdown?
- Does the proximity of the fluorophore to the peptide vector influence the binding and specificity of the agent (spacer efficacy)?
- Is the Osmium complex a viable *in vivo* fluorophore?
- Is the peptide vector specific enough for unambiguous *in vivo* tumour recognition?
- Is the tissue penetration sufficient for potential human local *in vivo* imaging e.g endoscopic investigation?
- Are there any obvious toxicological issues?

2.9.3 The experimental protocol

c-Myc and c-raf type transgenic mice of various ages expressing tumour burdens specifically of the lung or liver were subjected to Isoba isofluran/O_2 inhalation anaesthesia

prior to 100 µL intravenous injections of fluorescent agent or control peptide at 0.5 mg/mL concentration via the dilated tail vein. Whole body fluorescence emission images were recorded at 5, 30, 60, 120 and 240 minutes post injection.

Mice used in organ imaging experiments were anaesthetised and injected as described above, and a whole body scan taken at five minutes. The mouse was then sacrificed at 30 minutes post injection and the organs excised and imaged. Lungs were imaged both flaccid, and perfused with preserving formaldehyde solution. As there was no diminishing or interference of fluorescence due to the perfusion it was decided to always attempt perfusion as this allowed more accurate tumour site identification as illustrated in **Figure 72** where the enhancement of emission specificity in the right hand lobes of the perfused lung is revealed.
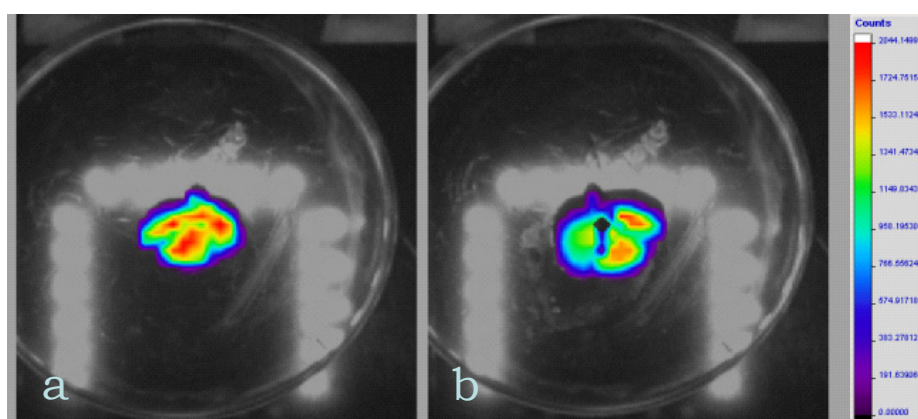
Figure 72: Excised transgenic lung. **a**) non-perfused; **b**) perfused. Red indicates areas of highest concentration of fluorescent signal. Three lobes in the perfused lung are clearly visible indicating potentially greater accumulation of fluorescent agent in the lobes of the right hand lobes of the organ than in the left lobe, whereas the non-perfused lung image offers less local information.

Lung tumour transgenic mouse models were predominantly utilised throughout this study due to this being the primary interest and expertise of the host's organisation. These solid tumours exhibit high level expression of the $\alpha_v\beta_3$ integrin due to accumulation of endothelial and smooth muscle cells triggered by the the pre-explained process of angiogenesis. An advanced liver cancer c-myc mouse model was also employed as a comparison and as a measure of the versatility of this science.

The transgenic mice models are briefly explained below. The mice were all bred in-house at ITEM Hannover and were of varying ages depending on the aggression, and thereafter progression of tumourigenesis, which is directly related to the lifetime of the mice.

2.9.4 Transgenic mice

The creation of transgenicity involves the integration of genes into a germline and their subsequent successful expression in a host. This provides an opportunity to study the

function of a certain gene in the initiation and propagation of disease.[100] Transgenic mice are useful as they can be employed to mirror human disease. They are generated by microinjection of mutated linear DNA into the male pronuclei of fertilised eggs from hybrid female mice. Viable eggs are then transferred into the oviduct of pseudopregnant surrogate recipient mice. The newborn transgenic founder mice are grown and confirmed by tail biopsy DNA dot blot assays, before being mated for propagation of transgenics.[100-103]

2.9.4.1 SP-C raf

This was the predominant transgenic line used in these studies. SP-C is a lung-specific Surfactant Protein 'C' promoter that contributes to the reduction of surface tension in the lung and the facilitation of gas exchange, it is one of a family of four known surfactant proteins. The expression of SP-C is restricted to AT-II cells whereas the other three are secreted by other cells too, therefore the promoter regions of this surfactant protein gene are appropriate candidates for the construction of lung specific mutations.[101]
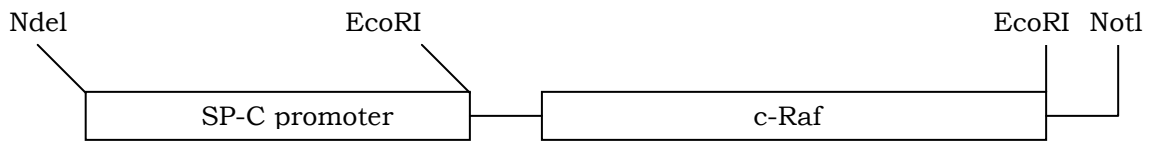
Figure 73: Cartoon showing the construction of an example linear DNA strand delivered to the germline for the initiation of oncogenesis. Notl and Ndel are restriction endonucleases.

The SP-C promoter is used to generate transgenic mouse lines specifically overexpressing the oncogene c-Raf. The Raf protein bears a terminal NH_2 regulatory domain which mediates the interaction with Ras and a COOH terminal kinase domain. c-Raf expression is a critical parameter in tumour development. Overexpression of c-Raf kinase sensitises cells for Ras transformation where the kinase becomes activated and triggers a phosphorylation signalling cascade that leads to continual activation and proliferation of cells - cancer.[100]

2.9.4.2 AAT-myc

This gene construct is genetically predisposed to the formation of hepatocellular carcinoma (HCC). This liver specificity is directed by α_1 alpha antitrypsin (AAT) promoter.[102] Interestingly this construct has been shown to cooperate with a second transgene, ALB-DS4, to vastly accelerate HCC formation.[103]

c-Myc is a member of a group of regulatory proteins involved in controlling cell cycle entry, progression and differentiation. This oncogene is again frequently overexpressed in human adenocarcenomas, suggesting a direct involvement in carcinogenesis. Mice of this line are useful authentic models to study the disease.

2.9.4.3 SPC-myc

This transgenic mouse line again exhibits lung carcinomas as directed by the SP-C promoter, but the oncogene is the c-myc protein. In this instance the mice were a black furred strain and the fur had to be removed to prevent interference with the long wavelength excitation and emission.

2.9.4.4 raf-cox

This transgenic mouse line was only used in one instance during this study and it was within a different white-haired mouse species.

2.9.5 Fluorescence imaging

A GE Healthcare eXplore Optix MX small animal fluorescence spectrometer was employed for the *in vivo* imaging of the anaesthetised mice. In this study the machine was set up to utilise three lasers with specific discrete excitation wavelengths of 635, 670 and 785 nm. Pulse durations were set to 100 ps and the maximum optical power was 800 µW. The detection system operated by means of a time-correlated single photon counting system with a photomultiplier tube detecting emission within a spectral range of 450 – 900 nm. eXplore Optix with automated data 2D image processing and display image analysis software was the operating system.
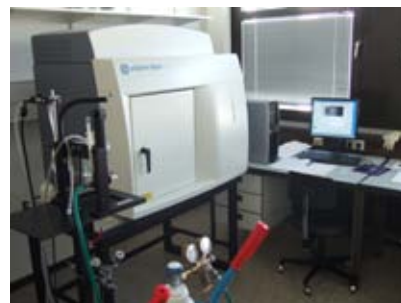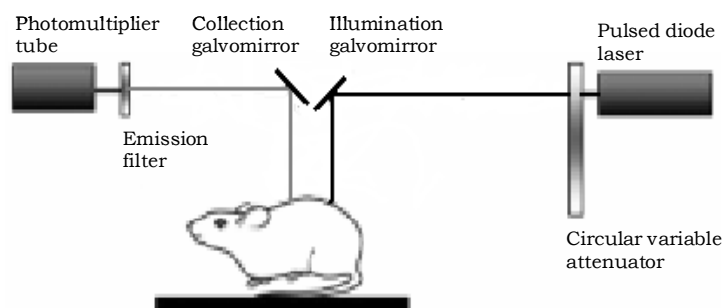
Figure 74: A highly simplified cartoon illustrating the important components of the imaging machine (**right**).

Put simply the machine functions by illuminating the subject with a short subnanosecond laser pulse and then a fast detector measures the distribution of the photons emitted as a function of time relative to the specific location of excitation. This is known as time-domain (TD) imaging.[104] It is important to consider that the laser light pulse propagating through different tissue types is broadened and attenuated due to background scattering and absorption, then having reached an embedded fluorophore, the resulting fluorescence is similarly affected by these factors although this is minimised through the use of near-IR fluorophores.

The ambiguity and complexity of live mammalian and specifically *lung* imaging became all too apparent in this study. Some of the problems encountered are well known.[105-108] A mammalian body is deemed a 'dirty' system where metabolic and excretory pathways and the complexities of the immune system, especially when hyperactive, combine to make the physiology and genetics of the mice from subtly dissimilar, to completely different from one another. This in some ways makes mouse testing a thorough and quite pseudohuman experiment, however for these initial studies, as will be discussed forthwith, the interpretation of images and development of a practical understanding of optimal machine settings were challenging. The results are interpreted somewhat tentatively and for a more conclusive study a greater sample would be required. The following results were obtained in less than a week of testing.

2.9.6 Results and discussion

2.9.6.1 Visibility and stability: NIR797

The NIR797 cyanine dye proved a very promising *in vivo* fluorophore. The excellent detectable emission intensity combined with great versatility in simple synthetic coupling protocol to primary amines, renders this a very useful organic fluorophore. NIR797 is also ideal when considering the capabilities of the optical spectrometer settings and available laser excitation wavelengths, 785 nm excitation proved perfect for the purposes of this work.
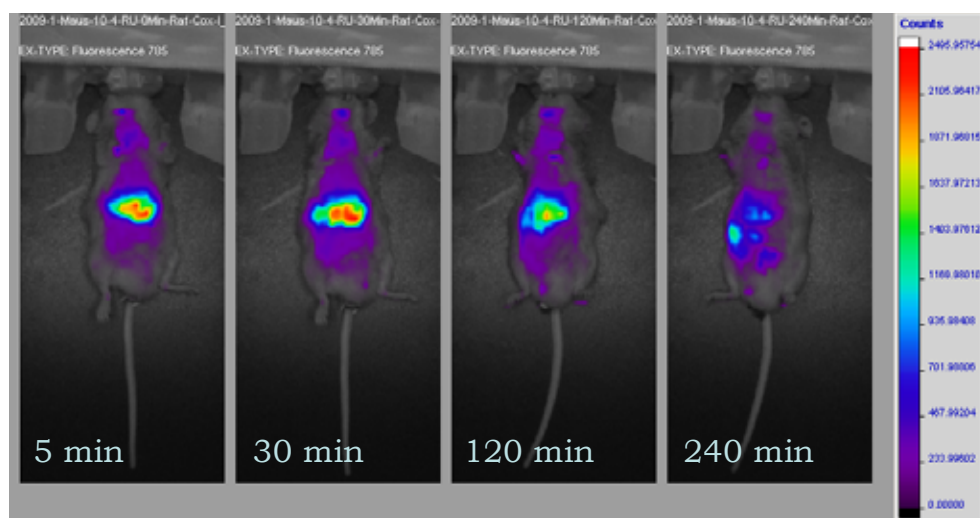
Figure 75: Early images of a cPep-NIR797 (**61**) injected raf-cox transgenic mouse showing what appears to be rapid accumulation in the renal region. The tumour-bearing lungs are higher up, above the forelegs, and show low level (purple) fluorescence. The fluorescence intensifies in what appears to be the liver over the first 30 minutes and then reduces as the metabolic excretion pathways eliminate the probe through the digestive tract.

The stability of cPep-NIR797 was immediately thrown into question (**Figure 75**). The first few images showed what was thought to be either preferential strong binding to the liver reticular endothelial system or efficient macrophage scavenging and metabolising of either the whole or broken down probe. The stability was tested by injection of cPep-NIR797, a 30 minute incubation period, mouse sacrifice and liver excision. The liver was then frozen,

ground and extracted with Et_2O and MeOH in an attempt to retrieve any unbound probe. MALDI-TOF mass spectrometry revealed a trace of whole compound at the correct mass suggesting the compound is stable.

2.9.6.2 Tumour specificity and masking

Generally speaking the whole body fluorescent images although showing excellent fluorescence intensity (with NIR797 fluorophore), showed far higher signal intensity from the liver region than from the tumour site to which the cyclic peptide probes are targeted. This issue of abdominal signal masking the tumour uptake has been reported previously with some RGD $\alpha_v\beta_3$-targeted probes such as the ^{99m}Tc conjugate applied in SPECT imaging,[109] and in the RGD-Cy5.5 conjugate, which also showed only moderate integrin binding affinity as a monomer, but showed great improvement as a multimeric peptide conjugate probe.[110]

This will be discussed further in due course with relation to our results. The lack of immediate tumour visualisation when our NIR797-cyclic peptide conjugates were injected into mice could initially be viewed with a measure of disappointment. However, the evaluation and interpretation of the results has led to some important discoveries with regards to the penetrative requirements of the fluorescence excitation and emission and the capabilities of the machinery employed with relation to the anatomy and physiology of the transgenic mice.

The weak to seemingly non-existent fluorescence observed from the chest cavity where the lung tumours are located is likely to be due to the net amount of time that the lung is within range of the excitation laser and emission detector with relation to the abdominal region. Indeed, **Figure 76** casts some light upon the situation as the excised perfused lung shows strong emission, whereas the whole body image, pre-sacrifice, shows negligible fluorescence from the chest region and the signal is concentrated in the abdominal region. This clearly suggests that the lung is completely out of range of the detection.
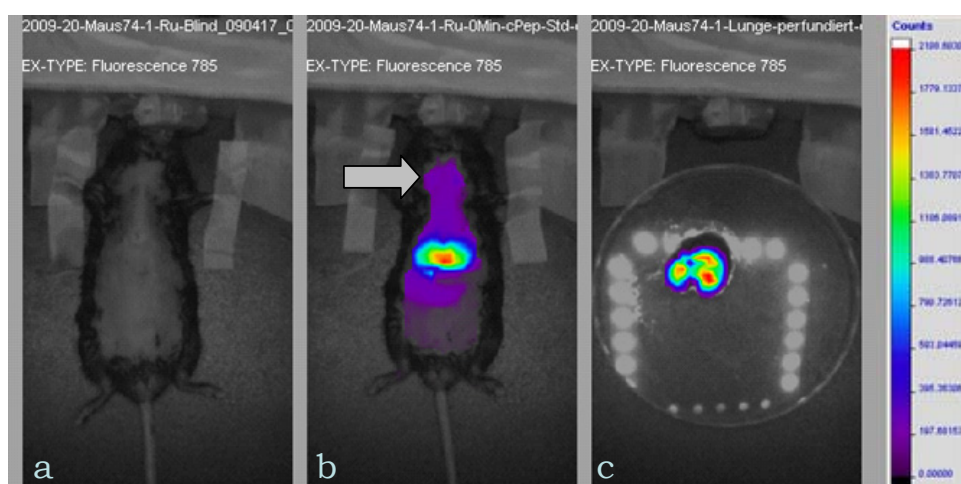
Figure 76: SP-C raf transgenic mouse with cPep-NIR injection: **a**) Blind; **b**) 30 minute post injection scan; **c**) The excised lung from the same mouse. The arrow highlights the chest cavity - the lung location in the live mouse.

This phenomenon could be explicable by considering the relative tissue penetrations of the excitation and emission. The excitation laser penetration and the subsequent maximum depth from which the near-IR emission penetrates back out through tissue and bone and is detected from the lung may well be intermittent as the lung is likely to be going in and out of range with each respiratory cycle, albeit a relatively rapid cycle of ~0.4 seconds.[114] Therefore it would seem plausible that the lung may or may not be within range of the laser excitation at any given time relative to the immediate position of the expanding and contracting organ. Alternatively, the laser excitation penetration could be sufficiently deep enough to excite the lung but the emission is experiencing quenching as a result of the depth of the lung, the tissue penetration of the emission and the inaccessibility of the rear right hand lobes and the left lobe where the tumour may be located relative to the detector as this is a two dimensional technique.

It is therefore important to consider the anatomy of the mouse and consider the optimum and maximum depth of tissue penetration for efficient specific fluorescence relative to the depth of the tumour bearing lungs.
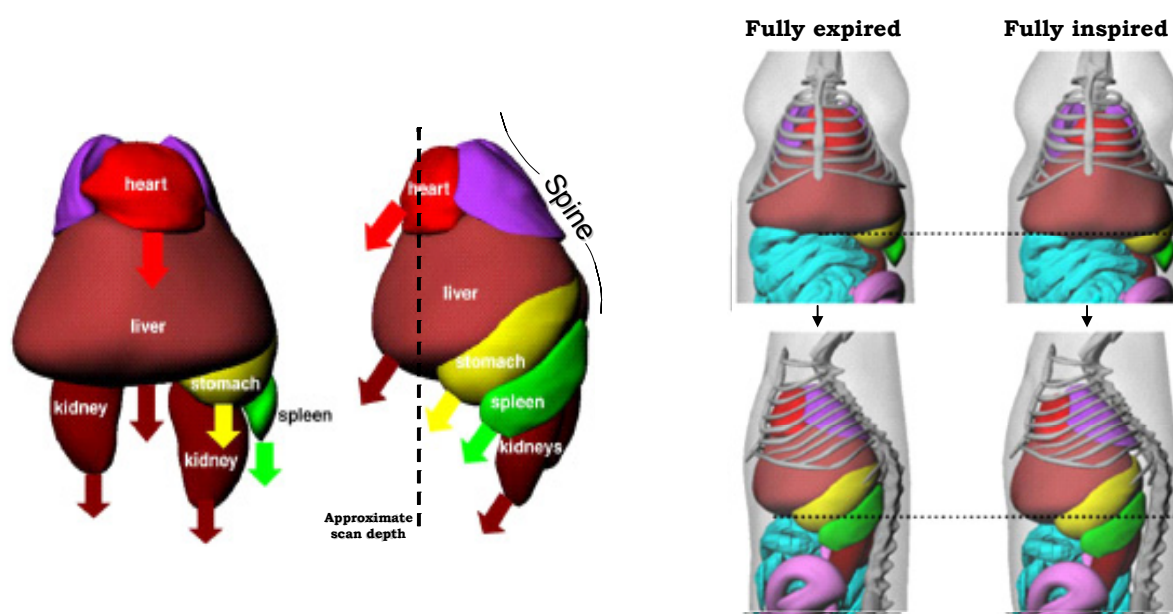
Figure 77: Adapted cartoons of a 4-D digital mouse phantom. **Left**) A representation of the direction of movement of organs during **lung** expansion, with an approximate depth of scan (5mm) used in our fluorescence imaging studies. **Right**) Anatomical images showing the position of the **lungs** to the rear and right of the thoracic cage. The dotted line indicates the relative positions of the diaphragm. Adapted from Ref. 111.

The lungs in a regular healthy mouse (**Figure 77**), are clearly predominantly to the rear and right hand side of the chest cavity even when fully inflated (typical volume change of ~0.15 mL).[111] It is envisaged through study of this universal model that only the middle lobe of the right lung may be within range (the machine was set at approximately 5 mm excitation depth with 100 µW power). It is also interesting to see the sheer bulk and
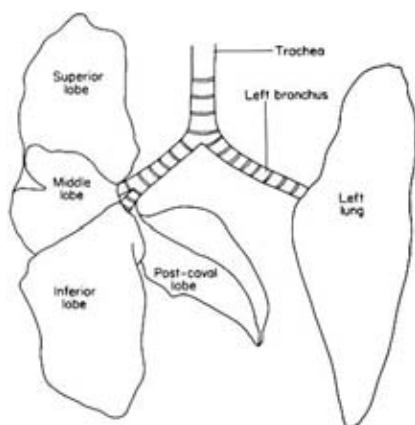
Figure 78: Diagram of the deflated lobes of mouse lungs.[112]

accessibility of the liver in relation to any other organ and hence any binding within, or excretion through this organ should be highly visible.

Given that the time between excitation pulse and detectable emission, the impulse response, is of 400 picosecond duration, it is therefore impossible that between excitation and emission the lung has contracted to a location too deep within the chest cavity for the emission to be detectable. It is therefore a fair approximation that the tumour bearing lung is either out of excitable range, or if excited, is partially or totally out of detectable emission range for potentially over half of the total scan time and even then perhaps only the right hand lobes are within range. The left lobe is deeper and more encompassed by the heart than the right lobes.

Research accomplished using an identical spectrometer investigated to some extent the effect of depth. **Figure 79** offers a simplistic idea of the difficulties associated with scan depth and penetration. The cartoon shows that if the probe is evenly spread throughout the sample and the laser fired to different depths, clearly the strongest signal is obtained at the shallowest depth. This is expected but serves to give a measure of maximum penetration. Future work would need to consider this as we have proved the power of our NIR797 probe.
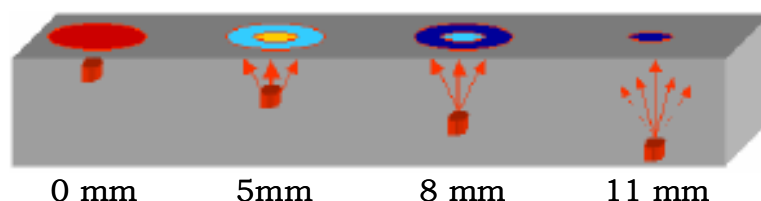
Figure 79: Cartoon illustrating penetrative issues associated with in vivo fluorescence imaging.[113]

Respiratory gating is a cunning technique employed in most micro-CT and MRI small animal scanning protocol. It very simply involves gating out the periods in the respiratory cycle when the lung perfusion is minimal and hence the signal is achieved at specific periods when the lungs are at maximum expansion. **Figure 80** below shows a graphical representation of respiratory gating and a comparison of micro-CT images between a non-gated (**a**) and gated (**b**) mouse lung slice. Unfortunately the spectrometer we employed in these experiments did not have respiratory gating capabilities.
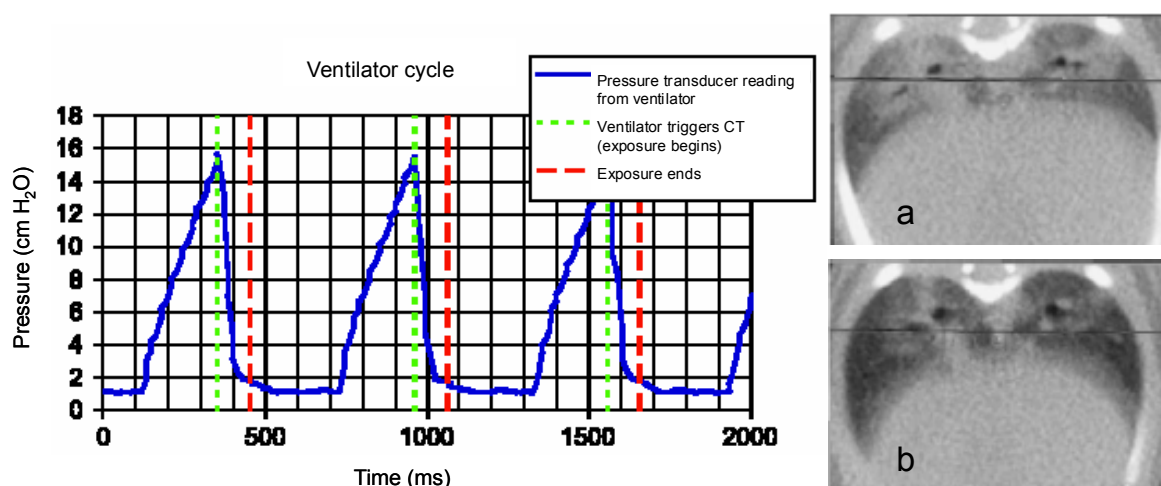
Figure 80: **Left**) Respiratory gating: graph depicting the breathing cycle of a mouse and the triggering of the micro-CT acquisition when fully inspired. The micro-CT acquires signals for 100 msec and the average breath cycle time is 600 ms. **Right**) The resulting micro-CT images **a**) normal breathing (no gating); **b**) respiratory gated. Adapted from Ref. 114.

It is important however to consider that the lungs of a transgenic mouse, depending on the extent of tumour progression, are generally of far greater size and mass than the lungs of a healthy mouse. **Figure 81** shows this enlargement due to tumour burden. In the diseased lung (**b**), the tumours are white specks. Therefore it is plausible that this engorgement could work in our favour and that the depth issue may not be quite as pronounced as the images would suggest.
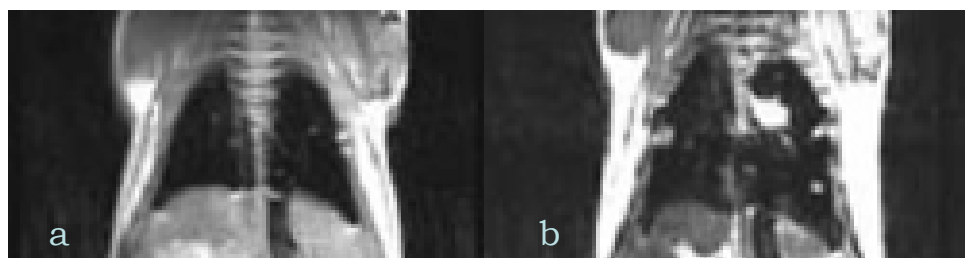
Figure 81: MRI size comparison between (**a**) healthy and (**b**) diseased lungs.[115]

However MRI and micro-CT images (**Figure 82**) of mouse lung adenomas still confirm the suspicions that the tumours are predominantly beyond the optimal depth of scan. These images show the dramatic contrast in lung depth depending on the level of the slice through the rib cage. The left hand image shows the lower lung and the large liver is clearly visible whereas the right hand image shows the lung filling the chest cavity but for the heart shielding the left hand lobe. The lower slice (left) highlights that even when the lung is perfused and bloated with tumour burden it is still completely out of scan range whereas the higher slice (right) shows clearly that one or more right hand lobes are within range. However promising this may appear it must be considered that the vast majority of the

lung is still out of range and hence the probability of visualising a tumour is slim unless deeper laser excitation and tissue penetration can be achieved to build up a three dimensional image.
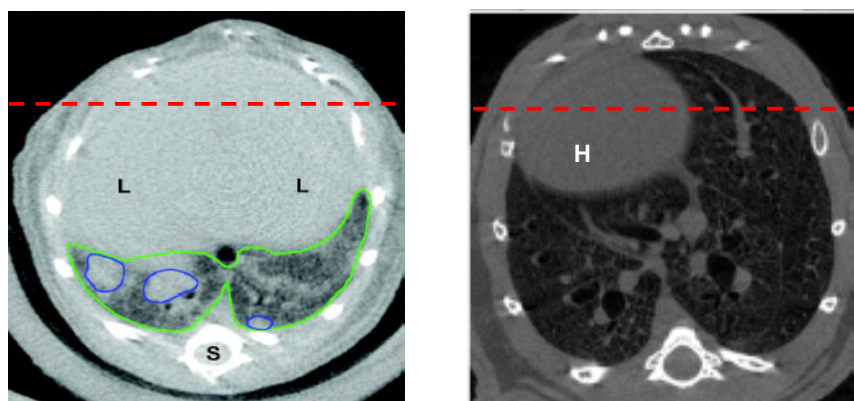
Figure 82: **Left**) Respiratory gated micro-CT image showing a lower chest slice with the lungs highlighted in green and tumours in blue. L = liver, S = spine. Adapted from Ref. 168. **Right**) Respiratory gated mid-lung slice. H = heart. Adapted from Ref. 169. Dotted lines in both images represent an approximate 5mm scan depth.

The results show that the stationary and accessible abdominal region reveals consistently far stronger fluorescence intensity than the chest. This has the effect of masking any lung fluorescence as shown in **Figure 83**, where an upper-limit has been set on the abdominal fluorescence intensity allowing any emission from the chest to become easier to see.
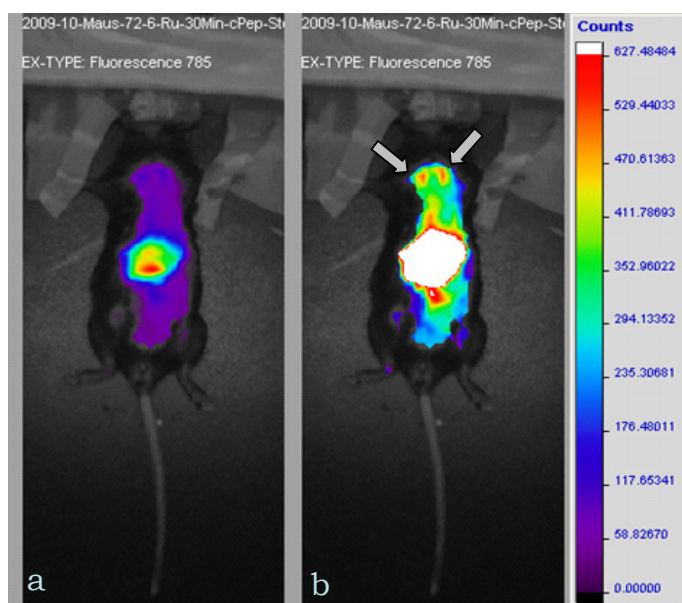
Figure 83: Transgenic mouse 30 minutes post injection of cPep-NIR797. **a**) No upper limit; **b**) The same mouse with an upper limit set at 130 counts with arrows highlighting perceived lung fluorescence.

All this discussion of the lungs being out of range and the implication that this may have been an inappropriate tumour model in which to test the synthesised fluorescent probes,

may at first seem all quite speculative. A measure of proof however, lies in the images of the lung whilst still functioning inside the mouse compared with the same lung excised. **Figures 76 & 84** reveal that the counts observed for the excised lung are comparable with those observed from the abdominal region in the full body image prior to sacrifice.
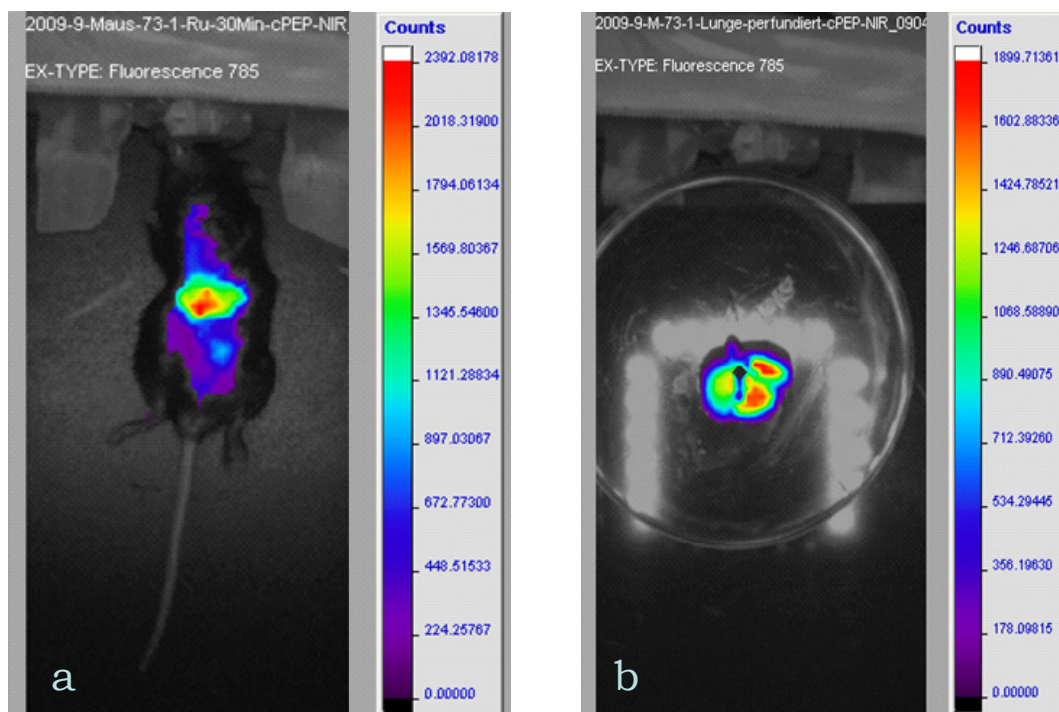
Figure 84: **a**) SP-C raf transgenic mouse injected with cPep-NIR797 (**61**), 30 minute scan; **b**) Excised lung from the self-same mouse.

The AAT-myc transgenic mouse liver tumour model serves to reinforce this proposition that there are problems associated with lung depth. It also offers a beam of hope for this technology.

Figure 85 reveals the whole body image of the advanced distended liver tumour. The intensity of the fluorescence from the tumour mass is clearly visible suggesting that not only is this a good mouse model to be studying with good binding and high levels of $\alpha_v\beta_3$ integrin but that the masking of tumour fluorescence by the strong abdominal signals an be overcome with careful imaging of the area of interest.
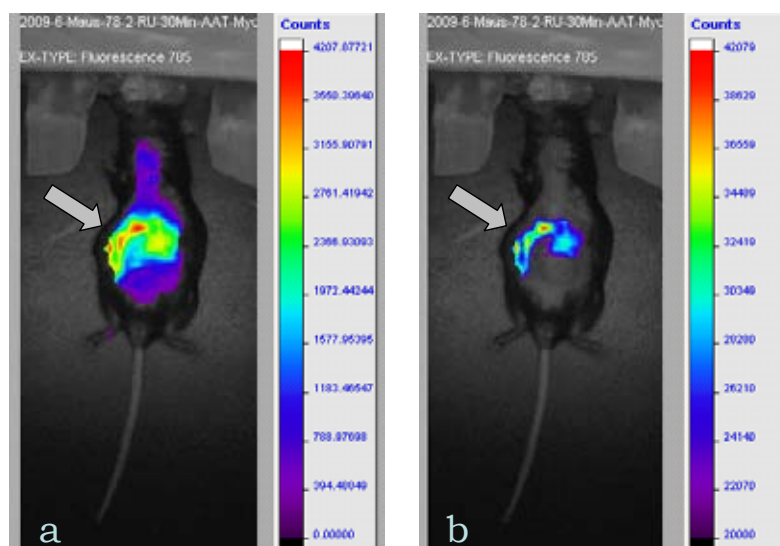
Figure 85: AAT-myc liver tumour mouse with cPep-NIR797 (**61**) injection. 30 minute images. **a**) No lower limit; **b**) Lower limit applied. Arrows highlight tumour site.

Many questions arise as to the nature and kinetics of the metabolism and waste management pathways and processes of transgenic mice in comparison to healthy mice.

Figure 86 shows a more body-wide spread of fluorescence emission throughout a healthy non-transgenic mouse. In a transgenic mouse the expulsion process seems much more efficient and the rapid accumulation and concentration of signal in the abdominal region suggests a hyperactive foreign body recognition and eviction mechanism is in place. This seems plausible as the mouse would be expected to mount a huge immune response to the malignancy.[127] These studies are therefore in accord with this evidence that there is an enhancement of immune system kinetics and metabolic pathways. Another possible contributing factor in lung tumour mice is that the swollen lung can impart restrictive pressure upon the heart; subsequently the heart beats faster to allow the normal respiratory reoxygenation mechanism to cope. This also enhances the rate of other

Figure 86: Healthy mouse cPep-NIR injection.

systems within the body including metabolism and the dispersion of fighting antibodies and scavenging macrophages.[128]

2.9.6.3 3-Dimensional reconstructions

Surgery is the cure or partial cure for many types of solid tumours. 3D imaging techniques offer far more accurate diagnostic options and the most recent optical spectrometers have 3D capabilities. Perhaps a more novel approach that has recently been shown to hold

potential is a dual modal imaging agent which combines an MRI active component for non-invasive 3D diagnosis, with a fluorescent component for accurate surgical excision. Another option could involve a combination of a PET-CT active isotope such as fluorine-18 with an optical fluorophore could be employed for similar purposes. SPECT is also available to be combined with the above advanced diagnostic imaging tools. It is envisaged that through a combinational approach more sophisticated and accurate imaging could be achieved. These techniques were discussed in greater depth in the introduction.

2D image   3D reconstruction
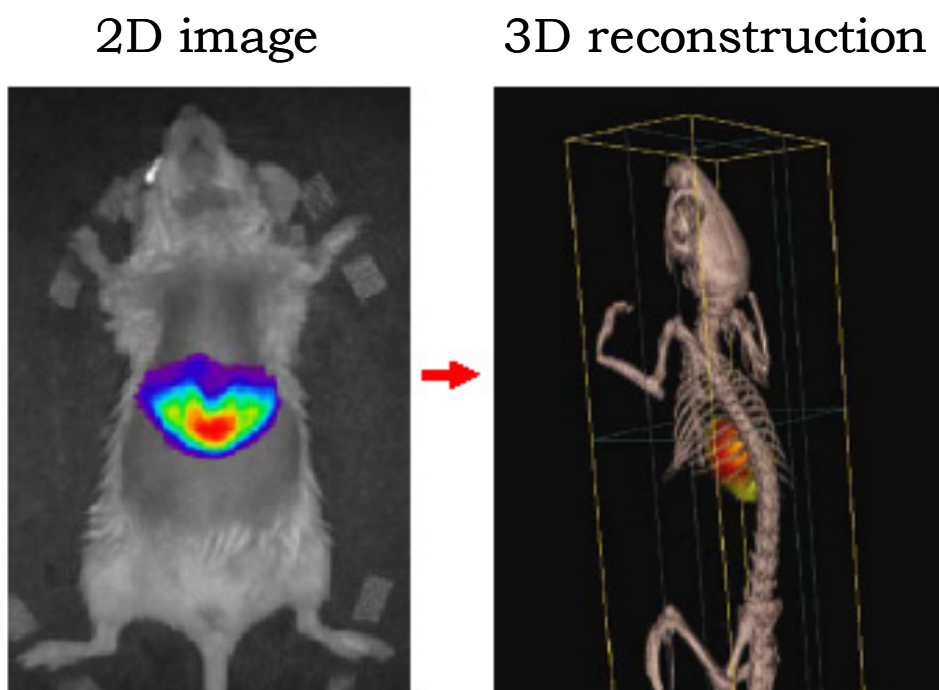
Figure 87: The potential of 3D optical imaging.[113]

2.9.6.4 Blocking studies

Ideally these studies would have been carried out using the same mouse with sufficient time between experiments for all traces of probe to have been excreted and the mouse to have recovered from the procedure. However due to time constraints the experiment was performed using different mice of the same transgenic class. The mouse was injected with non-labelled masking peptide and after an interval, during which the peptide is designed to block the integrin binding sites, the luminescent peptide was injected.

The blocking studies accomplished appear to have been successful due to the magnitude of the fluorescence intensity of blocked mice being far lower than that of a mouse not pre-injected with non-labelled cyclic blocking peptide, especially the 30 minute example (**Figure 88**). The incubation period is obviously variable and in this study a very swift five minute blocking period was initially attempted as firstly, the fluorescence is clearly visible after this time in previous experiments and secondly, it can be rationalised that even after five minutes there should be competition for binding sites and therefore there should be

reduced tumour-specific fluorescence. Literature blocking studies on different transgenic mouse strains have allowed one hour before introducing a probe to the system.[118]
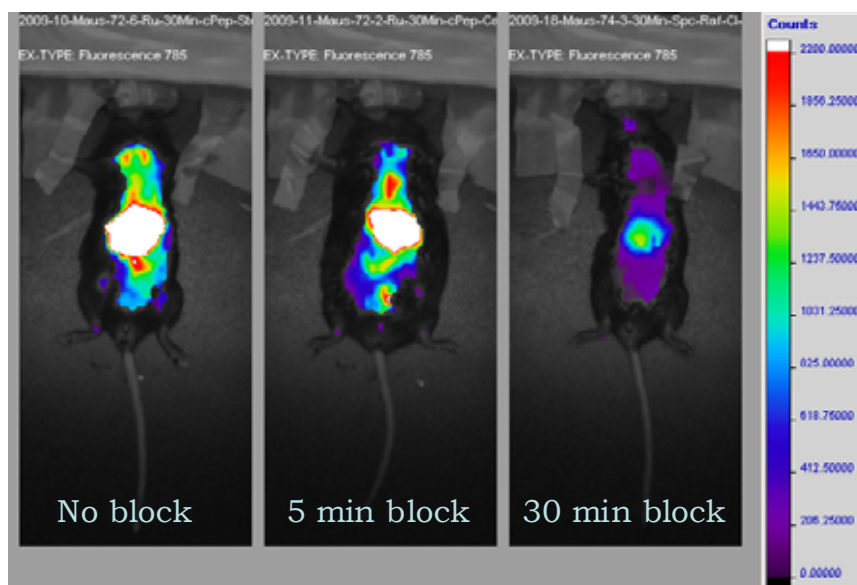
Figure 88: Three mice involved in blocking studies. All scans at 30 minutes post injection of cPep-NIR797.

It is abundantly clear that the efficacy of the standard blocking peptide is greatly enhanced when a longer time delay is applied. This confirms that the targeted cyclic peptide vector is an effective fluorophore delivery vehicle, but also suggests that the binding is strong and the kinetics of a desirable timescale.

Another possible factor of importance is dose. An overdose could easily lead to lower tumour specificity and a more body-wide distribution and therefore a greater accumulation throughout the excretion system. A much lower dose is likely to result in much greater economy of agent and therefore enhancement of specificity.

However, in this blocking study this is of less relevance as each mouse was given an identical dose. It could be argued that the intensity is also dependent on the extent of tumour development as each mouse will be different as described previously.

2.9.7 Appendix of all mouse images

1: AAT myc liver transgenic mouse with cPep-NIR797 (**61**) injection
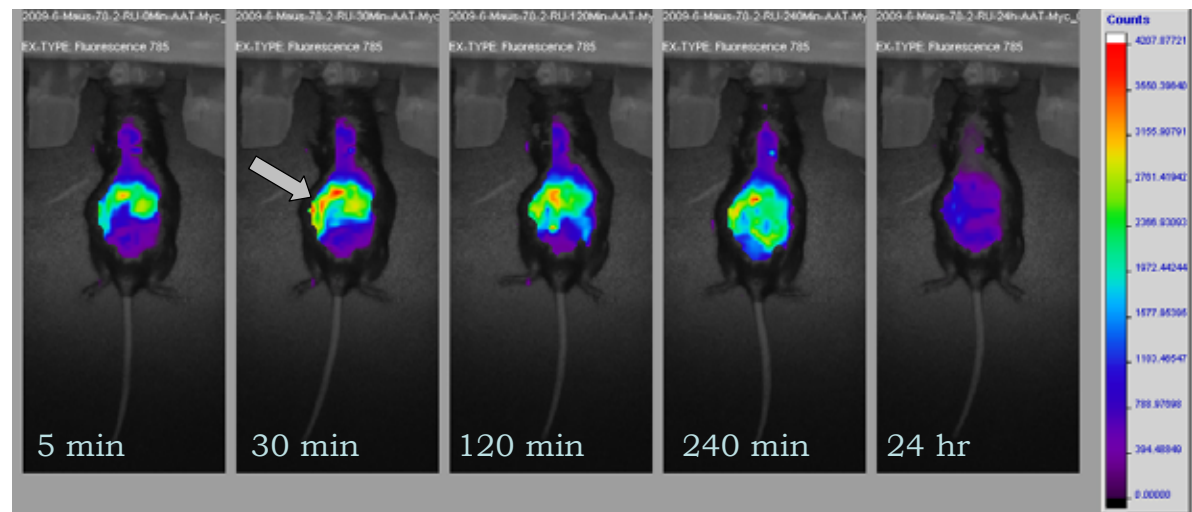
This is a very interesting and useful range of images from a specificity and kinetics perspective. Optimal tumour image was again 30 minutes post injection and it would appear that most of the fluorophore had been excreted within 24 hours. Compare with the image set **below**, where using the software a lower limit has been applied to the emission counts in an attempt to show the specificity of our luminescent probe to the tumour.
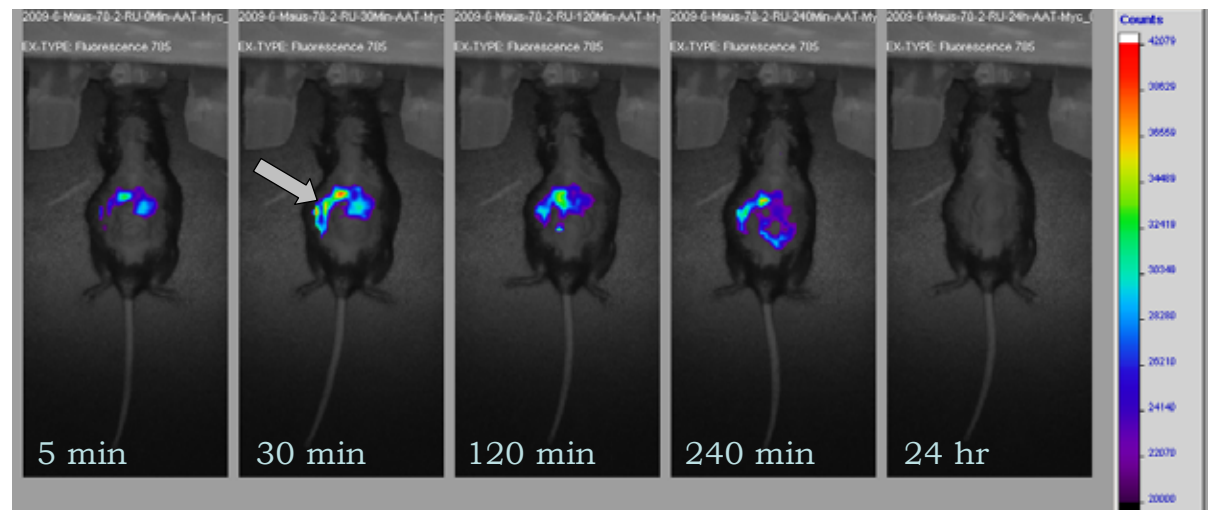
Compare and contrast these images with the non tumour-bearing AAT myc mouse below where a bodywide spread of imaging agent is shown with far less localisation.
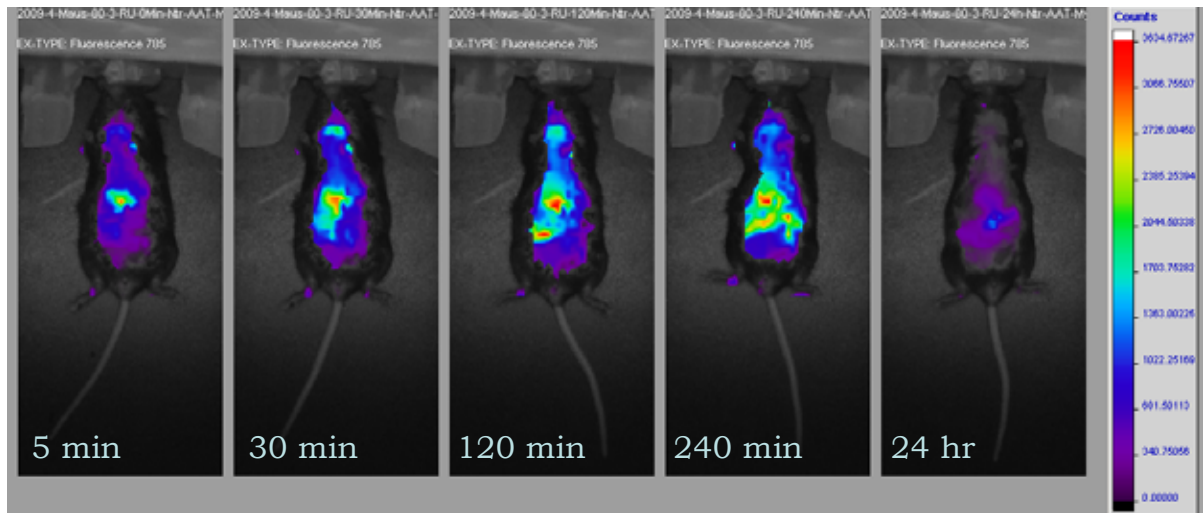
2: SP-C raf transgenic mouse blocking study: cPep std (**59**) injection, 30 minute wait, cPep-NIR797 30 minute scan, sacrifice, excise and image organs.
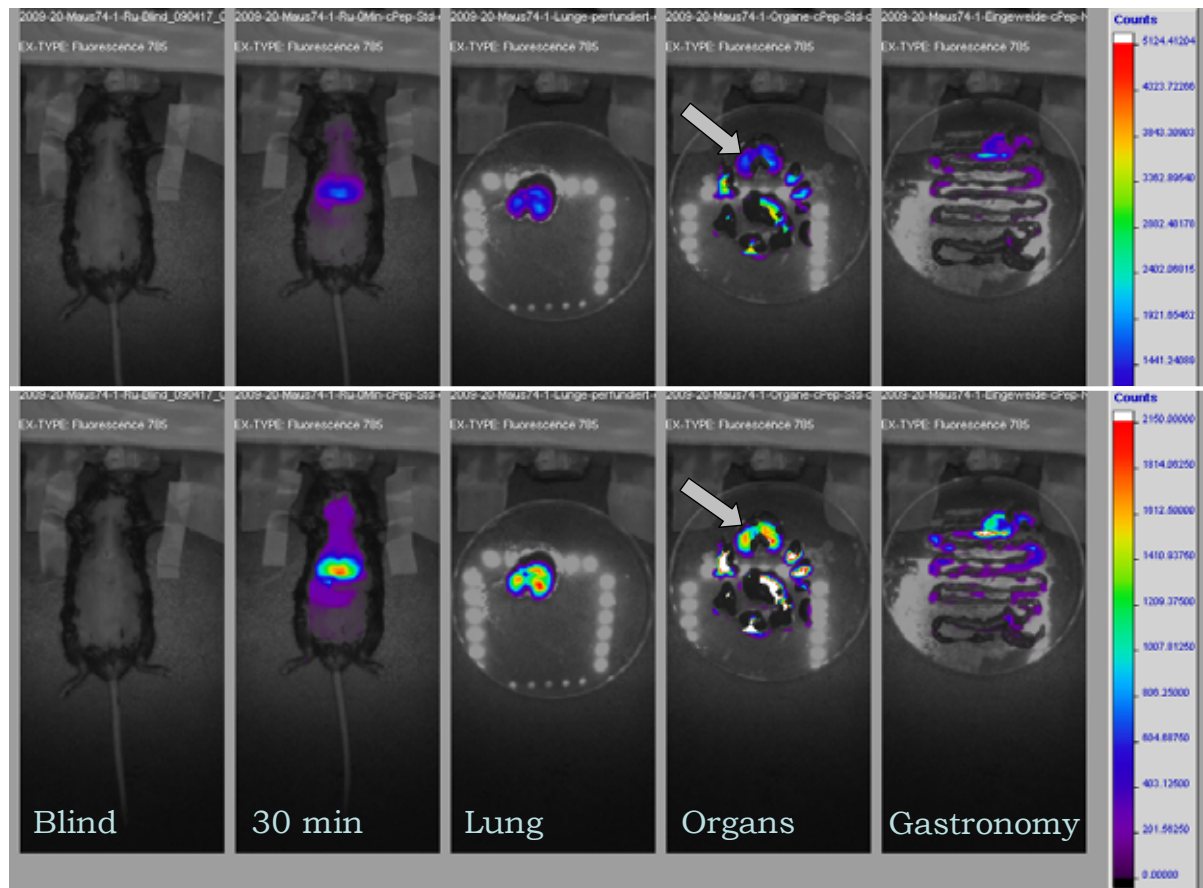
Again, these are two identical sets of images but the lower range has an upper limit of 2150 counts applied revealing the accumulation of fluorescent probe in the perfused lung. The same lung has been arrowed on the organs plate (the liver was dissected to carefully eliminate the gall bladder which in previous excisions had dominated the signal). However, part of the liver still dominates the signal intensity and this is clear as the upper limit cuts

it out to reveal lung florescence in the lower series. The kidneys also give strong signal and in the gastric plates there is fluorophoric presence in the stomach and small intestine. Importantly the black areas on the organs plates correspond to areas untouched by the laser either from being too shallow or too deep. This confirms to some extent the hypothesis that the depth of scan is absolutely key to accurate representation of the probe concentration in the tissues being imaged and that to build an accurate '3-D' picture perhaps the depth of scan needs to be adjusted several times as in fluorescence tomography.

3: SP-C raf transgenic mouse with NIR797 (**61**) injection.
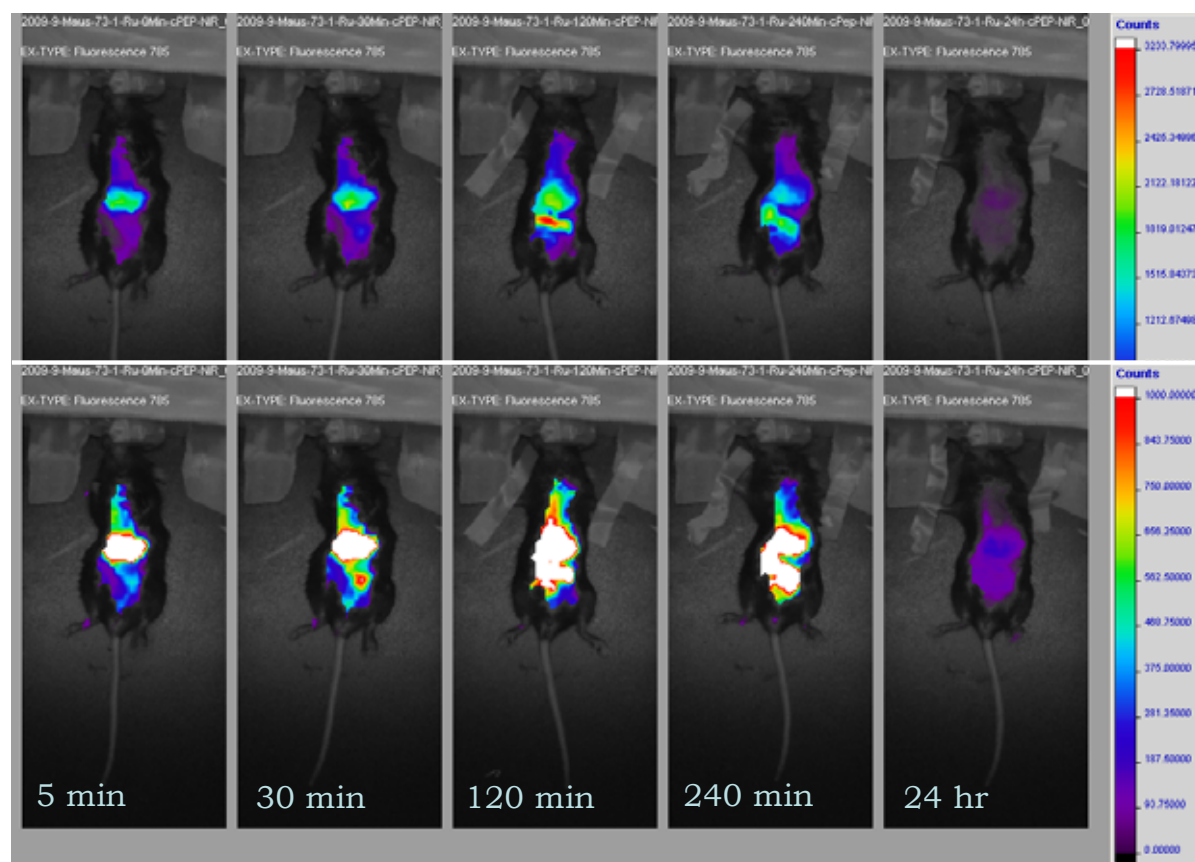
These image series of a lung tumour-bearing transgenic mouse are very telling not least because even with an upper limit there is no visible fluorescence from the lung region. Admittedly the front paws of the mouse have moved and are no longer held back but even so, the images, **below**, of the excised lungs show very strong fluorescence from the perfused and non-perfused excised organ adding evidence to the pre-discussed issues relating to the lung being out of range.

Also most apparent from this image is the importance of thorough hair removal as the dark pigmentation completely blocks the signal. Perhaps the position of the mouse also needs more consideration.

3b: Transgenic mouse with cPep-NIR (**61**) injection. Organs excised after 30 minutes. Lung images shown. **Left**) No upper limit set; **right**) upper limit set to reveal strong lung fluorescence in both perfused and non-perfused lungs.
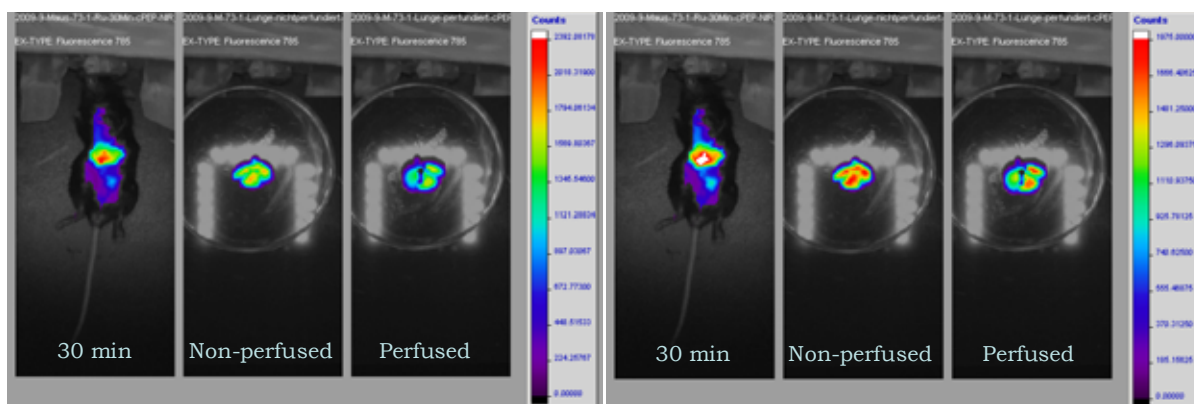
Organs

The significance of the excised organs images below is twofold; firstly the liver on the organs plate is completely black suggesting, initially, that it is not the liver emitting strongly in the abdominal region. The actual reason for the blackness of the liver is again due to depth of scan being set too low and the liver fluorescence being invisible. This is rationalised by the fact that the gall bladder fluoresces strongly as does the upper gastrointestinal tract on the gastric plate. Secondly, the masking effect of small regions of localised concentrated fluorescence easily leads to misinterpretation of images where it may seem there is no tumour specificity at first glance. This again calls for a 3D technique or the ability of the machinery to measure relative fluorescence counts at varying depths.

Organs: SP-C raf transgenic mouse with cPep-NIR797 (**61**) injection & sacrificed at 30 mins.
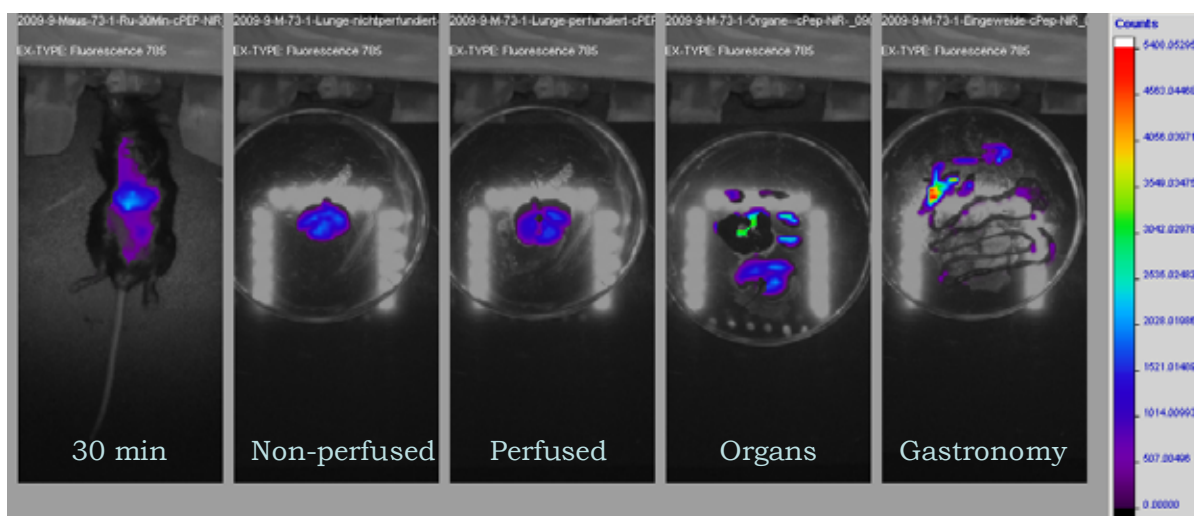
4: AAT myc Ntr transgenic mouse with cPep-NIR (**61**) injection.
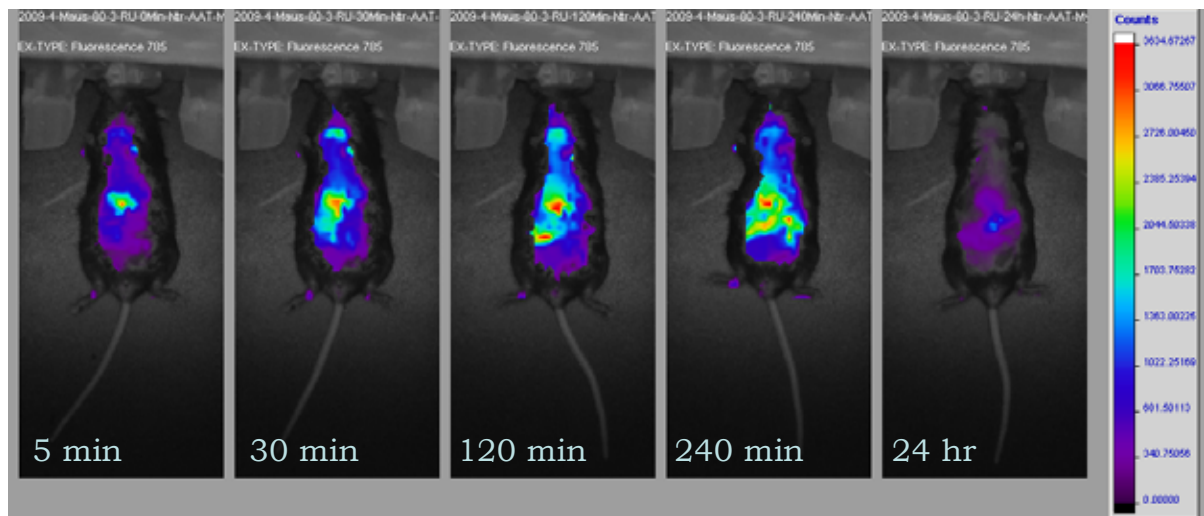
This is an example of a non tumour-bearing transgenic mouse where employing an upper limit reveals enhanced fluorescence from the chest cavity. Interestingly, due to a consistent administered dose, it is also obvious when compared with mouse '**1**' that the relative intensity has reduced suggesting less localisation and more of an even spread throughout the body, most of which is out of range.

Below is the same mouse with the upper limit of 2200 counts applied. It is uncertain as to what exactly is highlighted in the chest region (arrowed). Due to the aforementioned spread throughout the whole mouse it is envisaged that perhaps this is the heart rather than the lungs. This is in accord with the fact that tumour bearing lungs are enlarged and thus are perceived as being more visible, or within range, whereas healthy lungs are far smaller and predominantly located to the rear of the ribcage out of range. Therefore this is testament to the fact that this is likely a non tumour-bearing, or lowly expressing transgenic mouse.

4b: AAT myc Ntr transgenic mouse with cPep-NIR (**61**) injection with upper limit applied.
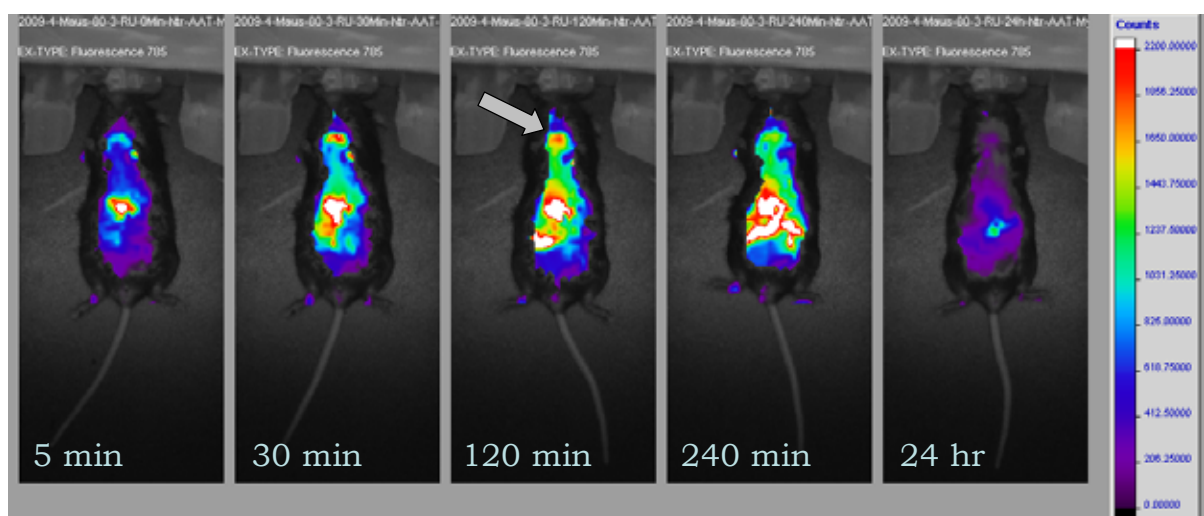
5: SP-C raf transgenic mouse cPep-Cap-NIR797 (**67**). **l**) No upper limit; **r**) Upper limit set.
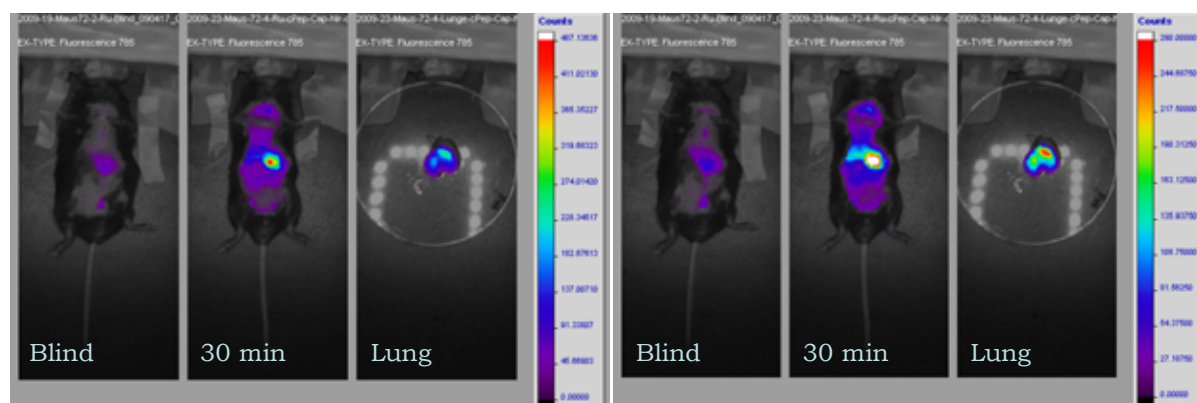
Firstly these images confirm that there is negligible difference in specificity when a caproic spacer is applied between the cyclic peptide vector and the fluorescent chromophore. Secondly, as previously discussed, excision of the lung reveals a far higher concentration of probe than the full body images would suggest.

6: Raf cox transgenic mouse 10.6, cPep-NIR797 (**61**).
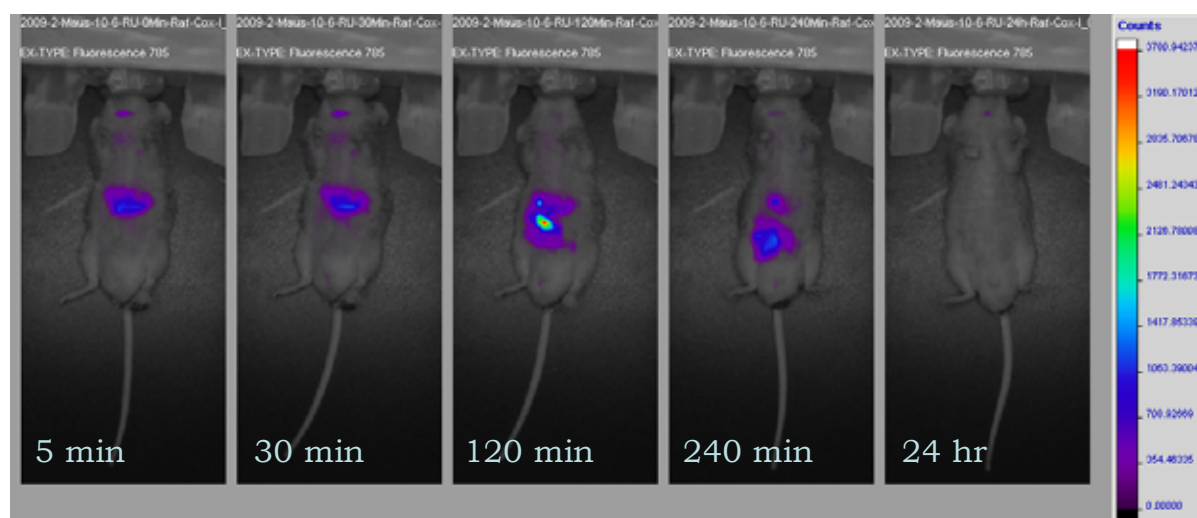
It is reasonable to assume that this mouse has none or relatively young tumours as excretion occurred completely within 24 hours. Emission intensity is reduced in comparison to tumour-bearing mice suggesting more even distribution throughout the vasculature with less localisation.

7: raf cox transgenic 10.4, cPep-NIR 797 (**61**).
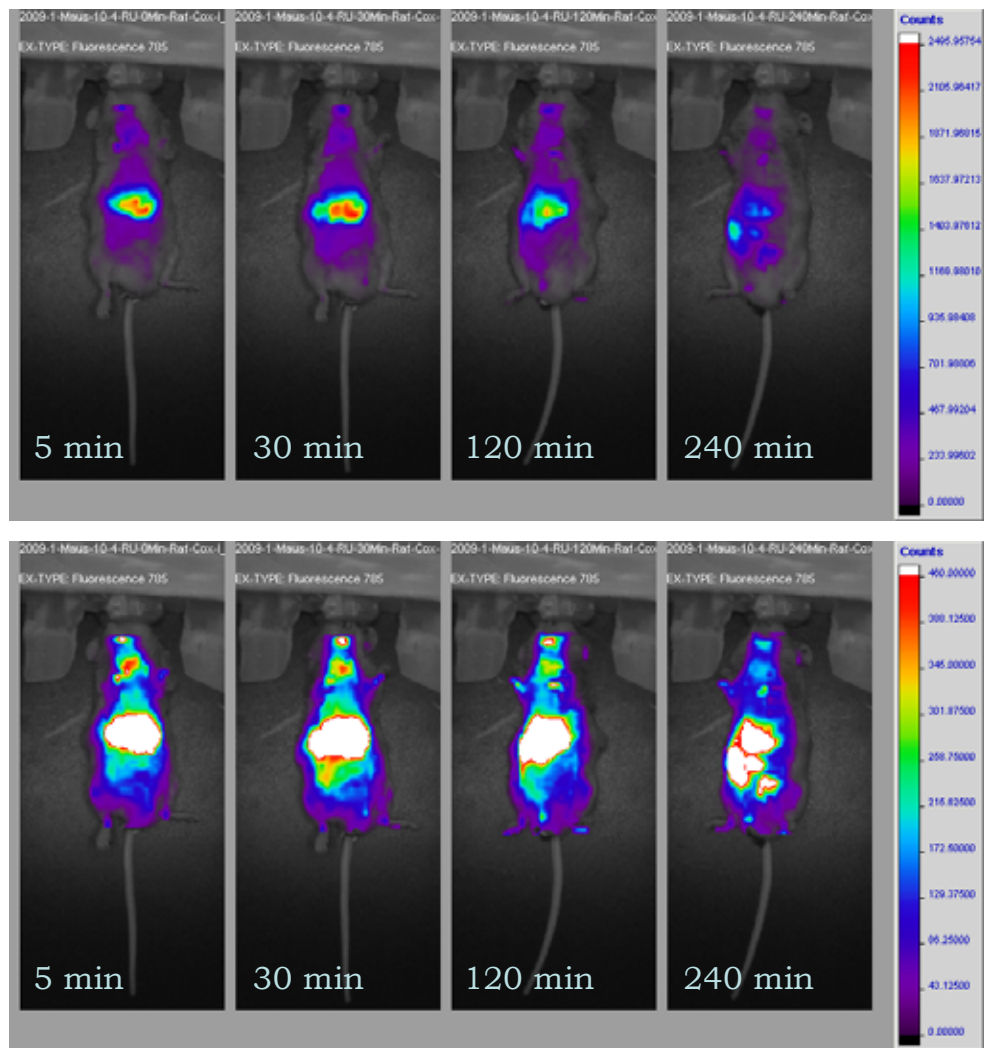
8: SP-C myc transgenic mouse cPep-NIR797 (**61**).
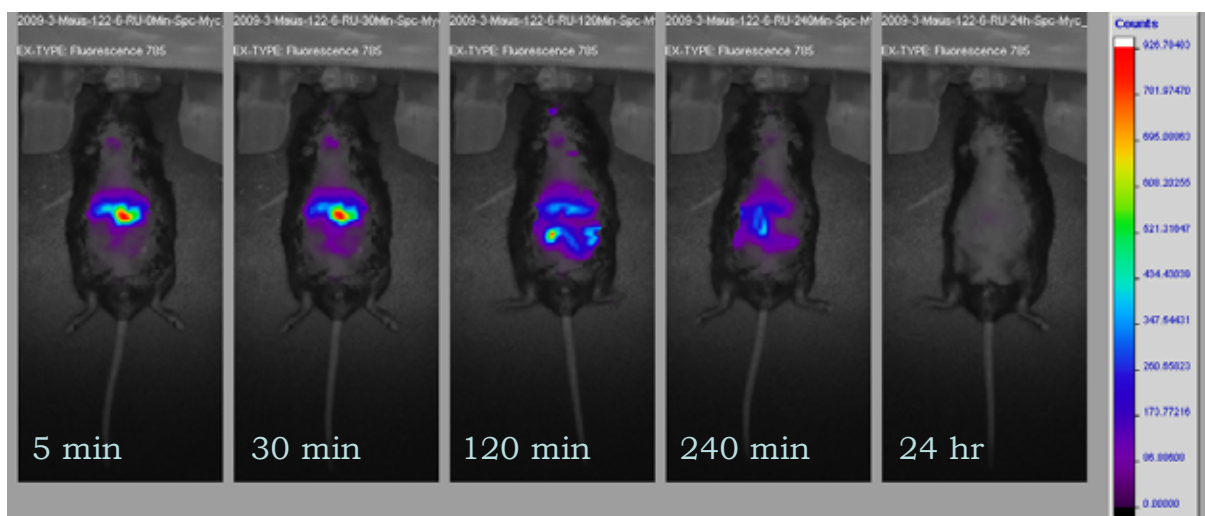
8b: SP-C myc transgenic mouse cPep-NIR797 (**61**), with upper limit.
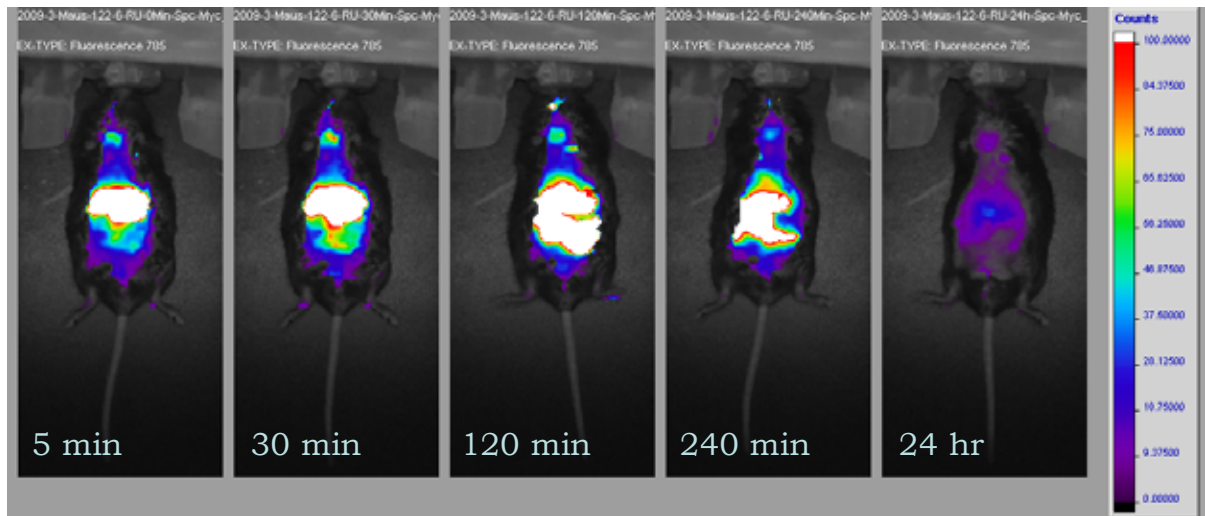
9: SP-C raf transgenic cPep-NIR797 (**61**). **Above**-no upper limit; **below**-upper limit set.
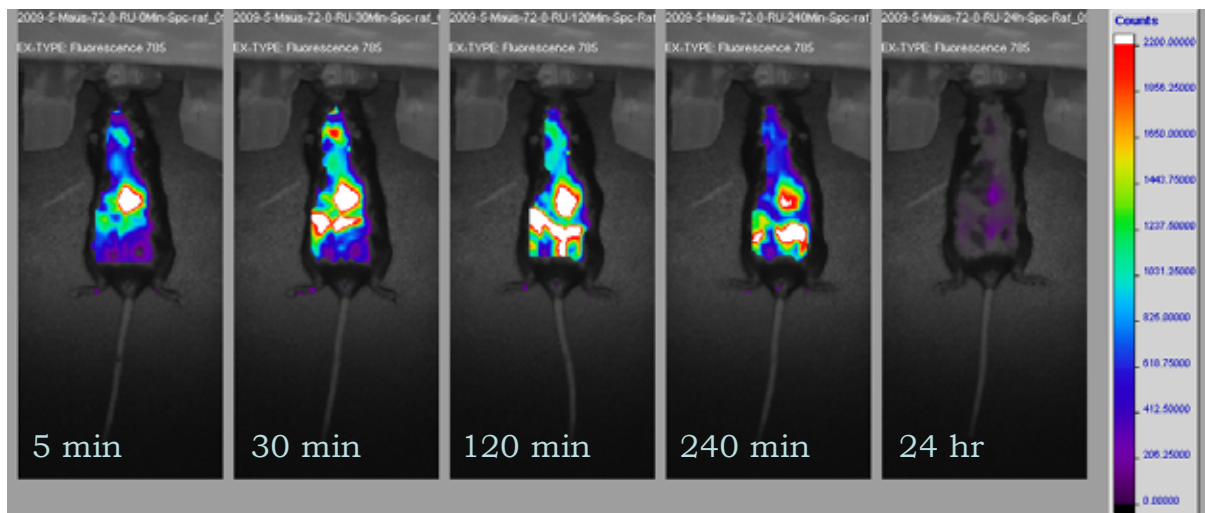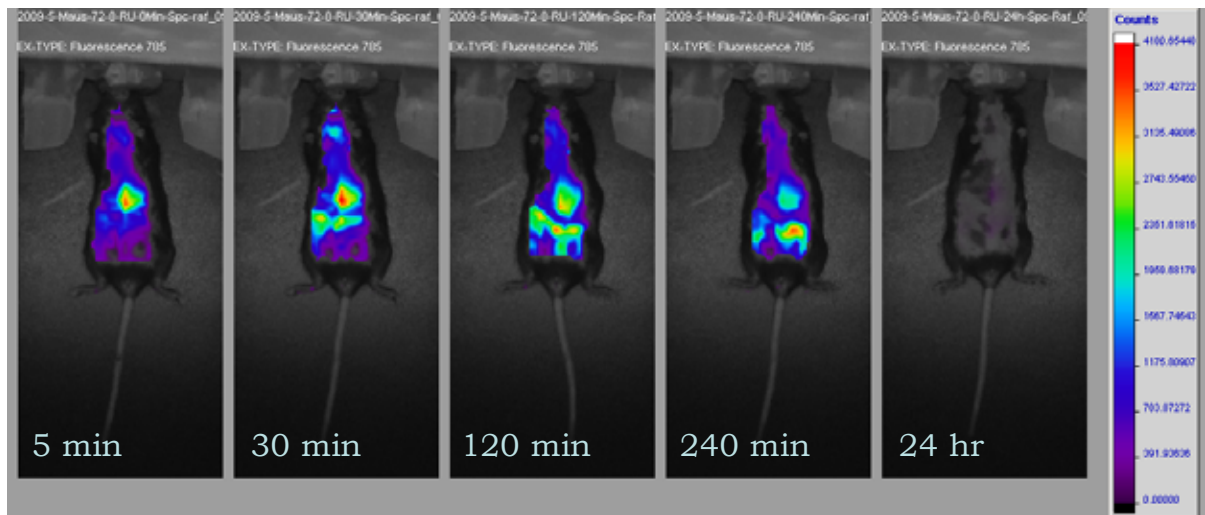
These images show very rapid metabolism and movement through the digestive tract but very little thoracic localisation. The heart is perhaps visible in the lower 30 minute image.

Osmium complex **68**

This metal based system is reliant upon intramolecular energy transfer as explained in the introduction. It proved inefficient as an *in vivo* imaging agent. The compound showed no detectable emission even when a maximum laser power was applied with alternative laser excitation wavelengths (see appendix 4). There was no suggestion of compound breakdown or issues with toxicology.

10: SP-C raf transgenic mouse injected with cPep-Cap-Os (**68**).
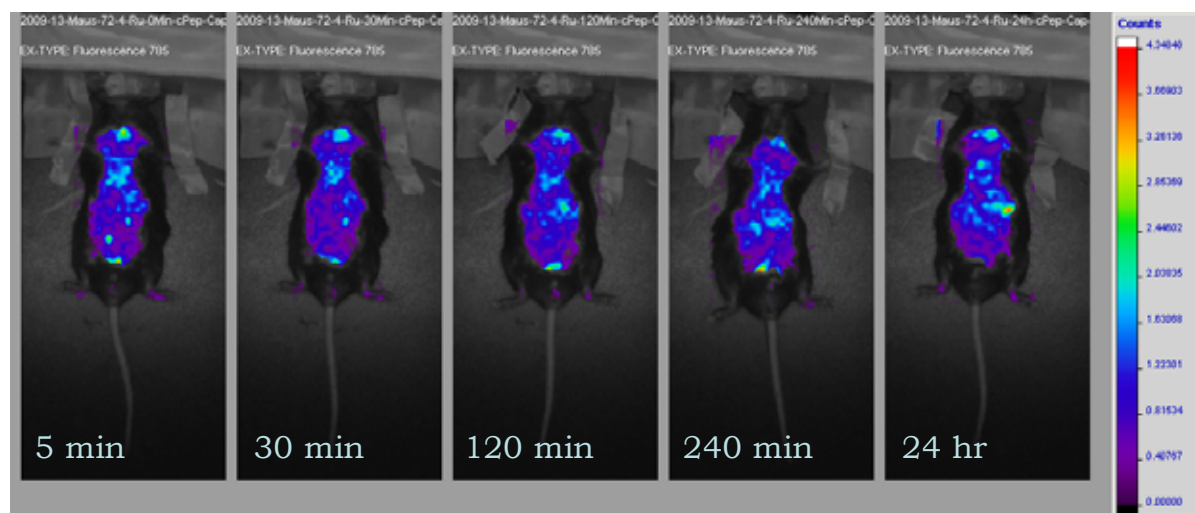
These images illustrate the inefficacy of the Osmium metal complex as a fluorescence imaging agent *in vivo*. Although the absorption and emission spectra of this probe were very promising due to excitation and emission showing in the near-IR region, this study confirms the realisation that the energy transfer mechanism of fluorescence renders the quantum yield and extinction co-efficient too low for observable *in vivo* imaging. This scan is identical to a pre-injection blind. Note the magnitude of the count scale.

The cPep-Cap-Os systems remain very promising for cellular imaging studies and possible mass spectrometric imaging using laser ablation inductively coupled plasma exchange (LA-ICP-MS) technology.

11: Different transgenic mice types at 24 hr post injection of cPep-NIR797.
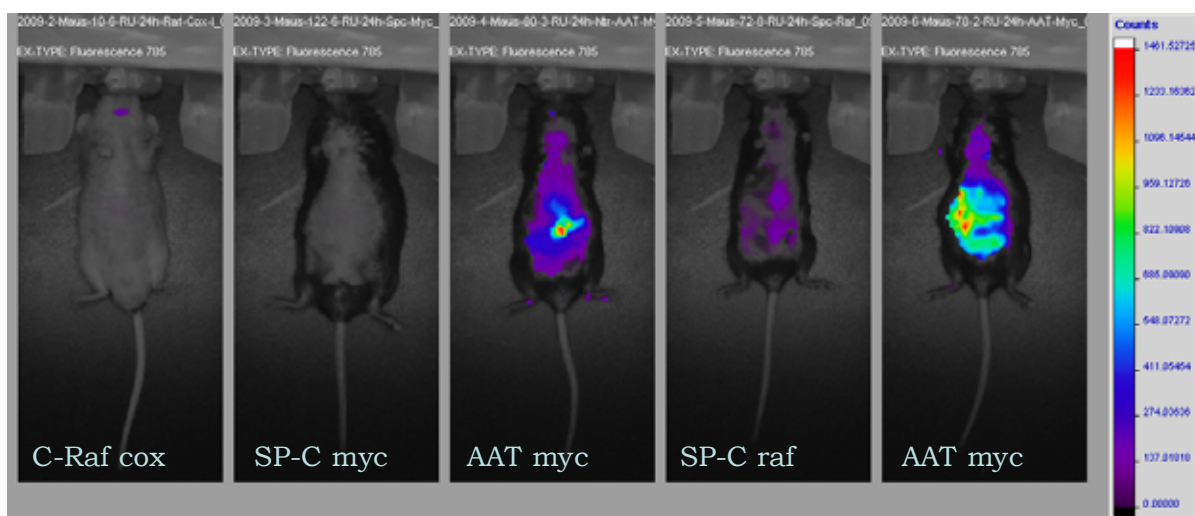
This is a very interesting comparison between different transgenic mouse types at 24 hours post injection showing clearly that AAT myc transgenic mice retain the fluorescent probe in vivo for a longer duration. The SP-C promoter lung cancer mouse models obviously have an enhanced metabolism, or, the binding in AAT myc transgenic mouse models is far stronger suggesting a greater prolificacy of the $\alpha_v\beta_3$ integrin.

Summary and conclusion

In summary the technique seems very dependent on the depth of the 2D plane of excitation and as such is unreliable for specific tumour imaging as other organs of the mouse also show uptake. This raises a questionmark over the tumour specificity of the cyclic peptide vector or the required dose (it is deemed that an overdose was applied in these studies), and the optimum time delay after injection before commencement of imaging. To answer the specificity question, unambiguous blocking study results need to be obtained on a greater sample quantity of mice. Optimal dose studies need to also be accomplished as, although 50 and 100 µL injections were compared from a fluorescence intensity perspective, the relative specificity was not studied. However, on a much more positive note, the system is obviously highly fluorescent in the near-IR region, and tissue penetration is more than sufficient for cellular, tissue slice and oncological biopsy studies. Different types of tumour (as was shown nicely by the liver tumour mouse), hold promise, especially if they are more accessible than the lungs. There are also questions over the accessibility of lung carcinoma integrins as there is very little extracellular matrix in lung tumours. The system does however hold great potential for 3-dimensional technology. Both the NIR797 label and the Osmium complex were non-toxic to the mice and easy to handle, store and were relatively stable. The lethargic solubility of the NIR797 cyclic peptides in saline is perhaps something that will be addressed through the incorporation of a PEG spacer unit.

In conclusion this has been a very interesting and worthwhile experience. In hindsight perhaps more shrewd design of the mouse experiments taking into consideration the anatomy of the mouse and perhaps altering its subsequent imaging position could have been applied. However, the results were very useful in determining the direction that the project should now take and being the first *in vivo* studies accomplished in the group it will serve as a good lesson to those continuing the research. It was a case of taking advantage of a great offer even though it wasn't perhaps what the compounds had been designed for.

Chapter 2: Experimental

This section incorporates all experimental data gathered for compounds successfully synthesised during this PhD. The compounds are presented in numerical order as referred to when discussed previously in the prose of the synthetic plan and results & discussion sections. The compounds have been split into six general categories as shown below.

This experimental section is followed by the appendices wherein lies a record of some of the most important or most significant data for unambiguous proof of characterisation, and to display data too prolific for the results and discussion section.

- Ugi experimental
- Linkers experimental
- Osmium experimental
- Peptide experimental
- Miscellaneous experimental

2.10 General information

2.10.1 Chromatography

Thin layer chromatography (TLC) was performed of pre-coated glass-backed silica gel plates: (silica gel 60, F_{254} s, 0.25mm thickness, supplied by Merck). Reversed phase TLC was performed using identical glass plates with RP-C18 F_{254} s silica coating. Visualisation was achieved using a variety of methods including UV light (254 / 365 nm lamp), basic $KmnO_4$ solution, acidic α-naphthol and acidic *p*-anisaldehyde.

Flash column chromatography was carried out using laboratory grade solvents on silica gel 40-63µ 60A, supplied by Fluorochem, UK. The eluent solvents and TLC visualisation technique are specified for each compound. Columns were run either under gravity, or with gentle pressure applied using hand bellows.

High performance liquid chromatography (HPLC) was performed on Dionex summit HPLC systems with Chromeleon 6.11 software. Analytical separation and purifications were acquired with the aid of a Summit p580 quaternary low pressure gradient pump with built in vacuum degasser. Similarly, for semi-preparative and preparative HPLC a high pressure gradient pump was employed and the identical multi-channel detector with a preparative flow cell. Helium degassed HPLC grade solvents were used throughout.

Phenomenex luna 10u C18 (2) 100A 10 micron columns were used of the following dimensions: Analytical 250 x 4.6 mm; Semi-preparative 250 x 10 mm with phenomenex 10

x 10 mm security guard cartridge system; Preparative 250 x 21.2 mm with phenomenex 21.2 x 60 mm security guard cartridge system.

A variety HPLC methods and solvent programmes (see below) were employed to maximise separation and streamline purification. The method used is specified for each compound and, where necessary, in the appendices.

2.10.2 HPLC methodology

Method	Solvent A	Solvent B	Run time
A	H_2O + 0.05 % TFA	MeCN + 0.05 % TFA	0 – 100 % B over **40** minutes followed by 20 minute 100 % B wash
B	H_2O + 0.05 % TFA	MeCN + 0.05 % TFA	0 – 100 % B over **50** minutes followed by 20 minute 100 % B wash
C	H_2O + 0.05 % TFA	MeCN + 0.05 % TFA	0 – 100 % B over **60** minutes followed by 20 minute 100 % B wash
D	H_2O + 0.05 % TFA	MeOH + 0.05 % TFA	0 – 100 % B over **40** minutes followed by 20 minute 100 % B wash
E	H_2O + 0.05 % TFA	MeOH + 0.05 % TFA	0 – 100 % B over **50** minutes followed by 20 minute 100 % B wash
F	H_2O + 0.05 % TFA	MeOH + 0.05 % TFA	0 – 100 % B over **60** minutes followed by 20 minute 100 % B wash

2.10.3 NMR spectroscopy

[1]H nmr spectra were recorded on Bruker AV300 and AVIII300 spectrometers. [13]C nmr spectra were recorded on Bruker AC300, AV400 and AVIII400 machines sing a pendant pulse sequence. Variable temperature (VT) ^{1}H nmr spectra were recorded on a Bruker AC500 spectrometer. COSY-90, HSQC and HMBC spectra were recorded on the aforementioned AVIII400 and AC500 machines.

Coupling constants (J) are given in hertz (Hz) and the multiplicities of spectral signals recorded as follows: 's' singlet; 'd' doublet; 'dd' doublet of doublets; 'ddd' doublet of double doublets; 't' triplet; 'q' quartet; 'm' multiplet; 'br' broad; 'stack' overlapping peaks arising from different protons. Chemical shifts (δ) are expressed in parts per million (ppm) downfield from trimethyl silane (TMS). Individual spectra are referenced relative to the residual solvent signal according to the values reported by Gottlieb.[244]

Off-line processing was accomplished using TOPSPIN, WINNMR and MestreC software.

2.10.4 Mass spectrometry

Positive and negative ion electrospray (ES+ & ES-) were measured on a Micromass LCT spectrometer using a methanol mobile phase. High resolution mass spectra (HRMS) were achieved using an appropriate lockmass.

Matrix assisted laser desorption ionisation time of flight (MALDI-TOF) mass spectra were recorded on a Bruker Biflex IV instrument and the matrix employed was gentisic acid with 0.1 % TFA unless otherwise stated.

Electron impact (EI) mass spectra were recorded on a VG Prospec spectrometer with a DS-90 data system. High resolution EI spectra were measured using perfluorokerosene (PFK) as an internal calibrant.

2.10.5 Infrared spectroscopy

A Perkin Elmer Paragon 1600 FTIR spectrophotometers was employed. Wavenumbers (v) are reported in cm^{-1} and the conditions are specified in the script.

2.10.6 Melting points

Melting points were determined in open ended glass capilliaries using Stuart scientific SMP1 apparatus and reported in degrees centigrade (˚C).

2.10.7 XRD

Single crystal X-ray diffraction data was recorded on a Rigaka R-axis 11c diffractometer equipped with a molybdenum rotating anode source and an image plate detector system. Alternatively a Bruker smart 6000 diffractometer with CCD detector and a sealed Cu tube source was employed.

2.10.8 Ultraviolet spectroscopy

Absorption spectra were recorded on a Shimadzu UV-3101PC UV-vis-NIR spectrometer using matched pairs of 1 cm path length quartz cuvettes.

2.10.9 Fluorescence spectroscopy

Yb - The sample was dissolved in the stated solvent and excited using the fourth harmonic from a pulsed Nd:YAG laser (266 nm) operating at 10 Hz. Light emitted at 90° to the excitation beam was focused onto the Spex TRIAX320 monochromator slits from which the appropriate wavelength was selected. The growth and decay of luminescence at selected wavelengths was detected using a germanium photodiode (Edinburgh instruments, EI-P) and recorded using a digital oscilloscope (Tektronix TDS220) for computational analysis.

Tb/Eu/Fluorescein – The samples were dissolved in the stated solvent and excited at selected wavelengths using a Perkin Elmer LS55 fluorometer or a PTI L-210M fluorescence system with a 75W xenon arc lamp as illumination source. Detection was achieved with a Shimadzu R298 PMT in a PTI model 814 analogue/photon counting multiplier. The emission monochromator was equipped with interchangeable 500 nm and 750 nm blazed

gratings. PTI Felix fluorescence analysis software for Windows produced the spectra. Excitation and emission spectra were recorded in solution with the stated solvent at ambient temperature. A 1 x 1 cm path length quartz cuvette with four transparent polished faces was used as the detection system was at 90º to the excitation source. Appropriate long-pass filters were used in each experiment.

2.10.10 Optical rotation spectroscopy

This data was obtained with the compound dissolved in either $CDCl_3$ or CD_3OD as stated, and inserted into an Optimal Activity PolAAr 2001 automatic polarimeter. $[\alpha]_D$ values are reported in units of $10^{-1}degcm^2g^{-1}$.

2.10.11 Reactions

Solution phase reactions were carried out, whenever appropriate, in flame or oven-dried (150 ˚C) glassware using laboratory grade solvents. Anhydrous solvents and reagents were either purchased or rigorously dried in the correct manner and stored under argon over 3Å molecular sieves.

2.11 Ugi experimental

Preparation of 1,4,7-tris-(ethoxycarbonylmethyl)-1,4,7,10-tetraazacyclododecane (**01**).

$C_{20}H_{38}N_4O_6$
Mol. Wt.: 430.54
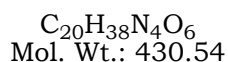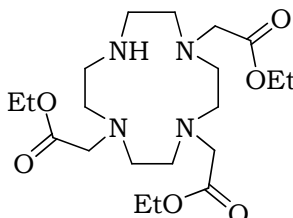
Ethyl bromoacetate (6.08 mL, 54.9 mmol), dissolved in dry CHCl$_3$ (20 mL), was added dropwise to a mixture of 1,4,7,10-tetraazacyclododecane (cyclen) (2.70 g, 15.7 mmol) and sodium hydrogen carbonate (13.17 g, 157 mmol) in dry CHCl$_3$ (40 mL). The resulting mixture was stirred vigorously at room temperature for 72 hours.

The thick white precipitate was removed by gravity filtration, washed with chloroform (3 x 10 mL) and the solvent evaporated under reduced pressure to leave the product as a viscous pale brown oil.

Purification by gradient flash column chromatography (silica, CH$_2$Cl$_2$ to CH$_2$Cl$_2$ / MeOH / NH$_{3(aq)}$ 9 / 1 / 0.05, R$_f$ = 0.59 in CH$_2$Cl$_2$ / MeOH / NH$_{3(aq)}$ system) yielded the pure product as a pale brown oil (4.77 g, 71%); ν_{max} (CHCl$_3$) / cm^{-1} 1735 (C=O); δ_H (300 MHz, CDCl$_3$) 0.96 (9H, t, J 7.1, 3 x OCH$_2$CH_3), 2.59, 2.62, 2.71, 2.81 (16H, 4 x s, 8 x CH_2 from cyclen), 3.11 (2H, s, NCH_2COOEt), 3.22 (4H, s, 2 x NCH_2COOEt), 3.86 (6H, q, J 7.1, 3 x OCH_2CH$_3$), 9.55 (1H, s, NH); δ_C (300 MHz CDCl$_3$) 12.1 (OCH$_2$$CH_3$), 45.2, 46.1, 47.1, 49.2 (4 x cyclen CH$_2$ envir.), 55.0 (NCH$_2$COOEt), 58.5 (OCH$_2$CH$_3$), 168.2, 169.1 (2 x C=O envir.); m/z (TOF ES+) 453.2689 ([M+Na]$^+$, 100 %, C$_{20}$H$_{38}$N$_4$O$_6$Na requires 453.2686).

Preparation of 1,4,7-tris-(ethoxycarbonylmethyl)-10-(*tert*-butoxycarbonylmethyl-1,4,7,10-tetraazacyclododecane (**02**).

$C_{26}H_{48}N_4O_8$
Mol. Wt.: 544.68

Tris-ethyl ester (**01**) (4.77 g, 11.1 mmol) was combined with dry Et$_3$N (7.79 mL, 55.4 mmol) in dry THF (20 mL). *Tert*-butyl bromoacetate (2.13 mL, 14.4 mmol) in dry THF (20 mL) was then added dropwise and the resulting mixture stirred at room temperature under N$_2$ for 72 hr.

The thick white precipitate was then removed by gravity filtration and washed with THF (3 x 10 mL). The solvent was then evaporated under reduced pressure and the flask subjected to high vacuum drying for 24hr to leave the product as a viscous pale brown oil (5.42 g, 89%); ν_{max} (CHCl$_3$) 1736 (C=O); δ_H (300MHz, CDCl$_3$) 1.20 (9H, t, *J* 7.1, OCH$_2$C*H$_3$*), 1.39 (9H, s, (C*H$_3$*)$_3$C), 2.77 (16H, s, 8 x C*H$_2$* from cyclen), 3.22 (2H, s, NC*H$_2$*COOR), 3.34 (6H, s, NC*H$_2$*COOR), 4.08 (6H, q, *J* 7.1, OC*H$_2$*CH$_3$); δ_C (300 MHz CDCl$_3$) 12.7 ((*C*H$_3$)$_3$C), 26.5 (OCH$_2$*C*H$_3$), 53.6, 53.9, 54.4 (cyclen *C*H$_2$ envir.), 59.9 (O*C*H$_2$CH$_3$), 80.8 ((CH$_3$)$_3$*C*), 171.7, 172.2 (2 x *C*=O); *m/z* (ES) 545 ([M+H]$^+$ 100%), 567 ([M+Na]$^+$ 11%). m/z (TOF ES+) 567.3370 ([M+Na]$^+$, 100 %, $C_{26}H_{48}N_4O_8$Na requires 567.3381).

Preparation of 1,4,7-(tris-ethoxycarbonylmethyl)-1,4,7,10-tetraazacyclododecyl ethanoic acid (**03**).

$C_{22}H_{40}N_4O_8$
Mol. Wt.: 488.57

To the tri-ethyl ester-mono *tert*-butyl ester derivative (**02**) (0.20 g, 0.37 mmol) was added TFA (3 mL) and CH_2Cl_2 (3 mL) at 0 °C. The mixture was then allowed to warm to ambient temperature and stirred for 16 hours. The solvent and excess TFA were then removed under reduced pressure and the viscous light brown oil dissolved and re-evaporated in CH_2Cl_2 (3 x 10 mL) in an attempt to remove all excess TFA. The crude product was then purified by preparative HPLC (Method E, t_r = 35.06 min) and lyophilised to afford the product as a fine white powder (0.17 g, 95%); ν_{max} (CHCl$_3$) 1746 (C=O), 3380 (COOH); δ_H (300MHz, CD$_3$OD) 1.25-1.34 (9H, m, 3 x OCH$_2$CH_3), 3.13 (8H, br s, 4 x cyclen CH$_2$), 3.46 (8H, br s, 4 x cyclen CH$_2$), 3.74 (6H, br s, 3 x NCH_2COOEt), 3.89 (2H, s, NCH_2COOH), 4.19 (6H, q, J 7.2, OCH_2CH$_3$), 4.25-4.32 (3H, q, J 7.2, OCH_2CH$_3$), 4.91 (1H, br s, COOH); δ_C (300 MHz CD$_3$OD) 11.9 (OCH$_2$$CH_3$), 48.2 (N$CH_2$COOEt), 51.1, 51.4, 53.2, 53.6 (4 x cyclen CH$_2$ envir.), 54.6 (NCH$_2$COOH), 59.4, 59.6 (2 x OCH$_2$CH$_3$ envir.), 167.9, 170.1 (2 x C=O envir.); m/z (TOF ES+) 489.2924 ([M+H]$^+$, 100 %, $C_{22}H_{41}N_4O_8$ requires 489.2918).
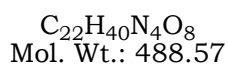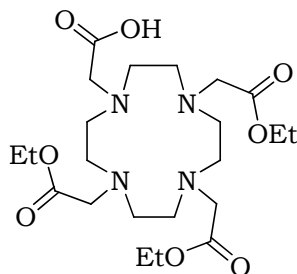
Preparation of Ugi 4-CCR bis-amide products: benzyl (**04**), naphthyl (**05**), pyrenyl (**06**) & indolyl (**07**).
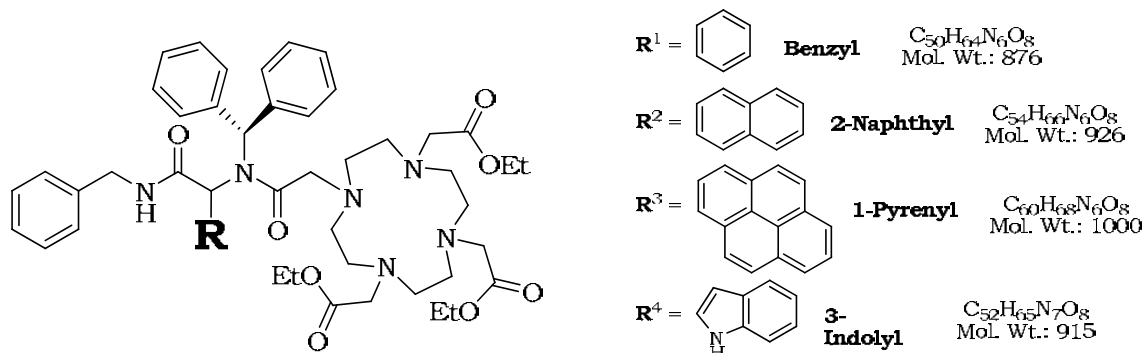
1,4,7-tris-(ethoxycarbonylmethyl)-1,4,7,10-tetraazacyclododecyl ethanoic acid (**03**) (0.14 g, 0.29 mmol) was dissolved in dry EtOH (5 mL) under nitrogen). Benzaldehyde (0.029 mL, 0.29 mmol), aminodiphenylmethane (0.049 mL, 0.29 mmol) and benzyl isocyanide (0.035 mL, 0.29 mmol) were then added simultaneously, and the mixture stirred vigorously at 65 °C for 72 hours. The solvent was then removed under reduced pressure and the crude product purified using gradient flash column chromatography (Silica, CH_2Cl_2 to CH_2Cl_2, MeOH, $NH_{3(aq)}$ 9 : 1 : 0.05, R_f = 0.43). The product was then further purified by preparative HPLC (Method B), and lyophilised to afford the product as a fine white powder.

04[$] (0.17 g, 68 %) R_f = 0.43; m.p. = 60-62 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1681, 1730, 1741 (3 x C=O), 3249, 3300 (NH); m/z (TOF ES+) 899.5683 ([M+Na]$^+$, 100 %, $C_{50}H_{64}N_6O_8Na$ requires 899.4695). Analytical HPLC t_r = 23.7 minutes.

05*[$] (0.22 g, 70 %) R_f = 0.40; m.p. = 66-68 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1683, 1736 (2 x C=O), 3246, 3296 (NH); m/z (TOF ES+) 927.5020 ([M+H]$^+$, 100 %, $C_{54}H_{67}N_6O_8$ requires 927.5001). Analytical HPLC t_r = 22.3 minutes.

06[$] (0.11 g, 36 %) R_f = 0.39; m.p. = 165-167 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1676, 1740 (2 x C=O), 3258, 3310 (NH); m/z (TOF ES+) 1023.4996 ([M+Na]$^+$, 100 %, $C_{60}H_{68}N_6O_8Na$ requires 1023.5010). Analytical HPLC t_r = 24.6 minutes.

07 (0.25 g, 86 %) R_f = 0.48; m.p. = 160-162 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1664, 1678, 1738, 1745 (4 x C=O), 3258, 3310 (NH); m/z (TOF ES+) 938.4792 ([M+Na]$^+$, 100 %, $C_{52}H_{65}N_7O_8Na$ requires 938.4769). Analytical HPLC t_r = 26.2 minutes.

*See **appendix 1** for XRD crystal structure proof of successful synthesis.

[$]Due to the conformational flexibility in the molecule the NMR spectra were too complex to assign. However see **appendix 2** for VT nmr studies.

Preparation of hydrolysed Ugi 4-CCR bis-amide products: benzyl (**08**), naphthyl (**09**), pyrenyl (**10**) & indolyl (**11**).
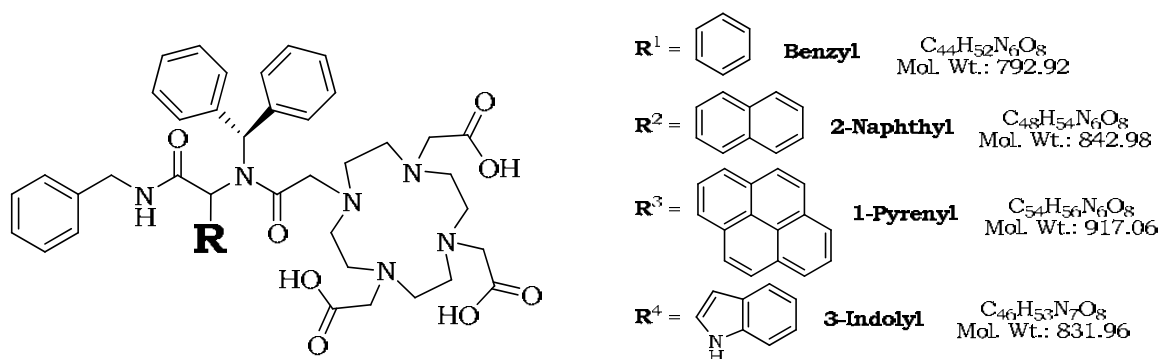
The Ugi products **04, 05, 06** & **07** (0.1 g) were dissolved in a 1:1 (v/v) mix of THF / H_2O (8 mL). LiOH (3.5 eq.) in H_2O (2 mL) was then added and the reaction mixture stirred for 24 hours at ambient temperature. Excess THF was then removed under reduced pressure and the mixture neutralised with 1M HCl. The product was purified using preparative HPLC (method D), and lyophilised to afford the product as a fine white powder.

08 (0.077 g, 85 %); ν_{max} (nujol)/cm^{-1} 1658, 1727 (2 x C=O); m/z (TOF ES+) 815.6831 ([M+Na]$^+$, 100 %, $C_{44}H_{52}N_6O_8Na$ requires 815.6822); m/z (TOF ES-) 791.3 (M-H), 100 %. Analytical HPLC t_r = 23.3 minutes.

09 (0.080 g, 88 %); ν_{max} (nujol)/cm^{-1} 1650, 1726 (2 x C=O); m/z (TOF ES+) 865.3901 ([M+Na]$^+$, 100 %, $C_{48}H_{54}N_6O_8Na$ requires 865.3910). Analytical HPLC t_r = 32.8 minutes.

10 (0.079 g, 86 %); ν_{max} (nujol)/cm^{-1} 1654, 1730 (2 x C=O); m/z (TOF ES+) 939.4057 ([M+Na]$^+$, 100 %, $C_{54}H_{56}N_6O_8Na$ requires 939.4048). Analytical HPLC t_r = 26.4 minutes.

11 (0.075 g, 83 %); ν_{max} (nujol)/cm^{-1} 1660, 1728 (2 x C=O); m/z (TOF ES+) 854.3853 ([M+Na]$^+$, 100 %, $C_{46}H_{53}N_7O_8Na$ requires 854.3867). Analytical HPLC t_r = 23.9 minutes.

Due to the conformational flexibility in the molecule, NMR spectra were too complex to assign.

Preparation of Europium complexed Ugi 4-CCR products: benzyl (**12**), naphthyl (**13**), pyrenyl (**14**) & indolyl (**15**).

$R^1 =$ Benzyl $C_{44}H_{49}EuN_6O_8$
Mol. Wt.: 941.86

$R^2 =$ 2-Naphthyl $C_{48}H_{51}EuN_6O_8$
Mol. Wt.: 991.92

$R^3 =$ 1-Pyrenyl $C_{54}H_{53}EuN_6O_8$
Mol. Wt.: 1066

$R^4 =$ 3-Indolyl $C_{46}H_{50}EuN_7O_8$
Mol. Wt.: 980.9
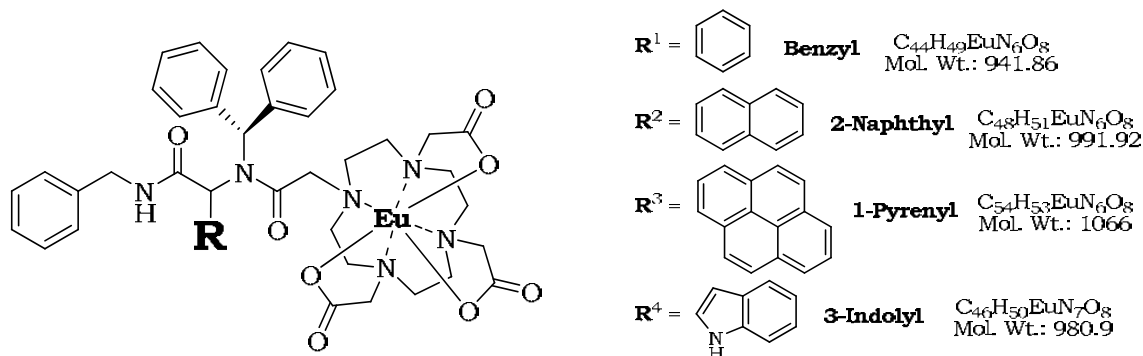
The carboxylic acids **08**, **09**, **10** & **11** (0.005 g) were dissolved in anhydrous MeOH (5 mL), and Eu(OTf)$_3$ (1.1 eq.) was added. The mixture was then stirred at 40 °C under argon for 3 days. The MeOH was then removed under reduced pressure and the product purified using HPLC (method D) to afford the products as fine white powders.

12* (0.005 g, 86 %); m.p. = >300 ˚C ; v_{max} (CHCl$_3$)/cm^{-1} 1595 br (COO-), 1658, 1679 (2 x C=O); m/z (TOF ES+) 965.2722 ([M+Na]$^+$, 100 %, C$_{44}$H$_{49}$N$_6$O$_8$EuNa requires 965.2733). Analytical HPLC t$_r$ = 37.5 & 40.7 minutes.

13 (0.005 g, 84 %); m.p. = >300 ˚C ; v_{max} (CHCl$_3$)/cm^{-1} 1599 br (COO-), 1659, 1670 (2 x C=O); m/z (TOF ES+) 1015.2879 ([M+Na]$^+$, 100 %, C$_{48}$H$_{51}$N$_6$O$_8$EuNa requires 1015.2865). Analytical HPLC t$_r$ = 35.0 & 37.8 minutes.

14 (0.005, 90 %); m.p. = >300 ˚C ; v_{max} (CHCl$_3$)/cm^{-1} 1588 br (COO-), 1670, 1682 (2 x C=O); m/z (TOF ES+) 1067.3216 ([M+H]$^+$, 100 %, C$_{54}$H$_{54}$N$_6$O$_8$Eu requires 1067.3239). Analytical HPLC t$_r$ = 43.1 minutes.

15 (0.006 g, 96 %); m.p. = >300 ˚C ; v_{max} (CHCl$_3$)/cm^{-1} 1600 br (COO-), 1655, 1678 (2 x C=O); m/z (TOF ES+) 1002.2817 ([M+Na]$^+$, 100 %, C$_{46}$H$_{50}$N$_7$O$_8$Na^{151}Eu requires 1002.2802). Analytical HPLC t$_r$ = 38.0 & 41.0 minutes.

Due to the conformational flexibility of the molecule and the shift effect of the lanthanide ion, the NMR spectra were too complex to assign. See appendix 5 for HPLC and m/z data.

*See appendix 1 for crystal structure.

Preparation of terbium complexed Ugi 4-CCR products: benzyl (**16**), naphthyl (**17**), pyrenyl (**18**) & indolyl (**19**).
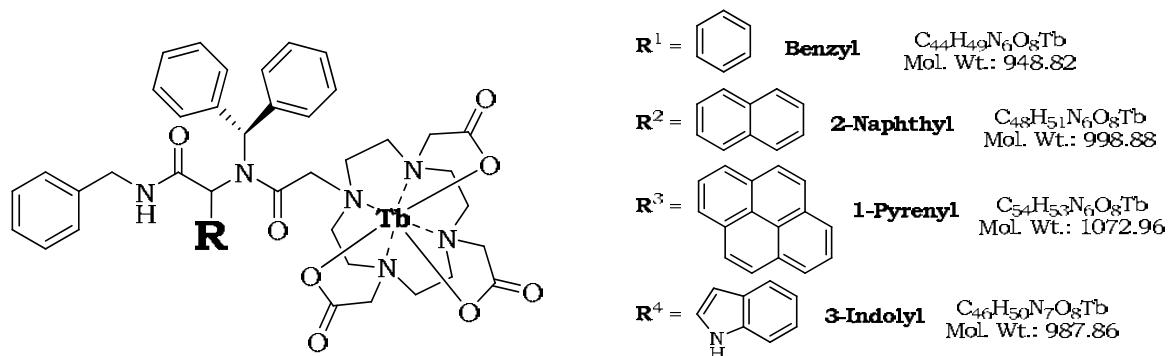
The carboxylic acids **08**, **09**, **10** & **11** (0.005 g) were dissolved in anhydrous MeOH and Tb(OTf)$_3$ (1.1 eq.) was added. The mixture was stirred at 40 °C under Ar$_{(g)}$ for 72 hours. The MeOH was then removed under reduced pressure and the product purified using HPLC method B.

16* (0.006 g, 92 %); m.p. = >300 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1592 br (COO-), 1674, 1698 (C=O); m/z (TOF ES+) 971.2763 ([M+Na]$^+$, 100 %, C$_{44}$H$_{49}$N$_6$O$_8$TbNa requires 971.2735). Analytical HPLC t$_r$ = 37.2 & 40.5 minutes.*

17 (0.005 g, 88 %); m.p. = >300 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1588 br (COO-), 1669, 1694 (C=O); m/z (TOF ES+) 1021.2920 ([M+Na]$^+$, 100 %, C$_{48}$H$_{51}$N$_6$O$_8$TbNa requires 1021.2952). Analytical HPLC t$_r$ = 38.7 & 41.6 minutes.

18 (0.004 g, 73 %); m.p. = >300 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1598 br (COO-), 1689, 1712 (C=O); m/z (TOF ES+) 1095.4239 ([M+Na]$^+$, 100 %, C$_{54}$H$_{53}$N$_6$O$_8$TbNa requires 1095.4221). Analytical HPLC t$_r$ = 41.7 minutes.

19 (0.005 g, 91 %); m.p. = >300 ˚C ; ν_{max} (CHCl$_3$)/cm^{-1} 1595 br (COO-), 1666 br (C=O); m/z (TOF ES+) 1010.2872 ([M+Na]$^+$, 100 %, C$_{46}$H$_{50}$N$_7$O$_8$TbNa requires 1010.2843). Analytical HPLC t$_r$ = 38.2 & 41.1 minutes.

Due to the conformational flexibility of the molecule and the shift effect of the lanthanide ion, the NMR spectra were too complex to assign. See **appendix 5** for HPLC and m/z data.

*See **appendix 1** for crystal structure.

Preparation of ytterbium complexed Ugi 4-CCR products: benzyl (**20**), naphthyl (**21**), pyrenyl (**22**) & indolyl (**23**).

$R^1 =$ **Benzyl** $C_{44}H_{49}N_6O_8Yb$ Mol. Wt.: 962.94

$R^2 =$ **2-Naphthyl** $C_{48}H_{51}N_6O_8Yb$ Mol. Wt.: 1012.99

$R^3 =$ **1-Pyrenyl** $C_{54}H_{53}N_6O_8Yb$ Mol. Wt.: 1087.07

$R^4 =$ **3-Indolyl** $C_{46}H_{50}N_7O_8Yb$ Mol. Wt.: 1001.97
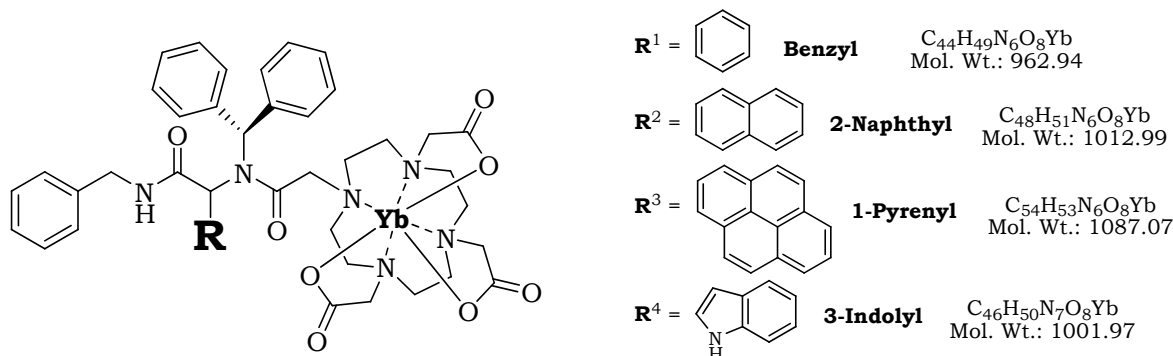
The carboxylic acids **08**, **09**, **10** & **11** (0.005 g) were dissolved in anhydrous MeOH (5 mL), and Yb(OTf)$_3$ (1.1 eq.) added. The mixture was stirred at 40 °C under N$_{2(g)}$ for 72 hours. The MeOH was then removed under reduced pressure and the crude product purified using HPLC method B to afford pure product as a fine white powder.

20* (0.005 g, 83 %); m.p. = >300 °C ; ν_{max} (CHCl$_3$)/cm^{-1} 1597 br (COO-), 1674, 1712 (C=O); m/z (TOF ES+) 986.2898 ([M+Na]$^+$, 100 %, C$_{44}$H$_{49}$N$_6$O$_8$YbNa requires 986.2921). Analytical HPLC t_r = 36.6 & 40.3 minutes.*

21 (0.005 g, 84 %); m.p. = >300 °C ; ν_{max} (CHCl$_3$)/cm^{-1} 1595 br (COO-), 1670, 1699 (C=O); m/z (TOF ES+) 1036.3044 ([M+Na]$^+$, 100 %, C$_{48}$H$_{51}$N$_6$O$_8$YbNa requires 1036.3055). Analytical HPLC t_r = 35.5 & 39.6 minutes.

22 (0.006 g, 92 %); m.p. = >300 °C ; ν_{max} (CHCl$_3$)/cm^{-1} 1587 br (COO-), 1673 br, 1728 (C=O); m/z (TOF ES+) 1110.311 ([M+Na]$^+$, 100 %, C$_{54}$H$_{53}$N$_6$O$_8$YbNa requires 1110.3243). Analytical HPLC t_r = 43.3 minutes.

Product discoloured rapidly.

23 (0.005 g, 78 %); m.p. = >300 °C ; ν_{max} (CHCl$_3$)/cm^{-1} 1598 br (COO-), 1671 br, 1723 (C=O); m/z (TOF ES+) 1025.3007 ([M+Na]$^+$, 100 %, C$_{46}$H$_{50}$N$_7$O$_8$^{174}YbNa requires 1025.3014). Analytical HPLC t_r = 38.1 & 40.8 minutes.

Due to the conformational flexibility of the molecule and the shift effect of the lanthanide ion, the NMR spectra were too complex to assign. See **appendix 5** for HPLC and m/z data.

*See **appendix 1** for crystal structure.

Preparation of ethyl ester caproic linker Ugi 4-CCR products: benzyl (**24**) & indolyl (**25**).
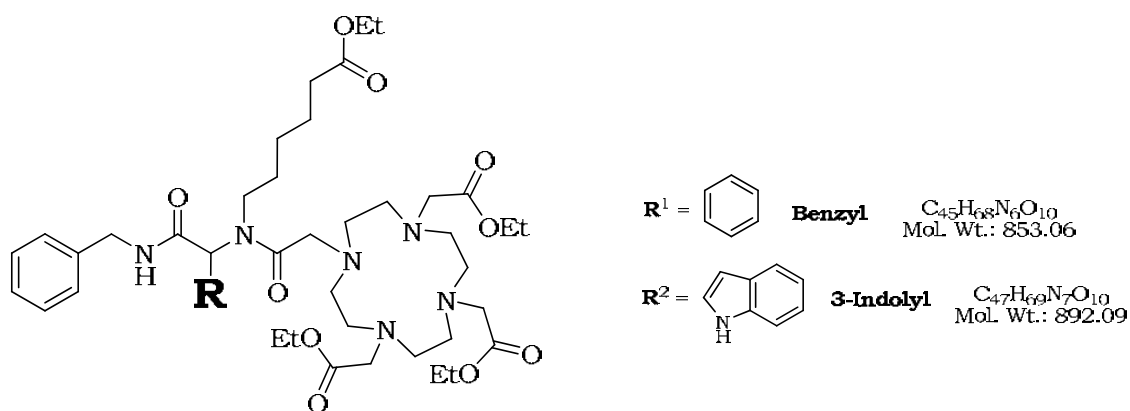
Aminocaproic ethyl ester hydrochloride **50** (0.092 g, 0.55 mmol) was combined with anhydrous Et_3N (0.076 mL, 0.55 mmol) and stirred in anhydrous EtOH (10 mL) for 5 minutes at ambient temperature under argon. Aldehyde (0.08 g, 0.55 mmol) was then applied and the mixture stirred for a further 30 minutes. 1,4,7-tris-(ethoxycarbonylmethyl)-1,4,7,10-tetraazacyclododecyl ethanoic acid **03** (0.55 mmol) and benzyl isocyanide (0.55 mmol) were then added simultaneously and the cocktail stirred at 65°C for 3 days under argon. The solvent was then removed under reduced pressure and the crude product purified using gradient flash column chromatography (Silica, CH_2Cl_2 100 % to CH_2Cl_2, MeOH, $NH_{2(aq)}$ 9.3 : 0.65 : 0.05). The product was then further purified by preparative HPLC (Method C) and lyophilised to afford the product as a fine white powder.

24 (0.27 g, 59 %); R_f = 0.42; v_{max} (CHCl$_3$)/cm^{-1} 1572, 1673, 1742 (C=O), 3249, 3300 (NH); m/z (TOF ES+) 875.1 ([M+Na]$^+$, 100 %); HRMS 875.4895 ([M+Na]$^+$, 100%, $C_{45}H_{68}N_6O_{10}Na$ requires 875.4887). Analytical HPLC t_r = 18.6 minutes.

25 (0.26 g, 53 %); R_f = 0.62; v_{max} (CHCl$_3$)/cm^{-1} 1565, 1654, 1739 (C=O), 3246, 3296 (NH); m/z (TOF ES+) 893.6 ([M+H]$^+$, 100 %), 915.5 ([M+Na]$^+$, 25 %); HRMS 892.5184 ([M+H]$^+$, 100%, $C_{47}H_{70}N_7O_{10}$ requires 892.5197). Analytical HPLC t_r = 19.7 minutes.

Preparation of hydolysed ethyl ester caproic linker Ugi 4-CCR products: indolyl (**26**).

$C_{37}H_{52}N_6O_{10}$
Mol. Wt.: 740.84
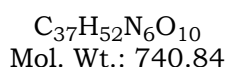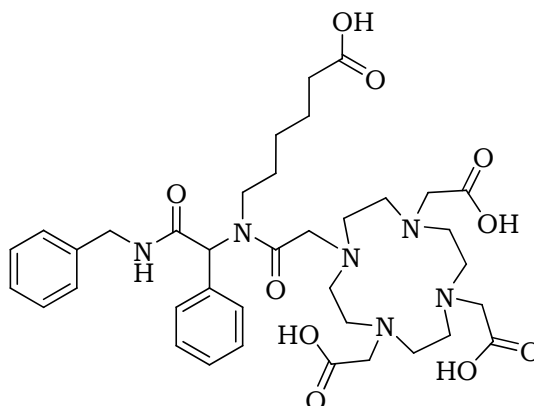
The Ugi product **24** (0.20 g, 0.24 mmol) was dissolved in a 1:1 mix of THF / H_2O. LiOH (3.5 eq.) in H_2O (2 mL) was then added and the reaction mixture stirred for 24 hours at ambient temperature. The excess THF was then removed under reduced pressure and the mixture neutralised with 1M HCl. The product was purified using preparative HPLC (method B), and lyophilised to afford the product as a fine white powder (0.15 g, 88 %). HPLC t_r = 13.6; m/z (TOF ES-) 739.62 ([M-H]-, 100 %).

Preparation of alternative protecting group strategy Ugi products. Tris-ethyl ester, caproic *tert*-butyl ester, benzyl (**27**) and 3-indolyl (**28**) Ugi reaction products.

R^1 = Benzyl $C_{47}H_{72}N_6O_{10}$ Mol. Wt.: 881.11

R^2 = 3-Indolyl $C_{49}H_{73}N_7O_{10}$ Mol. Wt.: 920.14
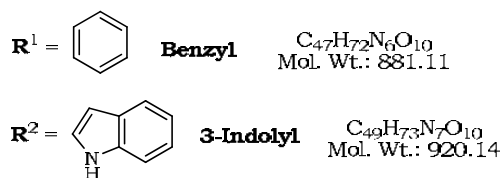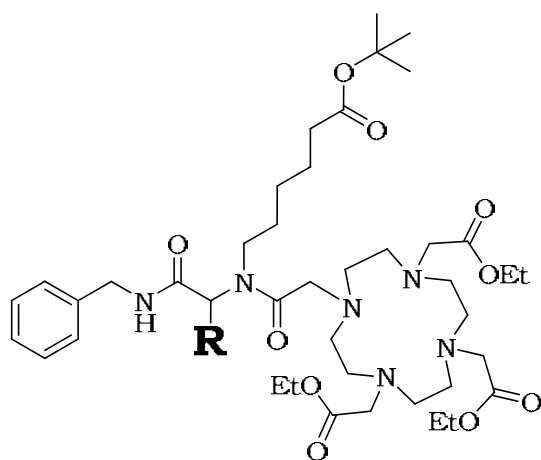
Amino caproic *tert*-butyl ester **49** (1.03 mmol) was combined with aldehyde (1.03 mmol) in anhydrous EtOH (5 mL) and stirred for 10 minutes under argon at ambient temperature.

1,4,7-tris-(ethoxycarbonylmethyl)-1,4,7,10-tetraazacyclododecyl ethanoic acid **03** (1.03 mmol) and benzyl isocyanide (1.03 mmol) were then added simultaneously and the cocktail stirred at reflux for 3 days under argon. The solvent was then removed under reduced pressure and the crude product purified using gradient flash column chromatography (CH_2Cl_2 to CH_2Cl_2 / MeOH / $NH_{2(aq)}$, 9 : 0.95 : 0.05). The product was then further purified by preparative HPLC (Method X) and lyophilised to afford the product as a fine white powder.

27 (68 %); R_f = 0.42; m/z (TOF ES+) 881.5388 ([M+H]$^+$, 100 %, $C_{47}H_{73}N_6O_{10}$ requires 881.5392); 903.5208 ([M+Na]$^+$, 100%, $C_{47}H_{72}N_6O_{10}Na$ requires 903.5203). Analytical HPLC t_r = 26.9 minutes.

28 (67 %); R_f = 0.45; m/z (TOF ES+) 942.5317 ([M+Na]$^+$, 100 %, $C_{49}H_{73}N_7O_{10}Na$ requires 942.5303). Analytical HPLC t_r = 31.9 minutes.

Preparation of ethyl ester hydrolysis products: benzyl (**29**), indolyl (**30**).

R^1 = Benzyl $C_{41}H_{60}N_6O_{10}$ Mol. Wt.: 796.95

R^2 = 3-Indolyl $C_{43}H_{61}N_7O_{10}$ Mol. Wt.: 835.99
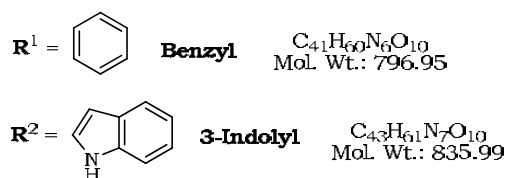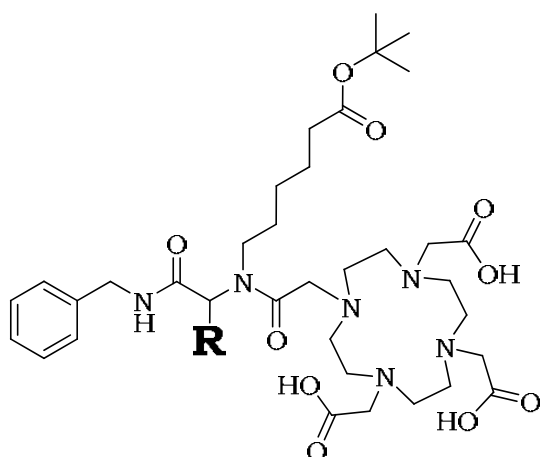
The Ugi products **27** & **28** were dissolved in a 1:1 (v/v) mix of THF / H_2O (8 mL). LiOH (3.5 eq.) in H_2O (2 mL) was then added and the reaction mixture stirred for 24 hours at ambient temperature. The excess THF was then removed under reduced pressure and the mixture neutralised with 1M HCl. The product was purified using preparative HPLC (method D), and lyophilised to afford the product as a fine white powder.

29 (0.03 g, 90 %); m/z (TOF ES+) 798.0 ([M+H]$^+$, 10%) 400.1 ([M/2+2H]$^+$, 100 %); Analytical HPLC t_r = 30.3 minutes.

30 (0.05 g, 88 %); m.p. = 66-68 °C; m/z (MALDI-TOF) 836.4 ([M+H]$^+$, 100 %), 858.4 ([M+Na]$^+$, 60 %). Analytical HPLC t_r = 34.3 minutes.

Preparation of terbium complexation Ugi products: benzyl (**31**), indolyl (**32**).

$R^1 =$ **Benzyl** $C_{41}H_{57}N_6O_{10}Tb$ Mol. Wt.: 952.85

$R^2 =$ **3-Indolyl** $C_{43}H_{58}N_7O_{10}Tb$ Mol. Wt.: 991.89
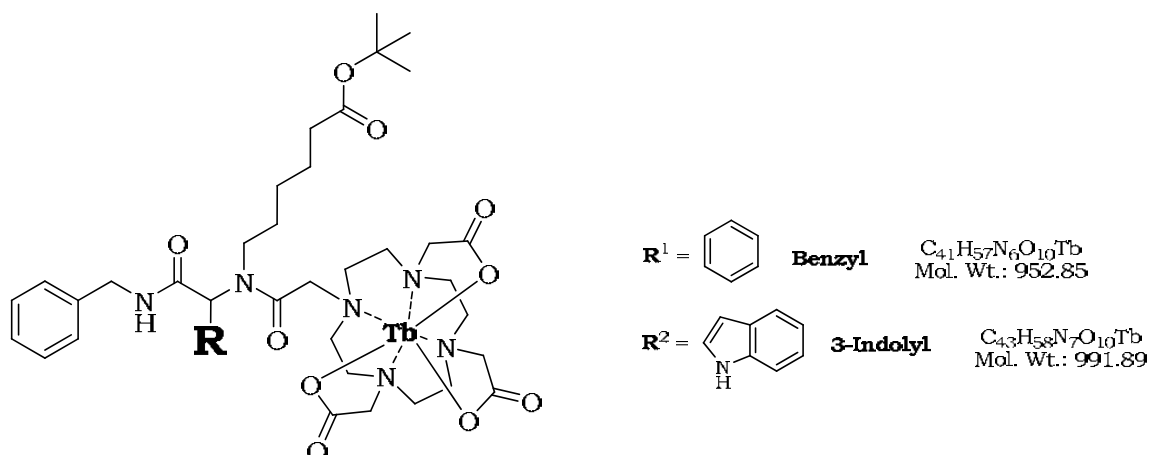
The carboxylic acids **29** & **30** were dissolved in anhydrous MeOH and Tb(OTf)$_3$ (1.1 eq.) added. The mixture was stirred at 40 °C under argon for 3 days. The crude mixture was then concentrated under reduced pressure and the product purified using HPLC method E.

31 (0.01 g, 48 %); m/z (TOF ES+) 953.5 ([M+H]$^+$, 40 %), 449.2 ([M/2+2H]$^+$, 100 %). Analytical HPLC t_r = 32.3 minutes.

32 (0.014 g, 57 %); m/z (MALDI-TOF) 992.4 ([M+H]$^+$, 90 %), 936.1 ([M-^{t}Bu]$^+$, 100 %). Analytical HPLC t_r = 35.7 minutes.

Preparation of terbium complexed caproic linker Ugi 4-CCR products: benzyl (**33**), indolyl (**34**).

$R^1 =$ **Benzyl** $C_{37}H_{49}N_6O_{10}Tb$ Mol. Wt.: 896.74

$R^2 =$ **3-Indolyl** $C_{39}H_{50}N_7O_{10}Tb$ Mol. Wt.: 935.78
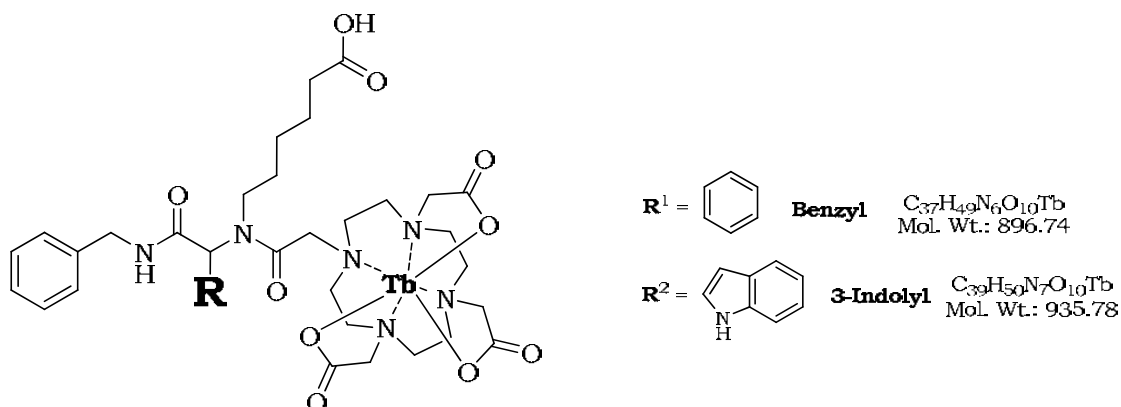
Preparation of **33**

The terbium complexed Ugi product **31** (0.016 g, 0.016 mmol) was taken up in CH$_2$Cl$_2$ (1 mL) and with stirring at ambient temperature a concoction consisting of TFA / TIPS / H$_2$O

(98 : 1 : 1 repectively) (1 mL) applied. The mixture was stirred for 2 hours and concentrated under reduced pressure repeatedly using co-concentration with CH_2Cl_2 (5 x 5 mL) and trituration with Et_2O to ensure complete removal of all waste products and the recovery of the product as a white powder.

33 (0.013 g, 88 %); m/z (TOF ES+) 897.2842 ([M+H]$^+$, 100%, $C_{37}H_{50}N_6O_{10}Tb$ requires 897.2813). Analytical HPLC t_r = 23.7 minutes.

Preparation of **34**

The tetra ethyl ester **25** (0.02 g, 0.022 mmol) was dissolved in a 1:1 mix of THF / H_2O. LiOH (3.5 eq.) in H_2O (2 mL) was then added and the reaction mixture stirred for 24 hours at ambient temperature. The excess THF was then removed under reduced pressure and the mixture neutralised with 1M HCl. The crude product was lyophilised and taken on without further purification.

The crude tetra acid (0.015 g, 0.0025 mmol) was then dissolved in anhydrous MeOH (5 mL), and Tb(OTf)$_3$ (0.09 g, 0.15 mmol) added. The mixture was stirred at 40 °C under argon for 72 hours. The solvent was then removed under reduced pressure and the crude product purified using HPLC (method D) and lyophilised to afford the product as a fine white powder.

34 (0.02 g, 18 %); v_{max} m/z (TOF ES+) 958.0 ([M+Na]$^+$, 100 %), 980 ([M-H+2Na]$^+$, 5 %); m/z (TOF ES-) 934.1 ([M-H]$^+$, 100 %), 956.2 ([M-H+Na]$^+$, 15 %). Analytical HPLC t_r = 31.2 minutes.

2.12 Linkers experimental

<u>PEG</u>

Preparation of *tert*-butoxycarbonylmethyl (ethylene glycol): tetra (**35**), hexa (**36**).

$$C_{14}H_{28}O_7$$
Mol. Wt.: 308.37

$$C_{18}H_{36}O_9$$
Mol. Wt.: 396.47

Hexa (ethylene glycol) (1.00 mL, 1.13 g, 4.00 mmol) was combined with *tert*-butylbromoacetate (0.65 mL, 4.39 mmol), silver (I) oxide (1.39 g, 6.0 mmol) and potassium iodide (0.13 g, 0.80 mmol) in dry CH_2Cl_2 at 0°C under argon. The mixture was allowed to warm to ambient temperature and stirred for 8 hours. The silver was removed by filtering the crude cocktail through celite and the solvent then removed under reduced pressure. Purification was achieved by flash column chromatography (9.5 CH_2Cl_2 : 0.5 MeOH) to afford the pure product as a colourless oil.

35 (0.94 g, 76 %). R_f = 0.43; ν_{max} (Neat)/cm^{-1} 3329 (OH), 2883 (CH), 1742 (C=O), 1101 (COC); δ_H (300 MHz; CDCl$_3$) 3.98 (2H, s, CH_2C=O), 3.55-3.69 (16H, stack, 8xOCH_2), 2.71 (1H, br s, OH), 1.42 (9H, s, C(CH_3)$_3$); δ_C (75 MHz) 169.6 (*C*=O), 81.4 (*C*(CH$_3$)$_3$), 72.4 (O*C*H$_2$), 70.6 (O*C*H$_2$), 70.5 (O*C*H$_2$), 70.2 (O*C*H$_2$), 68.9 (O*C*H$_2$), 61.6 (O*C*H$_2$), 28.0 (C(*C*H$_3$)$_3$); m/z (TOF ES+) 331.1733 ([M+Na]$^+$, 100%, $C_{14}H_{28}O_7$Na requires 331.1729).

36 (1.27 g, 80 %). R_f = 0.41; ν_{max} (Neat)/cm^{-1} 3332 (OH), 2880 (CH), 1740 (C=O), 1100 (COC); δ_H (300 MHz; CDCl$_3$) 3.97 (2H, s, OCH_2C=O), 3.55-3.68 (24H, stack, 12xOCH_2), 3.02 (1H, br s, OH), 1.42 (9H, s, C(CH_3)$_3$); δ_C (75 MHz) 169.5 (*C*=O), 81.3 (*C*(CH$_3$)$_3$), 72.3 (O*C*H$_2$), 70.4 (O*C*H$_2$), 70.2 (O*C*H$_2$), 69.9 (O*C*H$_2$), 68.8 (O*C*H$_2$), 61.5 (O*C*H$_2$), 27.8 (C(*C*H$_3$)$_3$); m/z (TOF ES+) 419.2257 ([M+Na]$^+$, 100%, $C_{18}H_{36}O_9$Na requires 419.2254).

Please see Ref 159.

Preparation of O-tert-butoxycarbonylmethyl-6-mesyl-(ethylene glycol): tetra (**37**), hexa (**38**).
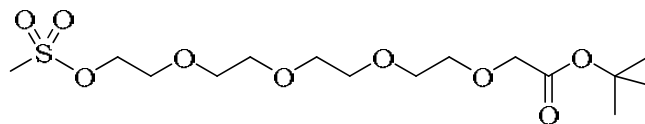
$C_{15}H_{30}O_9S$
Mol. Wt.: 386.46
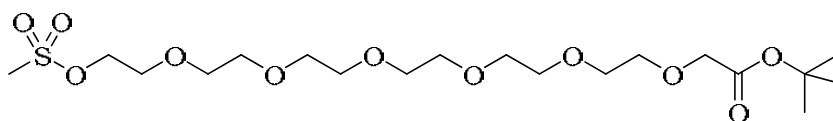
$C_{19}H_{38}O_{11}S$
Mol. Wt.: 474.56

Ester **36** (1.00 g, 2.52 mmol) was dissolved in dry CH_2Cl_2 (50 mL) and cooled to 0 °C under argon. Mesyl chloride (0.20 mL, 0.29 g, 2.52 mmol) and triethylamine (0.42 mL, 0.31 g, 3.03 mmol) were then added and the mixture stirred for a further hour at 0°C prior to allowing the return to ambient temperature and stirring overnight. The crude cocktail was then washed with 2% citric acid soln. (2 x 30 mL), sat. $NaHCO_3$ (2 x 30 mL), H_2O (2 x 30 mL) and brine (30 mL) and dried with anhydrous $MgSO_4$ prior to removing the solvent under reduced pressure. Purification by flash column chromatography (9.5 CH_2Cl_2 : 0.5 MeOH) yielded the pure product as a colourless oil.

37 (1.18 g, 94 %). R_f = 0.47; ν_{max} (Neat)/cm^{-1} 2867 (CH), 1747 (C=O), 1349 (S=O), 1100 (OCO); δ_H (300 MHz; CDCl$_3$) 4.33 (2H, m, CH_2OSO$_2$), 3.96 (2H, s, OCH_2C=O), 3.71 (2H, m, CH_2CH$_2$OSO$_2$), 3.58-3.68 (12H, stack, OCH_2), 3.03 (3H, s, SO$_2$CH_3), 1.42 (9H, s, C(CH_3)$_3$); δ_C (100 MHz) 169.8 (C=O), 81.1 (C(CH$_3$)$_3$), 70.3 (OCH$_2$), 70.2 (OCH$_2$), 70.1 (OCH$_2$), 69.8 (OCH$_2$), 69.3 (OCH$_2$), 68.9 (OCH$_2$), 37.4 (SO$_2$CH$_3$), 27.5 (C(CH$_3$)$_3$); m/z (TOF ES+) 409.2401 ([M+Na]$^+$, 100%, $C_{15}H_{30}O_9$SNa requires 409.2394).

38 (0.97 g, 81 %). R_f = 0.50; ν_{max} (Neat)/cm^{-1} 2865 (CH), 1745 (C=O), 1348 (S=O), 1102 (OCO); δ_H (300 MHz; CDCl$_3$) 4.28 (2H, m, CH_2SO$_2$), 3.91 (2H, s, OCH_2C=O), 3.67 (2H, m, CH_2CH$_2$SO$_2$), 3.50-3.60 (20H, stack, OCH_2), 2.99 (3H, s, SO$_2$CH_3), 1.37 (9H, s, C(CH_3)$_3$); δ_C (100 MHz) 169.5 (C=O), 81.3 (C(CH$_3$)$_3$), 70.5 (OCH$_2$), 70.4 (OCH$_2$), 70.3 (OCH$_2$), 70.2 (OCH$_2$), 69.2 (OCH$_2$), 68.8 (OCH$_2$), 37.5 (SO$_2$CH$_3$), 27.9 (C(CH$_3$)$_3$); m/z (TOF ES+) 497.2033 ([M+Na]$^+$, 100%, $C_{19}H_{38}O_{11}$SNa requires 497.2018).
Please see Ref 159.

Preparation of O-*tert*-butoxycarbonylmethyl-6-azido-(ethylene glycol): tetra (**39**), hexa (**40**).
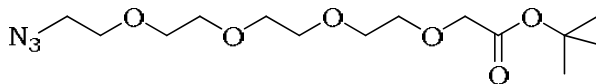
$C_{14}H_{27}N_3O_6$
Mol. Wt.: 333.38
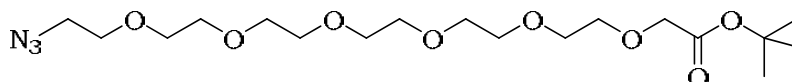
$C_{18}H_{35}N_3O_8$
Mol. Wt.: 421.49

Mesylate **38** (0.80 g, 1.69 mmol) was combined with sodium azide (0.16 g, 2.53 mmol) in dry DMF (40 mL). The mixture was heated at reflux for 8 hours under argon and cooled to ambient temperature. EtOAc (100 mL) was added and the crude cocktail extracted with H_2O (5 x 100 mL), dried over anhydrous $MgSO_4$, and finally the solvent removed under reduced pressure to afford the product as a colourless oil. Analysis revealed the product to be of high enough pueity to proceed without further purification.

39 (0.52 g, 75 %); R_f = 0.45 (9.5 CH_2Cl_2 : 0.5 MeOH). ν_{max} (Neat)/cm^{-1} 2102 (N_3), 1739 (C=O), 1112 (COC); δ_H (300 MHz; CDCl$_3$) 3.88 (2H, s, CH$_2$C=O), 3.56-3.49 (14H, stack, 7 x OCH$_2$), 3.21 (2H, t, J 5.2, C*H*$_2$N$_3$), 1.33 (9H, s, C(C*H*$_3$)$_3$); δ_C (75 MHz) 168.2 (*C*=O), 79.7 (*C*(CH$_3$)$_3$), 69.4 (O*C*H$_2$), 69.3 (O*C*H$_2$), 68.7 (O*C*H$_2$), 67.6 (O*C*H$_2$), 49.3 (*C*H$_2$N$_3$), 26.8 (C(*C*H$_3$)$_3$); m/z (TOF ES+) 356.1798 ([M+Na]$^+$, 100%, $C_{14}H_{27}N_3O_6Na$ requires 356.1785.

40 (0.54 g, 77 %), R_f = 0.43 (9.5 CH_2Cl_2 : 0.5 MeOH). ν_{max} (Neat)/cm^{-1} 2104 (N_3), 1740 (C=O), 1110 (COC); δ_H (300 MHz; CDCl$_3$) 3.91 (2H, s, C*H*$_2$C=O), 3.50-3.60 (22H, stack, OC*H*$_2$), 3.28 (2H, t, J 4.9, C*H*$_2$N$_3$), 1.37 (9H, s, C(C*H*$_3$)$_3$); δ_C (100 MHz) 169.4 (*C*=O), 81.2 (*C*(CH$_3$)$_3$), 70.5 (O*C*H$_2$), 70.4 (O*C*H$_2$), 69.8 (O*C*H$_2$), 68.8 (O*C*H$_2$), 50.5 (*C*H$_2$N$_3$), 27.9 (C(*C*H$_3$)$_3$); m/z (TOF ES+) 444.2322 ([M+Na]$^+$, 100%, $C_{18}H_{35}N_3O_8Na$ requires 444.2321).

Please see Ref 159.

Preparation of O-*tert*-butoxycarbonylmethyl-6-amino-(ethylene glycol): tetra (**41**), hexa (**42**).
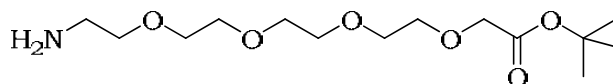
$$C_{14}H_{29}NO_6$$
Mol. Wt.: 307.38
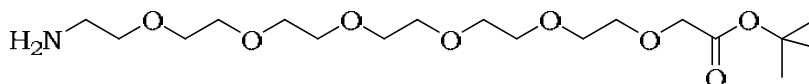
$$C_{18}H_{37}NO_8$$
Mol. Wt.: 395.49

Azide **39** (2.29 g, 5.23 mmol) was dissolved in dry THF (20 mL) under argon at 0 °C. PPh$_3$ (1.57 g, 5.98 mmol) was then added and after 30 minutes the ice bath was removed and the mixture stirred overnight at ambient temperature. H$_2$O (3 ml) was then added with stirring for 10 minutes prior to the solvent being removed under reduced pressure and the crude product purified by flash column chromatography (9.5 CH$_2$Cl$_2$: 0.5 MeOH to elute OPPh$_3$, 9.2 CH$_2$Cl$_2$: 0.75 MeOH : 0.05 NH$_{3(aq)}$ to elute product R$_f$ = 0.20) to reveal the product as a colourless viscous oil.

41 (1.35 g, 84 %); ν_{max} (Neat)/cm^{-1} 3368 (NH), 1739 (C=O), 1097 (OCO); δ_H (300 MHz; CDCl$_3$) 3.96 (2H, s, CH_2C=O), 3.65-3.55 (12H, stack, 6xOCH_2), 3.47 (2H, t, J 5.2, CH_2NH$_2$), 2.82 (2H, t, J 4.9, OCH_2), 2.30 (2H, br s, NH_2), 1.41 (9H, s, C(CH_3)$_3$); δ_C (100MHz) 169.6 (*C*=O), 81.5 (*C*(CH$_3$)$_3$), 72.7 (O*C*H$_2$), 70.5 (O*C*H$_2$), 70.4 (O*C*H$_2$), 70.1 (O*C*H$_2$), 68.9 (O*C*H$_2$), 41.5 (*C*H$_2$NH$_2$), 28.0 (C(*C*H$_3$)$_3$); m/z (TOF ES+) 308.2073 ([M+H]$^+$, 100%, C$_{14}$H$_{30}$NO$_6$ requires 308.2061).

42 (1.80 g, 87 %); ν_{max} (Neat)/cm^{-1} 3370 (NH), 1741 (C=O), 1100 (OCO); δ_H (300 MHz; CDCl$_3$) 3.69 (2H, s, CH_2C=O), 3.66-3.40 (22H, stack, OCH_2), 3.46 (2H, t, J 5.2, CH_2NH$_2$), 2.81 (2H, t, J 4.9, NH_2), 1.41 (9H, s, C(CH_3)$_3$); δ_C (100MHz) 169.5 (*C*=O), 81.5 (*C*(CH$_3$)$_3$), 73.0 (O*C*H$_2$), 70.6 (O*C*H$_2$), 70.4 (O*C*H$_2$), 70.1 (O*C*H$_2$), 68.9 (O*C*H$_2$), 41.6 (*C*H$_2$NH$_2$), 28.3 (C(*C*H$_3$)$_3$); m/z (TOF ES+) 396.2597 ([M+H]$^+$, 100%, C$_{18}$H$_{38}$NO$_8$ requires 396.2583).

Please see Ref 159.

Preparation of amino hexa (ethylene glycol) acetic acid **43**.

$$C_{14}H_{29}NO_8$$
Mol. Wt.: 339.38
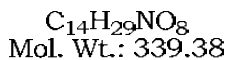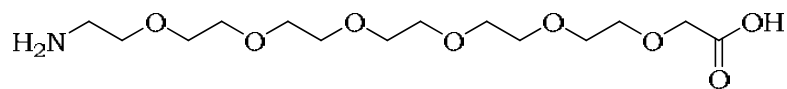
tert-Butyl ester hydrolysis was achieved by taking 1-amino-6-*tert*-butoxycarbonylmethyl-hexa ethylene glycol **42** (2.00 g, 5.06 mmol) up in a cocktail of CH_2Cl_2 and TFA (1 : 1, v / v, 20 mL) and stirred at ambient temperature for 2 hours. The reation was monitored by TLC (Rf = 0.1, 9.0 CH_2Cl_2 : 1.0 MeOH). Excess solvent and TFA was removed through cycles involving repeated dissolution in CH_2Cl_2 and removal under reduced pressure to afford the product as a very pale brown oil (1.70 g, 99 %); ν_{max} (Neat)/cm^{-1} 3000 (br COOH); δ_H (300 MHz; CDCl$_3$) 12.73 (1H, br s, COOH), 7.21 (2H, br s, NH_2), 4.16 (2H, s, CH_2C=O), 3.61-3.83 (22H, stack, OCH_2), 3.10 (2H, m, CH_2NH$_2$); δ_C (100MHz) 173.8 (C=O), 69.9 (OCH$_2$), 69.6 (OCH$_2$), 69.1 (OCH$_2$), 67.9 (OCH$_2$), 66.8 (OCH$_2$), 40.2 (CH$_2$NH$_2$); m/z (TOF ES+) 340.1971 ([M+H]$^+$, 100%, $C_{14}H_{30}NO_8$ requires 340.1958).

Preparation of N-*tert*-butoxycarbonyl-amino-hexa(ethylene glycol) acetic acid **44**.

$$C_{19}H_{37}NO_{10}$$
Mol. Wt.: 439.5
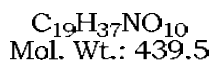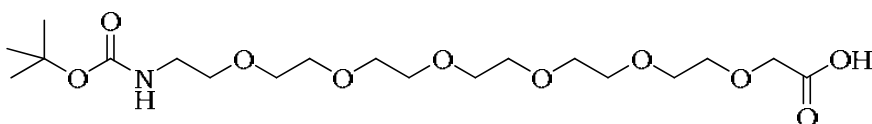
Amino acid **43** (1.20 g, 3.54 mmol) was dissolved in anhydrous CH_2Cl_2 (10 mL). Anhydrous DIPEA (1.23 mL, 7.07 mmol) was added and the mixture stirred at ambient temperature under argon. Di-tert-butyldicarbonate (0.93 g, 4.24 mmol) in anhydrous CH_2Cl_2 (10 mL) was then added dropwise over 40 minutes and the cocktail stirred for a further 2 hours with monitoring by TLC. The crude product was concentrated thoroughly (repeated co-concentration cycles with toluene) under reduced pressure and purified by gradient flash column chromatography to afford the pure product as a colourless oil (1.41 g, 91 %); ν_{max} (Neat)/cm^{-1} 2894 (br COOH), 1718 (C=O); δ_H (300 MHz; CDCl$_3$) 10.59 (1H, br s, COOH), 5.25 (1H, br s, NH), 3.87 (2H, s, CH_2C=O), 3.50-3.61 (22H, stack, OCH_2), 3.43 (2H, app t, J 5.1, CH_2NH), 1.33 (9H, s, C(CH_3)$_3$); δ_C (100MHz) 174.8 (COOH), 155.9 (C=ONH$_2$), 78.8

(C(CH$_3$)$_3$), 70.3 (OCH$_2$), 70.2 (OCH$_2$), 69.9 (OCH$_2$), 69.8 (OCH$_2$), 41.3 (CH$_2$NH), 28.2 (C(CH$_3$)$_3$); m/z (TOF ES+) 462.2315 ([M+Na]$^+$, 100%, C$_{19}$H$_{37}$NO$_{10}$Na requires 462.2297). Please see Ref 159.

Preparation of N-*tert*-butoxycarbonyl-3,6-dioxaoctane-1,8diamine, **45**
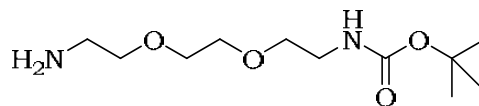
C$_{11}$H$_{24}$N$_2$O$_4$
Mol. Wt.: 248.32

Di-*tert*-butyl-dicarbonate (1.49 g, 6.85 mmol) was dissolved in CH$_2$Cl$_2$ (25 mL) and added dropwise to a mixture of tris-ethyleneglycol diamine (2.0 mL, 0.014 mol) and DIPEA (2.38 mL, 0.014 mol) at ambient temperature over 40 minutes. The cocktail was then stirred for a further 2 hours and monitored by TLC. Upon completion the mixture was evaporated thoroughly under reduced pressure and the crude residue purified by flash column chromatography Rf=0.33 (CH$_2$Cl$_2$ / MeOH / NH$_{3(aq)}$, 9.0 : 9.95 : 0.05) to afford the product as a viscous colourless oil (2.01 g, 50%). ν$_{max}$ (Neat)/cm^{-1} 3368, 3384 (NH), 1725 (C=O); δ$_H$ (300 MHz; CDCl$_3$) 5.20 (1H, br s, NH), 3.55 (4H, s, 2xOCH$_2$), 3.43-3.49 (4H, stack, 2xOCH$_2$), 3.24 (2H, mound, CH$_2$NH), 2.81 (2H, t, J 5.2, CH$_2$NH$_2$), 1.52 (2H, br s, NH_2), 1.37 (9H, s, (C(CH$_3$)$_3$); δ$_C$ (100MHz) 155.9 (C=O), 79.0 (C(CH$_3$)$_3$), 73.2 (OCH$_2$), 70.1 (2xOCH$_2$), 41.6 (CH$_2$NH$_2$), 40.2 (CH$_2$NH), 28.3 (C(CH$_3$)$_3$); m/z (TOF ES+) 271.1634 ([M+Na]$^+$, 100%, C$_{11}$H$_{24}$N$_2$O$_4$Na requires 271.1623).

Caproic

Preparation of mono-*tert*-butoxycarbonyl hexa ethylene diamine **46**.
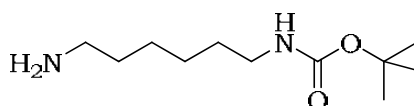
C$_{11}$H$_{24}$N$_2$O$_2$
Mol. Wt.: 216.32

Di-tert-butyldicarbonate (0.94 g, 4.30 mmol), was taken up in chloroform (10 mL) and added dropwise to a solution of hexamethylenediamine (2.50 g, 21.5 mmol) also in chloroform (90 mL) over 2 hours at 0˚C under argon. The reaction mixture was then allowed to warm to ambient temperature and stirred overnight. The precipitate was removed through filtration and the crude cocktail concentrated under reduced pressure.

The resulting heavy oil was taken up in EtOAc (100 mL) and extracted with H_2O (3 x 100 mL) and brine (3 x 50 mL) alternately before drying over $MgSO_4$. Removal of the solvent under reduced pressure afforded the product as a colourless oil (2.28 g, 49 %). ν_{max} (Neat)/cm^{-1} 3279, 3301 (NH), 1721 (C=O); δ_H (300 MHz; CDCl$_3$) 4.83 (1H, br s, NH), 2.98 (2H, m, H$_2$NCH_2), 2.56 (2H, t, J 6.9, CH_2NH), 2.29 (2H, br s, NH_2), 1.38-1.18 (17H, stack, 4xCH_2 & C(CH_3)$_3$); δ_C (100MHz) 155.7 (C=O), 78.2 (C(CH$_3$)$_3$), 41.4 (H$_2$NCH$_2$), 40.0 (CH$_2$NH), 32.8 (CH$_2$), 29.6 (CH$_2$), 28.0 (C(CH$_3$)$_3$), 26.1 (2xCH$_2$); m/z (TOF ES+) 217.1916 ([M+H]$^+$, 100%, $C_{11}H_{25}N_2O_2$ requires 217.1918).

Preparation of N-benzoyl amino caproic acid **47**.

$$C_{14}H_{19}NO_4$$
Mol. Wt.: 265.3
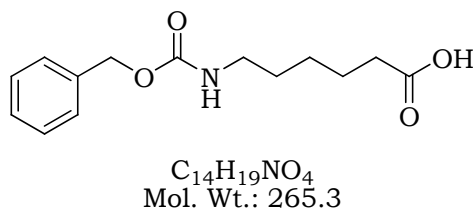
Benzylchloroformate (8.71 mL, mmol) was taken up in 1M NaOH (31 mL) and added dropwise to a mixture of 6-aminocaproic acid (8.00 g, 0.06 mol) in a further equivalent of 1M NaOH (31 mL) over 30 minutes at 0˚C. The cocktail was then allowed to warm to ambient temperature and stirred overnight. The aqueous solution was washed with Et$_2$O (2 x 50 mL) prior to acidification to pH = 1 with concentrated HCl. Again the aqueous solution was extracted with Et$_2$O (3 x 100 mL) and the organic phases combined, washed with brine (50 mL) and dried over $MgSO_4$ to afford the pure product as a fine white solid (14.18 g, 88 %). ν_{max} (Neat)/cm^{-1} 2853 (br COOH), 1694 (C=O); δ_H (300 MHz; CDCl$_3$) 10.96 (1H, br s, COOH), 7.26-7.35 (5H, stack, CHarom), 5.09 (2H, s, PhCH_2O), 4.98 (1H, mound, NH), 3.18 (2H, m, NHCH_2), 2.34 (CH_2COOH), 1.66 (2H, m, CH_2), 1.51 (2H, m, CH$_2$), 1.35 (2H, m, CH_2); δ_C (100MHz) 179.0 (COOH), 156.2 (C=ONH), 136.4 ($C_q$arom), 128.4 (Carom), 128.0 (Carom), 66.6 (PhCH$_2$O), 40.7 (NHCH_2), 33.8 (CH$_2$COOH), 29.5 (CH$_2$), 26.0 (CH$_2$), 24.2 (CH$_2$); m/z (TOF ES+) 288.1212 ([M+Na]$^+$, 100%, $C_{14}H_{19}NO_4Na$ requires 288.1205).

Preparation of benzyl 5-(tert-butoxycarbonyl)pentylcarbamate **48**.

$C_{18}H_{27}NO_4$
Mol. Wt.: 321.41
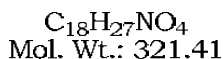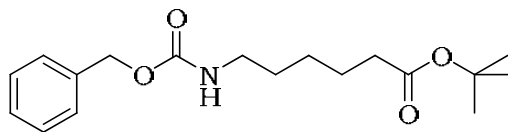
Under dry conditions N-benzoyl aminocaproic acid **47** (5.0 g, 0.019 mol) was combined with pyridine (3.05 mL, 0.038 mol) and ^tBuOH (9.012 mL, 0.095 mol) in anhydrous CH_2Cl_2 (50 mL) at 0˚C under argon. $POCl_3$ (1.76 mL, 0.038 mol) was then added dropwise and the mixture stirred overnight at ambient temperature. The crude mixture was washed with H_2O (2 x 30 mL), 5% $NaHCO_3$ (2 x 30 mL), H_2O (2 x 30 mL) and brine (20 mL) prior to drying with $MgSO_4$ and concentration under reduced pressure. The crude product was purified by careful gradient flash column chromatography R_f = 0.37 (100 % CH_2Cl_2 to 9.5 CH_2Cl_2 : 0.5 MeOH) and the product recovered as a fine white solid (4.60 g, 75 %). v_{max} (Neat)/cm^{-1} 1747, 1688 (C=O); δ_H (300 MHz; $CDCl_3$) 7.34-7.28 (5H, stack, C*H* arom), 5.06 (2H, s, PhC*H*$_2$O), 4.81 (1H, br s, N*H*), 3.16 (2H, dd, J 6.7, 13.2, NHC*H*$_2$), 2.18 (2H, t, J 7.4, C*H*$_2$C=O), 1.61-1.15 (15H, br stack, C(C*H*$_3$)$_3$ & 3xC*H*$_2$); δ_C (100MHz) 173.0 (CH$_2$*C*=O), 156.3 (*C*=ONH), 136.6 (*C*$_q$arom), 128.5 (*C*arom), 128.0 (*C*arom), 80.1 (*C*(CH$_3$)$_3$), 66.5 (PhC*H*$_2$O), 40.8 (NHC*H*$_2$), 35.3 (*C*H$_2$C=O), 29.6 (*C*H$_2$), 28.1 (C(*C*H$_3$)$_3$), 26.1 (*C*H$_2$), 24.6 (*C*H$_2$); m/z (TOF ES+) 344.1838 ([M+Na]$^+$, 100%, $C_{18}H_{27}NO_4Na$ requires 344.1846).

Preparation of tert-butyl 6-aminohexanoate **49**.

$C_{10}H_{21}NO_2$
Mol. Wt.: 187.28
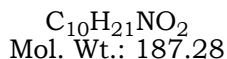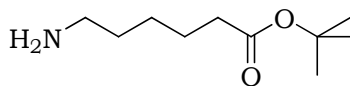
Under scrupulously dry conditions benzyl ester (1.05 g, 3.27 mmol) was taken up in anhydrous MeOH and stirred under argon with Pd/C (0.16 g, 15 % w/w of 10 %) at ambient temperature. The flask was well vented under high vacuum before allowing $H_{2(g)}$ into the flask with further venting 3 times to ensure complete evacuation of argon. The mixture was then stirred under balloon pressure of for $H_{2(g)}$ for 4 hours. The reaction was monitored by TLC, R_f = 0.03 (9.5 CH_2Cl_2 / 0.5 MeOH). Upon completion the mixture was

136

filtered through a pad of Celite and concentrated under reduced pressure to afford the pure product as a colourless oil (0.61 g, 99 %). ν_{max} (Neat)/cm^{-1} 3322, 3329 (NH), 1737 (C=O); δ_H (300 MHz; CDCl$_3$) 3.26 (2H, br s, NH_2), 2.64 (2H, t, J 7.1, CH_2NH$_2$), 2.14 (2H, t, J 7.4, CH_2C=O), 1.57-1.22 (15H, br stack, C(CH_3)$_3$ & 3xCH_2); δ_C (100MHz) 173.0 (*C*=O), 80.0 (*C*(CH$_3$)$_3$), 41.3 (*C*H$_2$NH$_2$), 35.4 (*C*H$_2$C=O), 32.2 (*C*H$_2$), 30.1 (C(*C*H$_3$)$_3$), 26.2 (*C*H$_2$), 24.7 (*C*H$_2$); m/z (TOF ES+) 210.1470 ([M+Na]$^+$, 100%, C$_{10}$H$_{21}$NO$_2$Na requires 210.1478).

Preparation of N-ethoxycarbonyl amino caproic acid **50**.

$$C_8H_{17}NO_2$$
Mol. Wt.: 159.23
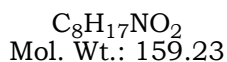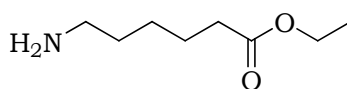
Thionyl chloride (5.56 mL, 76.2 mmol) was added dropwise to freshly distilled anhydrous EtOH (10 mL) at 0°C under argon over 20 minutes. Aminocaproic acid (5.0 g, 38.1 mmol) in anhydrous EtOH (10 mL) was then added over 5 minutes and the mixture heated at reflux for 1 hour. Upon cooling to ambient temperature the solvent was removed under reduced pressure and the residue redissolved in ice water. Saturated NaHCO$_3$ solution was added until the pH=10 and extracted with CHCl$_3$ (10 x 50 mL). The combined organic phases were then extracted with brine (50 mL) and dried over MgSO$_4$. Concentration under reduced pressure afforded the product as colourless oil which became a crystalline solid overnight (5.58 g, 92 %). ν_{max} (Neat)/cm^{-1} 3338, 3346 (NH), 1742 (C=O); δ_H (300 MHz; CDCl$_3$) 7.96 (2H, br s, NH_2), 3.98 (2H, q, J 7.1, OCH_2CH$_3$), 2.89 (2H, mound, H$_2$NCH_2), 2.19 (2H, t, J 7.2, CH_2C=O), 1.68 (2H, mound, CH_2), 1.50 (2H, m, CH_2), 1.30 (2H, mound, CH_2), 1.12 (3H, t, J 6.3, OCH$_2$CH_3); δ_C (100MHz) 173.5 (*C*=O), 60.2 (O*C*H$_2$CH$_3$), 39.6 (H$_2$N*C*H$_2$), 33.7 (*C*H$_2$C=O), 26.8 (*C*H$_2$), 25.7 (*C*H$_2$), 24.0 (*C*H$_2$), 13.9 (OCH$_2$*C*H$_3$); m/z (EI) 159.1259 ([M+Na]$^+$, 100%, C$_8$H$_{17}$NO$_2$ requires 159.1262).

2.13 Osmium experimental

Preparation of cis-dichloro-bis(2,2'-bipyridine)osmium(II) **51**.

$C_{20}H_{16}Cl_2N_4Os$
Mol. Wt.: 573.5
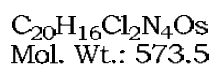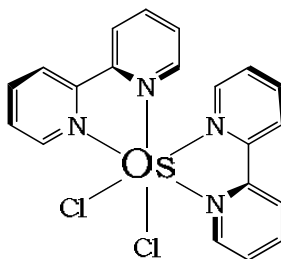
Lithium chloride (0.34 g, 8.09 mmol) and 2,2'-bipyridine (0.22 g, 1.42 mmol) were combined in degassed DMF (20 mL). Osmium trichloride (0.20 g, 0.67 mmol) was added and the mixture stirred under reflux for 2.5 hours. Upon cooling to ambient temperature Et_3N (0.50 mL) was added and the cocktail heated at reflux for a further hour and again allowed to cool. The crude solution was then poured into cold acetone (80 mL) and cooled at -5 °C for 2 days. The precipitated product was recovered by Büchner filtration and washed thoroughly with H_2O (100 mL) and dried scrupulously under high vacuum to reveal the product as a deep red-brown solid which was stored in the dark in a dessicator over silica gel (0.33 g, 85 %). m.p. = 186-188 °C; m/z (MALDI-TOF) 537.5 ([M-Cl+H]+, 30 %), 573.0 ([M+H]+, 80%), 656.1 (M + gentisic acid matrix (MW = 153) – Cl + H]+, 30 %), 691.0 ([M+matrix+H]+, 100 %).

Preparation of bis (2,2'-bipyridine)(4'-methyl-2,2'-bipyridine-4-carboxylic acid)-osmium(II) bis(hexafluorophophate) **52**.

$C_{32}H_{26}N_6O_2Os$
Mol. Wt.: 716.82
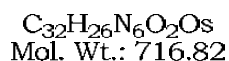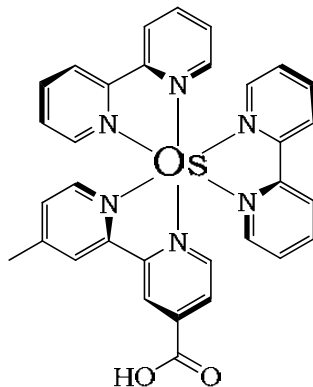
Os(bpy)$_2$Cl$_2$ **51** (0.064 g, 0.112 mmol) and 4'-methyl-2,2'-bipyridine-4-carboxylic acid (0.022 g, 0.102 mmol) were suspended in 70 % EtOH (10 mL) and heated at reflux for 8 hours. The EtOH was then removed under reduced pressure and the flask left to stand overnight. The precipitate was removed by filtration and washed with H$_2$O (50 mL). The filtrate was then acidified to pH=1 with HPF$_6$ and sat. NH$_4$PF$_{6(aq)}$ added dropwise until no further precipitation occurs. The precipitate was recovered by Büchner filtration and washed thoroughly with Et$_2$O (50 mL) and dried scrupulously under high vacuum to reveal the product as a deep green solid. Purification was achieved by HPLC (method C) t$_r$ = 13.6 minutes (analytically); t$_r$ = 19.7 minutes (semi-preparatively), to afford the pure product which was stored in the dark in a dessicator over silica gel (0.055 g, 68 %). ν_{max} (Neat)/cm^{-1} 3200-2600 (COOH); m.p. = 144-146 ˚C; δ_H (300 MHz; d3-MeCN) 2.59 (3H, s, CH$_3$), 7.13 (1H, app dd, J 1.0, 5.8, arom C*H*), 7.21-7.29 (4H, stack, arom C*H*), 7.40 (1H, app d, J5.8, arom C*H*), 7.55-7.58 (4H, stack, arom C*H*), 7.62 (1H, app dd, J 1.5, 5.8, arom C*H*), 7.72 (1H, d, J 5.8, arom C*H*), 7.77-7.84 (4H, stack, arom C*H*), 8.41-8.45 (4H, stack, arom C*H*), 8.50 (1H, app s, arom C*H*), 8.88 (1H, app s, arom C*H*); m/z (MALDI-TOF) 673.7 ([M-COO+H]$^+$, 75 %), 701.8 ([M-Me+H]$^+$, 20 %), 717.7 ([M+H]$^+$, 100 %), 739.2 ([M+Na]$^+$, 50 %).

Preparation of bis (2,2'-bipyridine)(4'-methyl-2,2'-bipyridine-4-carbonyloxy)-2,5-pyrrolidinedione)-osmium(II) bis(hexafluorophophate) **53**.

$C_{36}H_{27}N_7O_4Os$
Mol. Wt.: 811.87
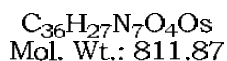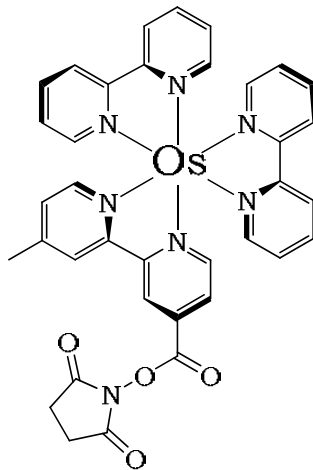
DCC (0.022 g, 0.105 mmol) was added to a stirring solution of carboxylic acid **52** (0.05 g, 0.070 mmol) and N-hydroxysuccinimide (0.009 g, 0.073 mmol) in anhydrous MeCN (2.5 mL) at ambient temperature under argon. This concoction was stirred for 8 hours during which a colour change to rusty-brown was observed. Filtration through a fine-porosity glass frit and removal of the solvent under reduced pressure was then accomplished prior to the crude mixture being taken up in 2-propanol (20 mL) and stored at -8 ˚C for 2 days. The precipitated product was recovered by filtration and thorough washing with Et_2O (30 mL) followed by prolonged drying under high vacuum. The product was identified by analytical HPLC (method) at R_t = 14.6 minutes but not further purified due to the product's susceptibility to hydrolysis. The product was afforded as a deep green/brown solid (0.05 g, 88 %). ν_{max} (Neat)/cm^{-1} 1739 (C=O); m.p. = 150-152 ˚C. Due to the rapid hydrolysis of the ester in the mass spectrometry methodology we were unable to see the product, only the precursor **52**. As described in the results and discussion, the product **53** was taken on without further action.

2.14 Peptide experimental

2.14.1 General solid phase reagents and equipment

Protected resin-bound glycine and all subsequent Fmoc protected amino acids were bought in fresh from Novabiochem or AGTC Bioproducts Ltd. Coupling reagent PyBop was also purchased from Novabiochem; anhydrous Hunig's base (DIPEA), piperidine and anhydrous TFA were sourced from Sigma-Aldrich; bead stain test reagent, picrylsulphonic acid (TNBSA), was supplied by Fluka as a 1% solution in DMF.

Dry solvents for coupling, capping, and washing protocols were all freshly distilled over appropriate drying agents and stored over molecular sieves, or purchased as dry, prior to the commencement of synthesis.

All solid-phase chemistry was carried out in plastic Luer vessels with replacable filters and agitation was achieved using a IKA-Vibrax-VXR machine. HPLC purification was performed using the methods listed in section 3.1.3. The ideal UV absorbance wavelength was found to be 254 nm and the most effective MALDI-TOF matrix was Gentisic Acid with 0.1 % TFA.

2.14.2 General solid phase synthetic protocol

Prior to initiation of the synthetic procedure the dry resin was always swollen by soaking in dry DMF for 10 minutes. Peptide synthesis was performed on a 0.33 mmol scale.

Fmoc removal

A 20% solution of piperidine in DMF (10 mL) was prepared and the deprotection enacted by shaking the resin-bound Fmoc protected peptide in 5 mL of this solution for 12 minutes. Following removal of solvent under reduced pressure and a rinse with fresh DMF (2 x 5 mL), the process was repeated a second time. Removal of the solvent under reduced pressure after the second shake was followed by a full washing protocol and a bead stain test.

Washing protocol

MeOH (2 x 5 mL) 1 minute shakes (capping is also effected in this washing step through the shrinking of the resin).

DMF (6 x 5 mL) 1 minute shakes.

CH_2Cl_2 (2 x 5mL) 1 minute shakes (only necessary prior to resin cleavage)

Bead stain test

A few resin beads were removed from the luer flask and placed in a small vial with DMF (2 drops), DIPEA (2 drops) and TNBSA (1 drop). A positive test was confirmed when upon gentle agitation for 1 – 2 minutes the beads turned an orange/red colour. Any uncoloured beads indicated an incomplete reaction and a further cleavage step was carried out.

Coupling protocol

When necessary a double coupling procedure was performed as follows: the protected peptide next in the amino acid sequence (2 eq) was pre-activated in a dry round-bottomed flask with PyBop (2 eq) & DIPEA (5 eq) in DMF (5 mL), and stirred under argon for 10 minutes. This mixture was then directly transferred into the Luer vessel containing the free amine of the swollen resin-bound peptide and shaken for 2 hours. The resin was then rinsed with DMF (2 x 5 mL, 1 min. shakes) and the coupling procedure repeated again. A washing protocol was then effected.

2-Chlorotrityl resin cleavage

A solution of 1% TFA in dry CH_2Cl_2 (15 mL) was prepared. A quantity (3 mL) of this solution was then added to the resin-bound peptide and a 1 minute shake repeated 3 - 5 times to effect full cleavage of the peptide from the acid labile 2-chlorotrityl resin. The naked resin was then washed thoroughly with CH_2Cl_2 and the washings all collected and combined with the TFA cleavage washings. The solvent and TFA were removed under reduced pressure and tenacious excess TFA then subject to thorough trituration with Et_2O to reveal the crude fully protected linear peptide.

Mini cleave protocol

The resin-bound peptide was suspended in CH_2Cl_2 and a few beads removed via a glass pipette. This was then transferred to a clean Luer vessel and the 2-chlorotrityl resin cleavage protocol performed. The crude peptide was then subject to HPLC and mass spectrometric analysis.

Individual protocols referred to in the following experimental pages are highlighted by italics.

2.14.3 Peptide synthesis

Solid phase preparation of fully protected linear peptide DfKRG.

Alloc **54**, IvDde **55**.

Fmoc-Asp(ᵗBu)-DPhe-Lys(Alloc)-Arg(pmc)-Gly   Fmoc-Asp(ᵗBu)-DPhe-Lys(IvDde)-Arg(pmc)-Gly

$C_{64}H_{83}N_9O_{15}S$
Exact Mass: 1249.57
Mol. Wt.: 1250.46

$C_{73}H_{97}N_9O_{15}S$
Exact Mass: 1371.68
Mol. Wt.: 1372.67
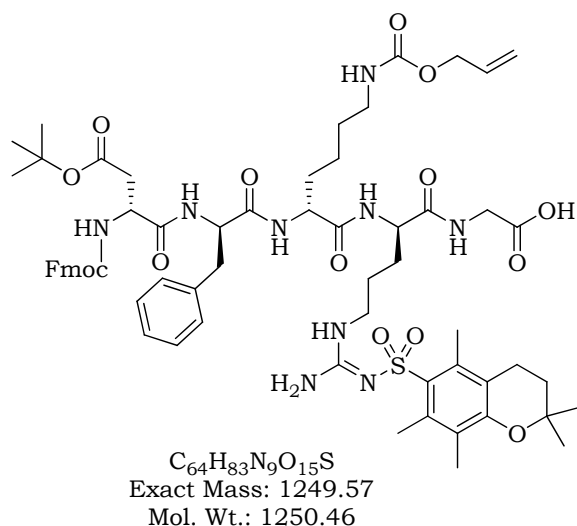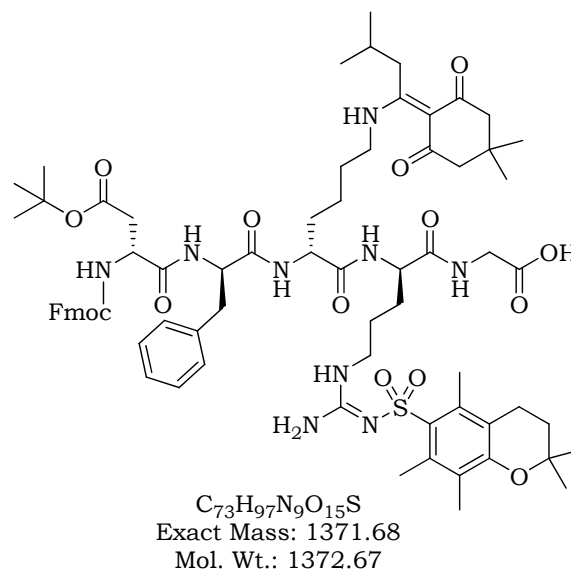
Both fully protected linear penta-peptides were constructed by hand using standard Fmoc solid phase peptide synthesis protocol. Scrupulous sequential deprotection, washing and coupling steps were observed.

Fmoc-Gly-2-Cltrl resin (0.50g, 0.33 mmol) was pre-swollen by soaking in DMF for 10 minutes prior to being subject to *Fmoc removal* and *washing protocols*. A *bead stain test* revealed the primary amine and a pre-described double *coupling protocol* effected using: Fmoc-Arg(pmc)-OH (0.44 g, 0.66 mmol), PyBop (0.34 g, 0.66 mmol) and DIPEA (0.29 mL, 1.65 mmol) in DMF (5 mL).

Washing, Fmoc removal and *bead stain test* protocols were then performed prior to the next protected amino acid being installed by an identical double *coupling protocol* using: Fmoc-Lys(IvDde)-OH (0.38 g, 0.66 mmol) or Fmoc-Lys(Alloc)-OH (0.30 g, 0.66 mmol), PyBop (0.34 g, 0.66 mmol) and DIPEA (0.29 mL, 1.65 mmol) in DMF (5 mL).

Washing, Fmoc removal and *bead stain test* protocols were then performed prior to the next protected amino acid being installed by an identical double *coupling protocol* using: Fmoc-DPhe-OH (0.26 g, 0.66 mmol), PyBop (0.34 g, 0.66 mmol) and DIPEA (0.29 mL, 1.65 mmol) in DMF (5 mL).

Washing, Fmoc removal and *bead stain test* protocols were then performed prior to the next protected amino acid being installed by an identical double *coupling protocol* using: Fmoc-Asp(ᵗBu)-OH (0.27 g, 0.66 mmol), PyBop (0.34 g, 0.66 mmol) and DIPEA (0.29 mL, 1.65 mmol) in DMF (5 mL). A final double washing protocol then completed the solid phase

synthesis leaving the crude fully protected penta-peptide on the resin stored at -8 °C under argon.

A *mini cleave* protocol was then carried out to confirm the success of the synthesis and to observe the purity of the product and identify any impurities:

54 Alloc peptide: Analytical HPLC (method A, t_r = 30.5 minutes); m/z (MALDI-TOF) 1272.9 (10 %, $[M + Na]^+$), 1250.9 (100 %, $[M + H]^+$), 1194.8 (5 %, $[M - {}^tBu + H]^+$), 984.9 (70 %, $[M - Pmc + H]^+$).

55 IvDde peptide: Analytical HPLC (method A, t_r = 28.6 minutes); m/z (MALDI-TOF) 1372.5 [10 %, $M + H]^+$), 1150.6 (100 %, $[M – Fmoc + H]^+$), 1104.4 (80 %, $[M - Pmc + H]^+$), 866 (30 %, $[M - Fmoc - Pmc + H]^+$).

See **appendix 5** for HPLC & m/z data

Preparation of cyclic fully protected peptides. Alloc **56**, IvDde **57**.

Cyc Asp(^{t}Bu)-DPhe-Lys(Alloc)-Arg(pmc)-Gly   **Cyc** Asp(^{t}Bu)-DPhe-Lys(IvDde)-Arg(pmc)-Gly

$C_{49}H_{71}N_9O_{12}S$
Exact Mass: 1009.49
Mol. Wt.: 1010.21

$C_{58}H_{85}N_9O_{12}S$
Exact Mass: 1131.6
Mol. Wt.: 1132.41
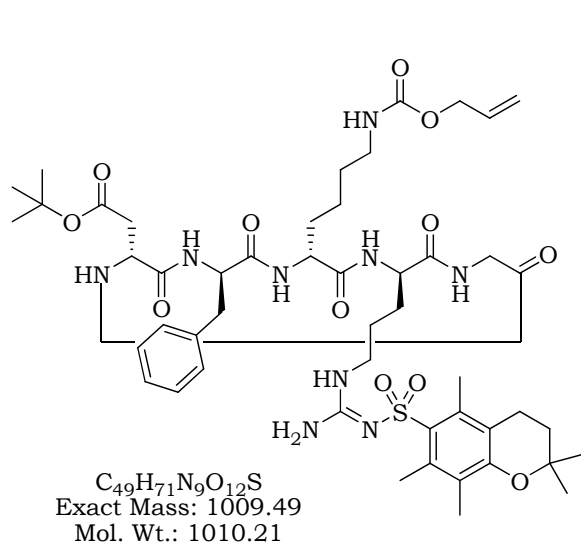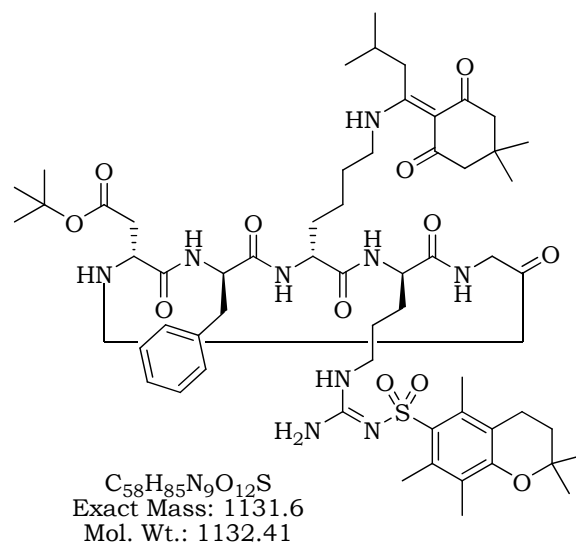
Cyclisation protocol

Resin bound fully protected linear peptide was subject to Fmoc removal and resin cleavage protocol (see general protocol section 3.4.2) followed by thorough trituration with Et_2O and dried overnight under high vacuum. The free peptide was then made up as a 0.5 mM solution in dry CH_2Cl_2. PyBop (1.2 eq) and DIPEA (3.0 eq) were added and the mixture stirred at ambient temperature for 5 hours with monitoring by analytical HPLC method D (230 nm). The crude mixture was then concentrated under reduced pressure and purified

by preparative HPLC. Lyophilisation revealed the pure cyclic peptide as a fine white powder.

56 Alloc peptide: Analytical HPLC (method A, t_r = 30.5 minutes); m/z (MALDI-TOF) 1010.9 (100 %, $[M + H]^+$), 954.8 (10 %, $[M - {}^tBu + H]^+$), 744.1 (70 %, $[M - Pmc + H]^+$).

57 IvDde peptide: Preparative HPLC (method D, t_r = 44.6 minutes); m/z (MALDI-TOF) 1133 $[M + H]^+$, 1077 $[M - {}^tBu + H]^+$, 867 $[M - Pmc + H]^+$.

See **appendix 5** for HPLC and m/z data

Deprotection to afford free lysine εNH₂ **58**.

Cyc Asp(^{t}Bu)-DPhe-Lys-Arg(pmc)-Gly

$C_{45}H_{67}N_9O_{10}S$
Exact Mass: 925.47
Mol. Wt.: 926.13
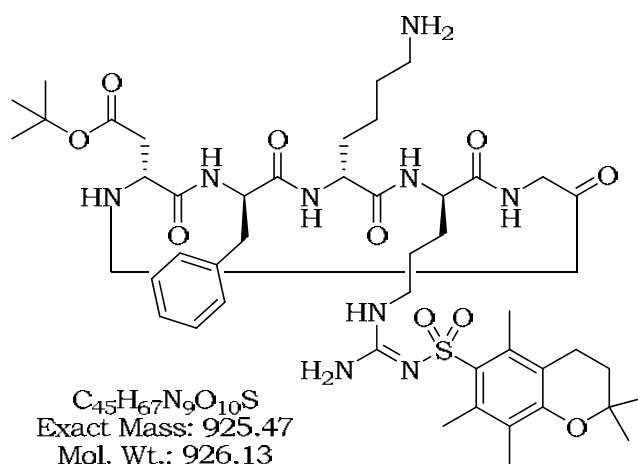
A. Alloc deprotection

Fully protected cyclised peptide **56** (0.015 g, 1.49×10^{-5} mol) was taken up in dry MeOH (5 mL) and stirred under argon at ambient temperature prior to addition of $Pd(PPh_3)_4$ (1.00 mol %). After 5 minutes stirring at ambient temperature, K_2CO_3 (0.0124 g, 8.91×10^{-5} mol) was added and the mixture stirred for a further 3 hours. The solvent was removed under reduced pressure and the crude mixture purified by semi-preparative HPLC (method A, t_r = 20.1 minutes) and lyophilised to give pure deprotected cyclic peptide **58** as a fine white powder (0.0125 g, 91 %). m/z (MALDI-TOF) 926 $[M + H]^+$, 870 $[M - {}^tBu + H]^+$, 660 $[M - Pmc + H]^+$.

B. IvDde deprotection

Fully protected cyclised peptide **57** (0.020 g, 1.77×10^{-5} mol) was dissolved in a solution of 2 % hydrazine hydrate in MeCN (2 mL). The mixture was stirred for 3 hours at ambient temperature, the solvent removed under reduced pressure and the crude product promptly

purified by semi-preparative HPLC (method E, t_r = 33.6 minutes). Lyophilisation yielded the pure deprotected product as a fine white powder (0.015 g, 92 %). m/z (MALDI-TOF) 926 [M + H]⁺, 870 [M - ᵗBu + H]⁺, 660 [M - Pmc + H]⁺.

See **appendix 5** for HPLC & m/z data

Preparation of fully deprotected cyclic peptide (blocking peptide) **59**.

Cyc Asp-DPhe-Lys-Arg-Gly

$C_{27}H_{41}N_9O_7$
Exact Mass: 603.31
Mol. Wt.: 603.67
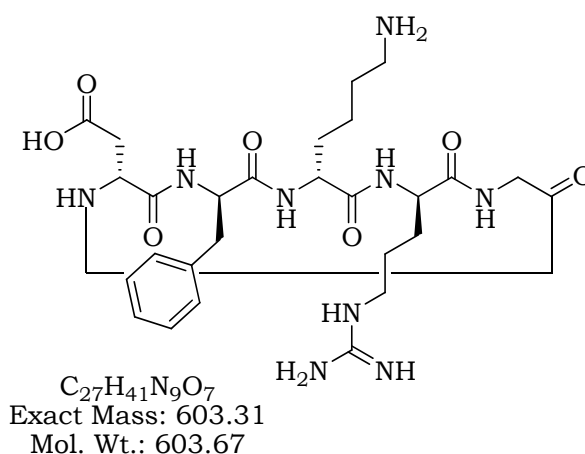
Alloc or IvDde deprotected cyclic peptide **58** (0.02 g, 2.16 x 10⁻⁵ mol) was dissolved in anhydrous CH_2Cl_2 (2 mL) and anhydrous TFA (2 mL) added. The mixture was then stirred at ambient temperature under $Ar_{(g)}$ for 4 hours and concentrated under reduced pressure. For free cyclic peptide to be then coupled to a fluorophore via the lysine εNH₂ the crude product was then repeatedly co-concentrated with CH_2Cl_2 and triturated with Et_2O before being taken on without further purification. Free cyclic peptide for blocking studies was purified by semi-preparative HPLC (method C, t_r = 23.08 minutes) and lyophilised to afford the pure product as a fine white powder (0.012 g, 95 %). m/z (MALDI-TOF) 604.2 [M + H]⁺, 626.2 [M + Na]⁺.

See **appendix 5** for HPLC and m/z data

Preparation of fluorescein labelled cyclic peptide **60**.

Cyc Asp(^{t}Bu)-DPhe-Lys(fluorescein)-Arg(pmc)-Gly

$C_{48}H_{53}N_{10}O_{12}S$
Exact Mass: 993.36
Mol. Wt.: 994.06
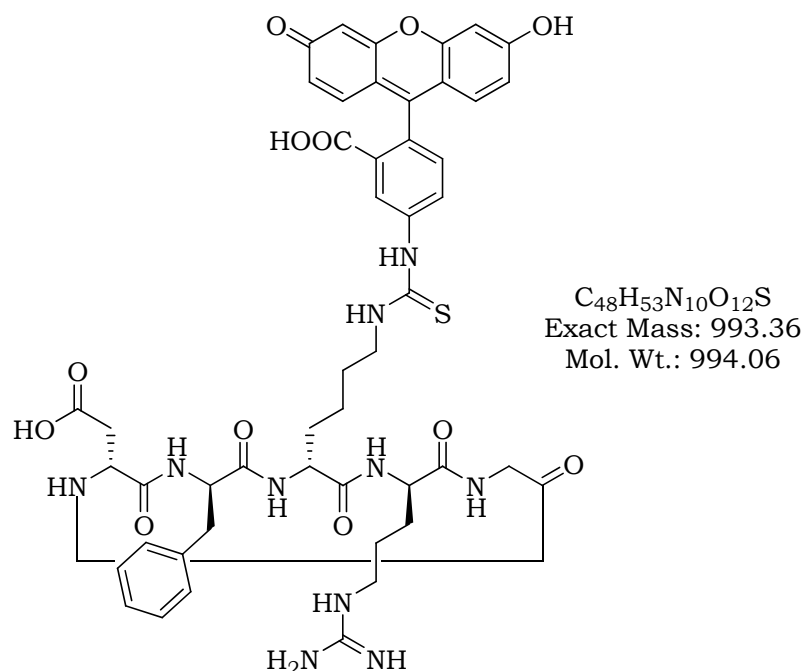
Pmc and ^{t}Bu protected cyclic peptide **58** (0.002 g, 1.08 x 10^{-5} mol) was dissolved in a solution of DMF / CH$_2$Cl$_2$ / Py (5 : 7 : 12 (v/v), 5 mL) and fluorescein isothiocyanate (0.002 g, 4.98 x 10^{-6}) added. The reaction was then monitored by analytical HPLC (method A, t_r = 25.1 minutes) M/z MALDI-TOF showed 994 [M + H]$^+$, 1260 [M + Pmc + H]$^+$, 1282 [M + Pmc + Na]$^+$ and was complete after stirring for 72 hours at ambient temperature in the dark. The crude product was concentrated under reduced pressure and immediately taken up in a 1:1 (v/v) mix of CH$_2$Cl$_2$ and TFA and stirred for a further 4 hours at ambient temperature. Again the crude mixture was analysed by analytical HPLC (method A), purified by semi-preparative HPLC (method B, t_r = 31.9 minutes), and lyophilised to afford the pure product as an orange coloured powder (0.009 g, 56 %). m/z (MALDI-TOF) 994 ([M + H]$^+$, 100 %).

See **appendix 4** for HPLC and m/z data
See **appendix 5** for HPLC and m/z data

Preparation of NIR797 labelled cyclic peptide **61**.

$C_{72}H_{91}N_{12}NaO_{13}S_4$
Exact Mass: 1482.56
Mol. Wt.: 1483.82
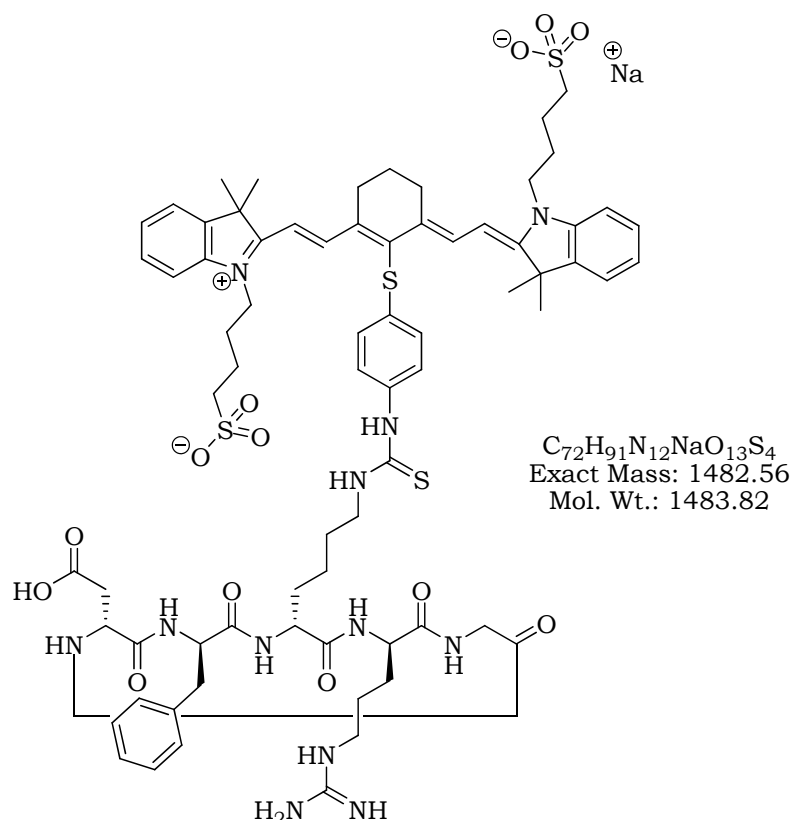
From 58

Pmc and ^tBu protected cyclic peptide **58** (0.001 g, 1.08 x 10^{-6} mol) was dissolved in pH 9.2 phosphate buffer solution (3 mL) and NIR797 isothiocyanate (0.0019 g, 2.16 x 10^{-5} mol) added. The reaction was then stirred for 72 hours at ambient temperature in the dark with monitoring by analytical HPLC (method B). The crude product was concentrated under reduced pressure, immediately taken up in a 1:1 (v/v) mix of CH_2Cl_2 and TFA and stirred for a further 4 hours at ambient temperature. The crude mixture was then concentrated and purified by semi-preparative HPLC (method C, t_r = 34.4 minutes), and lyophilised to afford the pure product as a deep green powder (0.001 g, 62 %).

From 59

Fully deprotected cyclic peptide **59** (0.001 g, 1.66 x 10^{-6} mol) was dissolved in pH 9.2 phosphate buffer solution (3 mL) and NIR797 isothiocyanate (0.003 g, 3.32 x 10^{-6} mol) added. The reaction was then stirred for 72 hours at ambient temperature in the dark. The crude product was then concentrated under reduced pressure, purified using semi-preparative HPLC (method C, t_r = 34.4 minutes) and lyophilised to afford the product as a deep green solid (0.0015 g, 63 %). m/z (MALDI-TOF) 1461 ([M – Na + H]$^+$, 100 %).

See **appendix 4** for fluorescence data

See **appendix 5** for HPLC & m/z data

Preparation of Os(bpy)$_2$(4-methylbpy) labelled cyclic peptide **62**.

C$_{59}$H$_{65}$N$_{15}$O$_8$Os
Exact Mass: 1303.48
Mol. Wt.: 1302.47
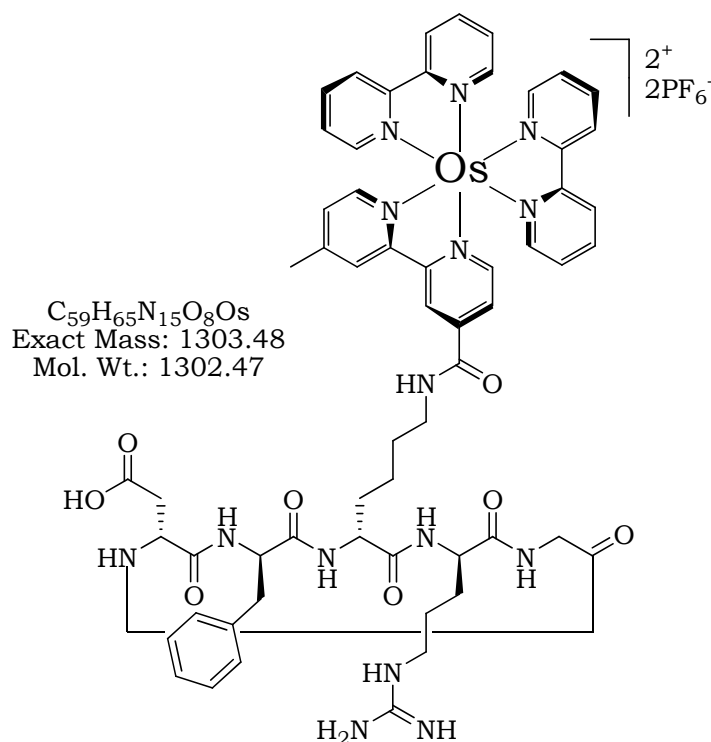
Pmc & ^{t}Bu protected cyclic peptide **58** (0.002 g, 2.16 x 10^{-6} mol) was dissolved in anhydrous MeCN (5 mL) and osmium bipyridyl succinic ester derivative MXXM (0.0018 g, 2.16 x 10^{-6}) added. The mixture was heated in the dark at 70 °C under Ar$_{(g)}$ for 48 hours. The product was revealed by analytical HPLC (method E, t_r = 30.3 minutes). The crude mixture concentrated under reduced pressure, purified using semi-preparative HPLC (method E, t_r = 39.3 minutes) and lyophilised to afford the product as a deep brown solid (0.002 g, 57 %). M/z MALDI-TOF 1624 [M + Pmc + ^{t}Bu + H]$^+$, 1938 [M + Pmc + ^{t}Bu + 2PF$_6{}^{2-}$ + Na + 2H]$^+$.

This intermediate product was then treated with a mixture of TFA in CH$_2$Cl$_2$ (1 : 1 (v/v), 4 mL) and stirred for 4 hours at ambient temperature. The crude mixture was then concentrated under reduced pressure with repeated co-concentrations with CH$_2$Cl$_2$ to remove all TFA and finally triturated with Et$_2$O to afford the product as a deep brown solid (0.0015 g, 96 %). m/z (MALDI-TOF) 1303 ([M + H]$^+$, 100 %) , 1617 ([M + 2PF$_6{}^{2-}$ + Na + 2H]$^+$, 60 %).

See **appendix 4** for fluorescence data
See **appendix 5** for HPLC & m/z data

Preparation of spacer appended cyclic peptides **63** & **64**.

Cyc Asp(tBu)-DPhe-Lys(Butoxycarbonyl aminocaproic acid)-Arg(pmc)-Gly

Cyc Asp(tBu)-DPhe-Lys(Butoxycarbonylamino hexaethylene glycol)-Arg(pmc)-Gly

63

$C_{56}H_{86}N_{10}O_{13}S$
Exact Mass: 1138.61
Mol. Wt.: 1139.41

64

$C_{64}H_{102}N_{10}O_{19}S$
Exact Mass: 1346.7
Mol. Wt.: 1347.62
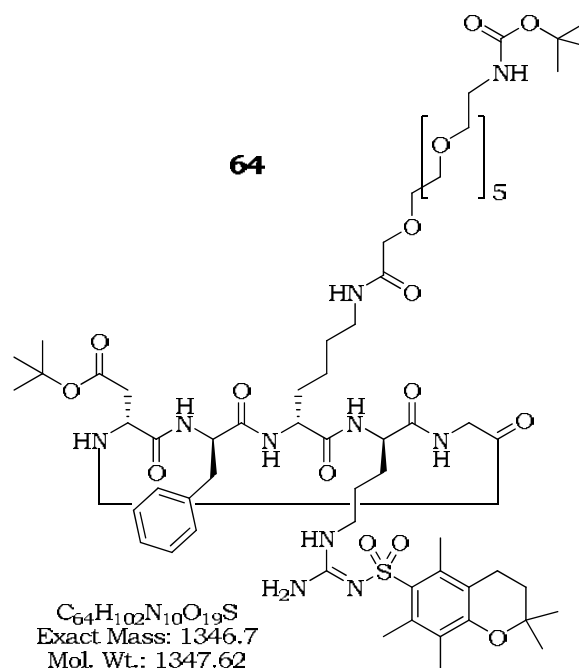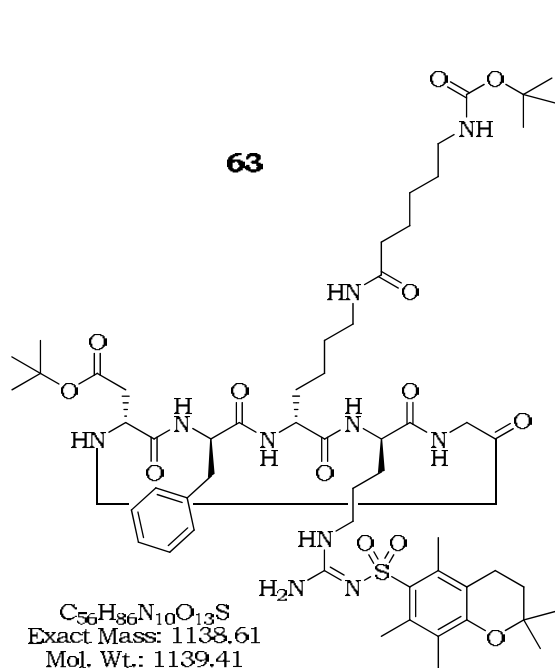
Pmc & tBu protected cyclic peptide **58** (0.02 g, 2.16 x 10^{-5}) was dissolved in anhydrous MeCN (5 mL). Boc-aminocaproic acid (0.0075 g, 3.24 x 10^{-5}) or N-Boc-amino hexa(ethylene glycol) acetic acid **44** (0.014 g, 3.24 x 10^{-5}) was combined with PyBop (0.022 g, 4.39 x 10^{-5}) and DIPEA (0.011 g, 6.48 x 10^{-5}) in anhydrous MeCN (5 mL) and stirred at ambient temperature under $Ar_{(g)}$ in a separate vessel for 10 minutes prior to addition to the peptide and stirring under $Ar_{(g)}$ for 72 hours. The product was identified by analytical HPLC (method A, t_r = 39.3 minutes (caproic)), (method E, t_r = 38.2 minutes (PEG)), concentrated under reduced pressure and purified by semi-preparative HPLC.

63 (0.017 g, 68 %) Semi-preparative HPLC (method A, t_r = 33.8 minutes); m/z (MALDI-TOF) 1161 [M + Na]+, 1139 [M + H]+, 1039 [M – Boc + H]+, 873 [M – Pmc + H]+.

64 (0.016 g, 54 %) Semi-preparative HPLC (method E, t_r = 32.5 minutes); m/z (MALDI-TOF) 1347 [M + H]+, 1291 [M – tBu + H]+, 1247 [M – Boc + H]+, 1191 [M – Boc – tBu + H]+, 1081 [M – Pmc + H]+.

See **appendix 5** for HPLC & m/z data

Deprotection of Boc, ^{t}Bu & Pmc to afford free extended cyclic peptides **65** & **66**.

65

NH$_2$

C$_{33}$H$_{52}$N$_{10}$O$_8$
Exact Mass: 716.4
Mol. Wt.: 716.83

66

NH$_2$

O

5

C$_{41}$H$_{68}$N$_{10}$O$_{14}$
Exact Mass: 924.49
Mol. Wt.: 925.04
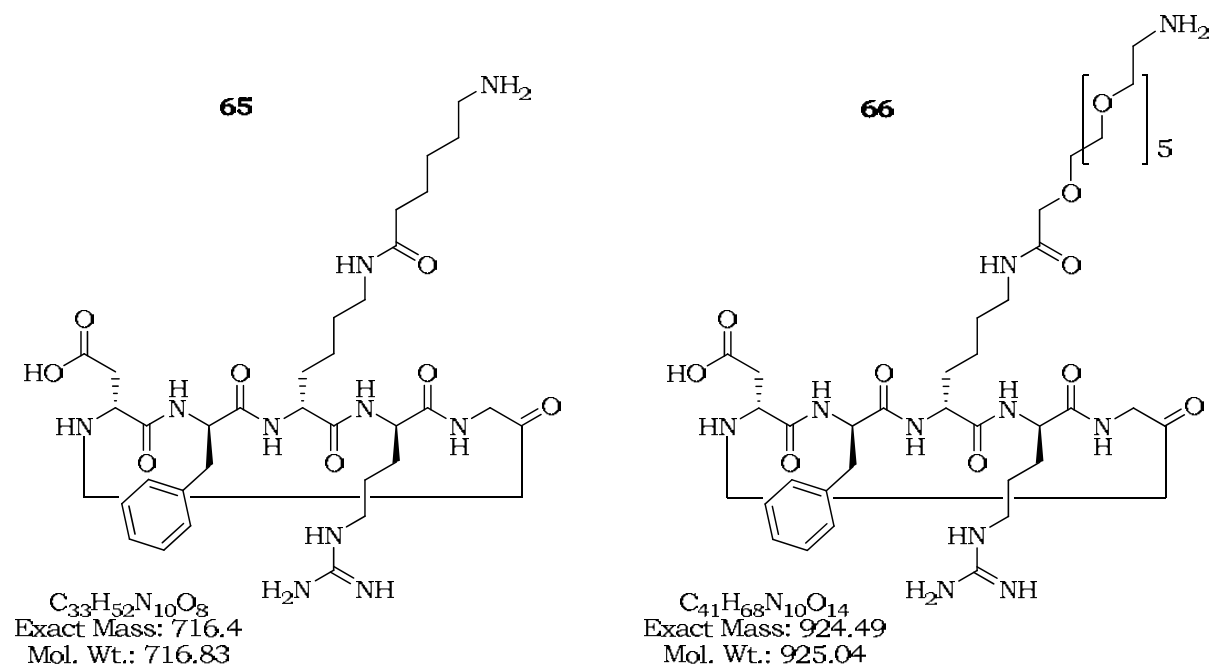
Cyclic peptides **63** (0.005 g, 4.39 x 10^{-6} mol) & **64** (0.005 g, 3.71 x 10^{-6} mol) were dissolved in anhydrous CH$_2$Cl$_2$ (2 mL) and TFA (2 mL) added. The mixture was then stirred at ambient temperature under Ar$_{(g)}$ for 4 hours before being concentrated under reduced pressure and repeatedly co-concentrated with CH$_2$Cl$_2$. The crude mixture was then triturated with Et$_2$O and dried to afford the products as either a fine white solid (**65**), or a colourless oil (**66**).

65 (0.003 g, 93 %), m/z (MALDI-TOF) 717.1 ([M + H]$^+$, 100 %)

66 (0.003 g, 95 %), m/z (MALDI-TOF) 926.4 ([M + H]$^+$, 100 %)

See **appendix 5** for m/z data

Preparation of NIR797 appended cyclic peptide **67**.

C$_{78}$H$_{102}$N$_{13}$NaO$_{14}$S$_4$
Exact Mass: 1595.64
Mol. Wt.: 1596.97
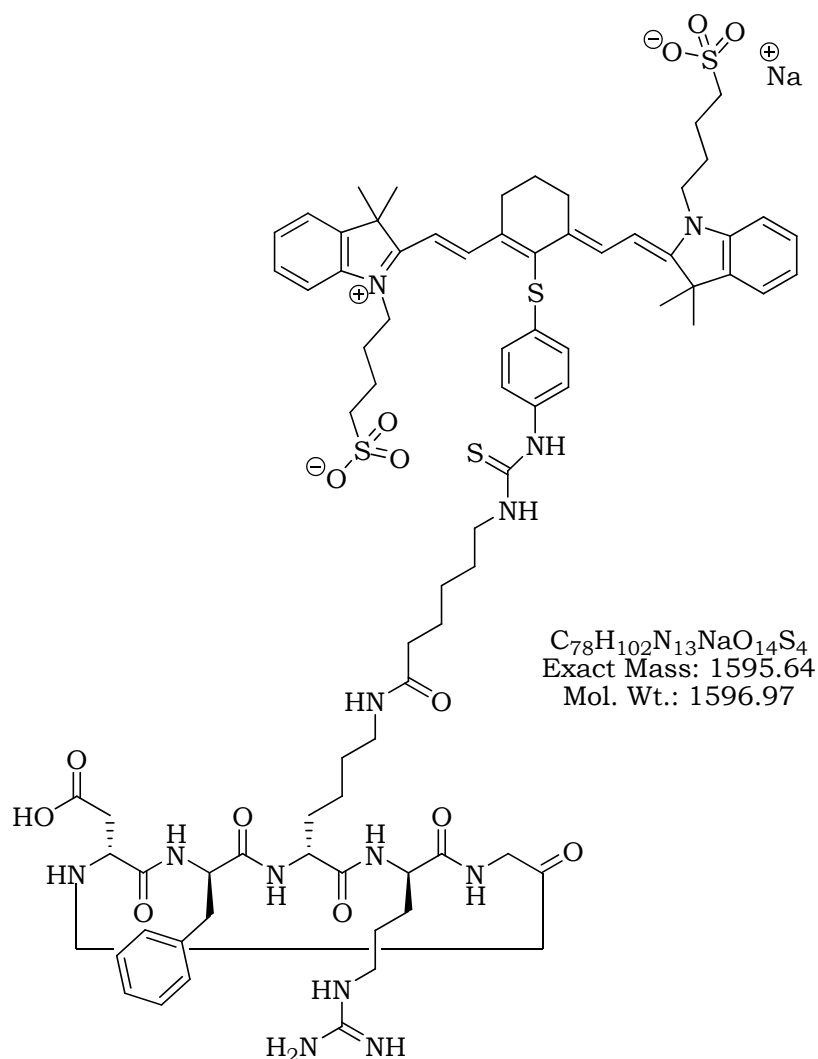
Extended cyclic peptide **65** (0.001 g, 1.40 x 10^{-6} mol) was dissolved in pH 9.2 phosphate buffer solution (5 mL) and NIR797 (0.0025 g, 2.79 x 10^{-6} mol) added. The mixture was then stirred at ambient temperature for 72 hours in the dark and monitored by analytical HPLC (method B, t$_r$ product = 26.9 minutes). The crude product was concentrated under reduced pressure, purified by semi-preparative HPLC (method B, t$_r$ = 27.9 minutes) and lyophilised to afford the product as a deep green solid (0.0015 g, 66 %). M/z (MALDI-TOF) 1631 [M + K]$^+$, 1596 [M + H]$^+$, 1574 [M – Na + H]$^+$.

See **appendix 4** for fluorescence data
See **appendix 5** for HPLC and m/z data

Preparation of osmium complex appended cyclic peptide **68**.

$C_{65}H_{76}N_{16}O_9Os$
Exact Mass: 1416.56
Mol. Wt.: 1415.63
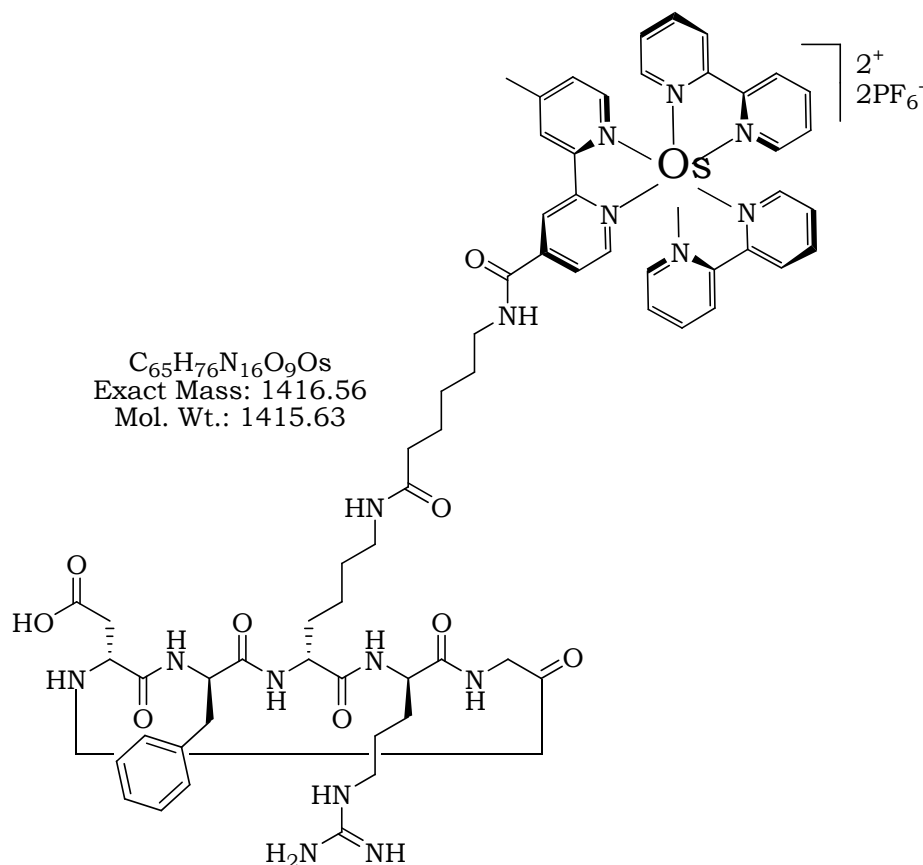
Caproic extended cyclic peptide **65** (0.002 g, 2.79 x 10^{-6} mol) was dissolved in anhydrous DMF (5 mL) and osmium succinic ester complex **53** (0.0027 g, 3.35 x 10^{-6} mol) added. The mixture was then stirred at ambient temperature under $Ar_{(g)}$ in the dark for 72 hours. Analytical HPLC revealed the product (method A, t_r = 15.2 minutes). The crude product was concentrated under reduced pressure, purified by semi-preparative HPLC (method A, t_r = 21.4 minutes) and lyophilised to afford the product as a deep brown solid (0.0018 g, 46 %). M/z (MALDI-TOF) 1417 $[M + H]^+$, 1439 $[M + Na]^+$, 1708 $[M + 2PF_6^{2-} + 2H]^+$.

See **appendix 4** for fluorescence data
See **appendix 5** for HPLC & m/z data

Preparation of Ugi complex appended cyclic peptide **69**.

$C_{82}H_{115}N_{15}O_{19}STb$
Exact Mass: 1804.75
Mol. Wt.: 1805.87
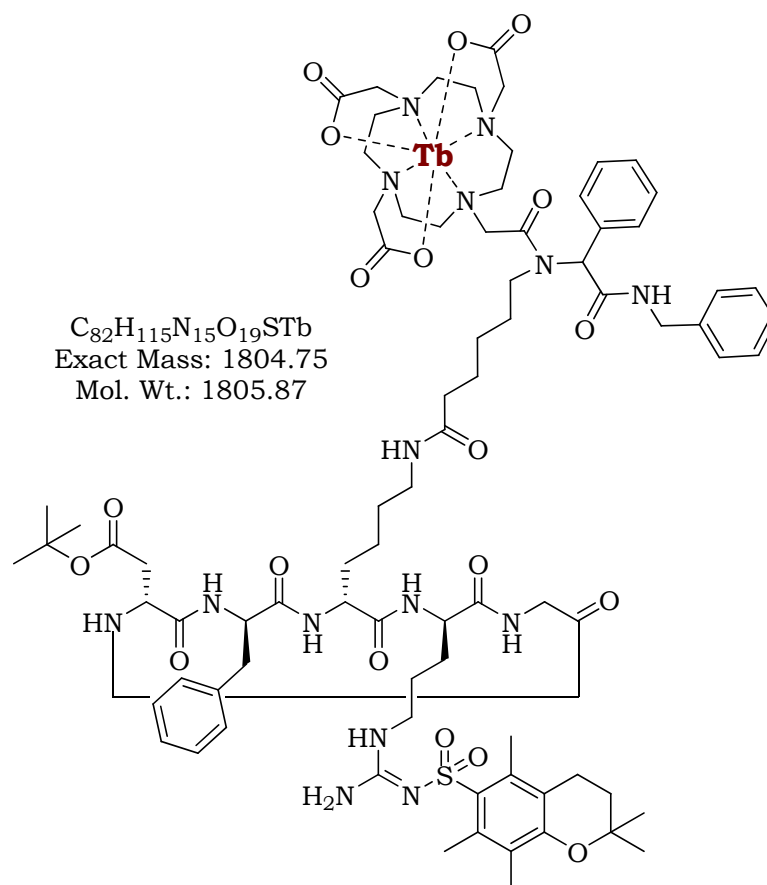
Protected cyclic peptide **58** (0.001 g, 1.08 x 10⁻⁶ mol) was dissolved in anhydrous DMF (2 mL). PyBop (0.0007 g, 1.30 x 10⁻⁶ mol), DIPEA (0.0004 mL, 2.16 x 10⁻⁶ mL) were prestirred for 10 minutes with Ugi terbium derivative **34** (0.001 g, 1.08 x 10⁻⁶ mol) in anhydrous DMF (2 mL) before addition to the peptide. The mixture was then stirred for 72 hours at ambient temperature and monitored by analytical HPLC (method A). Upon completion the crude product was purified by semi-preparative HPLC (method B, t_r = 37.2 minutes) and lyophilised to afford the product as a fine white powder (0.0008 g, 37 %). m/z (MALDI-TOF) 1805.0 ([M+H]⁺, 100 %).

2.15 Miscellaneous experimental

PEG Ugi test reaction: preparation of **92**.

$C_{40}H_{54}N_2O_{10}$
Mol. Wt.: 722.86
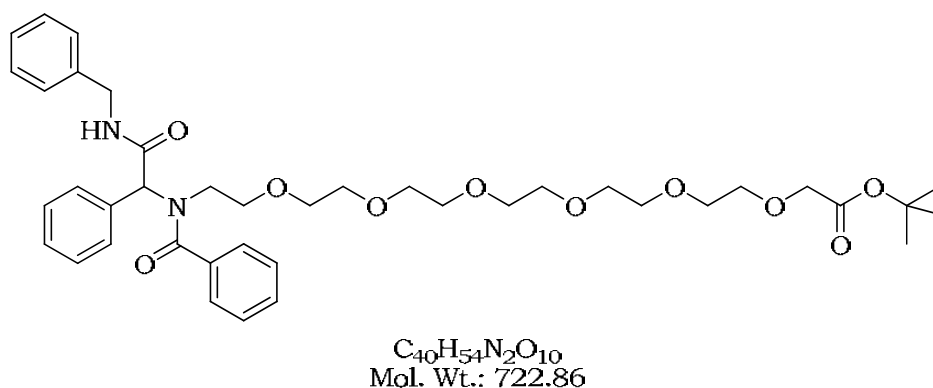
O-*tert*-butoxycarbonylmethyl-6-amino-(ethylene glycol) **42** (0.20 g, 0.51 mmol) was combined with benzaldehyde (0.05 mL, 0.51 mmol) and stirred for 10 minuted in anhydrous MeOH (8 mL) under argon at ambient temperature. Benzoic acid (0.06 g, 0.51 mmol) and benzyl isocyanide (0.06 mL, 0.51 mmol) were the added and the reaction stirred for a further 48 hours. Upon reaction completion the crude product was concentrated under reduced pressure and purified by flash column chromatography (CH_2Cl_2 / MeOH / $NH_{3(aq)}$ 85 : 14.9 : 0.1, R_f = 0.53) to afford the product as a colourless glassy solid. (0.34 g, 93 %). δ_H (300 MHz; CDCl$_3$) 7.60-7.22 (15H, stack, C*H* arom), 4.60 (1H, br s, N*H*), 4.02 (2H, s, C*H*$_2$C=O), 3.72-3.44 (26H, stack, OC*H*$_2$, PhC*H*$_2$), 1.48 (9H, s, C(C*H*$_3$)$_3$); δ_C (75 MHz) 169.4 (*C*=O), 169.1 (*C*=O), 168.2 (*C*=O), 138.4 (*C*$_q$ arom), 135.7 (*C*$_q$ arom), 135.4 (*C*$_q$ arom), 132.6 (*C* arom), 129.6 (*C* arom), 128.9 (*C* arom), 128.7 (*C* arom), 128.6 (*C* arom), 128.5 (*C* arom), 128.3 (*C* arom), 128.0 (*C* arom), 127.6 (*C* arom), 127.2 (*C* arom), 127.1 (*C* arom), 127.0 (*C* arom), 126.9 (*C* arom), 81.2 (*C*(CH$_3$)$_3$), 70.4 (O*C*H$_2$), 70.3 (O*C*H$_2$), 70.2 (O*C*H$_2$), 69.9 (O*C*H$_2$), 69.8 (O*C*H$_2$), 68.7 (Ph*C*H$_2$), 43.2 (N*C*H$_2$CH$_2$O), 28.3 (C(*C*H$_3$)$_3$); m/z (TOF ES+) 745.3676 ([M+Na]$^+$, 100%, $C_{40}H_{54}N_2O_{10}$Na requires 745.3670).

Preparation of N-*tert*-butoxycarbonyl-N'biotinyl-3,6-dioxaoctane-1,8diamine **93**.

$C_{21}H_{38}N_4O_6S$
Mol. Wt.: 474.61
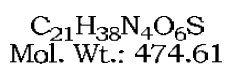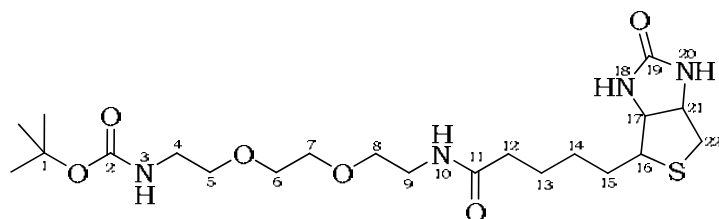
Biotin (0.50 g, 2.04 mmol), HBTU (0.67 g, 1.77 mmol) and DIPEA (0.38 mL, 2.18 mmol) were combined in anhydrous DMF (50 mL) and sonicated under argon at ambient temperature for 15 minutes. N-Boc-3,6-dioxaoctane-1,8-diamine **45** (0.34 g, 1.36 mmol) in DMF (10 mL) was then added dropwise over 20 minutes and the mixture stirred for a further hour at ambient temperature under argon. The crude cocktail was then concentrated under reduced pressure and purified either by gradient flash column chromatography R_f = 0.36 (CH_2Cl_2 to CH_2Cl_2 / MeOH / $NH_{3(aq)}$, 8.8 : 0.9 : 0.3), or preparative HPLC (method C) t_r = 31.9 minutes. The pure product was recovered as a fine white powder (0.57 g, 98 %). ν_{max} (Neat)/cm^{-1} ;δ_H (500 MHz; CDCl$_3$) 6.79 (1H, br s, N*H*-18), 6.72 (1H, br s, N*H*-10), 6.13 (1H, br s, N*H*-20), 5.12 (1H, br s, N*H*-3), 4.49 (1H, dd, J 4.9, 7.5, *H*-21), 4.30 (1H, dd, J 4.6, 7.5, *H*-17), 3.58 (4H, app s, 2 x *H*-6, 2 x *H*-7), 3.53 (4H, stack, 2 x *H*-5, 2 x *H*-8), 3.41 (2H, mound, 2 x *H*-9), 3.28 (2H, mound, 2 x *H*-4), 3.12 (1H, dd, J 7.2, 11.8, *H*-16), 2.87 (1H, dd, J 4.9, 12.8, *H*-22), 2.72 (1H, d, J 12.8, *H*-22), 2.21 (2H, t, J 7.4, 2 x *H*-12), 1.65 (4H, br stack, 2 x *H*-13, 2 x *H*-15), 1.41 (11H, app s, (C*H*$_3$)$_3$, 2 x *H*14); δ_C (125 MHz) 173.5 (*C*-11), 164.4 (*C*-19), 156.0 (*C*-2), 79.3 (*C*-1), 70.0 (*C*-5, *C*-6 & *C*-7), 69.8 (*C*-8), 61.9 (*C*-17), 60.3 (*C*-21), 55.5 (*C*-16), 40.4 (*C*-22), 40.3 (*C*-4), 39.1 (*C*-9), 35.8 (*C*-12), 28.3 (*C*(CH$_3$)$_3$), 28.1 (*C*-14), 27.9 (*C*-15), 25.5 (*C*-13); m/z (TOF ES+) 497.2410 ([M+Na]$^+$, 100%, $C_{21}H_{38}N_4O_6Na$ requires 497.2420).

This compound was unequivocally assigned using COSY-90, HSQC & HMBC 2D nmr experiments. **See appendix 6**.

Preparation of N-biotinyl-3,6-dioxaoctane-1,8diamine **94**.

$C_{16}H_{30}N_4O_4S$
Mol. Wt.: 374.5
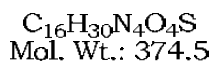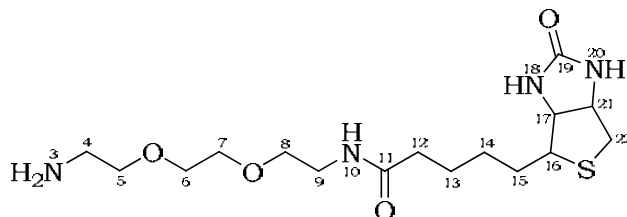
Boc protected compound **92** (0.074 g, 0.156 mmol) was dissolved in a mixture of TFA (2 mL) and anisole (0.02 mL). The mixture was then stirred at ambient temperature for 1 hour. By-products and TFA were removed under reduced pressure with repeated re-evaporation using CH_2Cl_2 (3 x 10 mL) and toluene (3 x 10 mL). Trituration with Et_2O afforded the pure product as a fine white powder (0.058 g, 99%). δ_H (300 MHz; d_4MeOH) 4.52 (1H, dd, J 4.6, 7.9, *H*-21), 4.32 (1H, dd, J 4.4, 7.9, *H*-17), 3.70 (2H, app t, J 5.2, 2x*H*-8 or 2x*H*-5), 3.65 (4H, app s, 2x*H*-6 & 2x*H*-7), 3.55 (2H, app t, J 5.2, 2x*H*-8 or 2x*H*-5), 3.37 (2H, app t, J 5.9, 2x*H*-9), 3.21 (1H, m, *H*-16), 3.13 (2H, mound, 2x*H*-4), 2.93 (1H, dd, J 4.9, 12.8, *H*-22), 2.71 (1H, d, J 12.8, *H*-22), 2.22 (2H, t, J 7.3, 2x*H*-12), 1.65 (4H, stack, 2x*H*-15 & 2x*H*-13), 1.44 (2H, m, 2x*H*-14); δ_C (300MHz) 176.3 (*C*-11), 166.1 (*C*-19), 72.1 (*C*-6 or *C*-7), 71.9 (*C*-6 or *C*-7), 71.4 (*C*-5), 68.6 (*C*-8), 64.1 (*C*-17), 62.3 (*C*-21), 57.7 (*C*-16), 41.7 (*C*-22), 41.3 (*C*-4), 40.8 (*C*-9), 37.4 (*C*-12), 30.4 (*C*-14), 30.2 (*C*-15), 27.5 (*C*-13); m/z (TOF ES+) 375.2066 ([M+H]⁺, 100%, $C_{16}H_{31}N_4O_4S$ requires 375.2072).
See **appendix 6** for nmr data

Preparation of N-*tert*-butoxycarbonyl-N'biotinyl-3,6-octane-1,8diamine **95**.

$C_{21}H_{38}N_4O_4S$
Mol. Wt.: 442.62
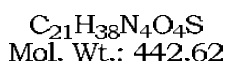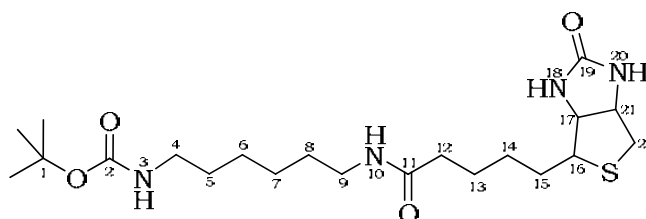
Biotin (0.39 g, 1.60 mmol), HBTU (0.53 g, 1.39 mmol) and DIPEA (0.30 mL, 1.70 mmol) were combined in anhydrous DMF (50 mL) and sonicated under argon at ambient temperature for 15 minutes. N-Boc-1,8-diamine **46** (0.23 g, 1.06 mmol) in DMF (10 mL)

was then added dropwise over 20 minutes and the mixture stirred for a further hour at ambient temperature under argon. The crude cocktail was then concentrated under reduced pressure and purified by gradient flash column chromatography R_f = 0.36 (CH_2Cl_2 to CH_2Cl_2 / MeOH / $NH_{3(aq)}$, 8.8 : 0.9 : 0.3) to afford the product as a fine white powder (0.36 g, 76 %). δ_H (300 MHz d4-MeOH) 4.49 (1H, dd, J 4.2, 7.9, *H*-21), 4.30 (1H, dd, J 4.4, 7.9, *H*-17), 3.22-3.14 (4H, stack, 2x*H*-4, 2x*H*-9), 3.02 (1H, app t, 6.9, *H*-16), 2.93 (1H, dd, J 4.9, 12.7, *H*-22), 2.70 (1H, d, J 12.7, *H*-22), 2.20 (2H, t, J 7.3, 2x*H*-12), 1.76-1.46 (25H, br stack, 8xCH_2 & C(C*H*$_3$)$_3$); δ_C (750 MHz, d6-DMSO) 170.9 (*C*-11), 162.6 (*C*-19), 155.5 (*C*-2), 77.2 (*C*-1), 60.1 (*C*-17), 58.2 (*C*-21), 54.5 (*C*-16), 38.3 (CH_2), 38.1 (CH_2), 35.2 (CH_2), 29.4 (CH_2), 29.1 (CH_2), 28.2 (C(CH_3)$_3$), 28.0 (CH_2), 26.1 (CH_2), 26.0 (CH_2), 25.3 (CH_2); m/z (TOF ES+) 465.2511 ([M+Na]$^+$, 100%, $C_{21}H_{38}N_4O_4SNa$ requires 465.2501).

Preparation of N'biotinyl-3,6-octane-1,8diamine, **96**.

$C_{16}H_{30}N_4O_2S$
Mol. Wt.: 342.5
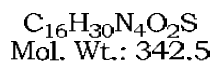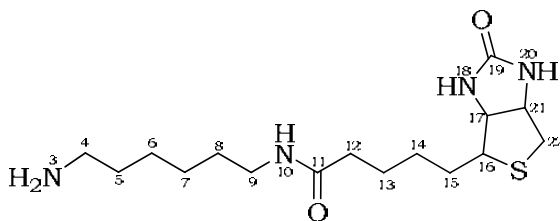
Boc protected compound **95** (0.078 g, 0.176 mmol) was dissolved in a mixture of TFA (2 mL) and anisole (0.02 mL). The mixture was then stirred at ambient temperature for 1 hour. By-products and TFA were removed under reduced pressure with repeated re-evaporation using CH_2Cl_2 (3 x 10 mL) and toluene (3 x 10 mL). Trituration with Et_2O afforded the pure product as a fine white powder (0.060 g, 99%) δ_H (300 MHz d4-MeOH) 4.49 (1H, dd, J 4.4, 7.9, *H*-21), 4.30 (1H, dd, J 4.4, 7.9, *H*-17), 3.20-3.14 (4H, stack, 2 x *H*-4, 2 x *H*-9), 3.02 (4H, stack, *H*-16, *H*-22, 2 x C*H*$_2$), 2.69 (1H, d, J 12.7, *H*-22), 2.19 (2H, t, J 7.3, 2 x *H*-12), 1.74-1.36 (16H, br stack, 8 x C*H*$_2$); δ_C (75 MHz, d6-DMSO) 173.4 (*C*-11), 162.9 (*C*-19), 155.8 (*C*-2), 77.6 (*C*-1), 60.8 (*C*-17), 58.7 (*C*-21), 55.0 (*C*-16), 38.9 (CH_2), 38.8 (CH_2), 35.7 (CH_2), 30.1 (CH_2), 29.5 (CH_2), 28.4 (CH_2), 26.7 (CH_2), 25.9 (CH_2), 25.4 (CH_2); m/z (TOF ES+) 343.2168 ([M+H]$^+$, 100%, $C_{16}H_{31}N_4O_2S$ requires 343.2174).

Preparation of benzyl formamide, **97**
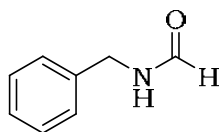
C$_8$H$_9$NO
Exact Mass: 135.07
Mol. Wt.: 135.16

Ethyl formate (17.77 mL, 0.22 mol) was added dropwise to benzylamine (27.30 mL, 0.25 mol) and stirred gently until the exothermic reaction ceased. The solution was then heated under reflux for 2 hours and the crude product distilled through a 25 cm Vigreux column. The pure product was a colourless viscous oil (27.9 g, 94 %); ν $_{max}$ (CHCl$_3$) 1688 (C=O), 3270, 3298 (NH); δ$_H$ (300MHz, CDCl$_3$) 4.42 (2H, d, J 5.88, CH_2), 6.32 (1H, s, NH), 7.24-7.32 (5H, stack, Arom CH), 8.18 (1H, s, COH); δ$_C$ (75 MHz, CDCl$_3$) 40.7 (CH_2), 126.8, 127.2, 128.3 (3 x Arom CH), 138.9 (Arom C_q), 160.9 (C=O); m/z (EI) 135 ([M]$^+$, 100 %), 106 ([M-COH]$^+$, 28 %), 91 ([M-NHCOH]$^+$).

Preparation of benzyl isocyanide, **98**
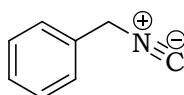
C$_8$H$_7$N
Exact Mass: 117.06
Mol. Wt.: 117.15

Benzyl formamide **97** (20.0 g, 0.15 mol) was mixed with pyridine (75.20 mL, 0.92 mol) in petroleum ether (100 mL) and stirred at 0°C in a three-necked round bottomed flask. POCl$_3$ (8.28 mL, 0.09 mol) was then added dropwise over 40 minutes and the resulting mixture stirred under reflux for 10 minutes. The cooled mixture was cooled and quenched with ice water until all solid material had dissolved. The solution was then extracted with pet. ether (3 x 60 mL) and the combined organic layers extracted with water (3 x 100 mL) before being dried over MgSO$_4$ and evaporated under reduced pressure. The crude product was then distilled under vacuum using a rotating glass oven (165 °C, 45 mbar) to yield the product as a pale brown pungent smelling oil (4.5 mL, 41 %); ν $_{max}$ (CHCl$_3$) 2132 (N≡C); δ$_H$ (300 MHz, CDCl$_3$) 4.64 (2H, s, CH_2), 7.34-7.44 (5H, stack, Arom CH); δ$_C$ (75 MHz, CDCl$_3$) 44.5 (CH_2), 127.3, 127.4, 128.0 (3 x Arom CH), 131.3 (Arom C_q), 148.9 (N≡C); m/z (EI) 117 ([M]$^+$, 100 %), 90 ([M-NΞC]$^+$, 72 %).

The following compounds were designed as inert smelling alternatives to benzyl isocyanide. The additional aromaticity was intended to provide added UV visibility and the potential for energy transfer in an alternative position to the aldehyde component in the Ugi 4-CCRs. This would then open up the possibility of using the Passerini reaction as the aldehyde could now be the linker position and the amine would have no function.

However, attempts at using these tosmic derivatives in multicomponent reactions failed.

Preparation of benzyl tosylmethylisocyanide, **99**
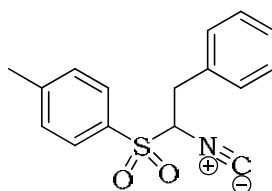
$C_{16}H_{15}NO_2S$
Exact Mass: 285.08
Mol. Wt.: 285.36

Tosylmethylisocyanide (TosMIC) (0.50 g, 2.56 mmol), tetrabutylammoniumiodide (TBAI) (0.19 g, 0.50 mmol) in CH_2Cl_2 (5 mL), and NaOH (1.50 g, 30 % in 5 mL H_2O) were combined, stirred vigorously for 5 minutes, and cooled to 0°C. Benzylbromide (0.31 mL, 2.56 mmol) was then added dropwise and the reaction stirred for 90 minutes at ambient temperature.

H_2O (25 mL) was added and the reaction mixture extracted with CH_2Cl_2 (3 x 15 mL). The combined organic layers were dried over $MgSO_4$ and the solvent removed under reduced pressure. The crude product was purified using flash column chromatography (Silica, Hexane / EtOAc 2 : 8, R_f = 0.48), to afford the pure product as a fine white crystalline solid (0.50 g, 98 %). Mp = 80-82 °C; ν_{max} (CHCl$_3$) 2138 (N≡C); δ_H (300 MHz, CDCl$_3$) 2.46 (3H, s, CH_3), 2.99 (1H, X of ABX, dd, J 12.0, 2.2, CH), 3.59 (1H, A of ABX, dd, J 12.0, 2.9, CH_2), 4.71 (1H, B of ABX, dd, J 8.5, 2.9, CH_2), 7.27-7.34 (5H, stack, 5 x benzyl CH), 7.42 (2H, d, J 8.1, tosyl CH), 7.91 (2H, d, J 8.1, tosyl CH); δ_C (75 MHz CDCl$_3$) 21.7 (CH_3), 34.5 (CH_2), 73.9 (CH), 129.9, 130.9, 131.2, 131.9, 132.1 (5 x Arom CH), 132.9, 135.1, 148.7 (3 x Arom C$_q$), 167.5 (N≡C); m/z (ES+) 308 ([M+Na]$^+$, 100 %), 326 ([M+K]$^+$, 15 %), HRMS (ES$^+$) Found: 308.0721, required for $C_{16}H_{15}NO_2NaS$: 308.0706.

Preparation of Di-benzyl tosylmethylisocyanide, **100**
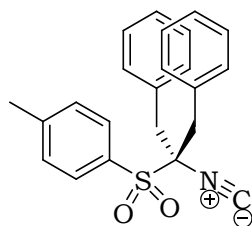
$C_{23}H_{21}NO_2S$
Exact Mass: 375.13
Mol. Wt.: 375.48

Tosylmethylisocyanide (TosMIC) (0.50 g, 2.56 mmol), tetrabutylammoniumiodide (TBAI) (0.19 g, 0.50 mmol) in CH_2Cl_2 (5 mL), and NaOH (1.50 g, 30 % in 5 mL H_2O) were combined, stirred vigorously for 5 minutes, and cooled to 0 °C. Benzylbromide (0.61 mL, 5.20 mmol) was then added dropwise and the reaction stirred for 24 hours at ambient temperature.

H_2O (25 mL) was then added and the reaction mixture extracted with CH_2Cl_2 (3 x 15 mL). The combined organic layers were dried over $MgSO_4$ and the solvent removed under reduced pressure. The crude product was purified using flash column chromatography (Hexane / EtOAc 1 : 8, R_f = 0.20) to afford the pure product as a fine white crystalline solid (0.80 g, 85 %). Mp = 88-91 °C; ν_{max} (CHCl$_3$) 2136 (N≡C); δ_H (300 MHz, CDCl$_3$) 2.47 (3H, s, CH_3), 3.29 (4H, d of ABq, J 14.2, 27.0, 2 x CH_2), 7.14-7.29 (10H, stack, benzyl arom CH), 7.36 (2H, d, J 8.1, tosyl CH), 7.85 (2H, d, J 8.1, tosyl CH); δ_C (300 MHz CDCl$_3$) 21.8 (CH_3), 39.7 (2 x CH_2), 81.8 (C$_q$), 127.9, 128.4, 129.8, 130.9, 131.2 (5 x Arom CH), 130.5, 132.5, 146.3 (3 x Arom C$_q$), 166.8 (N≡C); m/z (ES) 398 ([M+Na]$^+$, 100%); HRMS (ES$^+$) Found: 398.1191, required for $C_{23}H_{21}NO_2NaS$: 398.1195.

Preparation of 2-naphthyl tosylmethylisocyanide, **101**

$C_{20}H_{17}NO_2S$
Exact Mass: 335.1
Mol. Wt.: 335.42

Tosylmethylisocyanide (TosMIC) (0.50 g, 2.56 mmol), Tetrabutylammoniumiodide (TBAI) (0.185 g, 0.50 mmol) in CH_2Cl_2 (5 mL), and NaOH (1.50 g, 30 % in 5 mL H_2O) were

combined, and the mixture stirred vigorously. 2-Naphthylmethylbromide (0.57 g, 2.56 mmol) was then added dropwise and the reaction stirred for 90 minutes at ambient temperature.

H_2O (25 mL) was then added and the reaction mixture extracted with CH_2Cl_2 (3 x 15 mL). The combined organic layers were dried over $MgSO_4$ and the solvent removed under reduced pressure. The crude product was purified using flash column chromatography (Hexane / EtOAc 2 : 7, R_f = 0.36) to give the pure product as a fine white crystalline solid (0.53 g, 63 %). Mp = 102-104 °C; ν_{max} (CHCl$_3$) 2125 (N≡C); δ_H (300 MHz, CDCl$_3$) 2.43 (3H, s, C*H*$_3$), 3.16 (1H, X of ABX, dd, *J* 11.0, 2.2, C*H*), 3.76 (1H, A of ABX, dd, *J* 11.0, 2.9, C*H*$_2$), 4.82 (1H, B of ABX, dd, *J* 8.5, 2.9, C*H*$_2$), 7.32-7.95 (11H, stack, Arom C*H*); δ_C (75 MHz CDCl$_3$) 21.0 (*C*H$_3$), 34.1 (*C*H$_2$), 73.2 (*C*H), 125.6, 125.8, 125.9, 127.0, 127.8, 128.1, 129.4, 129.5 (8 x Arom *C*H), 129.9, 130.3, 132.0, 132.7, 146.0 (5 x Arom *C*$_q$), 165.0 (N≡*C*); *m/z* (ES) 358 ([M+Na]$^+$, 100 %), HRMS (ES$^+$) Found: 358.0878, required for $C_{20}H_{17}NO_2NaS$: 358.0872.

Preparation of Di-2-Naphtylmethyl tosylmethylisocyanide, **102**
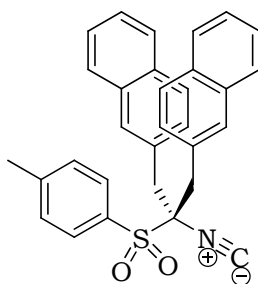
$C_{31}H_{25}NO_2S$
Exact Mass: 475.16
Mol. Wt.: 475.6

Tosylmethylisocyanide (TosMIC) (0.24 g, 1.25 mmol), tetrabutylammoniumiodide (TBAI) (0.092 g, 0.25 mmol) in CH_2Cl_2 (5 mL), and NaOH (1.50 g, 30 % in 5 mL H_2O) were combined, stirred vigorously, and cooled to 0 °C. 2-(bromomethyl) napthalene (1.13 g, 5.20 mmol) was then added dropwise and the reaction stirred for 24 hours at ambient temperature.

H_2O (25 mL) was then added and the reaction mixture extracted with CH_2Cl_2 (3 x 15 mL). The combined organic layers were dried over $MgSO_4$ and the solvent removed under reduced pressure. The crude product was then recrystallised from CH_2Cl_2 / MeOH (1 : 1), to give the pure product as a fine white crystalline solid (0.86 g, 72 %). Mp = 151-153 °C; ν_{max} (CHCl$_3$) 2122 (N≡C); δ_H (300 MHz, CDCl$_3$) 2.36 (C*H*$_3$), 3.42 (2H, d of ABq, A of AB, *J* 14.1, C*H*$_2$), 3.56 (2H, d of ABq, B of AB, *J* 14.1, C*H*$_2$), 7.19-7.79 (18H, stack, Arom C*H*); δ_C

(75 MHz, CDCl$_3$) 21.8 (CH$_3$), 40.1 (2 x CH$_2$), 81.9 (C), 127.7, 127.9, 128.1, 128.5, 130.3, 131.3 (8 x Arom CH), 129.7, 130.7, 132.9, 133.2, 146.3 (4 x Arom C$_q$), 167.3 (N≡C); m/z (ES) 498 ([M+Na]$^+$, 100 %), 514 ([M+K]$^+$, 8 %); HRMS (ES$^+$) Found: 498.1504, required for C$_{31}$H$_{25}$NO$_2$NaS: 498.1501.

Preparation of 9-anthracenyltosylmethylisocyanide, **103**
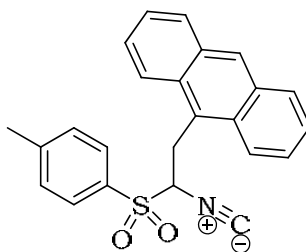
C$_{24}$H$_{19}$NO$_2$S
Exact Mass: 385.11
Mol. Wt.: 385.48

Tosylmethylisocyanide (TosMIC) (0.50 g, 2.56 mmol), Tetrabutylammoniumiodide (TBAI) (0.185 g, 0.50 mmol) in CH$_2$Cl$_2$ (5 mL), and NaOH (1.50 g, 30% in 5 mL H$_2$O) were combined, and the mixture stirred vigorously. 9- bromomethyl-anthracene (0.69 g, 2.56 mmol) was then added slowly over 10 minutes, and the reaction stirred for 90 minutes at ambient temperature.
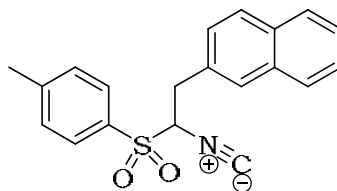
H$_2$O (25 mL) was then added and the reaction mixture extracted with CH$_2$Cl$_2$ (3 x 15 mL). The combined organic layers were dried over MgSO$_4$ and the solvent removed under reduced pressure. The crude product was purified using flash column chromatography (Hexane / EtOAc 2 : 7, R$_f$ = 0.36) to give the pure product as a fine yellow crystalline solid (0.53 g, 63 %); Mp = 163-165 °C; ν$_{max}$ (CHCl$_3$) 2130 (N≡C); δ$_H$ (300 MHz, CDCl$_3$) 2.49 (CH$_3$), 4.35 (1H, X of ABX, dd, J 12.0, 3.5, CH), 4.54 (1H, A of ABX, dd, J 12.0, 2.7, CH$_2$), 4.95, B of ABX, dd, J 8.4, 2.7, CH$_2$), 7.44-8.44 (13H, stack, Arom CH); δ$_C$ (75 MHz CDCl$_3$) 21.9 (CH$_3$), 26.3 (CH$_2$), 72.8 (CH), 123.3, 125.4, 127.1, 128.6, 129.7, 130.4 (6 x Arom CH), 124.5, 130.6, 131.2, 131.6, 147.0 (5 x Arom C$_q$), 167.0 (N≡C); m/z (ES) 408 ([M-CH$_3$]$^+$, 100 %), 440 ([M+Na]$^+$, 10 %).

Chapter 3: Galectin targeted fluorescent disaccharides for *in vivo* imaging

Cyclic peptide becomes disaccharide – change of target and delivery vehicle

3.1 Introduction to carbohydrates

Carbohydrates, sugars and saccharides, are all generic terms for the same type of chemical compound. Mono- and disaccharides are commonly referred to as sugars whereas longer more complex chains of these sugars are termed either poly- or oligosaccharides. All these come under the more general term of carbohydrate, which is the family of compounds possessing the empirical formula $C_n(H_2O)_n$.[119]

Glucose is the most well known sugar and indeed the most abundant organic molecule on the face of the planet. It is the monosaccharide unit most well known and researched for being a source of energy. It is also, however, the primary building block for some of Nature's most structurally important compounds such as cellulose and starch (**Figure 89**).
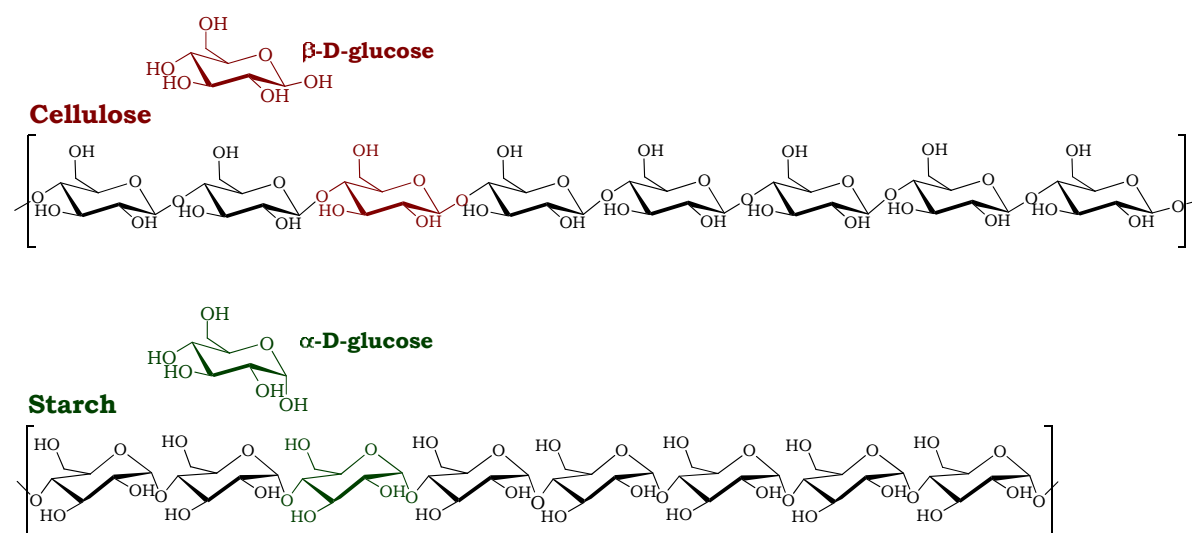
Figure 89: D-glucose in cellulose and starch showing the contrasting structures resulting from the two anomeric glycosidic bonds.

Being from several hundred to over ten thousand D-glucose repeat units in length, the strands are held together as layered fibres through van der Waals forces and hydrogen bonding. This creates a substance of great versatility and strength. Cotton is 90% cellulose and timber around 50%, this explains the status of cellulose as the most abundant substance on earth.[119b]

Cellulose and starch differ in their glycosidic bond geometry. Cellulose consisting of repeat β(1→4) linkages and starch α(1→4). The starch alpha bond is easily hydrolysed unlike the cellulose beta bond which explains the dramatically different properties of the two compounds.

Starch and cellulose are synthesised in plants through the process of photosynthesis where CO_2 is converted into glucose. Enzymes such as transferases and synthases then stimulate the build up of the polysaccharides which are subsequently broken down as required, to release energy through the citric acid cycle.

The best known properties and applications of sugars are in the food that we eat and the nourishment they provide. Sucrose, fructose and lactose are some well known sugars commonly found in the diet, and just like complex carbohydrates such as starch, are broken down through hydrolysis, by hydrolases, prior to metabolism. However, being in the simple sugar form they are a much more rapid source of energy.

3.1.1 Metabolism

The most important and rapidly available source of energy in biological systems is adenosine triphosphate (ATP).
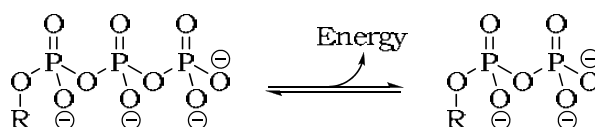
Scheme 27: Adenosine triphosphate (ATP) hydrolysed to adenosine diphosphate (ADP) releasing energy.

ATP is hydrolysed to ADP releasing 30.6 kJmol^{-1} of energy.[119] This energy is stored in the phosphodiester bond and is vital for the function of all processes in the aqueous physiological environment. However ATP needs to be regenerated and it is the oxidation of glucose that is the major contributor to this regeneration.

The citric acid cycle is a series of enzyme catalysed chemical transformations of great significance in all aerobic living cells. It is part of the metabolic process which converts carbohydrates, proteins and fats into water, carbon dioxide and usable energy.

The first metabolic process prior to the citric acid cycle is glycolysis where glucose it is split into two molecules of pyruvate through a ten step enzyme-catalysed reaction sequence. This involves several key steps including a rearrangement to form a five membered furanose and a retroaldol cleavage (**Scheme 28**).
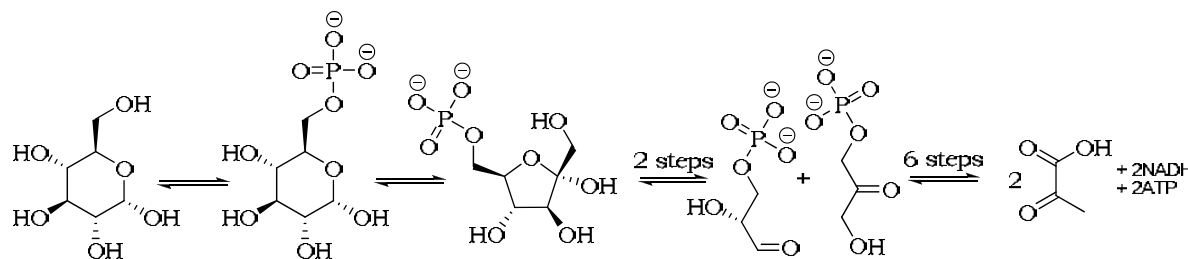
Scheme 28: An abbreviated mechanism of glycolysis leading to two molecules of pyruvic acid.

Before entering the citric acid cycle the pyruvate must be converted into an activated acetic acid thioester known as acetyl-SCoA. This basically imparts a good leaving group ready for initiation into the famous cycle (**Scheme 29**).[120]
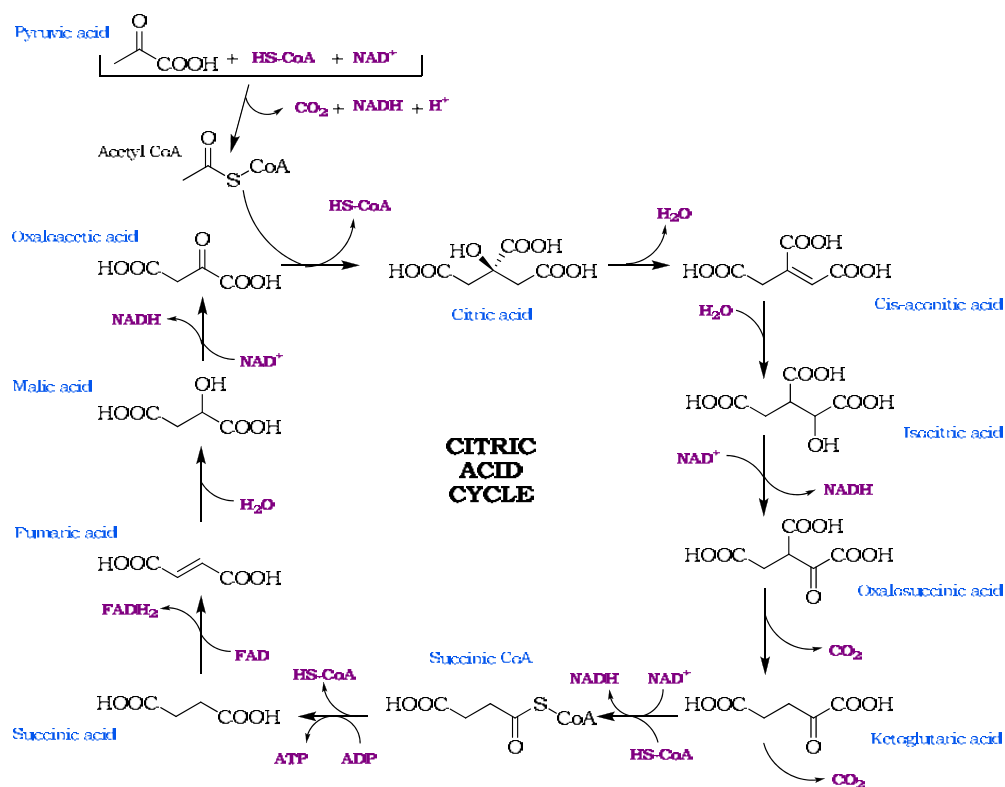
Scheme 29: The author's diagram of the Citric Acid Cycle.

The key to the citric acid cycle is the conversion of acetyl-SCoA into the reducing compounds NADH (the reduced form of nicotinamide adenine dinucleotide) and $FADH_2$ (the reduced form of flavin adenine dinucleotide). These are then oxidatively phosphorylated and the ten NADH and two $FADH_2$ nucleotides from one glucose molecule are converted into 32 molecules of ATP.

3.1.2 Carbohydrates and cancer

Thus far we have delved into a few the well known biological properties and processes involving carbohydrates. It is, however, some of the less well known properties of carbohydrates that are of more relevance and interest to this study. Over the past 25 years science has begun to reveal the involvement of carbohydrates, especially small sugar residues, in a vast array of other fundamental and precise biological processes both inside and outside of cells. It is becoming clearer and clearer that a variety of carbohydrates are key code molecules in biological communication events influencing the control of proceedings as diverse as conception, inflammation, infection and the growth and spread of cancer. [120b]

It is the latter function of carbohydrates within the field of oncology that will become the cornerstone of this discussion.

Most proteins synthesised in mammalian cells have carbohydrates attached to them. Carbohydrates are covalently bonded to these glycoproteins through either an 'N' linkage,

via asparagine, or through an 'O' linkage, via the amino acids serine or threonine. In fact carbohydrates effectively 'coat' the majority of cellular surface proteins, so that all interactions between a cell and its local environment are in some way affected by carbohydrates. These appended sugars can have a variety of influences upon the properties of the glycoproteins, including increasing or decreasing solubility and promoting resilience to attack and break down by peptidases and proteases.[120b]

A nice example of the carbohydrate residues on glycoproteins imparting extraordinary function to the molecule is in a class of biological antifreeze glycoproteins (AFGPs) isolated from the Arctic and Antarctic deep sea Teleost fish (**Figure 90**). These compounds allow the host to survive in water temperatures as low as -2 ˚C through the inhibition of ice crystal growth.[120c]
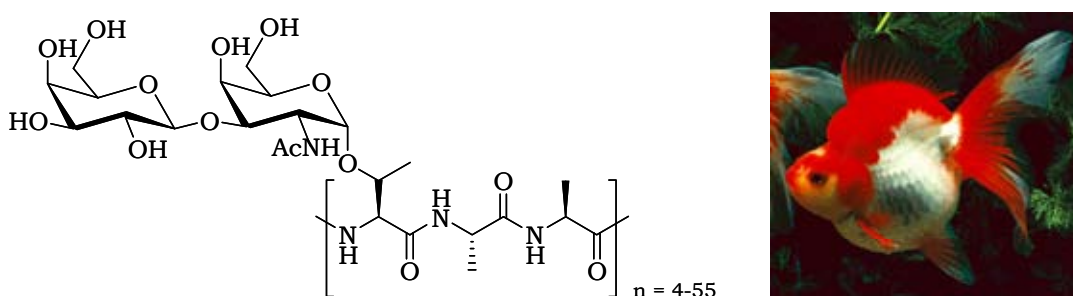
Figure 90: An example of a glycoprotein, an AFGP (antifreeze glycoprotein).[120c]

A vast array of branched sugars are also found appended to glycoproteins. These manifest the diverse range of variations not only in the sugar build-up of the oligosaccharides, but also in the position of glycosylation on the protein and the abundance of these sites. An example of an oligosaccharide on an 'N' linked glycoprotein in shown in **Figure 91**. Although this is not a true representation of conformation, it shows the branched 'claw-like' nature of the carbohydrate arm which is a perfect site for sugar specific proteins such as lectins to dock, bind and begin influencing signalling and cellular behaviour.[121]
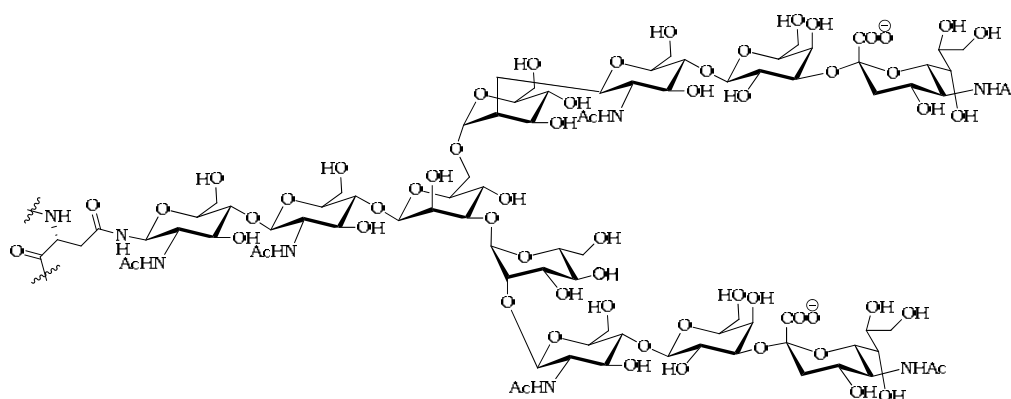
Figure 91: A branched glycoprotein oligosaccharide.[121]

Lectins bind to glycosylated structures with exquisite specificity even for complex, highly branched oligosaccharides, despite their typically very shallow binding sites, also known as carbohydrate recognition domains (CRDs).

3.1.3 Lectins

Mammalian lectins are a class of intra- and extracellular proteins that have a binding affinity for specific carbohydrates. They bind to a soluble carbohydrate residue or to a carbohydrate moiety on a glycoprotein or lipid. An undesirable function of these proteins is that viruses such as human influenza strains attach themselves to the cells of a host organism by binding to lectins during infection.[122a] Results have suggested that a lectin cross-linkage between the virus and the cellular surface is the most plausible mechanism for propagation of infection.[122b]

There are four types of lectins: C-type, P-type, pentraxins and galectins, each defined by the class of carbohydrates to which they bind, their structure, and their mode of action.[122c/d]

3.1.3.1 Galectins

To date there are over 15 known galectins and it is upon this group of lectins that our attention shall focus. Galectins are specifically β-galactoside binding lectins sharing significant amino acid sequence similarities in the CRD (**Figure 92**). They are divided into numerical categories with galectin-1 to galectin-4 being the best characterised and understood four galectins.

More and more of the mysteries of the galectin family are being unravelled. This is a family of proteins that grows larger as more are discovered, and yet their activity and far-reaching influence is still far from completely understood. Recent literature describes the galectin family as having a "diverse and bewildering, yet intriguing, array of activities both inside and outside of cells".[123] However, galectins are not only involved in extracellular biology through their binding to surface glycoproteins and the extracellular matrix (ECM), they have also been detected in subcellular locations such as the nucleus. Evidence suggests galectins play roles in a whole host of biological events including oncogenesis and the adhesion and proliferation of cells.

β-Galactosides are carbohydrate residues commonly located appended to cellular surface glycoproteins or within larger appended oligosaccharides which are involved trans-membrane cellular signalling processes.
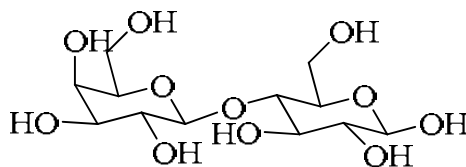
Figure 92: Lactose (β-D-galactopyranosyl-(1→4) β-D-glucopyranose), a **β-galactoside**, is a disaccharide common in dairy products. A molecule of D-galactose is bonded to a molecule of D-glucose through a β-(1'→4) glycosidic linkage.

For galectin recognition and binding the β-galactoside must possess the monosaccharide β-galactose moiety. Galectin binding specificity towards the β-galactose alone is low and does not effect good binding. Normally this monosaccharide unit is found incorporated into an oligosaccharide where the adjacent one or two monosaccharide building blocks also influence binding affinity and specificity. This is what differentiates between the different members of the galectin family as they each target and bind specific cellular surface protein oligosaccharide residues. This is what also makes them open to novel therapeutic targeting and the diagnostic imaging applications addressed in this study.

Lactose (**Figure 92**), is the best known example of a β-galactoside and indeed it binds to galectins, albeit with a low binding constant. This will be further investigated in the forthcoming discussion regarding the carbohydrate recognition domain (CRD) (page 174).

The galectins known thus far have either one or two characteristic CRDs within a single polypeptide chain. Galectins-1 and -2 are homodimers composed of two approximately 135 amino-acid carbohydrate-binding sub-units which fold up as a globular domain.[124,125] Galectin-3 possesses a unique structure in that it contains a single CRD followed by a flexible proline- and glycine- rich repeat unit of approximately 100 amino-acid residues, before terminating with an N-terminal domain of approximately 30 amino-acids. Galectin-4 is another dimer with a longer linking peptide between the two CRDs of approximately 30 amino-acid residues. **Figure 93** shows a simplified representation of these first four galectins and illustrates how they are structurally quite different despite possessing the required affinity for β-galactosides and CRD similarities.

Galectin-1 dimer

Galectin-3

Galectins 1,2,5,7,10,11,13,14 & 15

Galectin 3

Galectins 4,6,8,9 & 12
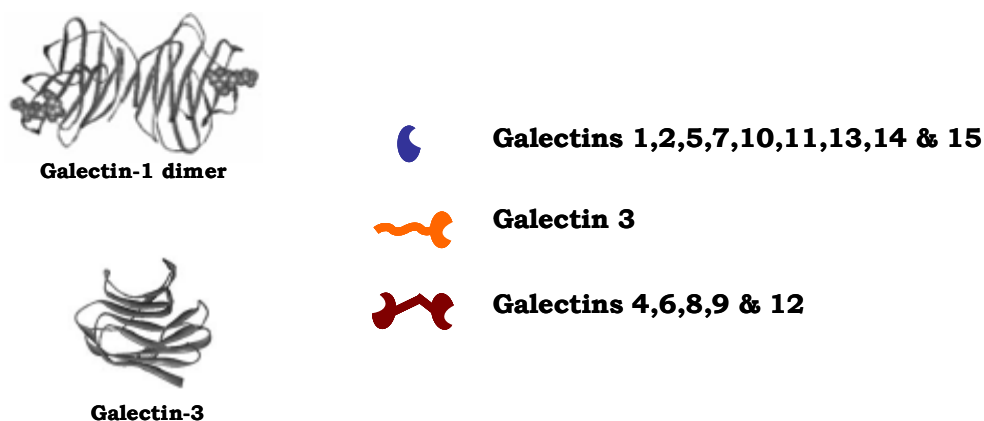
Figure 93: A cartoon illustrating the basic structure of galectins. For the tandem repeat unit (Gal-4 etc), it is important to recognise that the two CRDs are as different from each other as they are from their fellow galectins.[125]

3.1.4 Galectin-3

The emphasis of our discourse will now focus on galectin-3 for this is the target of the disaccharide probes designed and synthesised in this project.

Apart from the prediscussed structural differences and physiological localisation, it must be rationalised as to why this specific galectin is a good target for site-specific imaging.

Galectins are expressed in distinct but overlapping localities within mammalian tissues. For example, galectin-1 is abundant in skeletal, smooth and cardiac muscle, neurons, thymus, kidneys and placenta. In contrast, galectin-3 is most abundant in active macrophages, basophils, mast cells and some epithelial cells and sensory neurons. In many tissue types galectin-3 synthesis and propagation is only instigated under certain physiological conditions. It is the roles of Galectin-3, and its abundance in and around tumour cells, where it mediates many processes including intracellular signalling and intercellular adhesion, that makes it stand out as a potential vector for the site-specific delivery of diagnostic imaging agents.

Extracellularly, galectins can bind either di-valently or multivalently to the β-galactoside residues. The creation of these multivalent aggregates suggests that the mode of action of these proteins is to cross-link carbohydrate-containing chains on cellular surfaces and/or within the ECM. As in many other receptor-ligand systems, this can trigger a cascade of signalling across the cell membrane. This mechanism allows the modulation of processes including mitosis, apoptosis and cell-cycle progression.[126] The di- and multivalent binding properties of galectins also allows the aggregation of cells through intercellular bridging (**Figures 94 & 95**) as previously mentioned in the mechanism of some viruses (page 168).
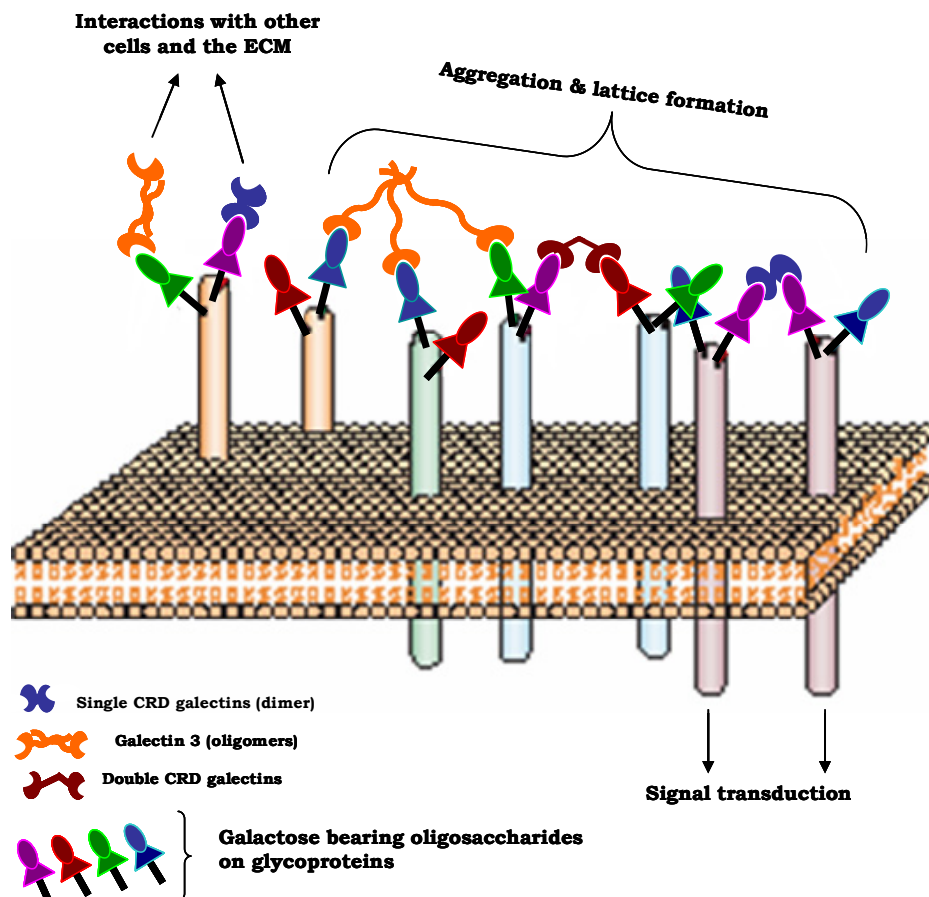
Figure 94: The author's diagram of the various processes involving galectins at the cell surface.

Galectin-3 forms oligomers when binding to multivalent carbohydrates and thereby is able to form an ordered array of agglomerated glycoconjugates as illustrated in **Figure 95**.
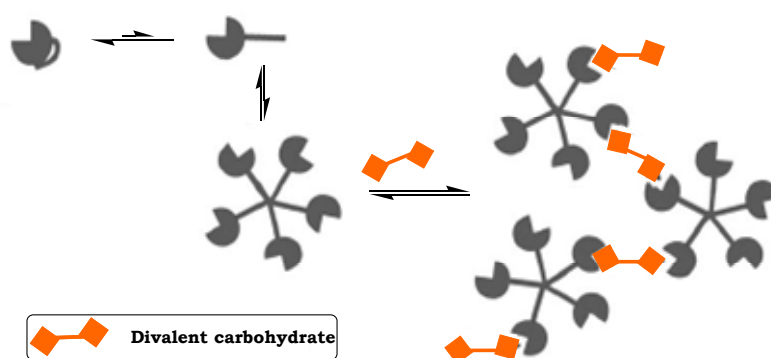
Figure 95: Galectin-3 oligomers forming a lattice. Note the free Gal-3 binding sites for targeting.

Some galectins are distributed in a wide variety of tissues whereas others are more specific. Distribution is modulated during the differentiation of individual cells and during the development and growth of organisms and tissues. As mentioned previously, this distribution varies according to physiological and pathological environmental conditions. Of

most relevance is the discovery that the expression of galectins is altered in and around tumour cells in comparison to their healthy counterparts. Galectins are often over-expressed in cancer cells especially in those cell types that normally show negligible galectin presence.[126-129] This is not ubiquitous, as some normal cells which naturally show good levels of galectins show a decrease in galectin expression upon the onset of malignancy. However, in many situations the increased expression of galectins correlates very clearly with an increase in the aggressiveness of tumours and the development of a metastatic phenotype.[130] This indicates that galectins may well coordinate cancer propagation. Studies have shown a weakly metastatic sarcoma cell line becoming highly metastatic through an over expression of galectin-3.[127] Further to this an intracellular migration of galectin-3 from the nucleus into the cytosol occurs during the transition from normal to healthy cells.[121]

There is unambiguous evidence that galectin-1 and galectin-3 expression is necessary for the initiation of the transformation of a benign or young tumour into an aggressively metastatic phenotype.[130,131] Inhibition studies of galectin-3 in breast and thyroid carcinoma cells revealed a reduction in the characteristic transformed phenotype, indicating a definite link. In an alternative study, introduction of galectin-3 into a healthy thyroid cell line induced the phenotype transformation and cells became cancerous.[132]

The mechanisms of galectin-3 involvement in cell transformation and oncogenesisis are not yet fully understood, but binding studies have revealed galectin-3 binding to the Ras oncogene which was discussed in the biological testing section of chapter 2 as the gene used to initiate some transgenic mice lines (**Page 88**).[100-103]

T-cells and macrophages are part of the immune response to cells identified as immunologically foreign. Once activated, T-cells divide rapidly and secrete cytokines which regulate and assist the immune response. Cytotoxic T-cells destroy tumour cells, however, the expression of cell-surface glycoproteins bearing the β-galactoside moiety allows the binding of galectins, and Galectin-3 can form multivalent complexes with T-cell receptors, restricting signal transduction, activation and motility, thereby downregulating the immune response to the cancer, promoting tumour cell survival and proliferation. Galectin-3 can also induce T-cell apoptosis.[127]

We are building up a picture of the diverse and widespread roles of galectins, especially galectin-3, in tumour development and progression. However it is the upregulation of galectin-3 in tumour sites that is of most interest and relevance to this study. The known reasons behind this tumour-specific accumulation of galectin-3 will be briefly discussed in order to justify its position as the site specific entity to which the probes will be targeted.

In chapter 2 the processes of angiogenesis and metastasis were introduced with respect to the upregulation of $\alpha_v\beta_3$ integrin bearing cells such as endothelial cells. Here we will see some overlap.

3.1.4.1 Galectin-3 in adhesion, angiogenesis and metastasis.

The spread of a cancer from a single primary site to other metastatic sites is a complex multi-step process involving cell-to-cell and cell-to-ECM adhesion and aggregation. Although the galectin family as a whole seems to play important specific roles, it is the fact that galectin-3 has been found clustered at sites of homotypic tumour cell-cell contact and also in heterotypic adhesion between carcinoma and endothelial cell monolayers that implicates its involvement in tumour cell clustering and then the break away and establishment of a tumour at an alternative site (**Figure 96**).[133]
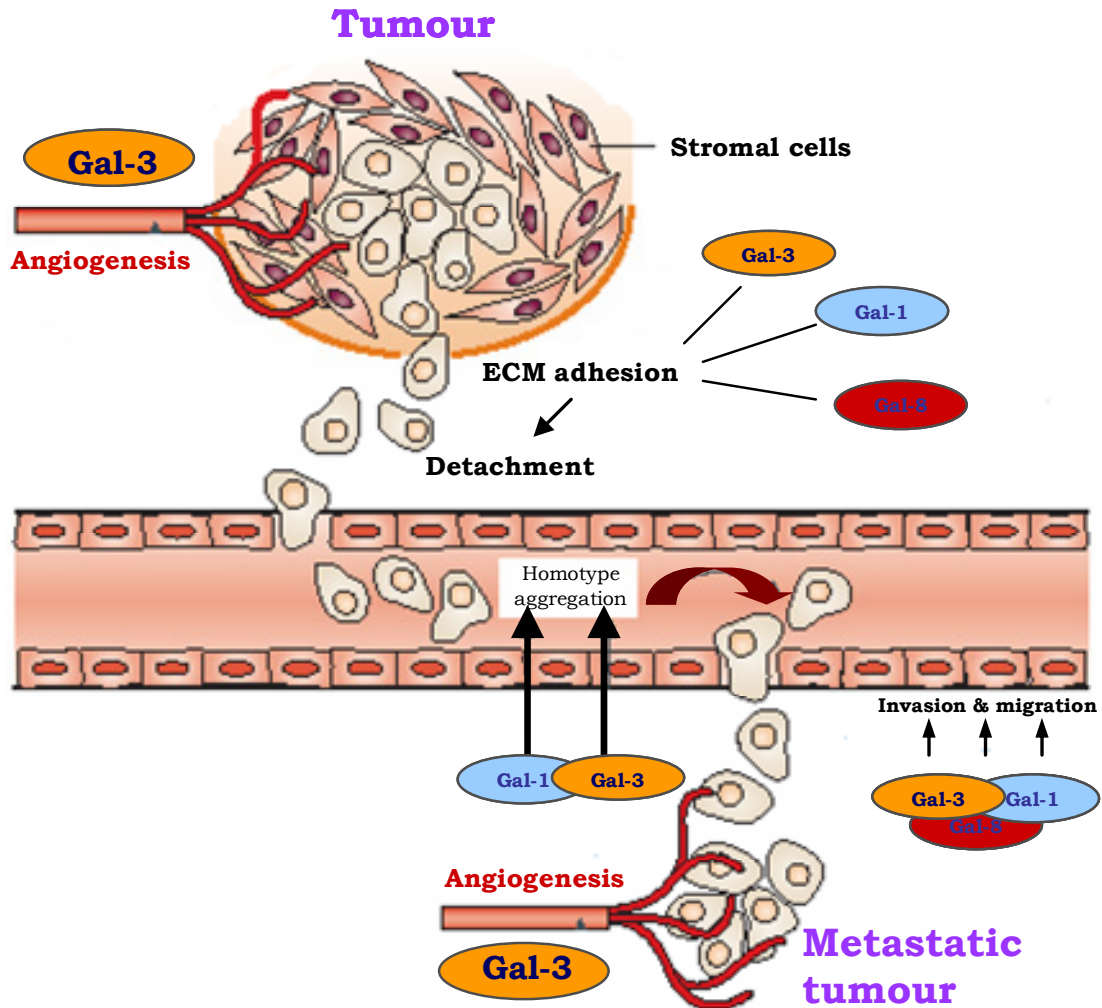
Figure 96: A simplified illustration of tumour metastasis highlighting some of the key roles of galectin-3. Adapted from Ref.127.

It is thought that galectin-3 encourages angiogenesis by promoting endothelial cell migration. Proof of this is reported where an increase in capillary density surrounding tumours, in comparison to controls, was observed in breast-tumour-bearing mice.[120]

This attempted explanation of the role of galectin-3 in metastasis is hugely simplified and it must be remembered that a vast number of simultaneous complex processes involving galectins, integrins, glycoproteins, endothelial cells, the ECM, the immune system and

many other biological entities such as growth factors, are each working together and against each other in complex mechanisms which are not yet fully understood. However, the important fact that galectin-3 is central and prolific in tumours is the prime reason that this protein will be our target in novel tumour imaging agents.

Having justified our target we must look more closely at the galectin-3 CRD and explore the binding-site with emphasis on the ensuing design of galectin-3-specific probes.

3.1.5 The carbohydrate recognition domain (CRD)

Whilst the diverse structure of the various members of the galectin family will influence their biological functions and physiological locations, it is first important to discuss the carbohydrate recognition domain in detail, as this will influence the design of β-galactoside probes for the imaging purposes of this study.

CRD binding is largely due to hydrogen bonding between the galectin backbone and side-chain carbonyl groups, and the hydroxyls of the sugar ligands. Hydrophobic interactions between hydrophobic galectin side-chains and hydrophobic areas on the glycoprotein carbohydrate residues also serve to increase binding affinity.
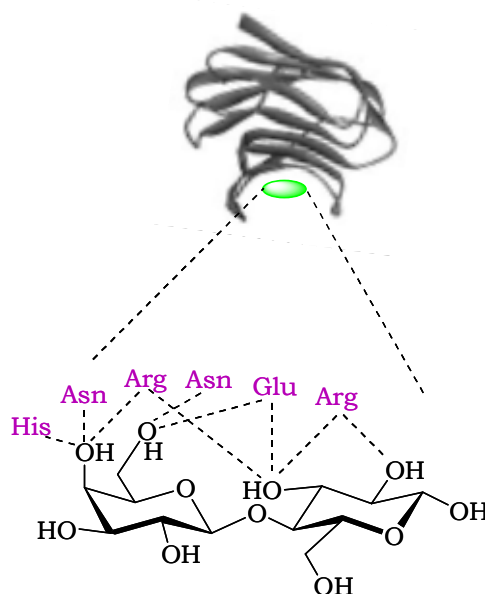
Figure 97: A schematic showing an example of the interactions binding lactose within the galectin-3 CRD.[127]

Known galectin carbohydrate ligands such as lactose and N-acetyllactosamine have been found to exhibit low inhibitory potency with IC_{50} = 0.2-1.0 mM. For effective galectin-3 inhibition concentrations over 10mM are required.[134]

The two specific monosaccharide building blocks of a β-galactoside have been found to profoundly influence the binding efficacy in the CRD. Indeed, albeit relatively shallow, galectins have a binding site able to accommodate species as large as tetrasaccharides.

The D-galactose monosaccharide is fundamental for all galectin binding but will still bind to the CRD without the anomeric-linked D-glucose. However, binding is enhanced by the

174

presence of D-glucose, suggesting that a combination of D-galactose with one, two or three extra monosaccharides would afford better alignment and orientation in the binding pocket.[135]

Crystal structure studies have shed light upon this binding pocket and **Figure 98** illustrates the D-galactose bound deeper within the CRD whereas the D-glucosamine is partially protruding.
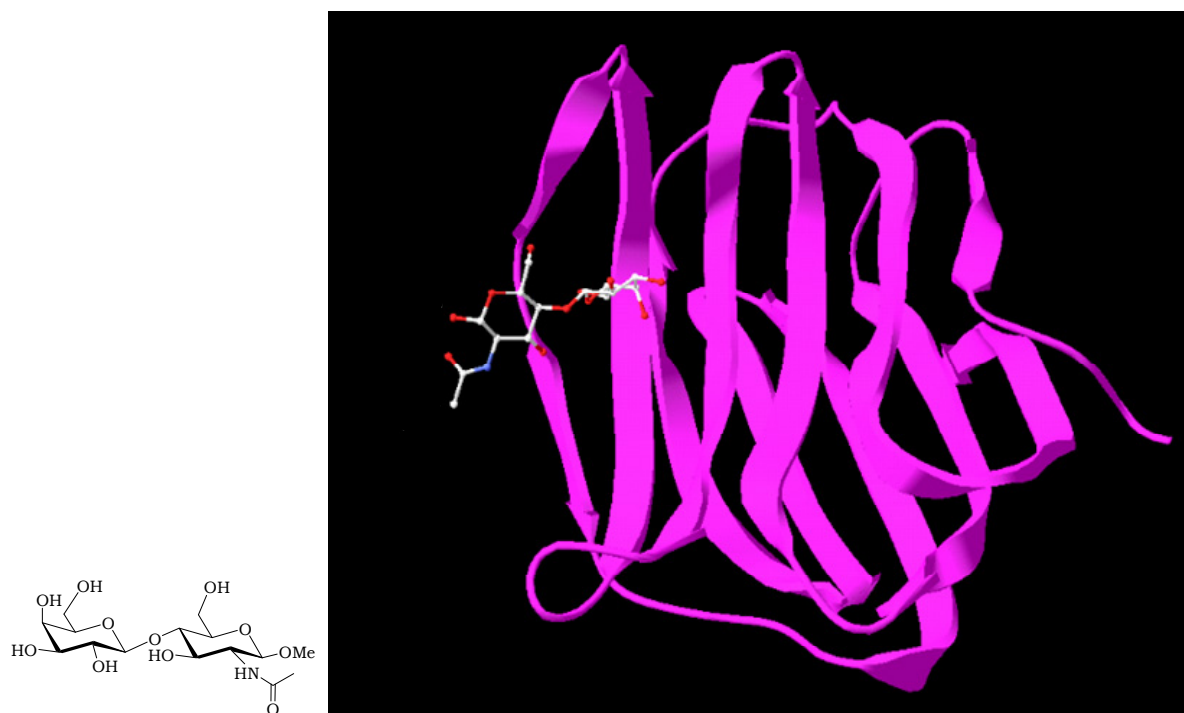
Figure 98: **Left**) N-acetyllactossamine, K_d = 69000 nM. **Right**) The protein crystal structure of gelectin-3 bound N-acetyllactosamine.[133] Pdb code 1L5G.

Galectin-3 crystal structure and modelling studies, combined with binding specificity measurements, have enhanced the understanding of this CRD to such a degree that careful scrutiny of the binding pocket has highlighted an area where there is an extended empty groove into the protein. Therefore an investigation was set up to establish the potential of this space for enhancing the binding of a β-galactoside through additional favourable interactions deeper within the CRD.[111] Diverse structural extensions of the D-galactose C-3' functional group nearest to the space in the binding site found vastly enhanced binding affinity and kinetics was possible. Derivatisation of this single position on the D-galactose moiety with aromatic groups improved the binding affinity by two orders of magnitude.[133-135]

Figure 99 shows images of the fine structure surrounding the CRD with N-acetyllactosamine in the binding site. The space for potential derivatisation is clear in the region of the C-3' hydroxyl.
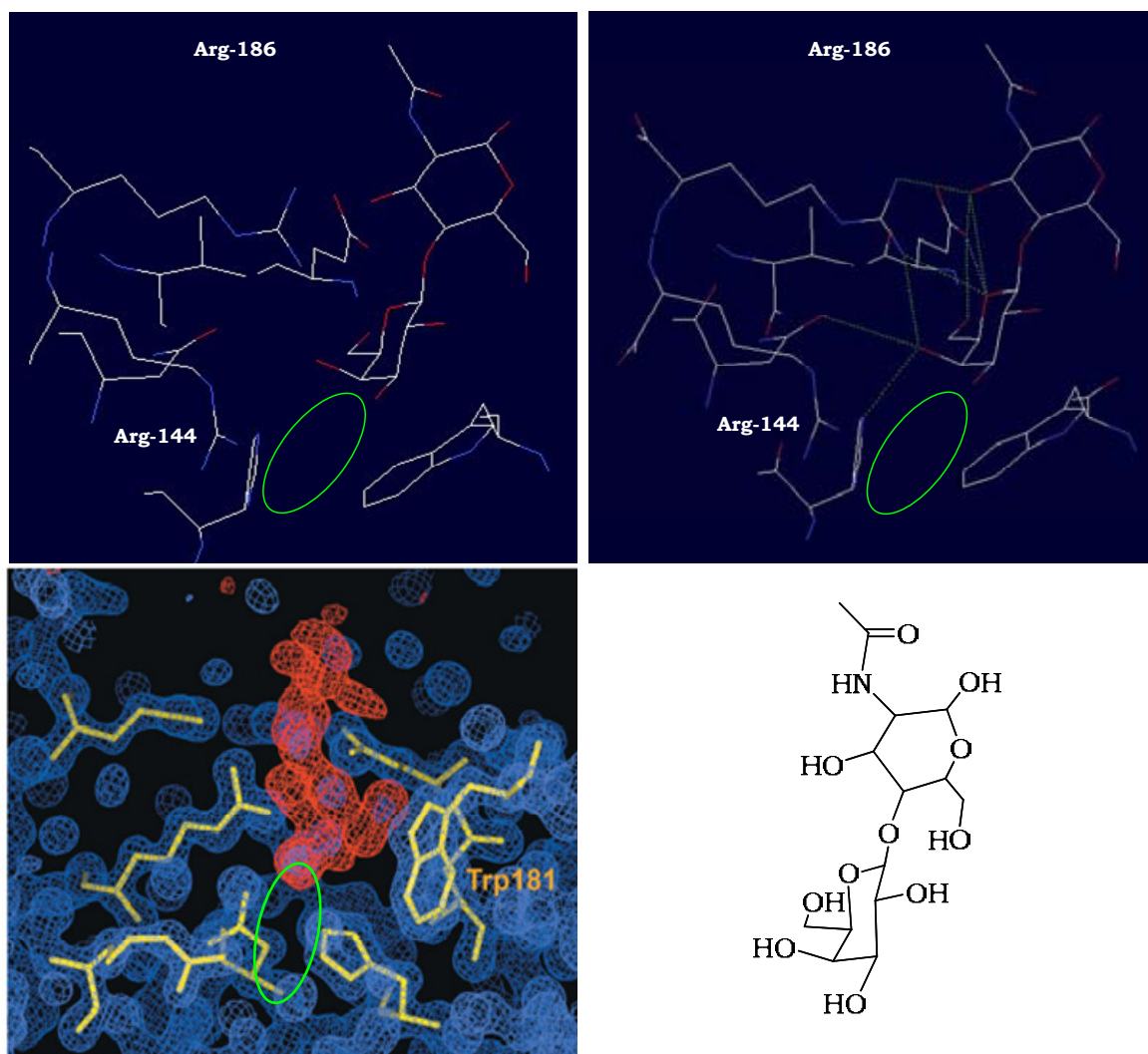
Figure 99: Computer models of N-acetyllactosamine within the galectin-3 CRD. The green rings show the space between residues Asn160, Arg144 and His158 for extension of the disaccharide. Adapted from refs. 111 & 114.

Nilsson *et al*[133] showed that the C-3' hydroxyl can be replaced with an azide which, upon reduction to the amine was successfully functionalised to form many different 3'-amino-N-acetyllatosamine derivatives. This required a complete overhaul of the synthetic protocol as will be discussed in due course and rendered the total synthesis longer through the necessity of starting from the unnatural monosaccharide L-gulose. It became abundantly clear through competitive binding studies of these amide derivatives against LacNAc that binding was enhanced by as much as threefold. Inhibitory activity was most significantly enhanced in aromatic amide derivatives (**Figure 100**).[133,134]
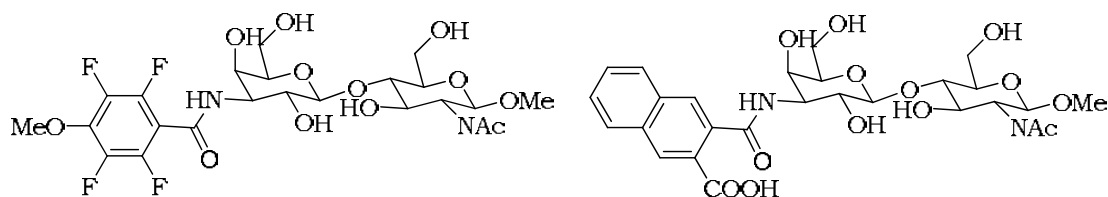
Figure 100: 3'-Amino-N-acetyllatosamine (LacNAc) derivatives found to exhibit enhanced binding to galectin-3.

This phenomenon was rationalised through electronic mapping and crystal structure studies revealing what is thought to be a key interaction between the Arg-144 residue and an aromatic ring. This enhances binding considerably, especially with 2-naphthamido derivatives, and the proposed mechanism involves the positively charged π-system of the guadinium ion (arginine side-chain) interacting through π-π stacking interactions with the naphthyl π system.[136,137]
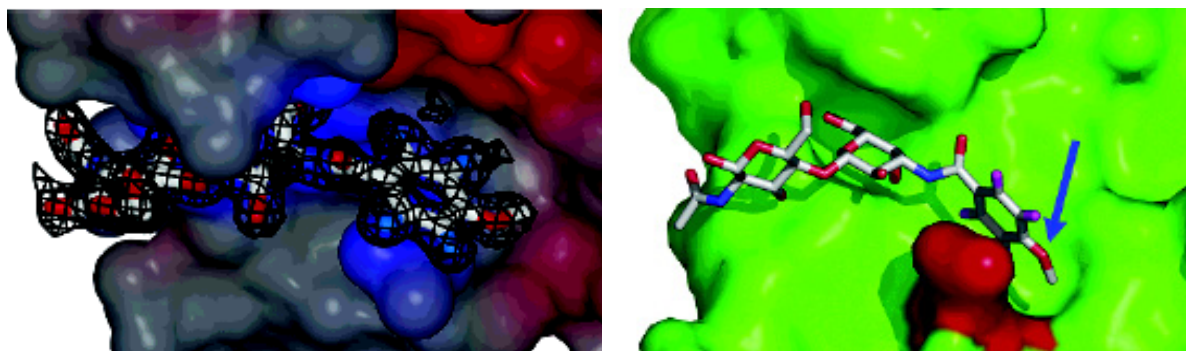
Figure 101: **Left**) Electron density map showing the blue protein surface as positive, and red as negative. **Right**) This clearer model shows the orientation of binding. Note the aromatic extension at C-3' (arrowed) bound deeper within the pocket and interacting with Arg-144.[133]

Although the binding efficiency is of great importance in this study and will be included in the synthetic strategy, it must be remembered that the emphasis of this study is not the *inhibition* of galectin-3 (which would be applicable to therapeutic type studies), but to develop galectin-3-targeted fluorescent probes to aid diagnostics.

To label the galectin-3 targeted disaccharide we must consider the availability of attachment sites that do not interfere with the binding pocket, and this means looking at the end of the galactoside which is protruding from the CRD.

Conveniently there are two possible sites available to append a fluorophore, and now our investigation must touch upon the literature precedent for derivatising the disaccharide in these positions.
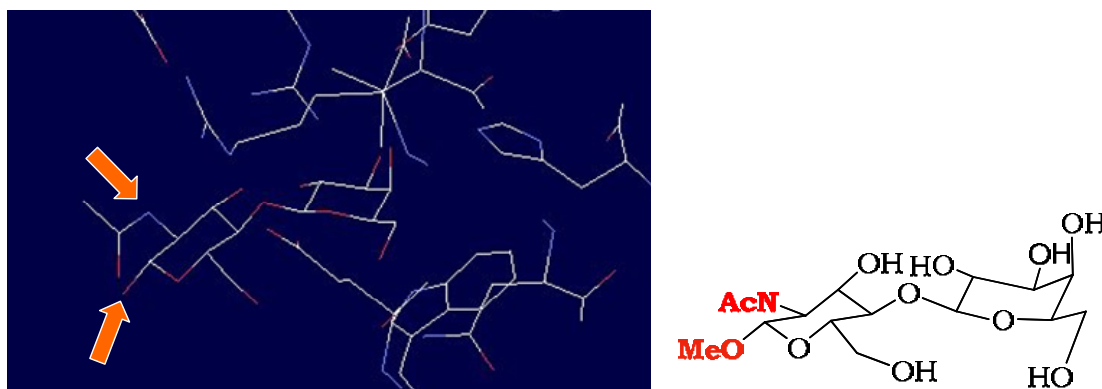
Figure 102: The two exposed positions for potential attachment of the fluorescent probe.

Literature precedent for the tagging or labelling of galactosides with organic fluorophores is quite scant. There has been work carried out to derivatise the glucosamine anomeric hydroxyl in many ways, especially in lipid chemistry where lipids have been attached to the anomeric position of saccharides as a hydrophobic tag predominantly for solubility and purification reasons.[136b-e]

However, surprisingly, there is little evidence of using the C-2 amino group as a point of attachment even though selective acetylation and other protecting group chemistry has been employed suggesting that this is a viable position for the incorporation of a fluorophore, via a spacer unit, to reduce the risk of interference with the binding site. The available literature in this area though, is of great interest to this study. There is precedent for attaching bulky groups to the amine, and again this research was carried out by the Swedish group of Nilsson, with the purpose of enhancing the inhibition of galectin-3.[133,134,137-139] The purpose of the study was to discover whether a second arginine residue (Arg-186) was within range of an aromatic moiety appended to the amine in the C-2 position of the disaccharide. As previously discussed an aromatic motif interacts almost as effectively with a guadinium positive charge, as a standard ionic interaction. Nilsson synthesised bis-amides such as the thiogalactoside below (**Figure 103**), and revealed enhanced inhibition. This β(1'→1) thiogalactoside is of quite different conformation in comparison to the β(1'→4) linkage of N-acetyllactosamine as the amine is actually in the C-3 rather than the C-2 postion. Hence it is difficult to draw comparisons in the context of the availability of the amine for derivatisation with a species not involved with binding such as a fluorophore.
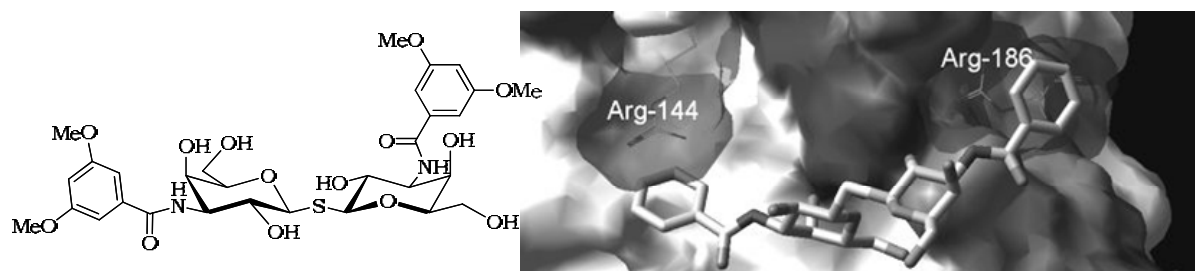
Figure 103: **Left**) A high affinity 3,3'-bisamido-thiogalactoside derivative with K_d = 50 nM (cf. C-3 NAc K_d = 1100 nm). **Right**) An energy minimised molecular model of the active site showing the two arginine residues interacting favourably with aromatic groups.[137]

The rationale behind the construction of a thio glycosidic 1'→1 linkage was to effect a more streamlined synthesis by effectively coupling together the two identical monosaccharides into a symmetrical species. This was a very efficient synthetic protocol, and as the thioglycoside had been shown previously to bind to β-galactosides comparably if not slightly better than LacNAc, it was a good strategy. To further selectively investigate the binding at Arg-186 a study discovered that a simple C-2 lactose phenyl ester bound with

considerably greater galectin CRD affinity than lactose, without any further functionality deeper in the pocket interacting with Arg-144.[137]

This must be taken into consideration in our synthetic planning but bears good omens as this work has revealed that a bulky group on the glucosamine moiety does not interfere with binding; on the contrary, in the case of an appended aromatic unit, binding is enhanced. The implication here is that there is room enough for a spacer unit to emerge from the binding pocket to an appended chromaphore. Indeed, it even opens the door to investigating an aromatic linker or the possibility of superior binding with the fluorophore at the free hydroxyl position.

We must now introduce the more hands-on side of working with carbohydrates and devise and rationalise a synthetic approach to building a galectin-3 targeted probe.

3.1.6 Carbohydrate chemistry

Thus far we have discussed the role of carbohydrates and how they are involved in widespread biological processes. We have explored the function and availability of the carbohydrate residues in glycoproteins and how they are used as binding sites for carbohydrate specific proteins such as lectins. This brought us on to why galectin-3 is a good target due to its up-regulation and involvement in tumour sites; and the great specificity of this protein towards unique carbohydrate residues. Finally we must consider the design and synthesis of galectin-3 specific probes and to begin with we must touch upon the chemistry of carbohydrates.

The Fischer convention is commonly used to represent structures with several stereogenic centres such as sugars or furanoses and pyranoses. A Fischer projection simply gives a representation of D and L enantiomers according to whether the OH-4 is to the left or right of the central column.[119,120]
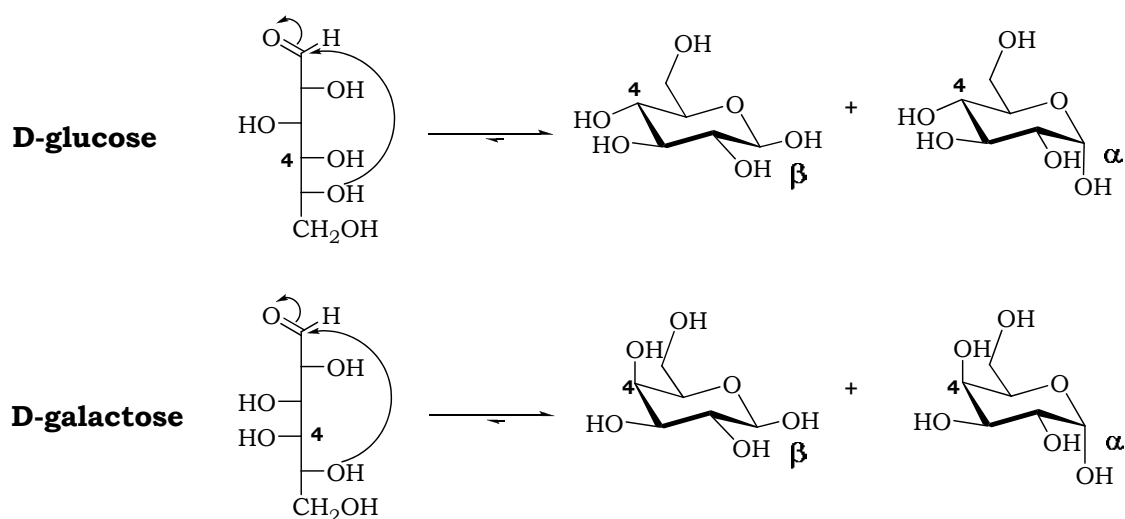
Scheme 30: The epimers D-glucose and D-galactose and their Fischer and chair conformations.

The abundance of open chain saccharide in solution, or in solid form, is negligible and the position of the equilibrium lies almost exclusively on the side of the cyclised structure (**Scheme 30**). However, due to the prolificacy of hydroxyls, there is of course the possibility of different sized lactol rings.

The position of equilibrium is rationalised by considering that intramolecular cyclisation is thermodynamically and kinetically favoured over intermolecular cyclisation due to entropic factors. With regards to ring size it is important to remember that five and six membered rings are thermodynamically more stable than more strained four and seven membered rings. The six membered ring is most stable as it adopts an essentially strain free chair conformation when all the components are equatorial, this is the lowest energy conformation and in water D-glucose consists of 38 % α-anomer and 62 % β-anomer.[119] However,

Figure 104: The anomeric effect

with an electronegative substituent at the anomeric position (C-1), such as an iodide, the stability is reversed and the axial α-anomer becomes more stable due to the anomeric effect (**Figure 104**).

Despite the α-position being sterically more hindered, it is the thermodynamically favoured anomer when the substituent is electronegative. The explanation comes down to electronic reasoning and the fact than an oxygen lone pair and the C-X bond are adopting an antiperplanar conformation allowing the delocalisation of electron density into the antibonding σ* orbital of the C-X bond, thus having a stabilising effect.[119,120] This has many implications which shall be discussed further in relation to the synthetic protocol employed in this study.

The presence of an oxygen substituent alpha to the anomeric centre means that any carbocation formed there is stabilised. Fischer glycosylation is the derivatisation of the anomeric centre to form an acetal (**Scheme 31**). Selective glycosylation is fundamental to sugar chemistry and synthetic design must always consider the anomeric selectivity and the use of the phenomenon that is neighbouring group participation (NGP).
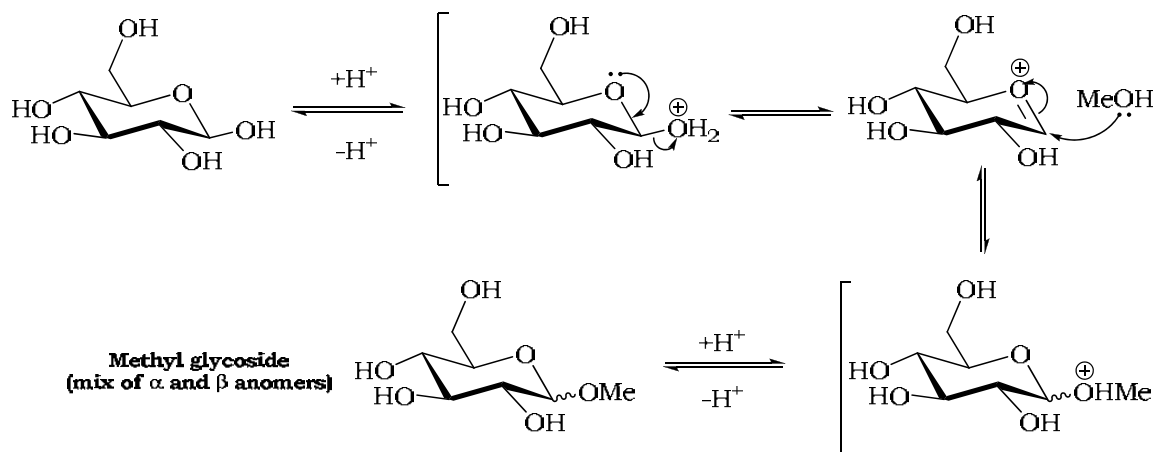
Scheme 31: Fischer glycosylation: the formation of a methyl glycoside.

A source of Bronstead acid such as SOCl$_2$ (thionyl chloride) with MeOH is used to drive the equilibrium through to completion. The result in this instance is the preferential formation of the α-product, as rationalised previously through the anomeric effect.

However, through shrewd protecting group chemistry at the C-2 hydroxyl, the stereochemical outcome of the reaction can be controlled at the anomeric centre to yield exclusively the β-anomer, (**Scheme 32**).
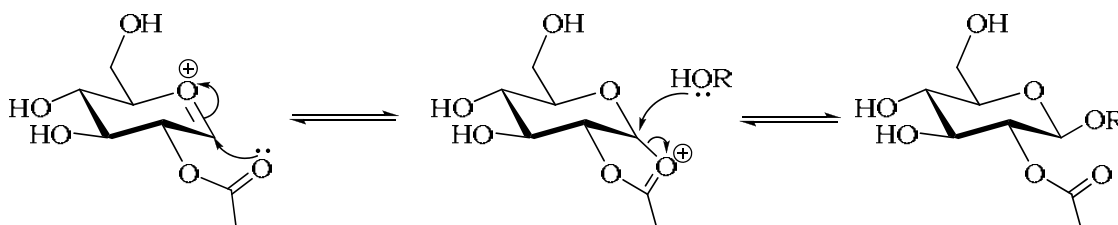
Scheme 32: Neighbouring group participation (NGP) controlling the anomeric centre.

This will be of importance in our synthetic strategy both in the construction of the monosaccharide building blocks **70** and **77**, and then in the crucial glycosylation step to form the β-galactoside **78** (**Figure 105**).
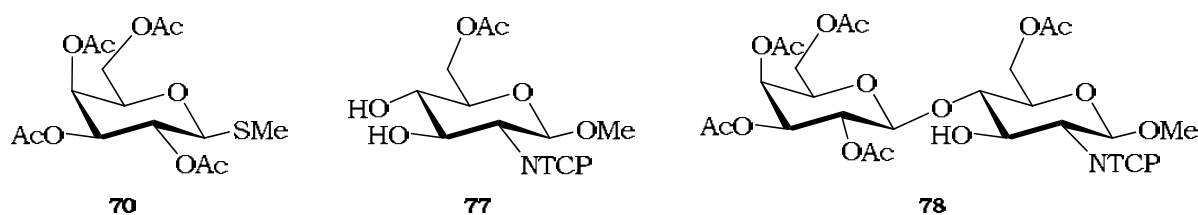
Figure 105: The three compounds where NGP will be fundamental in the synthesis.

3.1.7 The design of Galectin-3 specific carbohydrates

To exploit the unique galectin-3 binding pocket, and effectively mimic the binding of galectin-3 specific carbohydrates, it is important to consider the options available to synthetically rise to the challenge.

Literature work has concentrated on the binding affinity of various disaccharides. This research has not involved fluorescent labelling of the galectin-3 specific probes but has involved more of an investigative approach to explore the binding site and how to derivatise the disaccharides to optimise binding, solubility, stability and pharmacokinetics. **Figure 106** below, shows three of many reported disaccharides that have been tested against a galectin-3 rich cell line.
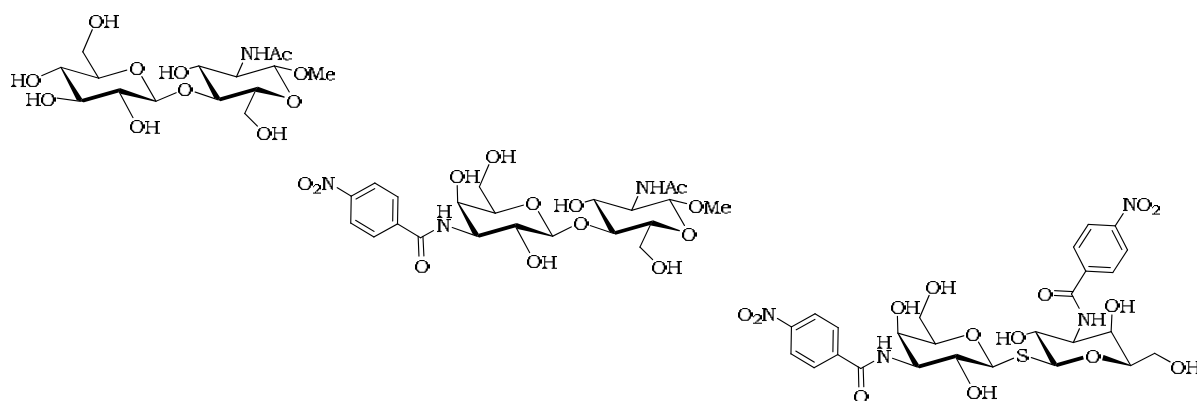
Figure 106: Three disaccharides investigated for galectin-3 inhibition.[138]

The ultimate intended application of these disaccharide systems has been with the intention of developing galectin-3 inhibitors for therapeutic application.

In the body enzymes are used to catalyse one-step glycosylations of exquisite specificity and complexity without the need for a protecting group strategy or even the isolation of glycosyl donor and acceptor. The conditions are mild and the yield quantitative.

It is estimated that for each glycosylation performed by synthetic chemists, an additional six associated protection and deprotection steps are required.

Enzymes that perform glycosylations are known as glycosyltransferases and the glycosyl donor leaving groups are either mono- or di-phosphates. The specific enzyme for the formation of our disaccharide β-galactoside is β-(1,4) galactosyltransferase. Three perfectly shaped binding sites ensure that the enzyme only catalyses the reaction between the correct substituents for the formation of a specific stereoselective and regioselective bond (**Figure 107**). The leaving group is also very important as it is often involved in the binding and hence the selectivity of the correct enzyme.
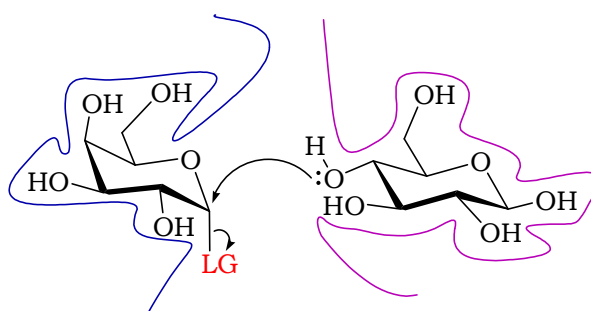
Figure 107: A simple illustration of enzymes regio- and stereoselectively controlling reactions.

In metabolism it is the reverse reaction, glycosyl hydrolysis, which breaks down carbohydrate glycosidic bonds. This process is catalysed by glycosylhydrolases, or glycosidases.

Ideally enzyme catalysed reactions would be employed in the laboratory as much as possible to increase the efficiency of synthesis. However, although there are numerous

reported examples in the literature of enzymes being employed for elegant organic synthetic protocol, there are problems with the isolation and stability of many enzymes.

It was however within our overall strategy to employ more rigorous synthetic methodology in the laboratory to develop an economical and reproducible route to construct galectin-3 specific fluorescent probes.

3.1.8 The proposed fluorophores

As explained in the results and discussion of chapter 2 (page 77), it was decided to initially move away from sensitised lanthanide chromophoric systems and attempt first to label the disaccharides with either commercially available or easily synthesised organic fluorophores (**Figure 108**). The reason for this was to enable the development of an efficient reaction pathway and then test the efficacy of the binding prior to adjusting the strategy to incorporate novel lanthanide systems as the fluorescent label.
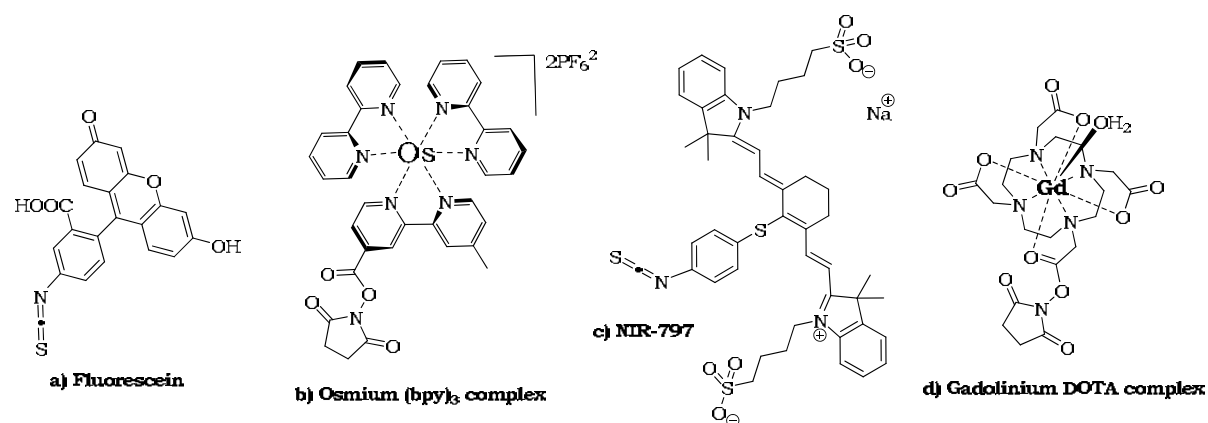
Figure 108: **a**) Fluorescein-the gold standard; **b/c**) near-IR emitting fluorophores used for peptide $\alpha_v\beta_3$ integrin targeted imaging (chapter 1), and **d**) an MRI sensitive gadolinium contrast probe.

3.2 Synthetic plan

The overall plan was to report an efficient and improved synthesis of β(1'→4) galactoside probes targeted to the galectin-3 protein. Cellular testing would then reveal the efficacy of CRD binding and subsequent visualisation through a fluorophore attached at the amine C-2 position of glucosamine through a 6-aminocaproic acid spacer unit.
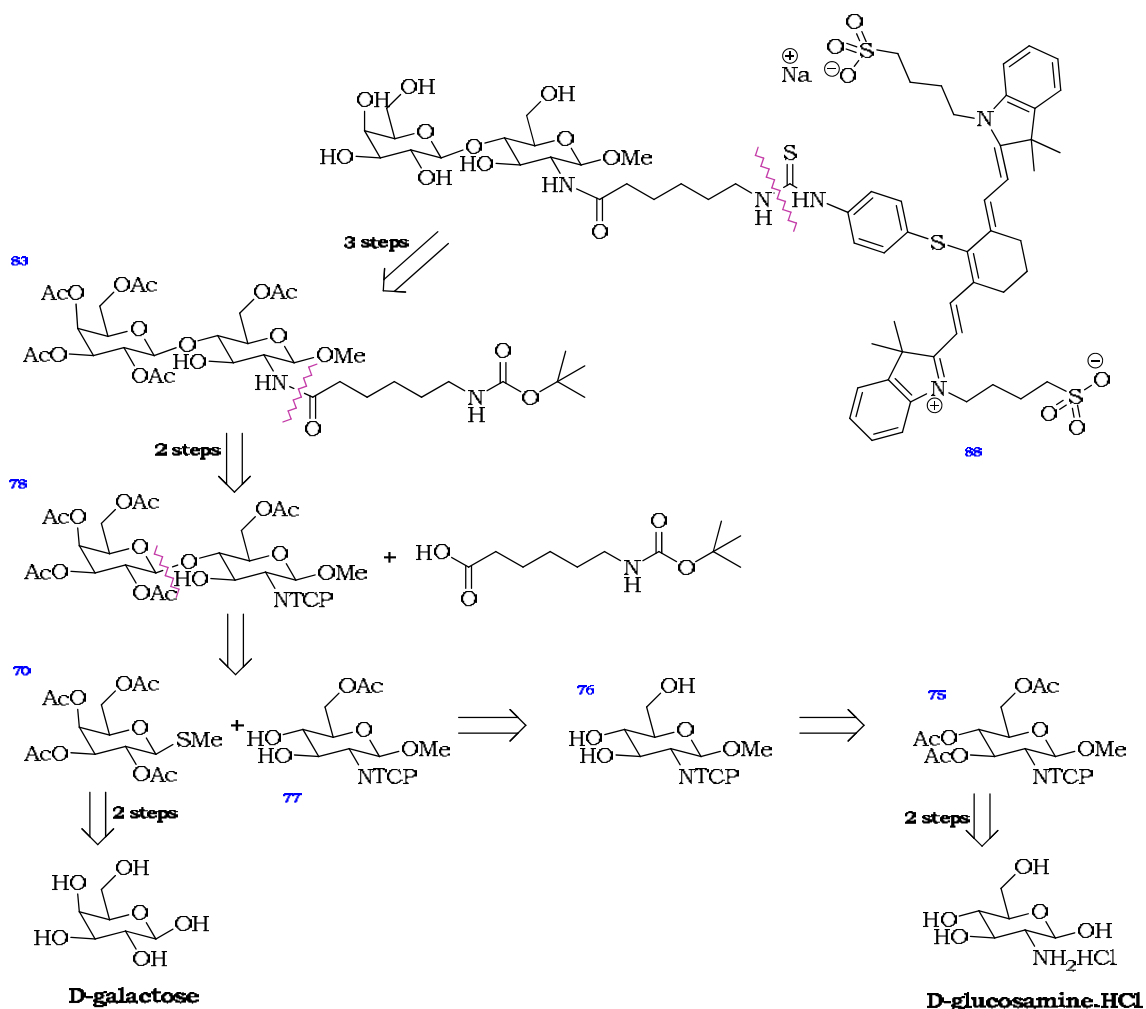
Scheme 33: Initial retrosynthetic analysis of NIR797 fluorescent labelled β-lactosamine.

The proven galectin-3 binding β(1'-4)-galactoside, N-acetyllactosamine, would be a good model disaccharide to synthesise and test the availability and potential of the amine for the purpose of fluorophoric tagging. The retrosynthetic analysis (**Scheme 33**) shows the proposed breakdown of the synthesis into the simple and readily available mono-saccharide starting materials D-galactose and D-glucosamine hydrochloride.

The forward synthesis of the glycoside donor **70** was deemed to be relatively straightforward with well precedented per-acetylation with Ac$_2$O/py followed by manipulation at the anomeric position.[140,141] This should proceed by first forming an α-iodide with I$_2$/Et$_3$SiH, followed by displacement with thiourea to form a β-thioglycoside and final methylation with MeI/Et$_3$N to give the product. This chemistry is well precedented and

so it was envisaged that as the glycosyl acceptor could be more challenging to synthesise with a tricky monoacetylation at the primary C-6 hydroxyl, it was wise that this should be conquered first.
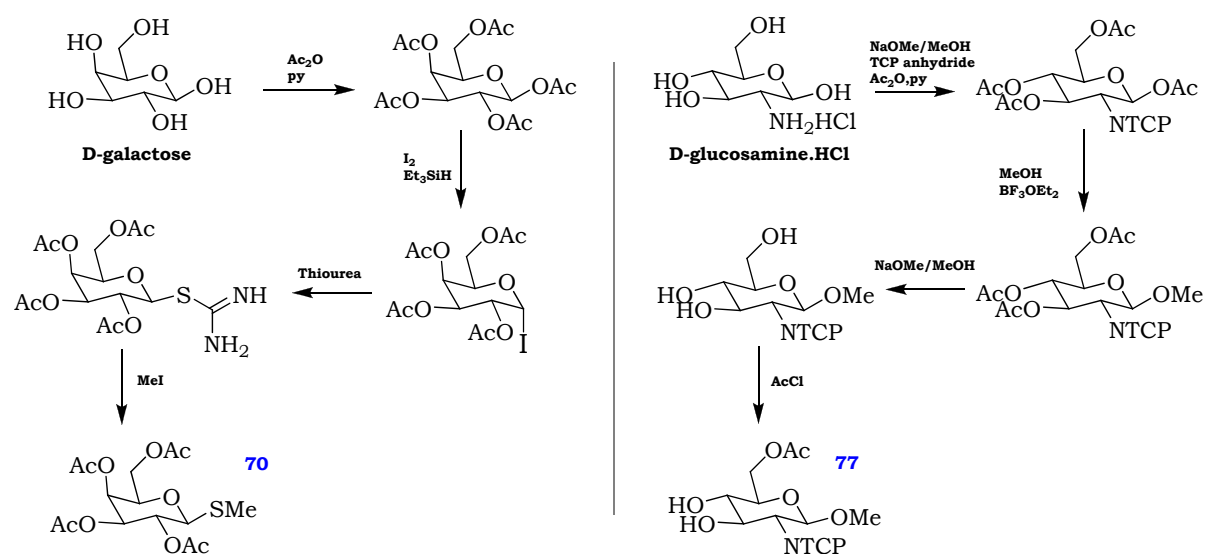
Scheme 34: The anticipated forward synthesis of the glycosyl donor **70,** and acceptor **77**.

Synthesis of the acceptor **77**, should proceed from D-glucosamine hydrochloride, which, upon freeing up the amine with an equivalent of base, would be ready for amine protection. It is wise to use a protecting group that is durable to mildly basic and strongly acidic conditions in this synthesis and the tetrachlorophthalic anhydride (TCP) amine protecting group fully meets these criteria.[142,143] This step requires Ac_2O to perform the role of completing the TCP protection by closing the anhydride, however it is reportedly feasible with an excess of Ac_2O and pyridine to peracetylate the hydroxyls in the same step.[143] The anomeric position would then be manipulated to selectively install a β-methyl ether under the control of the participating TCP neighbouring group as previously discussed. Deacetylation with NaOMe then sets the molecule up for monoacetylation at the primary C-6.[134] Precedent for this key step involves taking advantage of the slight increase in nucleophilicity of the primary alcohol, and greater steric accessibility over its tertiary counterparts OH-3 & OH-4. A protocol involving acetyl chloride and the mild base 2,4,6-trimethylpyridine (sym collidine) with careful control of the temperature at -42°C should afford glycosyl acceptor **77**.
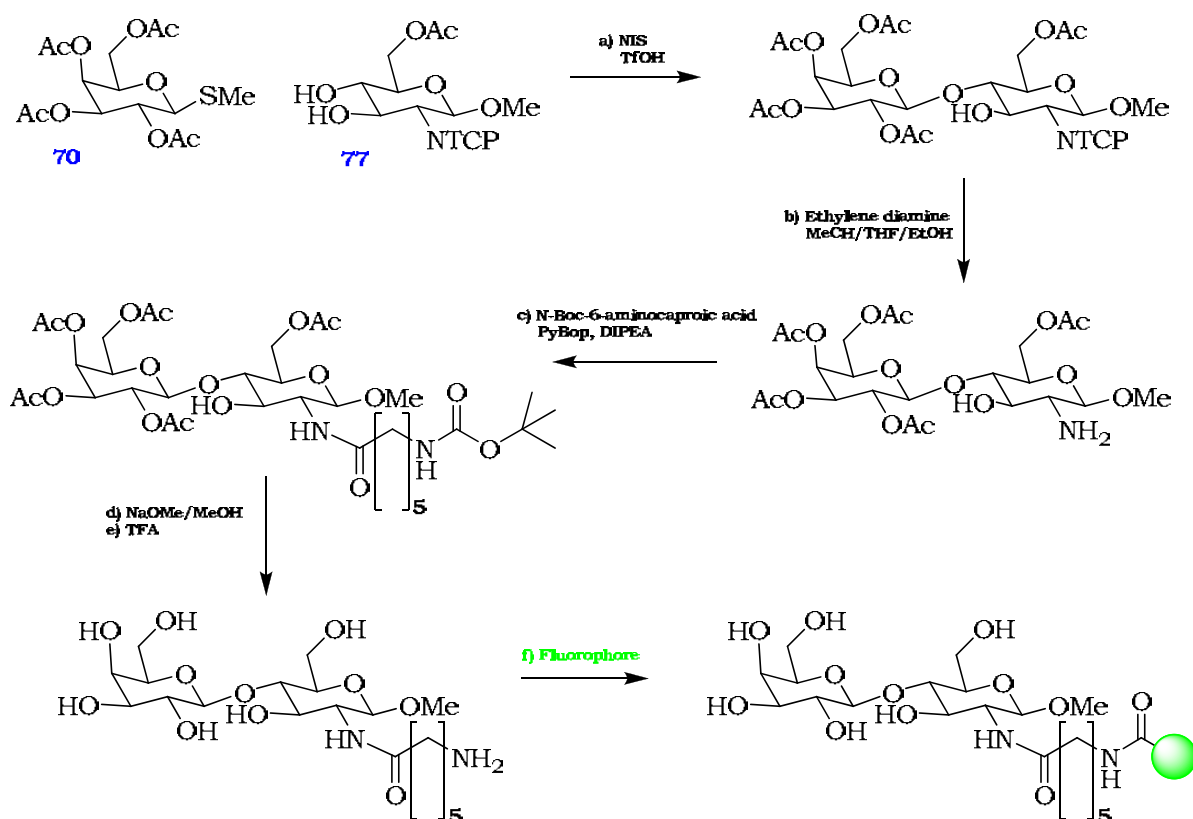
Scheme 35: The proposed continued forward synthesis: **a**) Glycosylation; **b**) TCP cleavage; **c**) Spacer attachment; **d/e**) Deprotection; **f**) Fluorophoric tagging.

The β(1'→4) glycosylation step is also a key step in the synthesis as there is the possibility of the formation of regioisomers at the C-3 and C-4 free secondary alcohols of the glycosyl acceptor **77**. This step is precendented and was found, under certain conditions, to be 1'-4 selective, probably due to OH-3 being more hindered and slightly less nucleophilic due to electron withdrawal from the NTCP group.[134,144] The conditions which will be attempted in our protocol to mediate the glycosylation is a NIS/TfOH combination.

Following disaccharide formation and characterisation to confirm the regioselectivity, TCP cleavage will be accomplished to afford the free amine for derivatisation. Precedent for this deprotection has often involved quite harsh conditions such as reduction using NaBH$_4$.[104] The preferred conditions must be mild so as to not hydrolyse the glycosidic linkage. Reported testing of various amine and solvent combinations revealed that ethylene diamine in a combination of MeCN/THF/EtOH was most effective for similar disaccharides.[103,120] However, the solvents are dependant on substrate solubility and EtOH alone may serve our purpose.[120] Interestingly most reported TCP deprotection chemistry protocols involve trapping the resulting free amine by immediate acetylation with Ac$_2$O in the same pot, this would afford us LacNAc. This suggests that either the free amine is unstable or, more likely, that purification is hugely simplified with an acetamide as opposed to a reactive free

amine. Indeed for galectin-3 binding site competition and blocking studies some LacNAc will be required so a portion will be acetylated to give **82**.

The final steps of the sequence involve completely unprecedented chemistry and so the coupling reaction with Boc-amino caproic acid using the favoured coupling reagents for our peptide chemistry (PyBop/DIPEA) will be attempted to install the spacer unit. Deacteylation with NaOMe/MeOH and Boc cleavage with TFA should be straightforward. The final step is coupling to the fluorophores which will bear good terminal electrophilic functional groups such as isothiocyanates or succinic esters, as detailed above, for simple nucleophilic substitution reactions.

3.3 Sugars results and discussion

3.3.1 Synthesis of monosaccharide building blocks
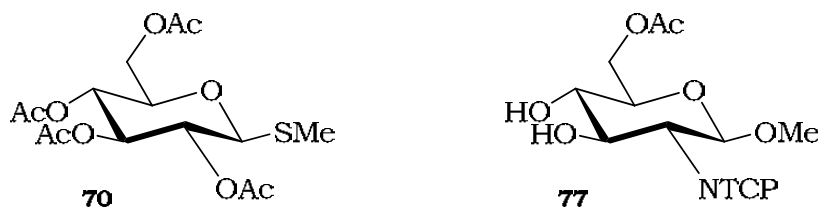
Figure 109: **Left**) D-galactose derived glycosyl donor **70**. **Right**) D-glucosamine derived glycosyl acceptor **77**.

It was envisaged that the construction of the two monosaccharide building blocks (**Figure 109**) would proceed relatively straightforwardly as there is literature precedent for many of the synthetic steps. As explained in the synthetic plan, methyl glycoside **77** requiring the greater number of synthetic steps, including a potentially tricky C-6 mono-acetylation, was attempted first.

3.3.1.1 Synthesis of glycosyl acceptor **77**
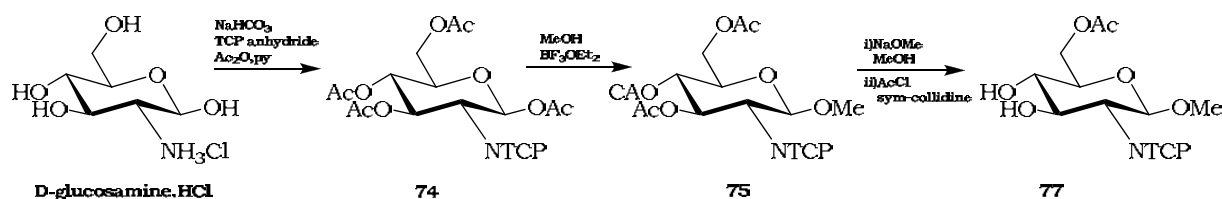
Scheme 36: The original strategy for the synthesis of glycosyl acceptor **77**.

The tetrachlorophthalic (TCP) protecting group was chosen for many reasons. It is tolerant to strongly acidic and mildly basic conditions, and offers keen involvement in neighbouring group participation (NGP) affording great β-selectivity. This is also reputedly due in part to the closest chlorine atom sterically hindering the α-position.[143] The electron withdrawing character induced by the four chlorine atoms also serves to increase the glycosyl donor properties relative to the phthaloyl protecting group.

Repeated attempts at the first step, tandem TCP protection of the amine with tetrachlorophthalic anhydride followed by per-acetylation with acetic anhydride/pyridine in the same pot, failed to produce a satisfactory yield of product. The attempted protocols followed those of Schmidt[143] and Kochetkov,[145] using NaOMe and NaHCO$_3$ as the bases respectively to first free up the amine from the D-glucosamine hydrochloride salt prior to the protective sequence with tetrachlorophthalic anhydride, Ac$_2$O and pyridine. Excess

acetic anhydride is applied to not only acetylate the hydoxyls, but also to ring-close the intermediate to the phthalimide (**Scheme 37**).
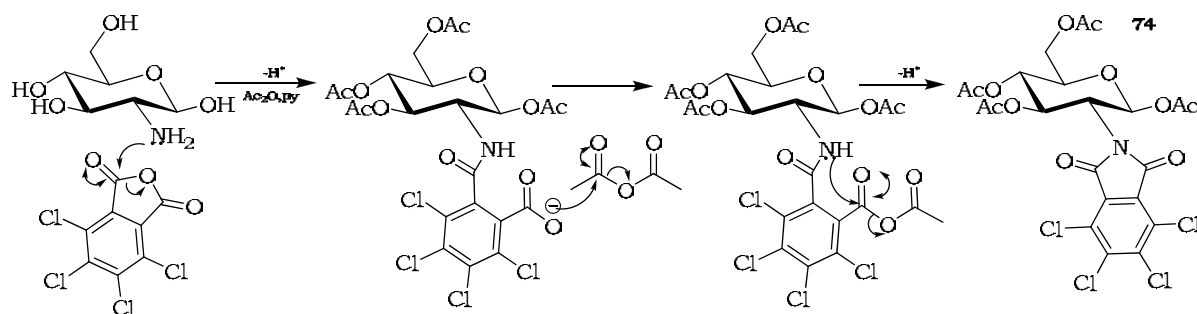
Scheme 37: A mechanism for TCP protection illustrating the role of acetic anhydride.

Both methods were attempted. The first involved D-glucosamine hydrochloride being taken up in aqueous $NaHCO_3$ and a slight excess of TCP anhydride applied to the mixture. After stirring for 30 minutes the solution was acidified and the precipitate dried over P_2O_5. This effectively isolates the crude TCP appended glucosamine as a carboxylic acid (**Scheme 38**). The dried intermediate product is then treated with Ac_2O and pyridine to afford the product. However this reaction failed and although the nmr clearly revealed the pyridinium acetic acid salt there was no sign of product.
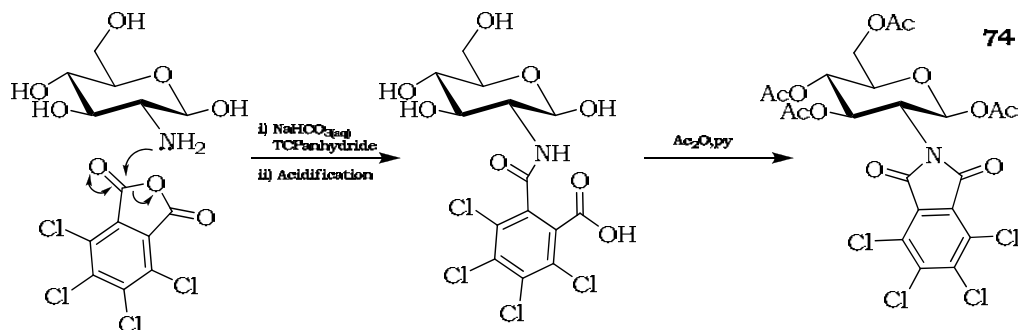
Scheme 38: The first attempted reported method.[145]

The second method employed an equivalent of NaOMe as base in dry MeOH with TCP anhydride, followed by further portions of NaOMe portionwise over 3-4 hours. This reputedly forms the N-acetyl intermediate prior to addition of Ac_2O/py to formulate the peracetylated N-TCP product.[143] Not only did the methodology sound dubious due to the continuous addition of excess NaOMe, but the reaction was irreproducible in our hands and failed to formulate the desired products even when attempted with alternative bases, such as $NaHCO_3$ from the first method (NaOMe is moisture sensitive).

It was decided to create glycosyl acceptor **77** through an alternative, albeit longer route, which whilst completing an efficient high yielding synthesis, allowed a good induction into the world of sugar chemistry.

Interestingly it would appear that other groups have experienced similar problems with the tandem TCP protection/acetylation step and opted for an alternative route too.[144,146] The alternative route we decided to follow involved temporary protection of the amine as an imine **71**, prior to per-acetylation **72**. Acid deprotection of the amine and reprotection with TCP anhydride then afforded **74** in excellent overall yield (**Scheme 39**).
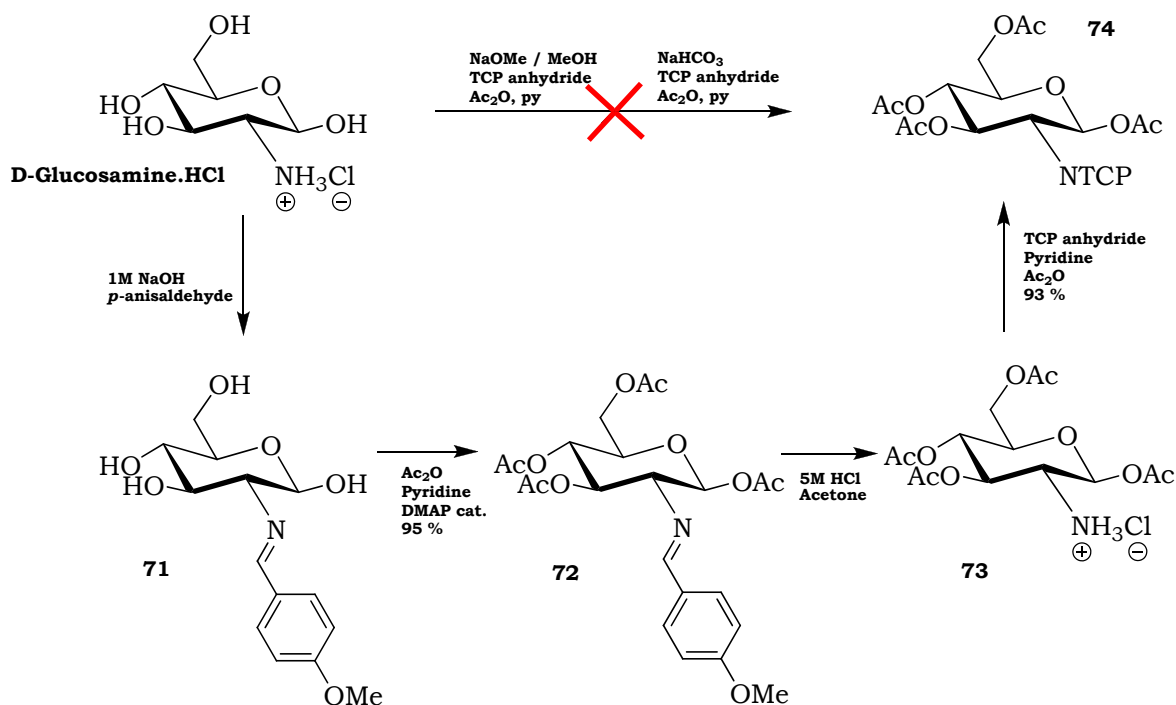
Scheme 39: The alternative route to formation of peracetylated TCP protected glucosamine **74**.

This route was accomplished without difficulty and proved scalable and reliable. The imine was formed through condensation of the amine with *p*-anisaldehyde under basic conditions. The reaction was very clean, the product precipitated out of the aqueous solution and was filtered and washed to afford crude **71** which was sufficiently pure to be taken on directly. Per-acetylation was performed under the standard Ac_2O/py conditions, except a sub-stoichiometric quantity of DMAP was also applied to act as a nucleophilic catalyst in the reaction. The reaction was simple to follow, with the much less polar product being purified on a straightforward silica flash column. Yield was again excellent.

At this point it is worth noting a disparity in our proton nmr spectral assignments in comparison to those in the literature for the imine **71**.[144] COSY and HSQC 2D experiments were thus performed and confirmed our nmr assignments were correct. It was assumed when assigning the spectrum that the less well resolved broader looking peaks of lower shift than the OMe at ~3.8 ppm were hydroxyls (**Figure 110**), and indeed this is what the literature suggested. However, peracetylated glucosamine derivative **72** showed a significant migration of peaks leaving just the H-2 doublet of doublets below the OMe and a

D$_2$O shake confirmed the disappearance of the hydroxyls. 2D-nmr spectroscopy put this matter of accurate assignment to rest revealing that upon acetylation the shift of the internal protons (H-1 to H-6) is increased considerably, proving the literature assignments wrong.
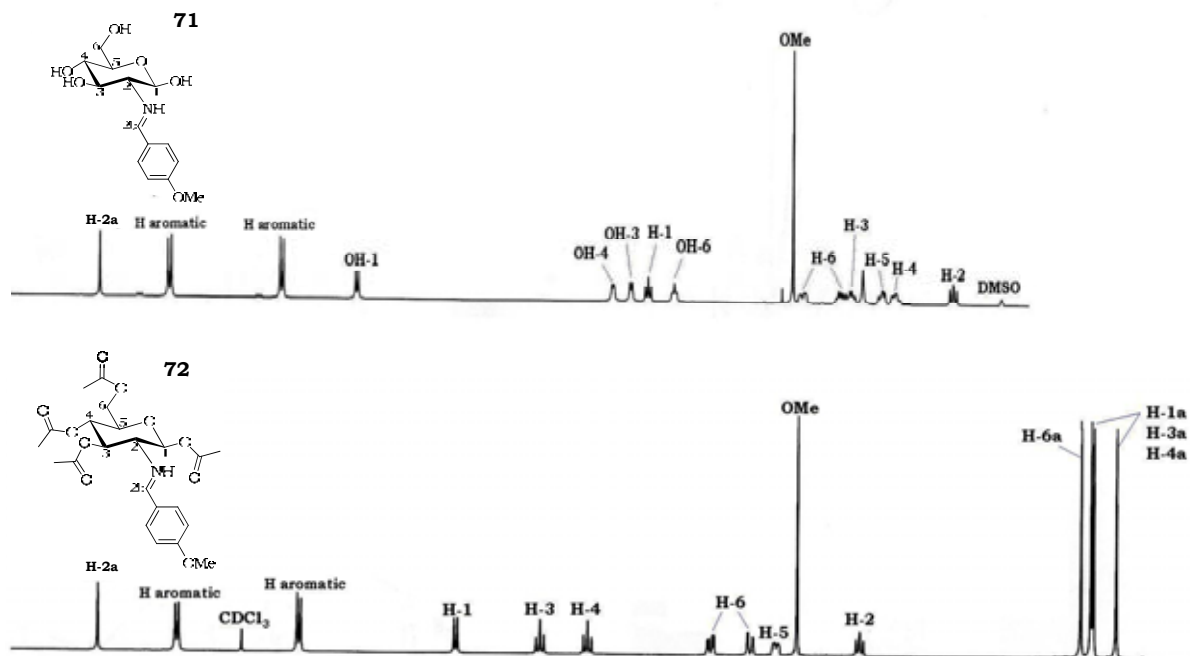
Figure 110: The ^{1}H nmr spectra for compounds **71** and **72**.

Returning to the synthesis, the amine was then rapidly and very cleanly deprotected with 5M HCl in refluxing acetone to reveal the hydrochloride salt which was again immediately taken on crude as the nmr was excellent. Amine protection with TCP anhydride, Ac$_2$O and pyridine in toluene proceeded smoothly and flash column chromatography revealed the product as a fine white powder. The aromatic quaternary TCP carbons gave a weak ^{13}C-nmr signal which was initially confusing but this was found to be a consistent feature. The mass spectrum revealed the characteristic isotopic signals for a tetrachloro- compound.

Synthesis of compound **75**
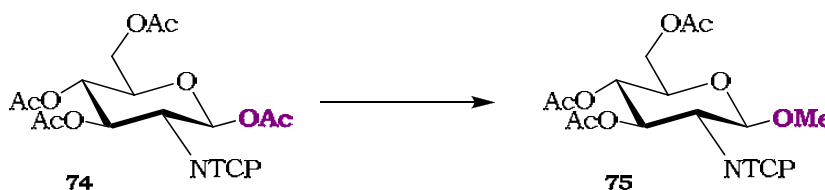
Scheme 40: Transformation at the anomeric centre.

The next step involved the transformation of the anomeric centre into a stable methyl glycoside, making this position unreactive to further chemistry; the anomeric substitution

has no involvement in galectin binding. The literature is rich in examples of anomeric manipulation with all manner of substituents including the formation of a β-methoxide, but not with TCP as the adjacent participating protecting group. However it was envisaged that following a protocol that has installed allyl or aryl glycosides in this position using a simple Lewis-acid-catalysed transformation with the corresponding alcohol would afford us the product.

The preliminary strategy involved a slight adaptation of the method of the work of Ellervik[144] who glycosylated with p-methoxyphenol (**Scheme 41**) using borontrifluoro diethyletherate (BF_3OEt_2), as catalyst.
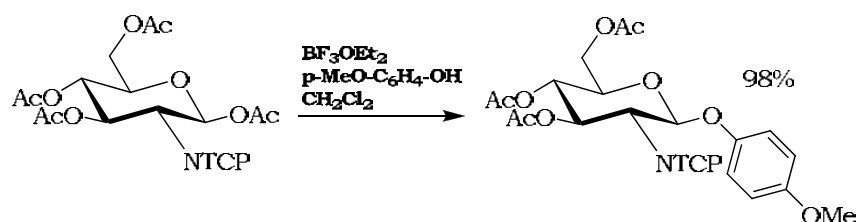
Scheme 41: Ellervik's glycosylation.[144]

Unfortunately this chemistry was not reproducible for us with methanol as the nucleophile, with no sign of product formation and recovered starting material. It was then decided to attempt the protocol followed by Schmidt.[143] This sequence was more complex and involved the regioselective C-1 deacetylation with hydrazinium acetate in DMF, prior to installing a trichloroacetimidate in a purely β-selective fashion through reaction with CCl_3CN in the presence of DBU. Tin triflate was then added as a Lewis acid to afford the glycosyl donor primed for nucleophilic attack by an alcohol. In this instance the chemistry is employed to actually construct β(1'→4) disaccharides (**Scheme 42**), but there are other examples of β-selective glycosylations under these conditions.
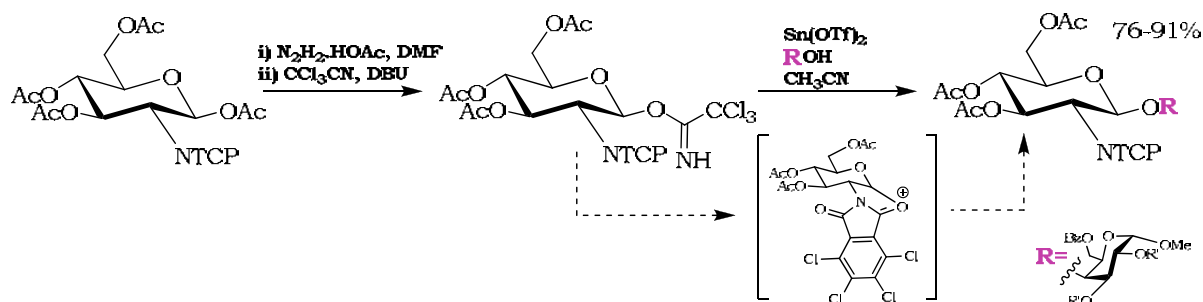
Scheme 42: Schmidt's glycosylation.[143]

Again this reaction failed repeatedly with methanol in place of the reported glycosyl donor. However, the crude nmr of the trichloroacetimidate intermediate did show successful completion of the first step so the reaction failed at the final hurdle. It was therefore time to attempt a final strategy to install the methyl glycoside and a paper by Debenham[142]

revealed a possible alternative route via the bromide, similar to the well known Koenigs-Knorr strategy (**Scheme 43**).
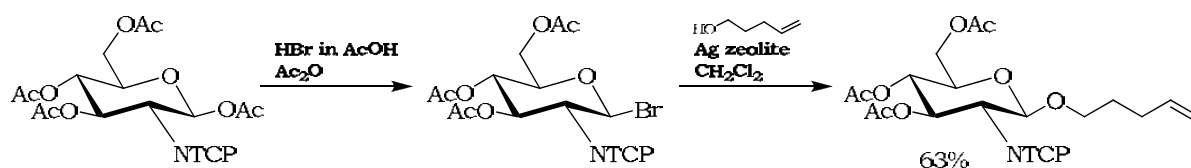
Scheme 43: Debenham's glycosylation.[142]

The bromide was formed readily and isolated but taken on crude after a thorough work up. Due to the unavailability of silver zeolite, it was decided to substitute this promoter with a stoichiometric amount of silver tetrafluoroborate (AgBF$_4$). On a very small scale, using an old batch of AgBF$_4$, this worked perfectly with an excess of MeOH in CH$_2$Cl$_2$ yielding the product in good yield (86 %). The synthesis was then continued and global deacetylation performed by applying a small quantity of sodium metal (ca. 10 mg) to the dilute substrate in dry MeOH, affording fresh NaOMe *in situ*. Completion of hydrolysis was confirmed by TLC and the pH adjusted to ~5 with Dowex 50W-X8 H$^+$ ion exchange resin to bind all sodium ions and protonate the hydroxyls prior to isolation of the product from the filtrate. However, upon scale-up of the glycosylation with a fresh batch of AgBF$_4$, an unexpected and fortuitous discovery was made. The reaction had previously been followed successfully by TLC but to our consternation there was no appearance of product spot and our sugar-specific α-naphthol TLC stain just revealed a product on the baseline. Change of eluent to a far more polar 10 % MeOH/CH$_2$Cl$_2$ mixture mobilised the product from the baseline and flash column chromatography isolated a pure product which analysis revealed, to our surprise, the fully deacteylated methyl glycoside **76**. We had effectively performed two steps in one (**Scheme 44**).
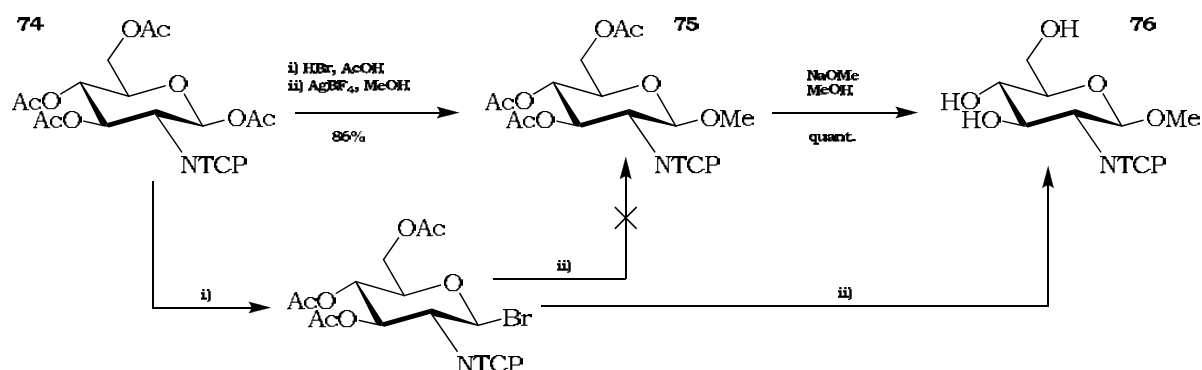
Scheme 44: The serendipitous transformation eliminating the NaOMe/MeOH hydrolysis step.

This useful discovery proved to be reproducible and reliable when repeated, and the most likely explanation is that a small quantity of HF is formed during the reaction hydrolysing

the acetates. This is plausible, and to reason as to why the first small scale attempt behaved as we anticipated and stopped at the peracetylated methyl glycoside **75**, it is suggested that the old AgBF$_4$ had degraded over time to AgO, Ag$_2$O, Ag$_2$CO$_3$ and/or BF$_3$, hence no HF had formed. Usefully this also proved how robust the TCP and OMe groups are to strong acid conditions. Characterisation of the intermediate bromide revealed a characteristic Br and Cl$_4$ isotope pattern in the mass spectrum indicative of the product. The product confirmed the loss of bromide and the appearance of the distinctive OMe singlet in the ^{1}H nmr at ~3.3 ppm with an integration of three, and of course the complete disappearance of signals from the acetates.

Synthesis of compound **77**
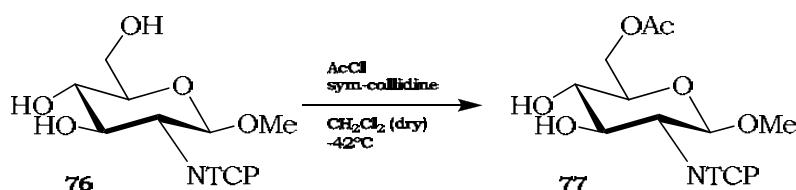
Scheme 45: Selective acetylation at the primary hydroxyl.

The final step is the tricky C-6 monoacetylation. Sorme,[134] offered an elegant route in which the more accessible and more reactive primary alcohol was selectively acetylated with a single equivalent of acetyl chloride with sym-collidine (2,4,6-trimethylpyridine) as base. The key to this transformation was the use of freshly distilled acetyl chloride under stringently dry conditions. It was also found that even a very slight excess of AcCl led to the acetylation of either the C-3 or C-4 hydroxyl giving the undesirable bis-acetylated product. The reaction proceeded as planned, with TLC clearly showing the formation of the product with a slightly higher R$_f$ than the diasappearing starting material. Following flash column chromatography, the ^{1}H nmr revealed a singlet at ~2.1 ppm with integration of three proving that monoacetylation had occurred and 2D-nmr confirmed it was at C-6. Thus the synthesis of the gycosyl acceptor was complete.

3.3.1.2 Synthesis of glycosyl donor **70**
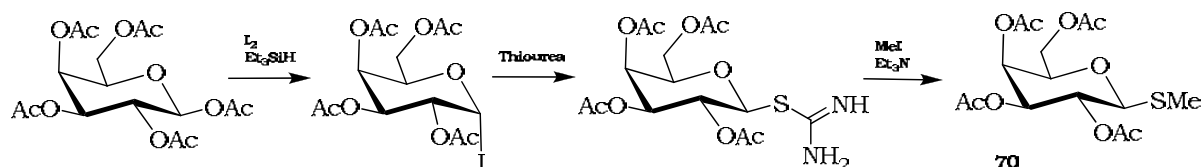
Scheme 46: The proposed synthetic pathway to thio-glycoside **70**.[140]

Anomeric thio groups can act as temporary protecting groups. They have traditionally been installed through treatment of the peracetylated sugar with the corresponding thiol under

Lewis-acid conditions.[140,146] However these protocols use harsh and sensitive reagents and afford modest yields of thio glycoside after lengthy experimentation. Iadonisi[140] and Field[147] have employed an alternative, much milder and more rapid strategy to access the desired anomeric thioalkyl derivatives. This proceeded via the creation of an α-glycosyl iodide intermediate followed by displacement of the iodide with thiourea to generate an isothiouronium species which was then readily alkylated with the required alkyl halide.

An identical protocol was attempted and it proved remarkably efficient. Peracetylated D-galactose (the literature precedent uses D-glucose) was treated with a slight excess of both I_2 and triethylsilane. Upon heating to reflux rapid conversion to the iodide was obvious through consumption of the starting material and the appearance of a less polar product. Basic work-up produced the crude iodide which was shown by nmr to include Si-containing impurities.

These impurities, however, are deemed not to interfere with further synthetic steps,[140] and the iodide, being susceptible to hydrolysis, was dried and the crude product taken on. This was then up in dry acetonitrile before a slight excess of thiourea was added. Again, upon heating at reflux the iodide disappeared and a polar product appeared on the baseline. Without further ado the reaction was cooled and at room temperature an excess of both methyl iodide and triethylamine were added sequentially. After stirring for 15 minutes the reaction was complete.

Flash column chromatography afforded the pure desired gycosyl donor in excellent yield and the synthesis proved to be reliable upon scale-up. As expected in the [1]H-nmr the SMe protons (~2.1 ppm) were further upfield than the OMe protons (~3.2 ppm) of the glycosyl acceptor **77**, and appear in the same 'bunch' as the OAc CH_3 protons so 5 singlets each of integration equal to three protons, were present from 2.00-2.25 ppm in $CDCl_3$.

So the two monosaccharide building blocks were accessed in excellent overall yield and improvements made in both reliability and yield when compared to the monosaccharide units reported in literature routes towards N-acetyllactosamine.[123]

3.3.2 Glycosylation: formation of the β(1'→4) glycosidic linkage to construct the disaccharide **78**
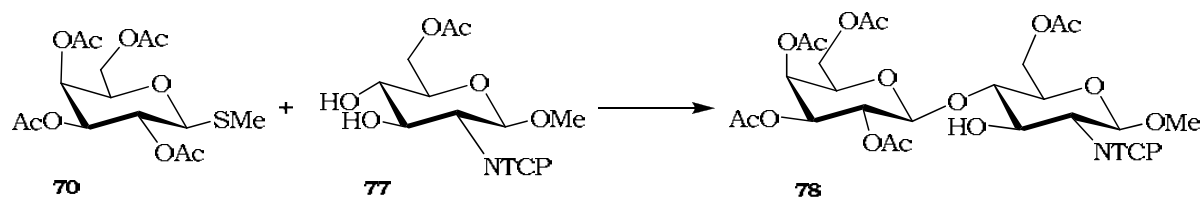
Scheme 47: Synthesis of fully protected lactosamine derivative **78**.

The key to this reaction was to perform the glycosylation in a β(1'→4) regioselective manner. Nilsson,[134,138] tested several catalytic systems to promote the reaction including MeSBr/AgOTf and Br$_2$/AgOTf. However the greatest yield and complete regioselectivity was achieved with an N-iodosuccinimide / triflic acid (NIS/TfOH) combination in a 1.5 : 0.15 equivalent molar ratio respectively. Thus we attempted this system and the reaction worked in our favour with the product being isolated as a fine white solid after purification by flash column chromatography.
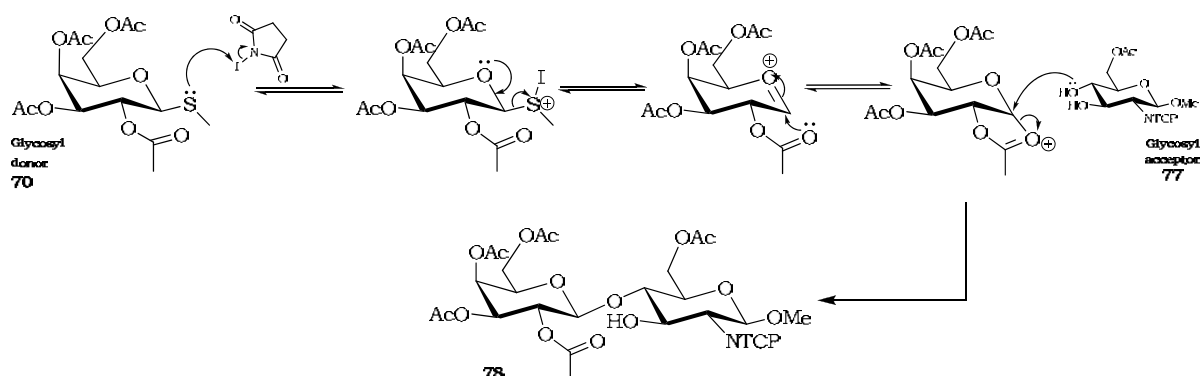
Scheme 48: Proposed mechanism for the β(1'→4) selective glycosylation reaction mediated by NIS/TfOH.

To fully assign the complicated but very well resolved and remarkably clean nmr spectra, it was necessary to perform COSY-90 and HSQC experiments. These afforded complete assignment of the ^{1}H and ^{13}C resonances but still shed no light upon the C-3/C-4 regioselectivity. To confirm the β(1'→4) product we obtained an HMBC spectrum which revealed the long range [3]J proton to carbon coupling; C-1' to H-4 and H-1' to C-4 (**Figure 111**).
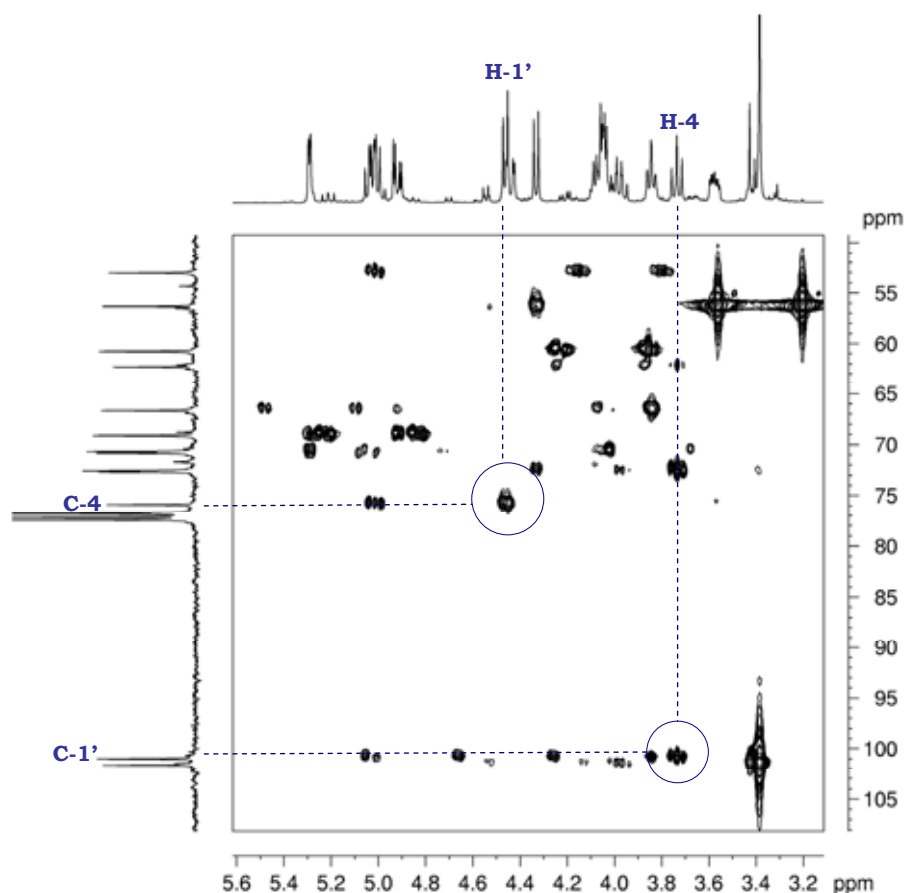
Figure 111: The 2D nmr confirmation of β(1'→4) regioisomer by HMBC.

Nilsson,[134] verified the regio product through acetylating the C-3 hydroxyl and subsequently performing a further reaction and purification. The result was a downfield shift of 1.1ppm for the doublet of doublets signal of H-3 in the proton nmr. However, as shall be explained more in due course, this was unnecessary as firstly the unambiguous couplings in the HMBC spectrum (**Figure 111**), but also as in the formation of standard N-acetyllactosamine for blocking and competition studies, we over-acetylated and the HMBC showed C-3 coupling to an acetyl group and an acetyl coupling to C-3 as well as the downfield shift described by Nilsson.

First though we must touch upon the TCP deprotection step which will afford the free amine ready for manipulation of this point which is not involved in binding in the galectin-3 carbohydrate recognition domain.

3.3.3 TCP deprotection

TCP deprotection is well precedented and a particularly useful study was accomplished by Fraser-Reid.[143] The N-tetrachlorophthaloyl group is particularly sensitive to strongly basic conditions and most deprotections reported are not purified and characterised as the free amine, but rather Ac_2O/py applied in the same pot to form the acetamide. Fraser-Reid

experimented with different deprotection conditions.[143] Amines such as hydrazine, ethylene diamine and methyl hydrazine and solvent combinations involving THF, MeCN, DMF, isopropanol and EtOH at various temperatures were investigated and an optimum TCP cleavage system was developed which was employed successfully by ourselves (**Scheme 49**).
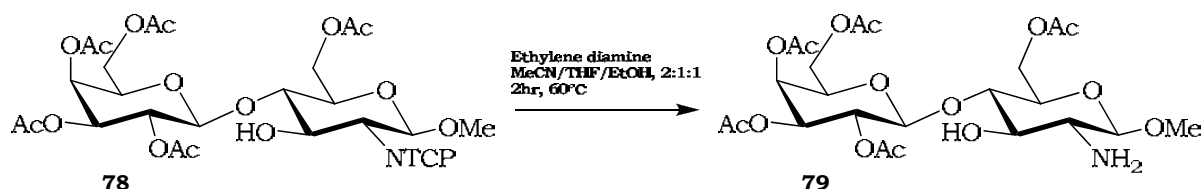
Scheme 49: Strongly basic TCP deprotection.

The reaction was simple to follow by TLC with the free amine predictably dropping close to the baseline. Upon repetition of this deprotection it was found that when purifying, for solubility reasons and optimum yield from the flash column, the crude mixture should not be concentrated under reduced pressure to dryness but rather should be concentrated to sensible volume for direct application to the column. This may seem counter intuitive but the residual ethylene diamine seems to reduce streaking, and by reducing the acidity of the silica renders the product more mobile and it is recovered much more quickly and in good yield.

The free amine was then divided; a portion was acetylated to afford the acetamide which after treatment with NaOMe/MeOH afforded the N-acetyllactosamine standard (**Scheme 50**), and a portion was coupled to a protected spacer unit for the attachment of the fluorophore.

Synthesis of N-acetyllactosamine derivative **82**
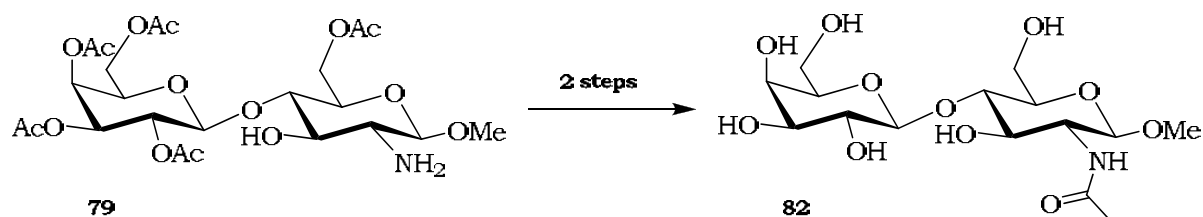
Scheme 50: Synthesis of β-1-methoxy-N-acetyllactosamine **82**.

This acetylation step is well documented as the final part of TCP deprotection protocol to immediately acetylate the amine as previously mentioned. However as we had isolated the free amine it was equally simple to proceed with an acetylation under standard conditions, and also gave confirmation of our nmr assignments. As the final step would involve hydrolsis of all the O-acetyls it was envisaged that an excess of Ac_2O would simply

acetylate the amine and OH-3 (**Scheme 51**), and then the acetates could all be promptly hydrolysed to afford the blocking disaccharide **82**. However, it was also important to be able to control the synthesis and selectively acetylate the more nucleophilic amine to form **81**, in case tagging at OH-3 were required at a later stage.
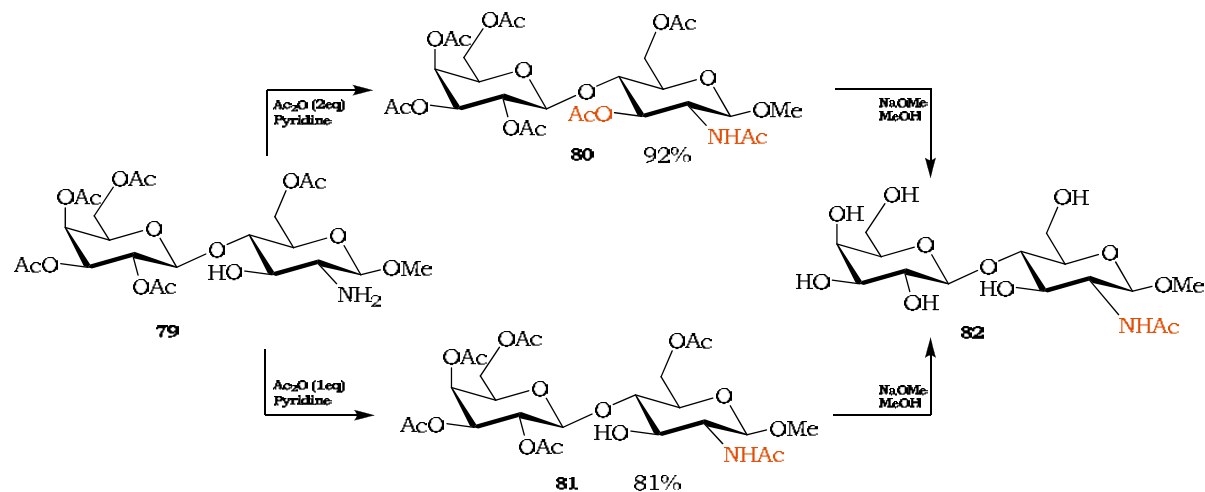
Scheme 51: The routes to the final product **82**.

The mono- and di- acetylations worked very effectively with the per-acetylated product **80** being isolated in improved yield. A new quite unusual chromatography eluent mixture of 40% acetone / 60% toluene was employed with very satisfactory results in comparison to the MeOH / DCM gradient eluent system used previously.

Long range nmr studies of the acetylation performed with excess Ac_2O **80,** proved the β(1'→4) regioisomeric product. HMBC analysis showed C-3 coupling to acetate protons, the reverse of this, and the acetate carbonyls to H-3 (**Figure 112**).
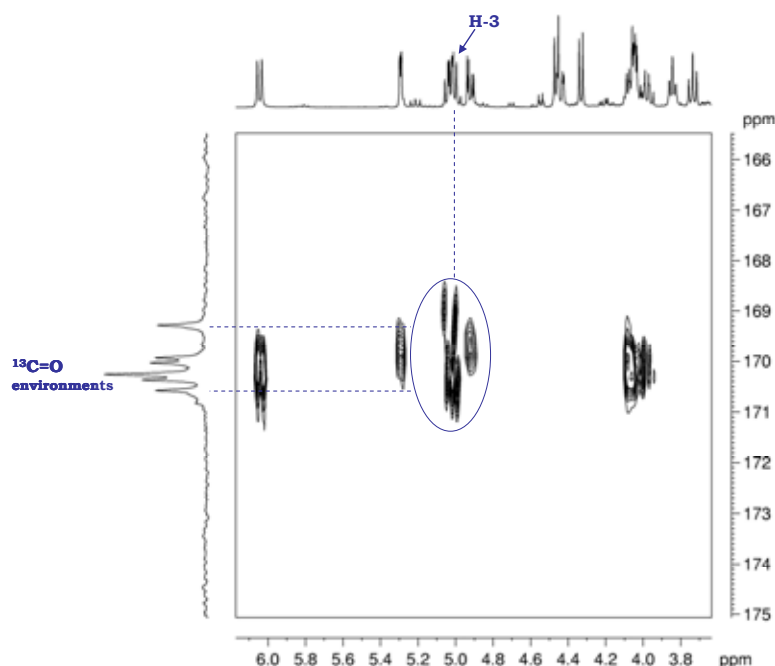
Figure 112: Detail of the compound **80** HMBC spectrum offering proof of acetylation in the 3-position.

3.3.4 Introduction of a spacer unit
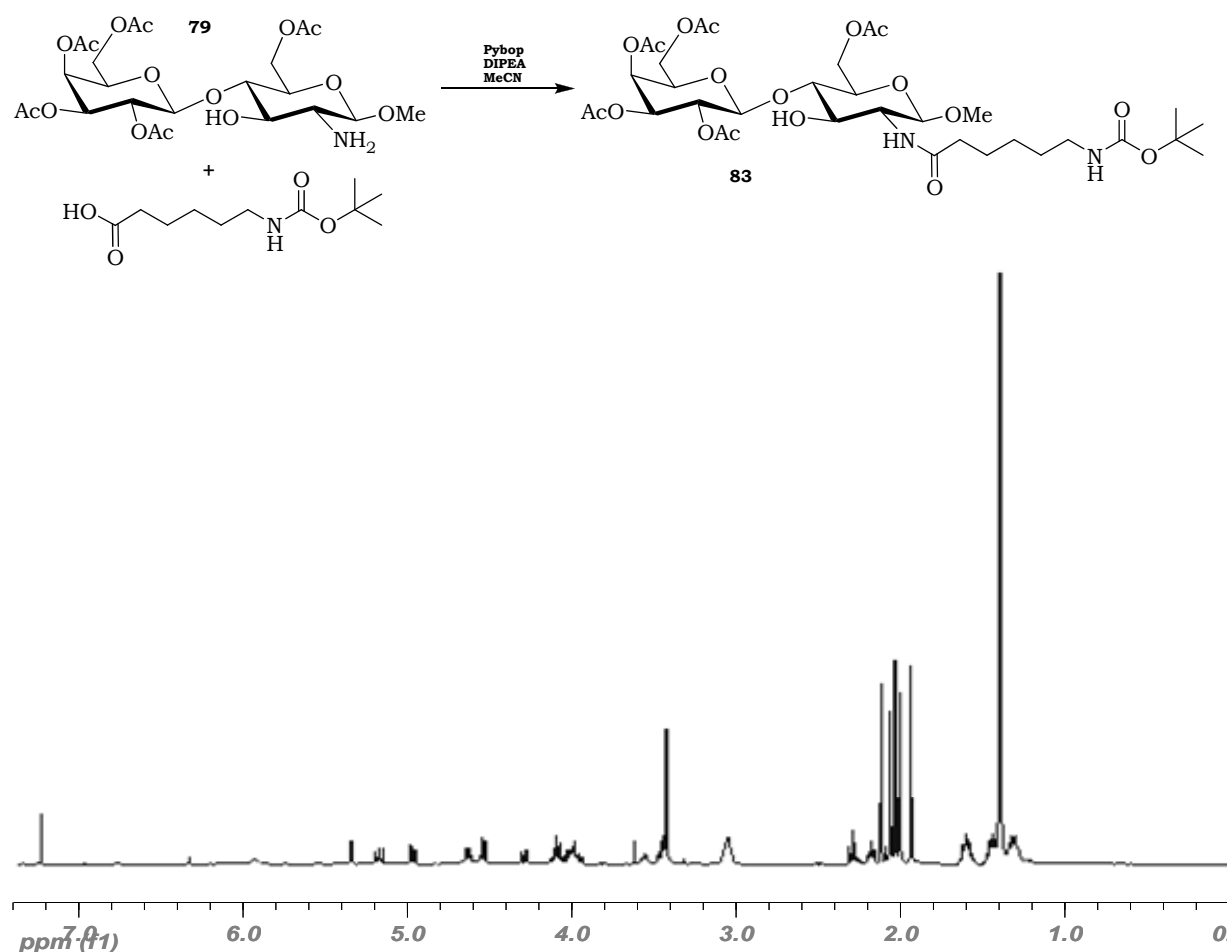
Figure 113: **Top**) The coupling of Boc-aminocaproic acid to the free amine of **79**. **Bottom**) The ^{1}H nmr of the product.

The purpose of the installation of a spacer unit was to distance the bulky fluorophore from the fundamental binding interactions that occur between the disaccharide, especially D-galctose, and the galectin-3 CRD. To this end the simple tert-butoxycarbonyl (BOC) protection of amino caproic acid was accomplished (details in the 'linkers' section page 129), and a well established coupling protocol applied using the reagents successfully employed in peptide experiments. These conditions are mild and high yielding. In hindsight the application of a stronger base such as NaOH could have saved us a step but would have brought with it the purification and characterisation problems which will be discussed forthwith.

The reaction was performed using a slight excess of Pybop and base, pre-mixed with a 1.5 equivalent excess of the carboxylic acid in dry acetonitrile prior to addition of the free amine.

The reaction was simple to follow by TLC with the clear appearance of a much more mobile product spot in comparison to the free amine on the baseline. After stirring at 40°C overnight and purification by careful flash column chromatography (by-products are of similar R_f), the product was afforded in an excellent 95 % yield. It was satisfying to note that the hydrophobic chain did not seem to reduce solubility, indeed the nmrs were achieved in CDCl$_3$ and, albeit now more complicated due to the methylene units, were fully assigned by COSY, HSQC and HMBC experiments.

2.3.5 Deprotection – synthesis of **86**
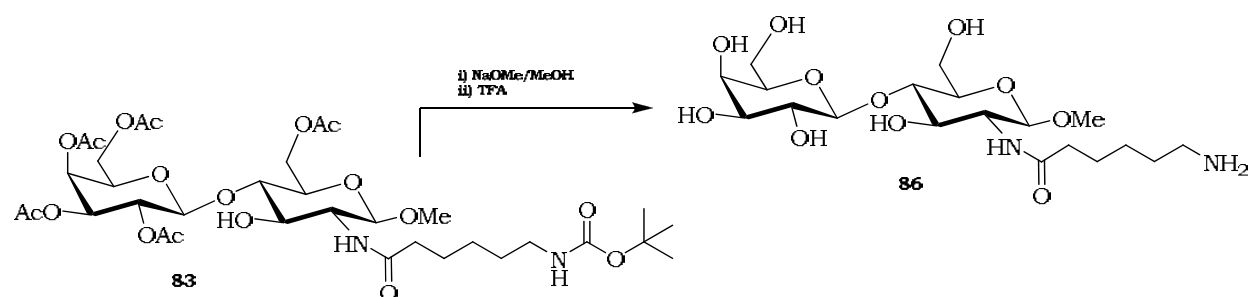
Scheme 52: The final deprotection steps.

The successful synthesis of disaccharide with Boc protected caproic linker **83** leads us to the final steps of the synthesis. At this stage however, the reaction route required some thought and investigation due to there being no similar procedures available in the literature and unforeseen purification issues.

Scheme 53 sheds a clearer light upon the situation. The options were either to deacetylate (NaOMe/MeOH), perform Boc cleavage (TFA) and then couple to the fluorophore in a simple final step, or, firstly cleave the Boc to uncover the free amine for fluorophore coupling and then hydrolyse the acetates as a final step.
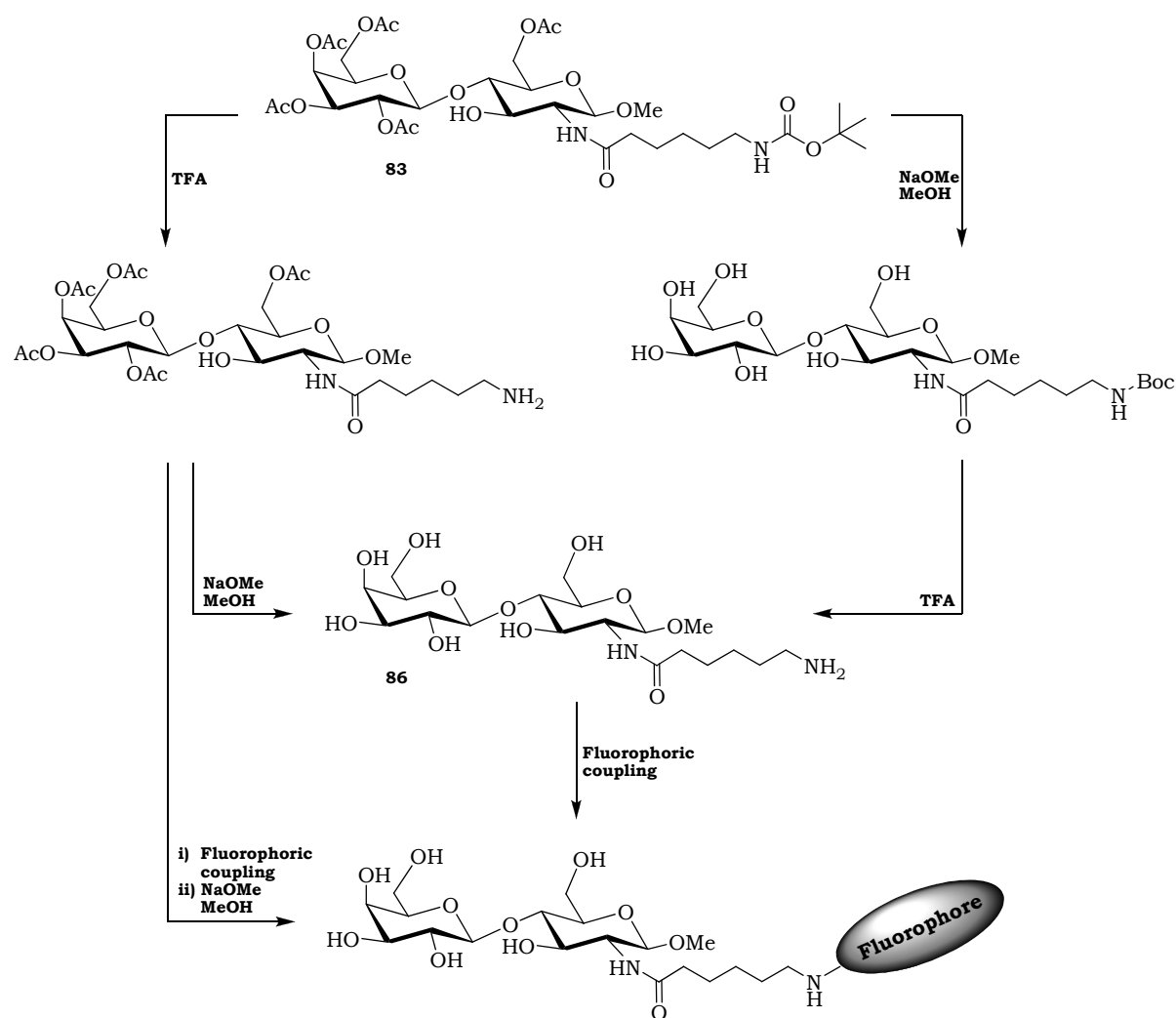
Scheme 53: The plausible routes from **83** to final labelled products

The favoured synthetic route was to create the fully deprotected product **86** prior to coupling to the fluorophore through nucleophilic attack of the primary amine as the final step. This was considered ideal as the complexity of the compound brought about by the incorporation of a bulky fluorophore means there is more potential for breakdown or undesirable side reactions in subsequent steps.

However, as the deacetylation step with NaOMe/MeOH rendered the disaccharide water soluble, the work-up and purification protocol became complicated. It was therefore decided to first perform Boc cleavage with TFA which is a very simple and quantitative reaction as the by-products are volatile and the product is recovered pure without further purification and taken on to the coupling step. The free amine intermediate could then be deacetylated or coupled with the fluorophore and deacetylated as the final step, contrary to our synthetic design.

The deacetylation products stick to silica, and although many eluents were attempted such as CH_2Cl_2 / MeOH/ $NH_{3(aq)}$, 79.95 : 20 : 0.05, there was no product progression from the TLC baseline. HPLC purification was attempted as reversed phase C18 silica TLC with

MeCN/H$_2$O as eluent revealed good product mobility. However the product is not UV active so normal HPLC protocol became impossible with the available equipment. Pulsed Amperometric Detection (PAD),[123] and Electronic Light Scattering (ELS),[122] detection capabilities were available on an analytical scale, but there was no option for scale up. After reaction completion which was observed clearly by the plummeting of R$_f$ on the TLC, neutralisation was then accomplished with ion exchange resin and the product recovered from the filtrate.

However, having proved both routes possible, it was decided to pursue our policy employed throughout the peptide synthetic protocol which was to make the fluorophore labelling the final step of the synthetic sequence, thereby eliminating potential problems associated with the fluorophores and final deprotection reagents. New purification methodology was discovered and again using Dowex-50WX8 H$^+$ resin as a scavenger,[151-154] we were able to afford clean deacetylated product without further purification.

This product was then subject to TFA cleavage to afford the pure fully deprotected product as the TFA salt in excellent yield after repeated co-concentrations and triturations with Et$_2$O to remove all excess TFA and other ether-soluble impurities.

Coupling was then performed on the crude using the pre-decided fluorophores (fluorescein, osmium complex & NIR797). As predicted, the fluorophores not only rendered the product visible by HPLC but also afforded excellent baseline separation for scale-up to semi-preparative quantities. The products seemed to stick to the reversed phase HPLC column when a methanolic solvent gradient was employed, but a mildly acidic acetonitrile (MeCN + 0.05 % TFA) gradient afforded the product in good yield.

Characterisation was no longer available by nmr as firstly the quantities of final product were approximately one milligram, and secondly, the added complexity incurred by the sophisticated fluorophore renders accurate assignment difficult. High resolution mass spectrometry, fluorescence spectroscopy and a clean HPLC trace were therefore the analytical techniques used to prove the identity and purity of the final products.

3.3.6 Coupling of Fluorescein & NIR797

- Fluorescein

This reaction was performed in a tried and tested solvent mixture of pyridine, DMF and CH₂Cl₂. The reaction was rapid and readily followed by analytical HPLC. The product was isolated in good yield and taken on to the cellular testing phase as an excellent standard with which to draw comparisons in further studies.
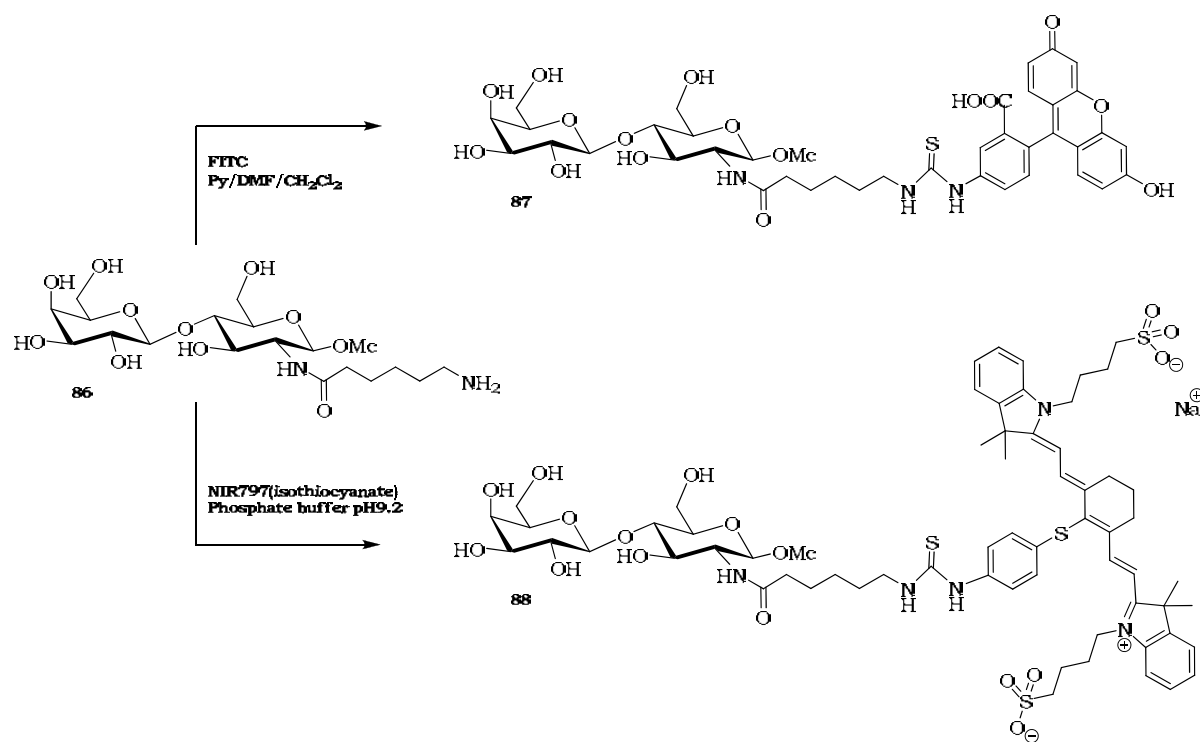
Scheme 54: The final products: 1-methoxy-N-caproic-fluorescein-lactosamine (**87**), and 1-methoxy-N-caproic-NIR797-lactosamine (**88**). See appendix 4 for fluorescence spectra & appendix 5 for HPLC & m/z data.

- NIR797

This reaction was performed in basic phosphate buffer conditions (pH = 9.2) to free up the amine from the trifluoroacetic acid salt and the appearance of the product peak was clear in the HPLC at 31 minutes (**Figure 115**).

Thus the synthetic phase was complete and the compounds ready for biological testing. Examples of a MALDI-TOF mass spectrum and HPLC trace of final product are shown in figures **114** and **115** overpage. The next step was to test the binding and fluorescence of these compounds in cells. The choice of cells was obviously important as the positive cell lines must express galectin-3 and the controls either must not, or the levels of gelectin-3 expression must be known. This work was carried out in the twilight of the research phase of this PhD and as such the results must be regarded as preliminary and at the time of writing are by no means conclusive.
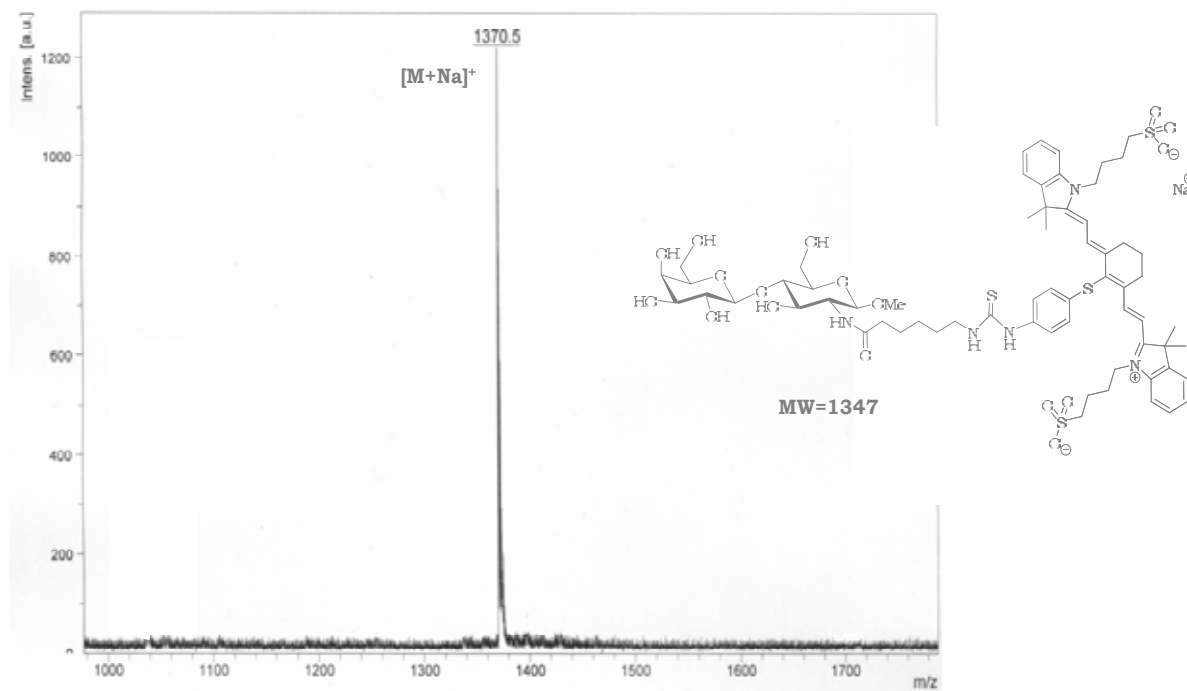
Figure 114: MALDI-TOF mass spectrum of the final product **88**.
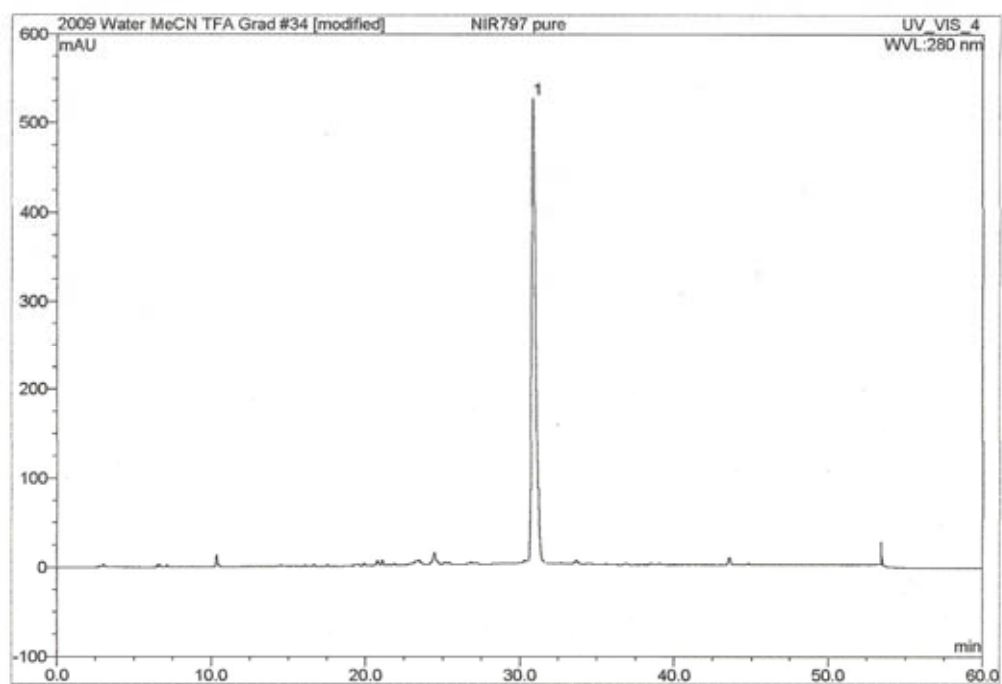
Figure 115: Analytical HPLC of the final product **88**.

3.4 Cellular studies

This work was accomplished in collaboration with Dr. Carmel McConville of the institute for cancer studies at the University of Birmingham and Queen Elizabeth Hospital Birmingham.

These studies are by no means conclusive and the work is continuing with both the fluorescent labelled cyclic peptides targeted towards $\alpha_v\beta_3$ integrins (chapter 2), and the labelled β-galactosides targeted to galectin-3.

Pure galectin-3 targeted compound **88** from HPLC purification was dissolved in phosphate buffer solution (PBS) and used in varying concentrations to stain cell lines either expressing or not expressing galectin-3 both inter- and extra-cellularly. The cell lines used initially were designed to develop reliable protocol with relation to both the growth and staining of the cells, and the microscope settings. The literature precedent for targeting galectin-3 predominantly involves the use of the U87-MG gliobastoma cell line,[52,54,56,156] which has a proven high galectin-3 expression. However this cell line is relatively slow growing and due to the time constraints and the ready availability of SHEP1 cells grown from an actual neuroblastoma biopsy sample, which also express galectin-3, it was decided to proceed with staining.

Cell lines:

SHEP1 neuroblastoma (positive Galectin-3 expressing cancer cell line); an 'S-type' (substrate adherent, Schwannian) cell line.

SH-SY5Y glioblastoma negative control (non/low galectin-3 expressing cancer cell line); an 'N-type' (neuronal) cell line.

Both are derived from the same patient and probably represent alternative pathways of differentiation from a neural crest derived precursor.

General synthetic protocol:

DCFl staining of SHEP1 and SH-SY5Y neuroblastoma cell lines.

1.   Innoculate 6.0 x 10^5 U87-MG, 7.5 x 10^5 SHEP cells or 10^6 SY5Y cells into 6 cm cell culture grade petri dishes and culture for 24 hours at 37°C/5%CO$_2$. (Culture medium is DMEM/F12 (Invitrogen) supplemented with non-essential amino acids (Invitrogen), 2 mM Glutamine and 15 % fetal bovine serum).

2.   Remove medium and wash with 4 mL PBS.

3.   Fix with 3 mL cold methanol for 10 min (pre-chill methanol at -20 °C for at least one hour), (methanol will both fix and permeabilize the cells).

4.   Rinse with 4 mL PBS, 3 x 5 min, OR, 4a. Rinse 3 x 5' with PBS/0.5 % bovine serum albumin (BSA) (Sigma aldrich A2153).

5.   Stain with 10 mM DCFl or cRGDfK in PBS (in the dark, room temp) for 30-40 min, OR, 5a. Stain with 10mM DCFl or cRGDfK in PBS/0.5% BSA (in the dark, room temp) for 30-40 min.

6.   Wash with PBS with gentle agitation, 3 x 15 min.

7.   Mount in Vectashield (Vector Labs) containing 0.5 ug/ml DAPI (DAPI: 4',6-diamidino-2-phenylindole is a fluorescent stain that binds strongly to DNA)

8.   View using a Nikon Eclipse E600 fluorescence microscope (ExEm filters 480/535 for fluorescein, 400/450 for DAPI), using x20 or x40 objective.

Results & discussion

Cell line: SHEP-1

Probe: Disaccharide-Caproic-Fluorescein, DCFl, compound **87**.
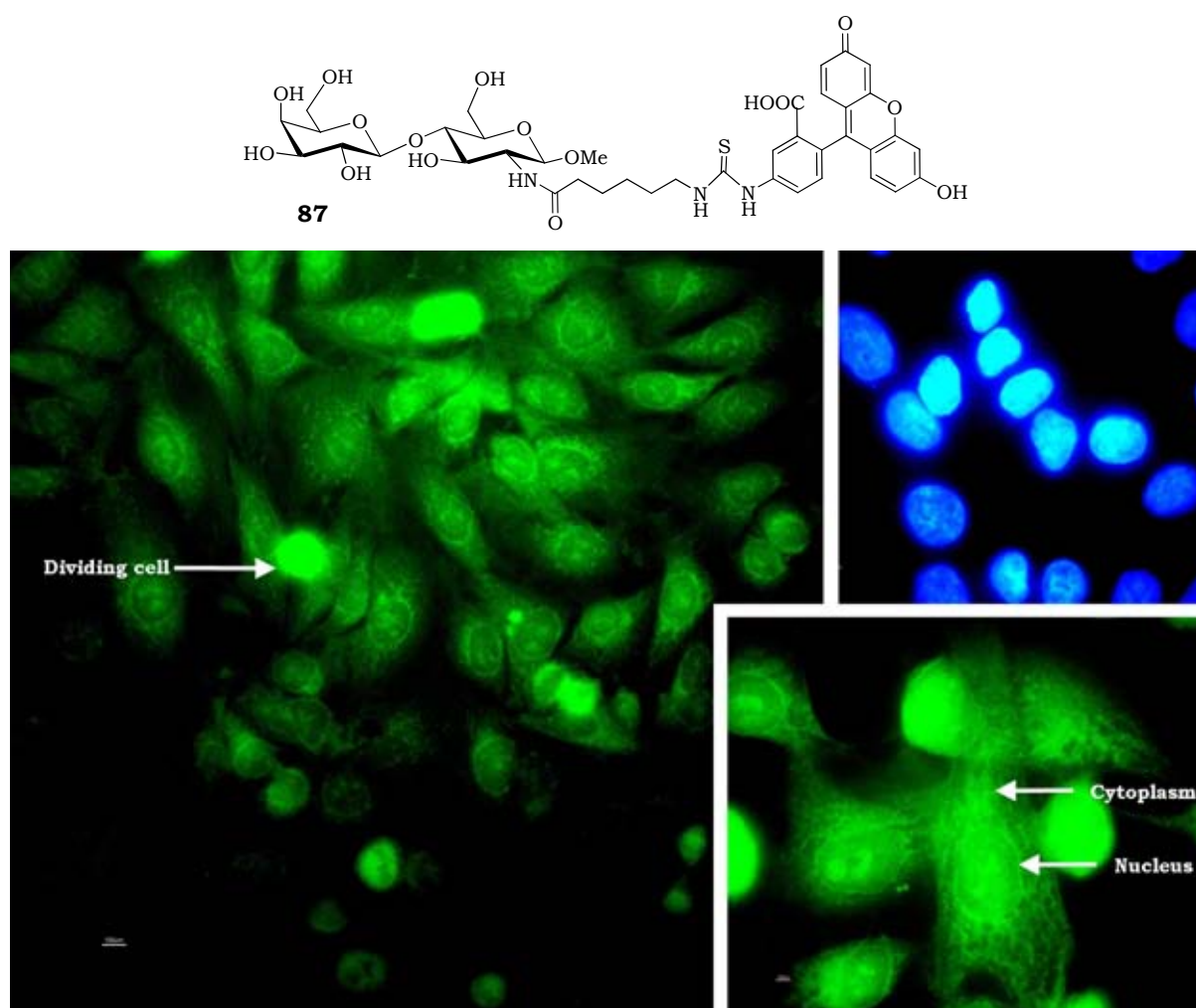
Figure 116: SHEP-1 cells stained with fluorescein labelled disaccharide **87**. Dividing cells show higher levels of fluorescence and appear a more intense green. The outlines of the nuclei are clearly visible and the less intense areas of emission are the cytoplasm. **Top right**) DNA binding of the DAPI blue stain showing clearly the location of the nuclei. The two stains are non competitive and fluoresce at different wavelengths and hence can be used simultaneously.

The images obtained were immediately aesthetically pleasing and definitely show a level of cellular binding. The cellular nuclei are well defined (**Figure 116**), with enhanced emission from the surface creating an outline. Interestingly, there is also emission from within the nucleus and the surrounding cytoplasm confirming the presence of galectins inside the cell. There is no fluorescence from the cellular surface or indeed from the extracellular matrix (ECM), as the cell membranes are not defined. This is somewhat disappointing as literature suggests that the majority of galectin-3 should be located extra-cellularly. However, although definitely expressing galectin-3, there is no report of the SHEP-1 cell line having ever been used for galectin-3 targeting studies and so comparisons cannot be drawn.
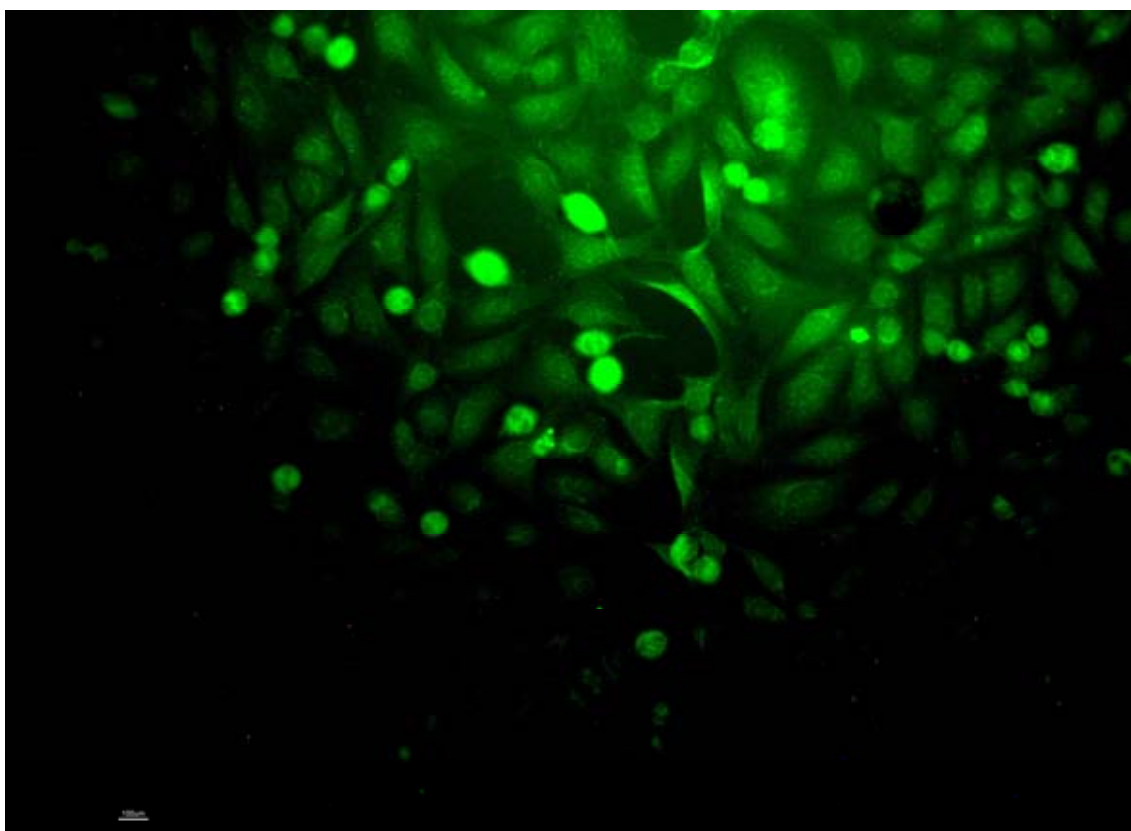
Figure 117: SHEP-1 neuroblastoma cells stained with probe **87**. Again cells of a brighter green are in the process of dividing. A greater degree of cytoplasm staining is shown on this slide.

It was found that the cells must be fixed and that permeabilising the cell membrane, in this case with methanol, allows the fluorescent probe to infiltrate the cytoplasm and bind to the glycoproteins therein expressing the β-galactoside moiety. This may also explain the reason why the cell membrane is not obvious in these images.

As a control it was again the ready availability and ease of cultivation that dictated our choice of cell line. SH-SY5Y was the negative control as it is reported to express very low levels of galectin-3 and should offer a good contrast in fluorescence intensity when

compared to the positive SHEP-1 cells under identical experimental conditions. It should also offer a measure of the binding specificity of the probe towards exclusively galectin-3. However, somewhat disconcertingly the SH-SY5Y cell line showed almost equivalent uptake and fluorescence intensity (**Figure 118**).
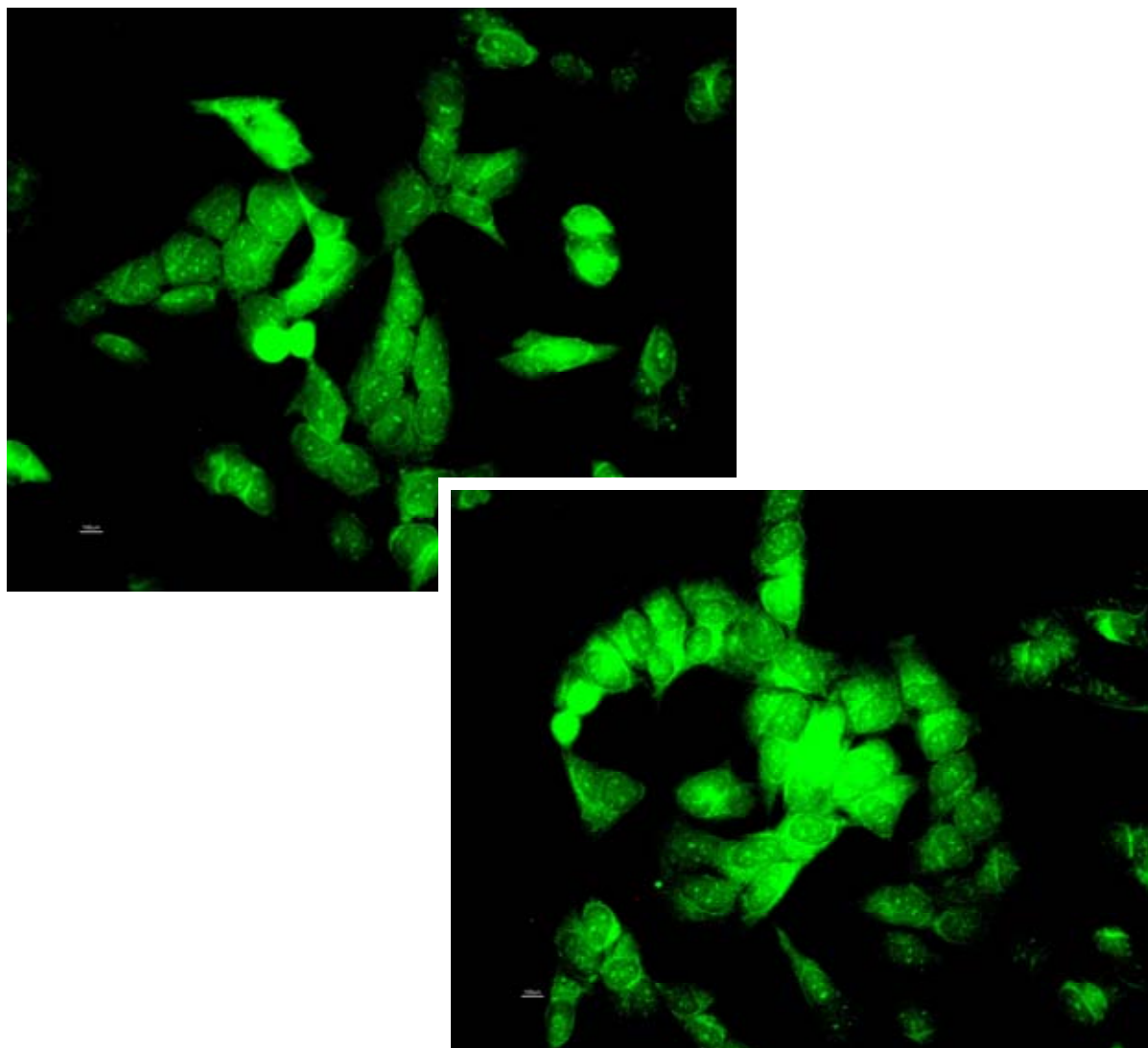
Figure 118: SH-SY5Y cells stained with **87**. DAPI nucleus staining confirmed the outlines of the cell nuclei.

The reason for this effect is likely to be the fact that both cell lines express galectin-1 in equal measure. The binding sites of galectin-1 and galectin-3 are very similar and hence it is plausible that the probes could also be binding to galectin-1.[157] Therefore any galectin-3 specific binding is being eclipsed by strong galectin-1 based emission.

Therefore our attempts to use the non-galectin-3 expressing cell line SH-SY5Y as a control could have been thwarted by the fact that the disaccharide vector of our targeted fluorescent probes were also binding to galectin-1.

Thus this study requires a cell line which expresses galectin-3 in the absence of galectin-1 as the positive; and a cell line expressing neither as the negative control.

The U87-MG cell line was acquired and the labelled disaccharides and cyclic peptides (U87-MG also expresses the $\alpha_v\beta_3$ integrin on the cell surface), are to be tested against these cells when they mature. This could also afford a great comparative study between cyclic peptide and β-galactoside targeting; chapters two and three of this thesis.

3.5 Summary & conclusions

The synthesis of the fluorescent labelled, fully deprotected, galectin-3 targeted probes involved a challenging 13 step reaction sequence.

The target was retrosynthetically broken down into two monosaccharides which would be coupled together in a regioselective β(1'→4) glycosylation reaction to form the required galectin-3 targeted β-galactoside. The challenge was then to develop an effective means of extending a spacer from the sugar to a distant fluorophore without interfering with the galectin carbohydrate recognition domain.

The synthesis of the D-glucosamine derived glycosyl acceptor started off on the wrong foot with the failure of the tandem acetylation and TCP protection step. This was rectified through employing an alternative route utilising a temporary amine protection with p-anisaldehyde to afford the imine followed by acetylation, hydrolysis and subsequent TCP protection. Although a lengthier reactive sequence this proved a very efficient alternative route with all yields higher than 93 %.

Our synthetic design centred around derivatisation of the acceptor C-2 amino position. It is clear from the crystal structures of reported N-acetyllactosamine derivatives that it is only this amine and the anomeric C-1 position that protrude from the binding site. Therefore the anomeric position was sealed off in the next step to prevent any further involvement in the synthesis. This involved the formation of a β-methyl glycoside which again proved tricky. The silver tetrafluoroborate system which was eventually implemented and optimised proved excellent. It proved not only high yielding, but also simultaneously hydrolysed the acetals which was anticipated to be a further step; two steps were thus achieved in one. Selective acetylation of the C-6 position then went smoothly with careful temperature control to afford the glycosyl acceptor in good yield.

The thio-glycosidic donor was synthesised straightforwardly from D-galactose following some elegant reported chemistry. The subsequent regioselective formation of the β(1'→4) galactoside was accomplished with great efficacy. The reaction sequence was then subject to some scrutiny as it became clear that after TCP cleavage, deacetylation at any stage renders the compound difficult to purify. Hence the amino caproic or ethylene glycol spacer was coupled to the amine, the terminus deprotected and the fluorophore coupled, prior to a final deacetylation to afford the products.

Initial cellular binding and microscopy studies with the fluorescein labelled β-galactoside revealed a promising uptake and emission from a SHEP-1 neuroblastoma cell line. However the negative control cell line also showed considerable fluorescence. As both cell lines express galectin-1 and literature suggests that our β-galactoside may well bind to both galectin-3 and galectin-1 this may well explain the observations. Therefore studies are continuing with a U87-MG glioblastoma cell line known to express exclusively galectin-3 and it is hoped that uptake and binding affinity can be measured accurately in due course.

3.6 Sugars experimental

Preparation of thioglycoside **70**.
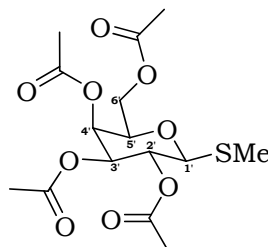
$C_{15}H_{22}O_9S$
Exact Mass: 378.1
Mol. Wt.: 378.39

D-galactose pentaacetate (1.00 g, 2.56 mmol) was taken up in dry CH_2Cl_2 (20 mL). Iodine (0.91 g, 3.59 mmol) and triethylsilane (0.58 mL, 0.42 g, 3.59 mmol) were added slowly (exothermic), and the mixture heated at reflux for five minutes. TLC analysis (hexane / EtOAc, 6 : 4) revealed the complete consumption of starting material and the reaction was allowed to cool to rt. The crude mixture was then diluted with CH_2Cl_2 (40 mL) and extracted with an aqueous mix of Na_2CO_3 and $Na_2S_2O_3$ (3 x 100 mL) and water (2 x 50 mL). The crude iodide was then dried over $MgSO_4$ and concentrated under reduced pressure. R_f = 0.60 (hexane / EtOAc, 6 : 4).

Dry crude iodide was then taken up in MeCN (10 mL), treated with thiourea (0.29 g, 3.84 mmol) and heated at 60 °C for 30 minutes until TLC analysis (hexane / EtOAc, 6 : 4) revealed a new baseline product and the disappearance of the iodide. The reaction was allowed to cool to rt and methyl iodide (0.31 mL, 0.71g, 5.12 mmol) and triethylamine (1.43 mL, 1.04 g, 0.01mol) were added sequentially. The mixture was then stirred at ambient temperature for 10 minutes until TLC analysis (hexane / EtOAc, 6 : 4) proved reaction completion. The crude product was concentrated under reduced pressure and purified by flash column chromatography (hexane / EtOAc, 3 : 7) to afford the product as a fine white solid (1.00 g, 98 %). R_f = 0.41; $[\alpha]^{22}_D$ = +2.8 (*c* 1.0 CHCl₃), *lit.*[147] +3.0 (*c* 1.0 CHCl₃) m.p = 110-112 °C, *lit.*[147] 112-113 °C; ν_{max} (Neat)/ cm⁻¹ 1701, 1733 (C=O); δ_H (300 MHz; CDCl₃); 5.42 (1H, dd, J 3.3, 6.7, H-4') 5.25 (1H, app t, J 9.9, H-2'), 5.04 (1H, dd, J 3.3, 9.9, H-3'), 4.38 (1H, d, J 9.9, H-1'), 4.07-4.16 (2H, stack, 2 x H-6'), 3.95 (1H, app t, J 6.7 H-5'), 2.18 (3H, s, SCH₃), 2.14(3H, s, OCOC*H*₃), 2.06(3H, s, OCOC*H*₃), 2.03(3H, s, OCOC*H*₃), 1.97(3H, s, OCOC*H*₃); δ_C (75 MHz; CDCl₃); 9.87, 19.13, 19.20, 19.32, 60.05, 65.11, 65.96, 70.42, 72.97, 81.86, 168.16, 168.55, 168.73, 168.90; *m/z* (TOF ES+) 401.0882 ([M+Na]⁺, 100% $C_{15}H_{22}O_9SNa$ requires 401.0873).

Preparation of imine **71**.
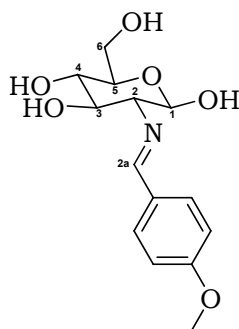
$C_{14}H_{19}NO_6$
Exact Mass: 297.12
Mol. Wt.: 297.3

D-Glucosamine hydrochloride (20.0 g, 0.093 mol) was dissolved in 1M aqueous sodium hydroxide (96 mL). *p*-Anisaldehyde (12.65 mL, 0.11 mmol) was then added dropwise to the mixture with vigorous stirring, forming a cloudy solution. After 15 minutes of stirring at ambient temperature, a thick white precipitate formed. Stirring was continued in an ice bath for one hour to ensure complete precipitation. The solid was then collected by Büchner filtration and washed with H_2O (2 x 100 mL) and a 1:1 mixture of MeOH and Et_2O (2 x 100 mL). The filter cake was dried further under high vacuum overnight, affording the product as a fine white solid. Analysis revealed the product to be sufficiently pure to be taken on without further purification (18.00 g, 66 %). $[\alpha]^{22}_D$ = +24.3 (*c* 0.5 D_2O); m.p = 158-160 ˚C (*lit.*[146] = 148-150 ˚C); ν_{max} (Neat)/cm^{-1} 3486, 3322 (br-OH), 2972-2897 (CH), 1638 (C=N), 1604 (Ar C=C); δ_H (400 MHz; d$_6$DMSO) 2.79 (1H, t, J 8.5, *H*-2), 3.10-3.18 (1H, m, *H*-4), 3.21-3.38 (1H, m, *H*-5), 3.41-3.46 (1H, m, *H*-3), 3.48-3.52 (1H, m, *H*-6), 3.71-3.79 (1H, m, *H*-6), 3.86 (3H, s, OC*H*$_3$), 4.57 (1H, t, J 5.7, O*H*-6), 4.69 (1H, t, J 7.2, *H*-1), 4.84 (1H, d, J 5.6, O*H*-3), 4.95 (1H, d, J 5.1, O*H*-4), 6.54 (1H, d, J 6.7, O*H*-1), 6.99 (2H, d, J 8.7, *H* aromatic), 7.69 (2H, d, J 8.7, *H* aromatic), 8.11 (1H, s, OC*H*$_3$); δ_C (75 MHz; DMSO-d$_6$) 54.91 (O*C*H$_3$), 60.92 (*C*-6). 70.01 (*C*-4), 74.24 (*C*-3), 76.49 (*C*-5), 77.83 (C-2), 95.27 (C-1), 113.54 (aromatic *C*H), 128.75 (aromatic *C*$_q$), 129.26 (aromatic *C*H), 160.69 (aromatic *C*OMe), 160.87(H*C*=N); m/z (TOF ES+) 320.1112 ([M+Na]$^+$, 100%, $C_{14}H_{19}NO_6Na$ requires 320.1110).

Per-Acetylation to afford tetra acetate **72**.
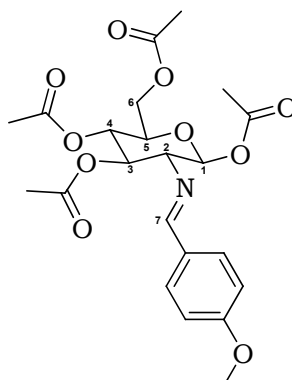
C$_{22}$H$_{27}$NO$_{10}$
Exact Mass: 465.16
Mol. Wt.: 465.45

Compound **71** (18.00 g, 0.06 mol) was combined with acetic anhydride (54.27 mL, 0.57 mol), pyridine (97.9 mL, 1.21 mol) and DMAP (1.8 g) and stirred at 0 °C. The solid slowly went into solution and the reaction mixture was stirred at ambient temperature overnight. The solution was then poured onto 300 mL of ice resulting in the formation of a thick white crystalline precipitate. These crystals were isolated by Büchner filtration, washed with water (2 x 50 mL) and ether (2 x 50 mL) and dried under vacuum to afford the product as a fine white solid which analysis revealed to be sufficiently pure to be taken on without further purification (19.54 g, 70 %). $[\alpha]^{22}_D$ = +19.4; m.p = 180-183 °C, (*lit.*[146], 168-172 °C); ν_{max} (Neat)/cm^{-1} 2977-2868 (C-H), 1749 (C=O), 1737 (C=O), 1637 (C=N), 1605 (Ar C=C); δ_H (300 MHz; CDCl$_3$); 1.84 (3H, s, COC*H*$_3$) 1.98 (3H, s, OCOC*H*$_3$), 1.99 (3H, s, OCOC*H*$_3$), 2.06 (3H, s, OCOC*H*$_3$), 3.41 (1H, t, *J* 8.5, *H*-2), 3.80 (3H, s, OC*H*$_3$), 3.91 (1H, ddd *J* 2.0, 4.6, 9.5, *H*-5), 4.08 (1H, s, *J* 2.0, 12.4, *H*-6), 4.34 (1H, dd, *J* 4.6, 12.4, *H*-6), 5.10 (1H, app t, *J* 9.5, *H*-4), 5.39 (1H, app t, J 9.5, *H*-3), 5.90 (1H, d, *J* 8.5, *H*-1), 6.87 (2H, d, *J* 8.7, *H* aromatic), 7.61 (2H, d, *J* 8.7, *H* aromatic), 8.12 (1H, s, C*H*=N); δ_C (100 MHz; CDCl$_3$); 23.23 (*C*H$_3$), 23.40 (*C*H$_3$), 23.48 (*C*H$_3$), 58.12 (O*C*H$_3$), 64.56 (*C*-6), 70.79 (*C*-4), 75.50 (*C*-2/5), 75.68 (*C*-2/5), 75.99 (*C*-3), 95.90 (*C*-1), 116.79 (aromatic *C*H), 131.05 (aromatic *C$_q$*), 132.97 (aromatic *C*H), 162.2 (aromatic *C*OMe), 166.97 (*C*-7), 168.7(*C*=O), 169.5 (C=O), 169.8 (*C*=O), 170.6 (*C*=O); m/z (TOF ES+) 488.1527 ([M+Na]$^+$, 100%, C$_{22}$H$_{27}$NO$_{10}$Na requires 488.1533).

Preparation of per-acetylated D-glucosamine hydrochloride salt **73**.
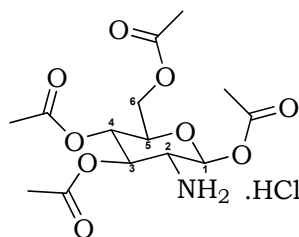
C$_{14}$H$_{21}$NO$_9$
Exact Mass: 347.12
Mol. Wt.: 347.32

Imine **72** (10.00 g, 0.021 mol) was dissolved in acetone (50 mL) and heated to reflux. To this solution was added 5M HCl (5 mL) dropwise and after five minutes a white thick precipitate formed. The system was then cooled to room temperature and the precipitate collected through Büchner filtration. The filter cake washed thoroughly with acetone (2 x 20 mL), and ether (2 x 50 mL) and the product dried further under high vacuum overnight to afford the product as a fine white solid. Analysis revealed the product to be sufficiently pure to be taken on without further purification (7.45 g, 98 %). [α]^{22}$_D$ = +33.2 (*c* 1.0 CHCl$_3$); m.p = >250 ˚C (decomposed at 215-218 ˚C), (*lit.*[146], >200 ˚C); ν$_{max}$ (Neat)/cm^{-1} 2830 (br N-H), 1757 (C=O), 1746 (C=O); δ$_H$ (300MHz; DMSO-d$_6$); 1.97 (3H, s, OCOC*H*$_3$), 2.00 (3H, s, OCOC*H*$_3$), 2.03 (3H, s, OCOC*H*$_3$), 2.17 (3H, s, OCOC*H*$_3$), 3.57 (1H, dd, J 8.8, 10.2, *H*-2), 3.96-4.09 (2H, stack, *H*-5 & *H*-6), 4.19 (1H, dd, J 4.3, 12.5, *H*-6), 4.93 (1H, app t, J 9.5, *H*-4), 5.34 (1H, dd, J 9.5, 10.2, *H*-3), 5.89 (1H, d, J 8.8, *H*-1), 8.70 (3H, br s, N*H*$_3$); δ$_C$ (75 MHz; DMSO-d$_6$); 20.94 (*C*H$_3$), 21.07 (*C*H$_3$), 21.48 (*C*H$_3$), 21.57 (*C*H$_3$), 52.77 (*C*-4), 61.86 (*C*-6), 68.41 (*C*-3), 70.92 (*C*-5). 72.19 (*C*-2), 90.66 (*C*-1), 169.22 (*C*=O), 169.90 (*C*=O), 170.33 (*C*=O), 170.56 (*C*=O); m/z (TOF ES+) 370.1110 ([M + Na]$^+$, 100%, C$_{14}$H$_{21}$NO$_9$Na requires 370.1114).

TCP protection to afford tetraacetate **74**.
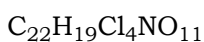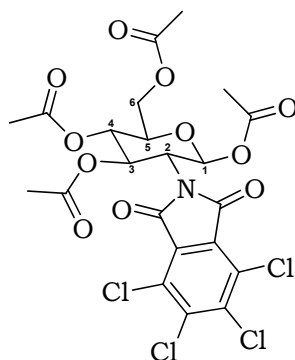
$C_{22}H_{19}Cl_4NO_{11}$
Exact Mass (4 x ^{35}Cl) : 612.97
Mol. Wt.: 615.2

Tetraacetate **73** (10.00 g, 0.026 mol) was dissolved in pyridine (50 mL). Tetrachlorophthalic anhydride (8.86 g, 0.030 mol) was then added and the mixture stirred at ambient temperature for 10 hours, cooled to 0 ˚C and Ac_2O (2.95 mL, 0.030 mol) added. The mixture was then allowed to warm to rt and stirred for a further hour. Concentration under reduced pressure was then performed with repeated co-concentrations with toluene. The crude product was purified by gradient flash column chromatography (hexane to EtOAc / hexane, 1 : 1) to afford the product as a fine white solid (7.20 g, 78 %). $[\alpha]^{23}_D$ = +58.3 (c 1.1 $CHCl_3$), lit.[145] +69.7 (c 0.9 $CHCl_3$); m.p = 165-167 ˚C, (lit.[145], 169-171˚C); ν_{max} (Neat)/cm^{-1} 2971 (C-H), 1749 (C=O), 1721 (C=O); δ_H (300 MHz; $CDCl_3$) 1.88 (3H, s, OCOCH_3), 2.00 (3H, s, OCOCH_3), 2.01 (3H, s, OCOCH_3), 2.09 (3H, s, OCOCH_3), 3.95 (1H, ddd, J 4.4, 9.0, 10.1, H-5), 4.11 (1H, dd, J 9.0, 12.5, H-2), 4.35 (1H, dd, J 4.4, 12.5, H-6), 4.25 (1H, dd, J 9.0, 10.1, H-6), 5.21 (1H, dd, J 9.0, 10.1, H-4), 5.77 (1H, dd, J 9.0, 10.1, H-3), 6.44 (1H, d, J 9.0, H-1); δ_C (75 MHz; $CDCl_3$); 22.25 (CH_3), 22.37 (CH_3), 22.53 (CH_3), 22.63 (CH_3), 56.13 (C-2), 63.22 (C-6), 69.72 (C-3), 72.51 (C-4), 74.44 (C-5), 91.44 (C-1), 128.63 (C$_q$),131.40 (C$_q$), 141.53 (phthalimide C=O) 170.42 (C=O), 171.14 (C=O), 172.29 (C=O), 172.40 (C=O), one C$_q$ not visible; m/z (TOF ES+) 635.9610 ([M+Na]$^+$, 100%, $C_{22}H_{19}NO_{11}^{35}Cl_4Na$ requires 635.9614).

Preparation of methyl glycoside **75**.
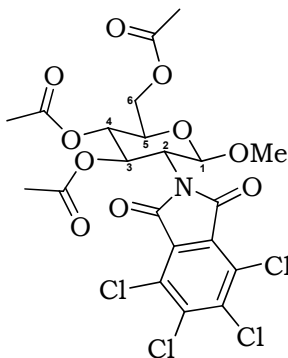
C$_{21}$H$_{19}$Cl$_4$NO$_{10}$
Exact Mass (4 x ^{35}Cl): 584.98
Mol. Wt.: 587.19

Tetraacetate **74** (0.20 g, 0.33 mmol) was combined with HBr in acetic acid (0.53 mL, 32 wt. %), and Ac$_2$O (0.08 mL, 0.88 mmol) and stirred in darkness at ambient temperature overnight. The reaction mixture was then diluted with CHCl$_3$ (10 mL) and poured onto ice (15 mL). Subsequent to the ice melting, the layers were separated and the organic portion washed with H$_2$O (2 x 15 mL) and NaHCO$_{3(sat.)}$ (2 x 10 mL), dried over MgSO$_4$ and concentrated under reduced pressure to afford the intermediate as a syrup. To this crude glycosyl bromide was promptly added anhydrous MeOH (0.2 mL, 4.93 mmol) and silver tetrafluoroborate (0.06 g, 0.32 mmol) in anhydrous CH$_2$Cl$_2$ (10 mL). The reaction was stirred in darkness for 36 hours at reflux and was then allowed to cool before filtration through a pad of Celite. Concentration of the filtrate under reduced pressure and purification by gradient flash column chromatography (CH$_2$Cl$_2$ to CH$_2$Cl$_2$ / MeOH, 9 : 1), afforded the product as a fine white solid (0.18 g, 92 %). R$_f$ = 0.48 (CH$_2$Cl$_2$ / MeOH, 9 : 1); [α]^{22}$_D$ = +32.2 (*c* 1.10 CHCl$_3$); m.p = 159-161 ˚C; ν$_{max}$ (Neat)/ cm^{-1} 1748 (C=O), 1721 (C=O); δ$_H$ (300MHz; CDCl$_3$) 1.87 (3H, s, OCOC*H$_3$*), 2.01 (3H, s, OCOC*H$_3$*), 2.09 (3H, s, OCOC*H$_3$*), 3.43 (3H, s, OC*H$_3$*), 3.79-3.84 (1H, m, H-2), 4.05-4.35 (3H, stack, H-5, 2 x H-6), 5.17 (1H, app t, J 9.7, *H*-4), 5.24 (1H, d, J 3.0, *H*-1), 5.68 (1H, app t, J 9.7, *H*-3); δ$_C$ (75 MHz; CDCl$_3$) 20.5 (*C*H$_3$), 20.6 (*C*H$_3$), 20.7 (*C*H$_3$), 21.0 (*C*H$_3$), 55.3 (*C*-2), 57.0 (O*C*H$_3$), 61.9 (*C*-6), 68.7 (*C*-3), 70.9 (*C*-4), 71.9 (*C*-5), 98.6 (*C*-1), 127.0 (*C*$_q$), 130.0 (*C*$_q$), 140.5 (phthalimide *C*=O), 169.3 (*C*=O), 170.5 (*C*=O), 170.6 (*C*=O), one C$_q$ was not visible; m/z (TOF ES+) 609.9645 ([M+Na]$^+$, 100%, C$_{21}$H$_{19}$NO$_{10}$^{35}Cl$_3$^{37}Cl requires 609.9631).

Deacetylation to afford triol **76**.
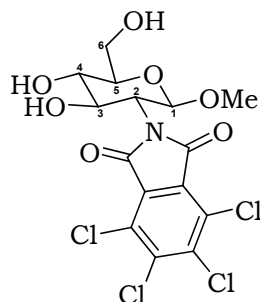
$C_{15}H_{13}Cl_4NO_7$
Exact Mass (4 x ^{35}Cl): 458.94
Mol. Wt.: 461.08

Triacetate **75** (0.15 g 0.25 mmol) was dissolved in anhydrous MeOH (10 mL) with stirring at ambient temperature under argon. Sodium metal (0.01 g) was then added and the reaction stirred for a further 15 minutes. Thin layer chromatography (hexane / EtOAc, 6.5 : 3.5) indicated reaction completion with a single strong baseline spot and the disappearance of the starting material. DOWEX-50WX8-200 ion exchange resin (pre-washed with 5M HCl then MeOH) was then added to the reaction mixture with stirring until the pH = 5. The crude mixture was filtered and concentrated under reduced pressure to afford the product as a fine white solid which analysis revealed to be sufficiently pure to be taken on without further purification (0.11 g, 99 %). $[\alpha]^{22}_D$ = +7.7 (*c* 1.0 CHCl$_3$), *lit.*[147] +3.0 (*c* 1.0 CHCl$_3$); m.p = 159-161 ˚C, no lit. m.p found; ν_{max} (Neat)/ cm^{-1} 3363 (OH), 1750 (C=O); δ_H (300 MHz; CDCl$_3$); 5.05 (1H, d, J 8.4, H-1), 4.20 (1H, dd, J 8.4, 11.3, H-6), 3.85-4.00 (2H, stack, H-6, H-3), 3.75 (1H, dd, J 5.10, 11.3, H-4), 3.38-3.45 (5H, stack, OCH$_3$, H-2, H-5); δ_C (75 MHz; CDCl$_3$) 56.0, 57.9, 60.7, 70.3, 70.4, 77.3, 98.3, 127.3, 127.6, 128.2, 148.7; m/z (TOF ES+) 481.9344 ([M+Na]$^+$, 100%, $C_{15}H_{13}NO_7{}^{35}Cl_4Na$ requires 481.9359).

Mono-acetylation to afford compound **77**.
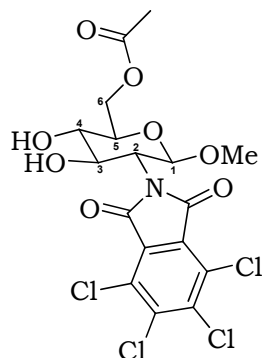
C$_{17}$H$_{15}$Cl$_4$NO$_8$
Exact Mass (4 x ^{35}Cl): 500.96
Mol. Wt.: 503.11

Triol **76** (0.20 g, 0.43 mmol) was taken up in dry CH$_2$Cl$_2$ (10 mL) and cooled to -42 °C by means of a CO$_{2(s)}$ / MeCN bath whilst stirring under argon. Sym-collidine (2, 4, 6-trimethylpyridine) (0.29 mL, 2.17 mmol) was then added to the reaction vessel followed by dropwise addition of acetyl chloride (0.034 mL, 0.48 mmol) over 10 minutes. The mixture was stirred at -42 °C for 3 hours and monitored by TLC. Further acetyl chloride (8.48 µL, 0.12 mmol) and sym collidine (0.13 mL, 0.95 mmol) were then added and the mixture allowed to warm to -20 °C over one hour. The reaction was then quenched with MeOH (10 mL) and diluted with CH$_2$Cl$_2$ (20 mL) prior to washing with 0.5 M HCl (2 x 30 mL) and saturated NaHCO$_3$ (2 x 30 mL). The combined organic phases were dried over MgSO$_4$ and concentrated under reduced pressure. The crude product was purified by gradient flash column chromatography (CH$_2$Cl$_2$ to CH$_2$Cl$_2$ / MeOH, 9 : 1), to afford the product as a fine white solid (3.97 g, 88 %). R$_f$ = 0.66 (CH$_2$Cl$_2$ / MeOH, 9 : 1); [α]$^{22}_D$ = -16.7 (*c* 1.0 CHCl$_3$), *lit.*[144] -18.4 (*c* 1.0 CHCl$_3$); m.p = 135-137 ˚C, no lit. m.p found; ν$_{max}$ (Neat)/cm^{-1} 3407 (OH), 1781 (C=O), 1718 (C=O); δ$_H$ (300 MHz; CDCl$_3$) 5.01 (1H, d, J 8.5, *H*-1), 4.35 (1H, dd, J 2.4, 12.0, *H*-6), 4.27 (1H, dd, J 5.2, 12.0, *H*-6), 4.14 (1H, dd, J 8.5, 12.0, *H*-3) 3.88 (1H, dd, J 8.5, 10.5, *H*-4), 3.60 (1H, ddd, J 2.4, 5.2, 10.5, *H*-5), 3.35-3.45 (4H, stack, OC*H*$_3$ & *H*-2), 2.06 (3H, s, OCOC*H*$_3$); δ$_C$ (75 MHz; CDCl$_3$) 19.7, 55.8, 57.0, 62.7, 70.2, 70.8, 73.6, 98.4, 117.0, 127.3, 139.2, 170.5; m/z (TOF ES+) 525.9 (1 x ^{37}Cl, 3 x ^{35}Cl isotope, [M+Na]$^+$, 100%). 523.9449 ([M+Na]$^+$, 100%, C$_{17}$H$_{15}$NO$_8$^{35}Cl$_4$Na requires 523.9453).

Glycosylation to afford disaccharide: 1-Methoxy-O-(2',3',4',6'-tetra-O-acetyl-β-D-galactopyranosyl)-(1'->4)-2-tetrachlorophthalimido-6-O-acetyl-3-deoxy-β-D-glucopyranoside **78**.

$C_{31}H_{33}Cl_4NO_{17}$
Exact Mass (4 x ^{35}Cl): 831.05
Mol. Wt.: 833.4
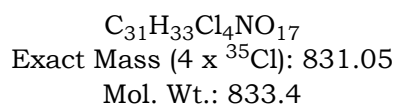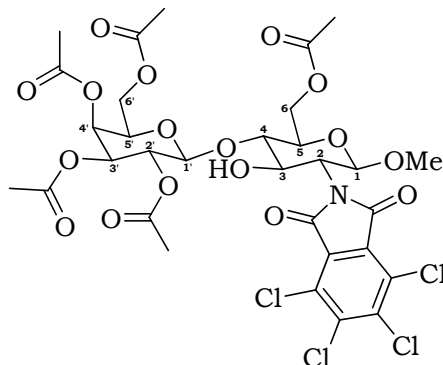
Diol **77** (0.20 g, 0.40 mmol) and thioglycoside **70** (0.17 g, 0.44 mmol) were combined in dry CH$_2$Cl$_2$ (15 mL) with pre-activated 3 Å molecular sieves (1.0 g). This mixture was stirred at ambient temperature under argon for 30 minutes before cooling to -42 °C. Catalytic triflic acid (5.27 µL, 0.06 mmol) and N-iodosuccinimide (0.13 g, 0.59 mmol) were then added to the reaction vessel and after 2 hours at -42 °C the crude cocktail was allowed to warm to room temperature, filtered, diluted with CH$_2$Cl$_2$ (35 mL) and washed with 10 % aqueous Na$_2$S$_2$O$_3$ (2 x 25 mL). Subsequent drying with MgSO$_4$ and removal of the solvent under reduced pressure revealed the crude product as a pale brown oil. Purification by flash column chromatography yielded pure product as a fine white powder (0.24 g, 72 %). R$_f$ = 0.50 (hexane / EtOAc, 3 : 2); $[\alpha]^{22}_D$ = +9.1 (c 1.0 CHCl$_3$), lit.[144] +7.6 (c 1.0 CHCl$_3$); mp = 204-206 ˚C; δ$_H$ (300 MHz; CDCl$_3$) 5.32 (1H, d, J 3.3, H-4'), 5.18 (1H, dd, J 8.3, 10.3, H-2'), 5.07 (1H, d, J 8.3, H-1), 4.97 (1H, dd, J 3.3, 10.3, H-3'), 4.54 (1H, d, J 8.3, H-1'), 4.34-4.26 (3H, stack, H-3, 2 x H-6'), 4.10-3.98 (4H, stack, H-4, 2 x H-6, H-5'), 3.68 (1H, dd, J 3.3, 9.7, H-5), 3.54 (1H, app t, J 8.3, H-2), 3.38 (3H, s, OCH$_3$), 2.09 (3H, s, OCOCH$_3$), 2.08 (3H, s, OCOCH$_3$), 2.04 (6H, s, 2 x OCOCH$_3$), 1.92 (3H, s, OCOCH$_3$); δ$_C$ (75MHz; CDCl$_3$); 20.3 (OCOCH$_3$), 20.4 (OCOCH$_3$), 20.5 (OCOCH$_3$), 20.8 (OCOCH$_3$), 21.0 (OCOCH$_3$), 56.3 (C-4 or C-5'), 56.8 (OCH$_3$), 61.6 (C-6), 62.6 (C-6'), 66.8 (C-4'), 67.2 (C-2'), 67.8 (C-3), 68.7 (C-3'), 69.5 (C-4 or C-5'), 70.6 (C-5), 83.2 (C-2), 98.8 (C-1), 101.9 (C-1'), 127.3 (C$_q$), 169.4 (C=O), 169.7 (C=O), 169.9 (C=O), 170.4 (C=O), 170.6 (C=O), two C$_q$ signals were not visible; m/z (TOF ES+) 854.0407 ([M+Na]$^+$, 100% C$_{31}$H$_{33}$NO$_{17}$Na ^{35}Cl$_4$ requires 854.0400).

See **appendix 6** for COSY 90 and HSQC 2D-nmr experiments confirming the assignments.

TCP cleavage to afford acetylated lactosamine: 1-Methoxy-O-(2',3',4',6'-tetra-O-acetyl-β-D-galactopyranosyl)-(1'->4)-2-amino-6-O-acetyl-3-deoxy-β-D-glucopyranoside, **79**.

$C_{23}H_{35}NO_{15}$
Exact Mass: 565.2
Mol. Wt.: 565.52
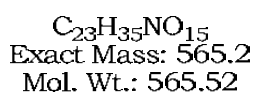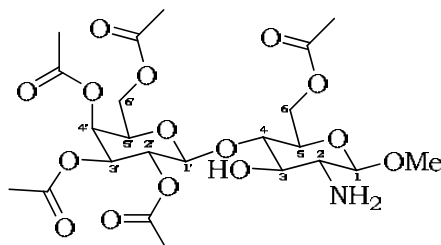
Disaccharide **78** (0.20 g, 0.24 mmol) was combined with ethylenediamine (0.032 mL, 0.48 mmol) in the three solvent blend MeCN / THF / EtOH (2 : 1 : 1, 12 mL) and stirred at 60 °C for 12 hours. The reaction was monitored by thin layer chromatography (CH_2Cl_2 / MeOH, 9.7 : 0.3). Upon disappearance of the TCP protected starting material **77** (R_f = 0.6), the solvents were removed under reduced pressure and the crude concoction purified by flash column chromatography (9.7 CH_2Cl_2 : 0.3 MeOH). To afford the product as a pale yellow film (0.125 g, 92 %). R_f = 0.10; $[\alpha]^{22}_D$ = +14.2 (c 1.0 $CHCl_3$); δ_H (500 MHz, $CDCl_3$) 5.33 (1H, d, J 3.3, *H*-4'), 5.17 (1H, dd, J 8.0, 10.4, *H*-2') 4.93 (1H, dd, J 3.3, 10.4, *H*-3') 4.48 (1H, d, J 8.0, *H*-1'), 4.25 (1H, dd, J 1.7, 12.8, *H*-6'), 3.95-4.15 (6H, stack, *H*-1, *H*-5, 2 x *H*-6 & 2 x *H*-6'), 3.46-3.56 (5H, stack, OC*H*₃, H-3, *H*-5'), 3.41 (1H, dd, J 8.3, 9.4, *H*-4), 2.71 (1H, dd, J 8.3, 9.4, *H*-2), 2.4 (2H, br, N*H*₂), 2.10 (3H, s, OCOC*H*₃), 2.03 (3H, s, OCOC*H*₃), 2.02 (3H, s, OCOC*H*₃), 2.00 (3H, s, OCOC*H*₃), 1.91 (3H, s, OCOC*H*₃); δ_C (75 MHz, $CDCl_3$) 170.7 (*C*=O), 170.4 (*C*=O), 170.0 (*C*=O), 169.9 (*C*=O), 169.5 (*C*=O), 104.1 (*C*-1), 102.0 (*C*-1'), 82.6 (C-4), 74.3 (C-5'), 71.9 (*C*-3), 71.4 (*C*-5), 70.9 (*C*-3'), 68.8, 66.9 (*C*-4'), 62.9 (*C*-6'), 61.8 (*C*-6), 57.3 (O*C*H₃), 56.6 (*C*-2), 20.8 (OCO*C*H₃), 20.6 (OCO*C*H₃), 20.4 (OCO*C*H₃), only three seen; *m/z* (TOF ES+) 588.1927 ([M+Na]+, 100% $C_{23}H_{35}NO_{15}Na$ requires 588.1904).

See **appendix 6** for COSY 90, HSQC and HMBC 2D-nmr experiments confirmind the assignments.

Preparation of per-acetylated N-acetyllactosamine: 1-Methoxy-O-(2',3',4',6'-tetra-O-acetyl-β-D-galactopyranosyl)-(1'->4)-2-acetamido-3,6-di-O-acetyl-2-deoxy-β-D-glucopyranoside **80**.
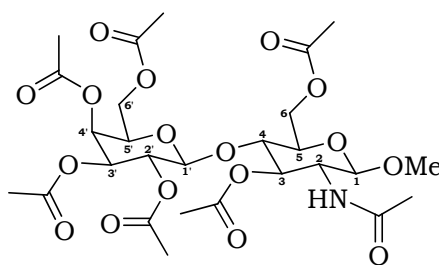
$C_{27}H_{39}NO_{17}$
Exact Mass: 649.22
Mol. Wt.: 649.60

Disaccharride **79** (0.05 g, 0.088 mmol) was dissolved in anhydrous pyridine (20 mL) at ambient temperature under argon. Acetic anhydride (0.017 mL, 0.18 mmol) was then added and the cocktail stirred overnight. Thin layer chromatography revealed the appearance of product at R_f = 0.63 (toluene / acetone, 3 : 7). Excess solvent was removed under reduced pressure and the crude product purified by flash column chromatography using the aforementioned eluent. The product was recovered as a fine white powder (0.049 g, 92 %). $[\alpha]^{22}_D$ = +18.2 (c 1.2 CHCl$_3$); m.p = 134-137 °C; δ_H (400 MHz, CDCl$_3$) 6.04 (1H, d, J 9.4, N*H*), 5.29 (1H, d, J 3.3, *H*-4'), 4.99-5.03 (2H, stack, *H*-2' & *H*-3), 4.92 (1H, dd, J 10.4, 3.3, *H*-3'), 4.41-4.47 (2H, stack, *H*-6 & *H*-1' (d, J 7.7)), 4.33 (1H, d, J 7.7, *H*-1), 4.04 (4H, stack, 2 x *H*-6', *H*-6 & *H*-2), 3.84 (1H, app t, J 3.3, *H*-5'), 3.73 (1H, app t, J 8.6, *H*-4), 3.57 (1H, m, *H*-5), 3.38 (3H, s, OC*H*$_3$), 2.08 (3H, s, OCOC*H*$_3$), 2.05 (3H, s, OCOC*H*$_3$), 2.01 (3H, s, OCOC*H*$_3$), 1.99 (6H, s, 2 x OCOC*H*$_3$), 1.90 (6H, s, 2 x OCOC*H*$_3$); δ_C (100 MHz, CDCl$_3$) 170.6 (*C*=O), 170.4 (*C*=O), 170.3 (*C*=O), 170.3 (*C*=O), 170.1 (*C*=O), 170.0 (*C*=O), 169.3 (*C*=O), 101.6 (*C*-1), 101.1 (*C*-1'), 75.9 (*C*-4), 72.7 (*C*-5 or *C*-3), 72.6 (*C*-5 or *C*-3), 70.8 (*C*-5' or *C*-3'), 70.7 (*C*-5' or *C*-3'), 69.1 (*C*-2'), 66.6 (*C*-4'), 62.3 (*C*-6), 60.8 (*C*-6'), 56.3 (O*C*H$_3$), 53.0 (*C*-2), 23.1 (OCO*C*H$_3$), 20.8 (2 x OCO*C*H$_3$), 20.5 (2 x OCO*C*H$_3$), 20.4 (2 x OCO*C*H$_3$); *m/z* (TOF ES+) 672.2116 ([M+Na]$^+$, 100% C$_{27}$H$_{39}$NO$_{17}$Na requires 672.2114).

See **appendix 6** for COSY 90, HSQC and HMBC nmr experiments confirming the assignments.

Preparation of O-(2',3',4',6'-tetra-O-acetyl-β-D-galactopyranosyl)-(1'->4)-2-acetamido-6-O-acetyl-3-deoxy-β-D-glucopyranoside **81**.
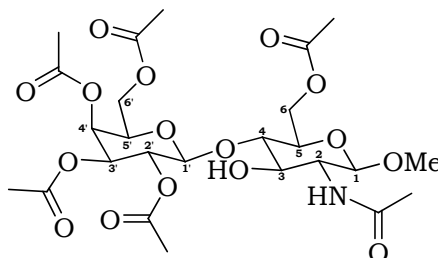
$C_{25}H_{37}NO_{16}$
Exact Mass: 607.21
Mol. Wt.: 607.56

Disaccharride **79** (0.04 g, 0.071 mmol) was dissolved in pyridine (15 mL) and stirred at ambient temperature. Acetic anhydride (0.067 mL, 0.071 mmol) was then added and the cocktail stirred for 12 hours. Upon reaction completion (according to thin layer chromatography) the crude mixture was concentrated under reduced pressure and the crude product purified by flash column chromatography (toluene / acetone, 4 : 6) to afford the product as a fine white powder (0.032 g, 81 %). R_f = 0.40; $[\alpha]^{22}_D$ = +12.6 (c 1.1 CHCl$_3$); m.p = 154-156 °C; ν_{max} (Neat)/ cm^{-1} 1739, 1667 (C=O); δ_H (400 MHz, CDCl$_3$) 5.63 (1H, d, J 8.0, N*H*), 5.39 (1H, d, J 3.3, *H*-4'), 5.22 (2H, dd, J 8.0, 10.4, *H*-2'), 5.00 (1H, dd, J 3.3, 10.4, *H*-3'), 4.67 (1H, d, J 8.0, *H*-1), 4.57 (1H, d, J 8.0, *H*-1'), 4.33 (2H, stack, *H*-6, *H*-6'), 4.35-3.98 (5H, stack, *H*-6, *H*-6', *H*-3, *H*-5' & O*H*), 3.61 (1H, stack, *H*-5), 3.51-3.42 (5H, stack, *H*-4, *H*-2), 3.45 (3H, s, OCH$_3$), 2.16 (3H, s, OCOC*H*$_3$), 2.10 (3H, s, OCOC*H*$_3$), 2.07 (3H, s, OCOC*H*$_3$), 2.06 (3H, s, OCOC*H*$_3$), 2.01 (3H, s, NHCOC*H*$_3$), 1.97 (3H, s, OCOC*H*$_3$); δ_C (100 MHz, CDCl$_3$) 170.7 (*C*=O), 170.5 (*C*=O), 170.1 (*C*=O), 170.0 (*C*=O), 169.6 (*C*=O), 164.5 (*C*=O), 101.8 (*C*-1'), 101.0 (*C*-1), 82.7 (*C*-4), 71.8 (*C*-5' & *C*-3), 71.3 (*C*-5), 70.8 (*C*-3'), 68.8 (*C*-2'), 66.8 (*C*-4'), 62.8 (*C*-6), 61.6 (*C*-6'), 56.8 (*C*-2 or OCH$_3$), 56.6 (C-2 or O*C*H$_3$), 23.7 (NHCO*C*H$_3$), 20.9 (OCO*C*H$_3$), 20.6 (2 x OCO*C*H$_3$), 20.5 (2 x OCO*C*H$_3$); m/z (TOF ES+) 630.2 ([M+Na]$^+$, 100 %), HRMS 630.2010 ([M+Na]$^+$, 100% $C_{25}H_{37}NO_{16}Na$ requires 630.2027).

See **appendix 6** for COSY 90 and HSQC 2D-nmr experiments confirming the assignments.

Deacetylation to afford blocking sugar 1-methoxy-N-acetyl-lactosamine **82**.

$C_{15}H_{27}NO_{11}$
Exact Mass: 397.16
Mol. Wt.: 397.37
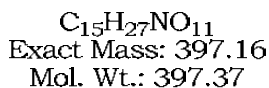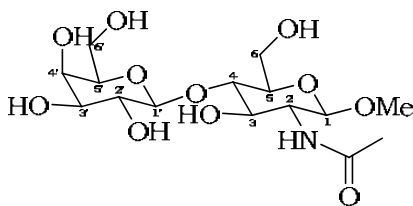
Per-acetylated disaccharide **80 (**or **81)** (0.04 g, 0.062 mmol) was dissolved in anhydrous MeOH (10 mL) with stirring at ambient temperature under argon. Sodium metal (0.005 g) was then added and the reaction stirred for a further 15 minutes. Thin layer chromatography (toluene / acetone, 3 : 7) indicated reaction completion with a single strong baseline spot and the disappearance of the starting material. DOWEX-50WX8-200 ion exchange resin (pre-washed with 5M HCl then MeOH) was then added to the reaction mixture with stirring until the pH = 5. The crude mixture was filtered and concentrated under reduced pressure to afford the product as a fine white solid which analysis revealed to be sufficiently pure to be taken on without further purification (0.023 g, 96 %). $[\alpha]^{22}_{D}$ = - 20.2 (*c* 0.5 H$_2$O); m.p = 220-223 ˚C, (*lit.*[158], 216-218 ˚C); δ_H (300 MHz, D$_2$O) 2.05 (3H, s, C*H*$_3$CO), 3.45 (3H, s, OC*H*$_3$), 3.48 (1H, m, *H*-5'), 3.58 (1H, m, *H*-5), 3.61-3.75 (3H, m, *H*-4, *H*-3 & *H*-2'), 3.70 (2H, m, *H*-6 & *H*-6'), 3.78 (1H, s, *H*-3'), 3.83 (2H, dd, J 12.2, 5.9, *H*-6 & *H*-6'), 3.93 (1H, d, J 1.9, *H*-4'), 4.37 (1H, d, J 5.9, *H*-1), 4.61 (1H, s, *H*-1'); δ_C (75 MHz, D$_2$O) 21.4 (CO*C*H$_3$), 53.8 (*C*-2), 55.9 (O*C*H$_3$), 59.3 (*C*-6'), 60.1 (*C*-6), 67.6 (*C*-2'), 68.0 (*C*-3'), 69.8 (*C*-4'), 72.0 (*C*-3), 73.9 (*C*-5'), 75.4 (*C*-5), 78.9 (*C*-4), 100.1 (*C*-1'), 101.3 (*C*-1); *m/z* (TOF ES+) 420.1482 ([M+Na]$^+$, 100% $C_{15}H_{27}NO_{11}Na$ requires 420.1477).

Coupling of Boc-aminocaproic acid to form disaccharide **83**.
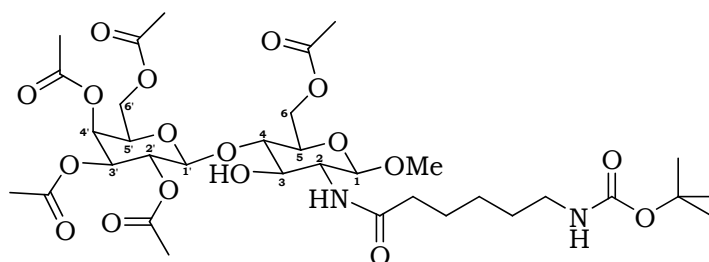
$C_{34}H_{54}N_2O_{18}$
Exact Mass: 778.34
Mol. Wt.: 778.8

tert-Butoxycarbonyl aminocaproic acid (0.12 g, 0.52 mmol) was dissolved in anhydrous MeCN (10 mL) and stirred under argon at ambient temperature. PyBOP (0.26 g, 0.52 mmol) and DIPEA (0.18 mL, 1.04 mmol) were added and the reaction stirred for 30 minutes. This mixture was then added to TCP cleaved disaccharide **79** (0.15 g, 0.25 mmol) in anhydrous MeCN (5 mL). The mixture was then stirred at reflux for 24 hours and TLC clearly showed the appearance of product R_f = 0.62 (toluene / acetone, 2 : 3). Concentration under reduced pressure with repeated co-concentration with toluene afforded the crude product as a brown oil which was purified by flash column chromatography afford the pure product as a transparent glassy solid. (0.16 g, 82 %). $[\alpha]^{22}_D$ = +10.3 (*c* 1.1 CHCl₃); ν_{max} (Neat)/ cm⁻¹ ; δ_H (400 MHz, CDCl₃) 5.98 (1H, br s, N*H*), 5.38 (1H, d, J 3.3, *H*-4'), 5.21 (1H, dd, J 8.1, 10.4, *H*-2'), 5.01 (1H, dd, J 10.4, 3.3, *H*-3'), 4.67 (1H, d, J 8.1, *H*-1), 4.58 (1H, d, J 8.1, *H*-1'), 4.33 (1H, dd, J 1.6, 10.0, *H*-6), 4.16 - 3.98 (5H, stack, 2 x *H*-6', *H*-5', *H*-3, *H*-6), 3.60 (1H, dd, J 3.3, 10.0, *H*-5), 3.50 (4H, stack, *H*-2 & OC*H*₃), 3.10 (3H, stack, *H*-4 & alkyl C*H*₂), 2.33 (1H, app t, J 7.4, alkyl C*H*₂), 2.21 (1H, app t, J 7.4, N*H*), 2.16 (3H, s, OCOC*H*₃), 2.10 (3H, s, OCOC*H*₃), 2.07 (3H, s, OCOC*H*₃), 2.05 (3H, s, OCOC*H*₃), 1.98 (3H, s, OCOC*H*₃), 1.63 (3H, stack, alkyl C*H*₂'s), 1.48 (2H, m, alkyl C*H*₂), 1.43 (9H, s, (C*H*₃)₃), 1.36 (2H, m, alkyl C*H*₂); δ_C (75 MHz, CDCl₃) 177.1 (*C*=O), 173.9 (*C*=O), 170.7 (*C*=O), 170.5 (*C*=O), 170.1 (*C*=O), 169.9 (*C*=O), 169.5 (*C*=O), 101.8 (*C*-1'), 101.0 (*C*-1), 82.7 (*C*q(CH₃)₃)), 71.8 (*C*-5), 71.3 (*C*-3 & *C*-5'), 70.9 (*C*-3'), 68.8 (*C*-2'), 66.8 (*C*-4'), 62.8 (*C*-6), 61.5 (*C*-6'), 56.8 (O*C*H₃), 56.5 (*C*-2), 40.3 (*C*-4), 36.5 (*C*H₂-alkyl), 33.7 (*C*H₂-alkyl), 29.6 (*C*H₂-alkyl), 28.4 (C(*C*H₃)₃)), 26.1 (*C*H₂-alkyl), 25.2 (*C*H₂-alkyl), 20.8 (OCO*C*H₃), 20.6 (OCO*C*H₃), 20.5 (2 x OCO*C*H₃), 20.4 (OCO*C*H₃); *m/z* (TOF ES+) 801.3263 ([M+Na]⁺, 100% C₃₄H₅₄N₂O₁₈Na requires 801.3269).

See **appendix 6** for COSY 90, HSQC and HMBC nmr experiments confirming the assignments.

225

Preparation of N-Boc-Caproic-lactosamine blocking sugar **84**.

$C_{24}H_{44}N_2O_{13}$
Exact Mass: 568.28
Mol. Wt.: 568.61

Pentaacetate **83** (0.001 g, 0.0013 mmol) was dissolved in anhydrous MeOH (10 mL) with stirring at ambient temperature under argon. Sodium metal (0.005 g) was then added and the reaction stirred for 15 minutes. DOWEX-50WX8-200 ion exchange resin (pre-washed with 5M HCl then MeOH) was then added to the reaction mixture with stirring until the pH = 5. The crude mixture was filtered and concentrated under reduced pressure to afford the product as a fine white solid which analysis revealed to be sufficiently pure to be taken on without further purification (0.0007 g, 96 %).

m/z (TOF ES+) 591.2 ([M+Na]$^+$, 100 %), HRMS 591.2741 ([M+Na]$^+$, 100% $C_{24}H_{44}N_2O_{13}Na$ requires 591.2767).

Preparation of free amine **85**.
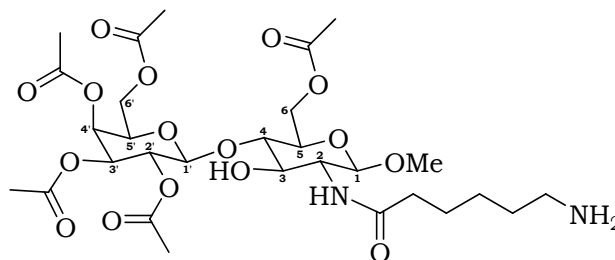
$C_{29}H_{46}N_2O_{16}$
Exact Mass: 678.28
Mol. Wt.: 678.68

Boc protected disaccharide **83** (0.03 g, 0.04 mmol) was dissolved in CH$_2$Cl$_2$ (5 mL) and TFA (2 mL) was added dropwise with stirring at ambient temperature. The Boc deprotection was followed by thin layer chromatography and the reaction was complete after 2 hours with a

single spot on the baseline (toluene / acetone, 4 : 6). Excess solvent and TFA was removed through cycles involving repeated dissolution in CH_2Cl_2 and removal under reduced pressure. The crude product was a light brown oil, pure product was recovered as a fine white solid by thorough trituration with Et_2O (0.026 g, 99 %). ν_{max} (Neat)/ cm^{-1} 3364, 3387 (NH_2), 1724, 1737 (C=O); δ_H (500 MHz, CDCl$_3$) 7.88 (2H, br s, NH_2), 6.01 (1H, br s, NH), 5.39 (1H, app s, H-10), 5.18 (1H, app t, J 9.0, H-8), 5.02 (1H, app d, J 10.2, H-9), 4.62 (1H, d, J 7.6, H-1), 4.38 (1H, d, J 6.4, H-7), 4.32 (1H, d, J 10.7, H-6), 4.18 (1H, stack, H-12), 4.08 (4H, stack, H-3, H-6, H-11 & H-12), 3.74 (1H, stack, H-5), 3.54 (1H, stack, H-2), 3.42 (3H, s, OCH_3), 2.96 (3H, app br s, H-4, CH_2NH$_2$), 2.32-2.18 (3H, stack, alkyl CH_2), 2.14 (3H, s, OCOCH_3), 2.09 (3H, s, OCOCH_3), 2.06 (3H, s, OCOCH_3), 2.01 (3H, s, OCOCH_3), 1.97 (3H, s, OCOCH_3), 1.71-1.54 (6H, mound, alkyl CH_2's), 1.38 (2H, app br s, alkyl CH_2); δ_C (100 MHz, CDCl$_3$) 176.8 (C=O), 174.9 (C=O), 170.9 (C=O), 170.2 (C=O), 170.0 (C=O), 169.8 (C=O), 169.5 (C=O), 101.8 (C-1'), 101.7 (C-1), 72.0 (C-5), 71.3 (C-3 & C-5'), 70.9 (C-3'), 68.8 (C-2'), 67.0 (C-4'), 62.8 (C-6), 61.3 (C-6'), 56.8 (OCH$_3$), 55.3 (C-2), 51.6 (C-4), 39.7 (CH$_2$-alkyl), 33.5 (CH$_2$-alkyl), 26.4 (CH$_2$-alkyl), 24.6 (CH$_2$-alkyl), 23.8 (CH$_2$-alkyl), 20.7 (2 x OCOCH$_3$), 20.6 (2 x OCOCH$_3$), 20.5 (OCOCH$_3$); m/z (TOF ES+) 679.2 ([M+H]$^+$, 100 %), 701.2 ([M+Na]$^+$, 75 %); HRMS 679.2926 ([M+H]$^+$, 100% $C_{29}H_{47}N_2O_{16}$ requires 679.2916).

Deacetylation to afford fully deprotected extended disaccharide **86**.

C$_{19}$H$_{36}$N$_2$O$_{11}$
Exact Mass: 468.23
Mol. Wt.: 468.50

Acetylated disaccharide **85** (0.003 g, 4.42 x 10^{-6} mol) was dissolved in anhydrous MeOH (10 mL) with stirring at ambient temperature under argon. Sodium metal (0.005 g) was then added and the reaction stirred for a further 15 minutes. Thin layer chromatography (toluene / acetone, 3 : 7) indicated reaction completion with a single strong baseline spot and the disappearance of the starting material. DOWEX-50WX8-200 ion exchange resin (pre-washed with 5M HCl then MeOH) was then applied to the reaction mixture until the pH = 5. The crude mixture was filtered and concentrated under reduced pressure to afford the product as a fine white solid which analysis revealed to be sufficiently pure to be taken

on without further purification (0.002 g, 96 %). *m/z* (TOF ES+) 469.1 ([M+H]+, 100%), 491.1 ([M+Na]+, 50 %); HRMS 491.2217 $C_{19}H_{36}N_2O_{11}Na$ requires 491.2228).

Coupling of FITC and deacetylation to afford labelled disaccharide **87**.
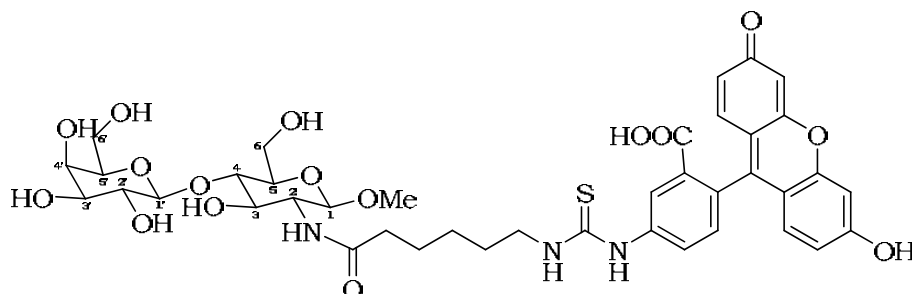
$C_{40}H_{47}N_3O_{16}S$
Exact Mass: 857.27
Mol. Wt.: 857.88

Acetylated disaccharide **85** (0.005 g, 7.37 x 10^{-6} mol) was dissolved in a combination of anhydrous DMF, CH_2Cl_2, and py (10 mL, 5 : 7 : 2) and FITC (0.0043 g, 1.11 x 10^{-5} mol) added. The reaction mixture was then stirred at ambient temperature under argon for 48 hours. The crude mixture was filtered and concentrated under reduced pressure prior to semi-preparative HPLC purification (method B, t_r = 29.1 minutes) and lyophilised to afford the product as a fine orange solid (0.007 g, 87 %). m/z MALDI-TOF 1068.0 ([M+H]+, 100 %), 1090.1 ([M+Na]+, 10 %)
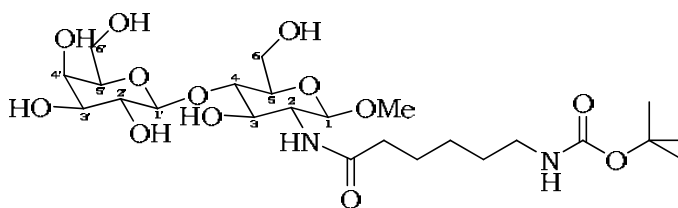
Fluorescein labelled acetylated disaccharide (0.006 g, 5.62 x 10^{-6} mol) was dissolved in anhydrous MeOH (10 mL) with stirring at ambient temperature under argon. Sodium metal (0.002 g) was then added and the reaction stirred for 15 minutes. DOWEX-50WX8-200 ion exchange resin (pre-washed with 5M HCl then MeOH) was then applied to the reaction mixture until the pH = 5. Filtration and concentration under high vacuum afforded the pure product as a fine orange powder (0.0046 g, 96 %). Analytical HPLC t_r = 23.2 minutes; m/z MALDI-TOF 858.1 ([M+H]+, 100 %), 880.1 ([M+Na]+, 30 %)

See **appendix 4** for Fluorescence spectra.
See **appendix 5** for HPLC and m/z data.

Preparation of NIR797 labelled disaccharide **88**.

$C_{64}H_{86}N_5NaO_{17}S_4^-$
Exact Mass: 1347.48
Mol. Wt.: 1348.64
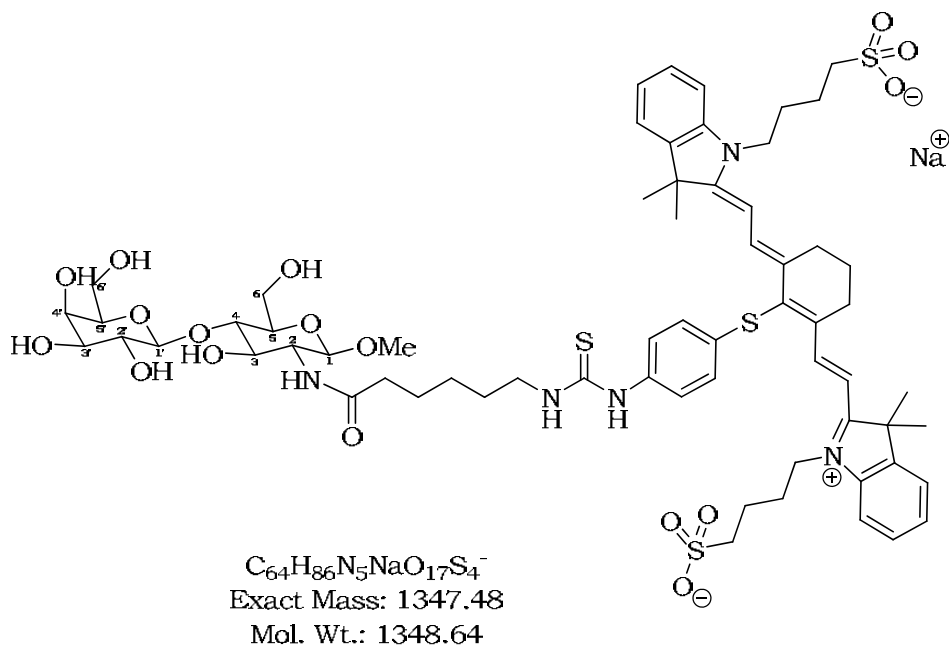
Fully acetylated N-caproic disaccharide **85** (0.003 g, 4.42 x 10^{-6} mol) was dissolved in a mixture of MeCN (5 mL) and acetone (0.5 mL) and stirred at ambient temperature. NIR797 isothiocyanate (0.006 g, 6.40 x 10^{-6} mol) in pH = 9.2 phosphate buffer (4.5 mL) was then added and the mixture stirred at ambient temperature for 72 hours. Analytical HPLC (method A, t_r = 22.4 minutes) revealed the product which was then purified by semi-preparative HPLC (method A, t_r = 28.7 minutes) and lyophilised to afford the product as a deep green powder (0.006 g, 89 %). m/z MALDI-TOF 1580.4 ([M+Na]$^+$, 100 %).

NIR797 labelled acetylated disaccharide (0.005 g, 3.16 x 10^{-6} mol) was dissolved in anhydrous MeOH (10 mL) with stirring at ambient temperature under argon. Sodium metal (0.002 g) was then added and the reaction stirred for 15 minutes. DOWEX-50WX8-200 ion exchange resin (pre-washed with 5M HCl then MeOH) was then applied to the reaction mixture until the pH = 5. The crude product was concentrated under reduced pressure and purified by semi-preparative HPLC (method A, t_r = 23.7 minutes) to afford the product as a deep green solid (0.003 g, 68 %). m/z MALDI-TOF 1370.5 ([M+Na]$^+$, 100 %).

See **appendix 4** for Fluorescence spectra.
See **appendix 5** for HPLC and m/z data.

Preparation of osmium complex labelled disaccharide **89**.

$C_{61}H_{70}N_8O_{17}Os$
Exact Mass: 1378.45
Mol. Wt.: 1377.48
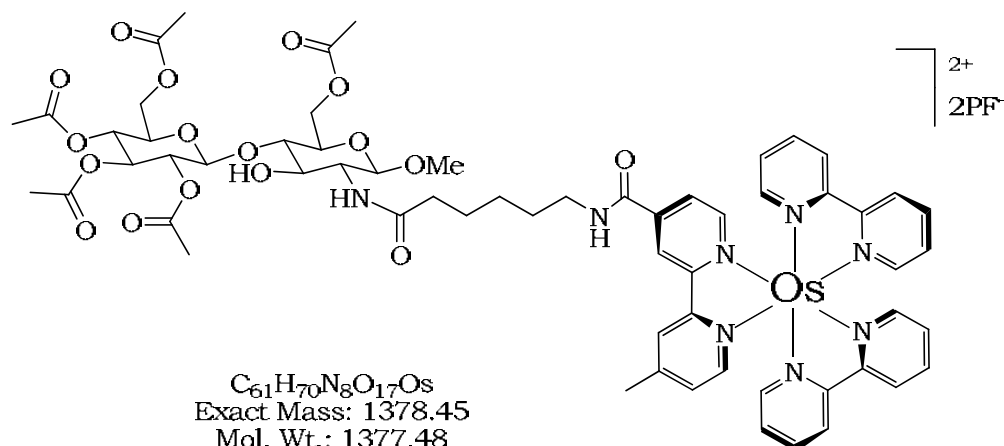
Fully acetylated N-caproic disaccharide **85** (0.002 g, 2.95 x 10^{-6} mol) was dissolved in anhydrous DMF (8mL) and Osmium succinic ester **53** (0.0036 g, 4.42 x 10^{-6} mol) added to the mixture which was then stirred in the dark at ambient temperature under argon for 10 minutes. Anhydrous pyridine (1 mL) was then added and the cocktail stirred for a further 48 hours. Analytical HPLC revealed the product (method A, t_r = 17.1 minutes). The crude mixture was concentrated under reduced pressure prior to purification by semi-preparative HPLC (method B, t_r = 17.07 minutes), and lyophilisation to afford the product as a brown solid (0.003 g, 79 %). m/z MALDI-TOF 1378.5 ([M+H]$^+$, 100 %), 1693.5 [M+ 2PF$_6$ + Na]$^+$.

See **appendix 4** for Fluorescence spectra.
See **appendix 5** for HPLC and m/z data.

Preparation of hexa-PEG extended disaccharide derivative **90**.

$C_{42}H_{70}N_2O_{24}$
Exact Mass: 986.43
Mol. Wt.: 987
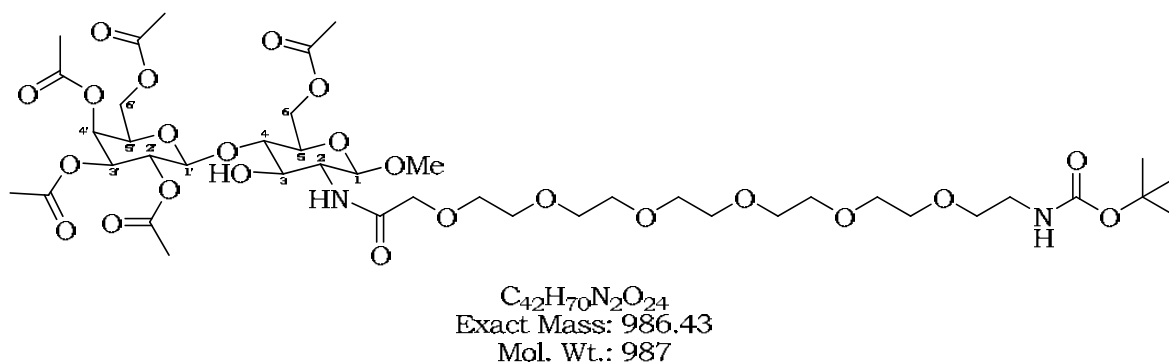
Butoxycarbonyl hexaPEG carboxylic acid **44** (0.035 g, 7.96 x 10^{-5} mol) was dissolved in anhydrous CH$_2$Cl$_2$ (5 mL) and stirred under argon at ambient temperature. PyBOP (0.055 g,

1.061 x 10^{-4} mol) and DIPEA (0.046 mL, 2.65 x 10^{-4} mol) were added and the cocktail stirred for 30 minutes. This mixture was then added to disaccharide **79** (0.03 g, 5.31 x 10^{-5} mol) which had been pre-dissolved in anhydrous CH$_2$Cl$_2$ (5 mL). The reaction was then stirred at ambient temperature for 48 hours with TLC monitoring R$_f$ = 0.33 (CH$_2$Cl$_2$ / MeOH, 9.7 : 0.3). The crude product was then concentrated under reduced pressure and purified by very slow gradient flash column chromatography (CH$_2$Cl$_2$ to CH$_2$Cl$_2$ / MeOH, 9.7 : 0.3) for afford the product as a pale brown oil (0.03 g, 58 %). m/z MALDI-TOF 1009.3 [M+Na]$^+$, 909 [M – Boc + Na]$^+$.

Nmr spectra were not sufficiently well resolved for accurate assignment.

Boc cleavage to afford free amine **91**.
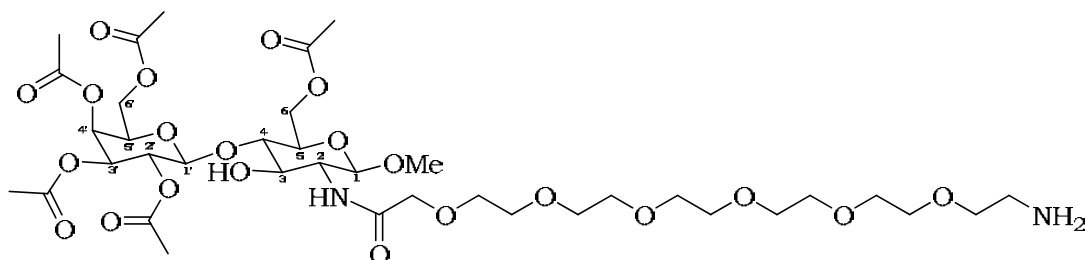
C$_{37}$H$_{62}$N$_2$O$_{22}$
Exact Mass: 886.38
Mol. Wt.: 886.89

PEG appended disaccharide **90** (0.02 g, 2.03 x 10^{-5} mol) was dissolved in CH$_2$Cl$_2$ (5 mL) and TFA (5 mL) was added dropwise with stirring at ambient temperature. The Boc deprotection was followed by thin layer chromatography and the reaction was complete after 2 hours with a single spot on the baseline (CH$_2$Cl$_2$ / MeOH, 9.7 : 0.3). Excess solvent and TFA was removed under reduced pressure with repeated co-concentrations with CH$_2$Cl$_2$. Trituration with Et$_2$O revealed the product was a light brown oil, and the product was taken on without further purification (0.017 g, 97 %). *m/z* (TOF ES+) 887.1 ([M+H]$^+$, 100 %), HRMS 887.3872 ([M+H]$^+$, 100% C$_{37}$H$_{63}$N$_2$O$_{22}$ requires 887.3868).

Nmr spectra were not sufficiently well resolved for accurate assignment.

Chapter 4: References

1.	Richards, M. A., *British Journal of Cancer*, **2009**, 101, S1-S4. b) Richards, M. A., *British Journal of Cancer*, **2009**, 101, S125-S129.
2.	Olesen, F.; Hansen, R. P.; Vedsted, P., *British Journal of Cancer*, **2009**, 101, S5-S8.
3.	Neal, R. D., *British Journal of Cancer*, **2009**, 101, S9-S12. b) Macleod, U.; Mitchell, E. D.; Burgess, C.; Macdonald, S.; Ramirez, A. J., *British Journal of Cancer*, **2009**, 101, S92-S101.
4.	Cancer statistics registration, England 2006, published **2008**, *The Office for National Statistics.* b) Cancer incidence and mortality in the UK, Statistical Bulletin, August **2009**, *The Office for National Statistics.*
3.	www.cancerresearchuk.org/cancerstats., updated **2009**.
4.	Vanderbilt University. www.laserfocusworld.com/display_article/36703/12
5.	Stiller, C.; Quinn, M.; Rowan, S., *The health of children and young people*, **2004**, Ch. 13, Office for National Statistics. b) Boyd-Kirkup, J., *Postnote*, The Parliamentary office of Science and Technology, **2009**, 330. c) Bosanquet, N.; Sikora, K., *The Lancet Oncology*, **2004**, 5, 9, 568-574.
6.	Fortner, J. G.; Mackean, B. J.; Kim, D. K.; Howland, W. S.; Turnball, A. D.; Goldiner, P.; Carlon, G.; Beattie, E. J., *Cancer*, **2006**, 47, 9, 2162-2166. b) Kurtz, J. M.; Amalric, R.; Brandone, H.; Ayme, Y.; Jacqiemier, J.; Pietra, J-C.; Hans, D.; Pollet, J-F.; Bressac, C.; Spitalier, J-M., *Cancer*, **2006**, 63, 10, 1912-1917.
7.	Licha, K., *Topics in Current Chemistry*: Contrast Agents for Optical Imaging, **2002**, Springer. b) Manning, C. H.; Goebel, T.; Thompson, R. C.; Price, R. R.; Lee, H.; Bornhop, D. J., *Bioconjugate Chem.*, **2004**, 15, 6, 1488-1495.
8.	Alty, J.; Hoey, E., *Practical Ultrasound*, **2006**, Hodder Education. b) Mathis, G.; Bitschnan, R.; Gehmacher, O.; Scheier, M.; Kopf, A.; Schwarzler, B.; Amann, T.; Doringer, W.; Hergan, K., *Ultraschall. Med.*, **1999**, 20, 2, 54-59.
9.	Corne, J.; Carroll, M.; Delany, D.; Brown, I., *Chest X-ray made easy*, **2002**, Churchill Livingstone.
10.	Charron, M.; Beyer, T.; Bohnen, N. N.; Kinahan, P. E.; Dachille, M.; Jerin, J.; Nutt, R.; Meltzer, C. C.; Villemagne, V.; Townsend, D., *Clin. Nucl. Med.*, **2000**, 25, 11, 905-910. b) Webb, R. W.; Brant, W. E.; Major, N., *Fundamentals of Body CT*, **1998**, W. B. Saunders Company, 2nd Sub Edn. c) Kalender, W. A., Computed Tomography: Fundamentals, System Technology, Image Quality, Applications, 2nd Edition, 2005, Wiley VCH; Gonzalez, A. B. de., *The Lancet*, **2004**, 9406, 345-351. d) Pelberg, R.; Mazur, W., *Cardiac CT Angiography Manual*, **2007**, 1st Edition, Springer.
11.	Valk, P. E.; Delbeke, D.; Bailey, D. L.; Townsend, D. W.; Maisey, M. N., *Positron Emission Tomography: Clinical Practice*, **2006**, Springer. b) Lindholm, P.; Minn, H.;

	Leskinen-Kallio, S.; Bergmann, J.; Ruotsulainen, U.; Joensuu, H., *J. Nucl. Med.*, **1993**, 34, 1, 1-6. c) Spaepen, K.; Stroobants, S.; Dupont, P.; Van Steenlieghen, S.; Thomas, J.; Vandenberghe, P.; Vanuytsel, L.; Bormans, G.; Balzerini, J.; De Wolf Peters, C.; Mortelmans, L.; Verhoef, G., *J. Clin. Oncology*, **2001**, 19, 2, 414-419. d) Knott, R. M.; Robertson, M.; Muckersie, E.; Forrester, J. V., *Biochem. J.*, **1996**, 318, 313-317. e) Ogawa, M.; Hatano, K.; Oishi, S.; Kawasumi, Y.; Fujii, N.; Kawaguchi, M.; Doi, R.; Imamura, M.; Yamamoto, M.; Ajito, K.; Mukai, T.; Saji, H.; Ito, K., *Nucl. Med. Biol.*, **2003**, 30, 1-9.
12.	Christian, P.; Waterstram-Rich, K., *Nuclear Medicine & PET-CT technology & techniques*, **2007**, Elsevier.
13.	Wernick, M. N.; Aarsvold, J. N., *Emission Tomography – the fundamentals of PET & SPECT*, **2004**, Elsevier Academic Press. b) Tany, R. H.; Da Silva, A. J.; Matthay, K. K.; Price, D. C.; Huberty, J. P.; Hawkins, R. A.; Hasegawa, B. H., *J. Nucl. Med.*, **2001**, 42, 237-247. c) Ding, H. J.; Huang, Y. F.; Tzeng, C. C.; Wei, L. M.; Yeh, S. J., *Bioorg. Med. Chem. Lett.*, **1999**, 9, 22, 3199-3202; Ebmeier, K. P., *SPECT in Dementia*, **2003**, Karger.
14.	a) Segars, W. P.; Tsui, B. M. W.; Frey, E. C.; Johnson, G. A.; Berr, S. S. *Molecular Imaging and Biology*, **2004**, 6, 3, 149-159. b) Hashemi, R. H.; Bradley, W. G.; Lisanti, C. J., *MRI: The Basics*, **2003**, Lippincott, Williams & Williams, 2nd Revised Edition; Hore, P., *Nuclear Magnetic Resonance*, **1995**, OUP Oxford; Hore, P.; Jones, J.; Wimperis, S., *NMR: The Toolkit*, **2000**, OUP Oxford. c) Ross, R.; Goodpaster, B.; Kelley, D.; Boada, F., *Ann. NY Acad. Sci.*, **2006**, 904, 12-17. d) Haase, A., *Magnetic Resonance in Medicine*, **2005**, 13, 1, 77-89. e) Tanabe, M.; Ohnuma, N.; Iwai, J.; Yoshida, H.; Takahashi, H.; Maie, M.; Etoh, T.; Kawamura, K., *Medical & Pediatric Oncology*, **2006**, 24, 5, 292-299. f) Kushner, B. H., *J. Nucl. Med.*, **2004**, 45, 7, 1172-1188. g) Caravan, P.; Ellison, J. J.; McMurry, T. J.; Lauffer, R. B., *Chem. Rev.*, **1999**, 99, 2293-2352.
15.	Norum, O-J.; Giercksky, K-E.; Berg, K., *Photochem. Photobiol. Sci.*, **2009**, 8, 758-762. b) Shabaan, S.; Ba, L. A.; Abbas, M.; Burkholz, T.; Denkert, A.; Gohr, A.; Ludger, A.; Florenz Sasse, W.; Weber, W.; Jacob, C., *Chem. Commun.*, **2009**, 4702-4704. c) Moseley, M., *Stroke*, **2004**, 35, 2632-2634.
16.	a) Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien R. Y., *Science*, **2006**, 312, 217-224. b) Tsien, R. Y., *Annu. Rev. Biochem.*, **1998**, 67, 509-544. c) Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y., Nature Methods, 2005, 2, 12, 905-909. c) Shu, X.; Royant, A.; Lin, M. Z.; Aguilera, T. A.; Lev-Ram, V.; Steinbach, P. A.; Tsien, R. Y., *Science*, **2009**, 324, 804-807.
17.	a) Tsien, R. Y., *Biochemistry*, **1980**, 19, 11, 2396-2404. b) Grynkiewicz, G.; Poenie, M.; Tsien, R. Y., *J. Bio. Chem.*, **1985**, 260, 6, 3440-3450. c) Kaltoyannis, N., *The f elements*, **1999**, OUP Oxford; SPEX Fluorescence group application notes –

	Timegated separation of Lanthanide Luminescence, Jobin Yvon Inc., Horiba Group, **2003**, F27.
18.	Cotton, S., *Lanthanide & Actinide Chemistry*, **2006**, WILEY. b) Kobayashi, S., *Lanthanides: Chemistry & use in Organic Synthesis*, **1999**, Springer.
19.	Faulkner, S.; Matthews, J. L., *Comprehensive Co-ordination Chemistry*, **2003,** 2nd Edn., Vol. 9, Elsevier, Oxford. b) R. Paschotta. www.rp-photonics.com/img/yb_cs.png.
20.	a) Bunzli, J-C. G., *Acc. Chem. Res.*, **2006**, 39, 1, 53-61. b) Barigelletti, F.; Flamigni, L.; Balzani, V.; Collin, J-P.; Sauvage, J-P.; Sour, A.; Constable, E. C.; Cargill Thompson, A. M. W., *Coord. Chem. Rev.*, **1994**, 132, 209; Beeby, A.; Faulkner, S.; Williams, J. A. G., *Dalton Trans.* **2002**, 1918-1922. c) Clarkson, I. M.; Beeby, A.; Bruce, J. I.; Govenlock, L. J.; Lowe, M. P.; Matthieu, C. E.; Parker, D.; Senanayake, K., *New J. Chem.*, **2000**, 24, 377; Förster, T., *Ann. Phys.* (*Liepzig*), **1948**, 2, 55. d) Burroughs, S. E.; Horrocks, W. D.; Ren, H.; Klee, C. B., *Biochemistry*, **1994**, 33, 10428.
21.	a) Faulkner, S.; Natrajan, L. S.; Perry, W. S.; Sykes, D., *Dalton Trans.* **2009**, 3890-3899; Lazarides, T.; Sykes, D.; Faulkner, S.; Barbieri, A.; Ward, M. D., *Chem. Eur. J.*, **2008**, 14, 9389-9399. b) Montgomery, C. P.; Murray, B. S.; New, E. J.; Pal, R.; Parker, D., *Accounts of Chemical Research*, **2009**, 42, 7, 925-937.
22.	Parker, D.; Dickins, R. S.; Puschmann, H.; Crossland, C.; Howard, J. A. K., *Chem. Rev.*, **2002**, 102, 1977-2010. b) Parker, D.; Beeby, A.; Clarkson, I. M.; Dickens, R. S.; Faulkner, S.; de Sousa, A. S.; Williams, J. A. G., *J. Chem. Soc. Perkin. Trans.*, **1999**, 2, 493. c) Sanchez, P.; Valero, E.; Galvez, N.; Dominguez-Vera, J. M.; Marinone, M.; Poletti, G.; Corti, M.; Lascialfari, A., *Dalton Trans.*, **2009**, 800-804; Petit, S.; Baril-Robert, F.; Pilet, G.; Reber, C.; Luneau, D., *Dalton Trans.*, **2009**, 6809-6815. d) De, M.; Rana, S.; Akpinar, H.; Miranda, O. R.; Arvizo, R. R.; Bunz, U. H. F.; Rotello, V. M., *Nature Chemistry*, **2009**, 1, 461-465.
23.	a) Domling, A.; Ugi, I., *Angew. Chem. Int. Ed. Engl.* **2000**, 39, 3168-3210. b) Dömling, A., *Chem. Rev.*, **2006**, 106, 17-89.
24.	Yehia, N. A. M.; Antuch, W.; Beck, B.; Hess, S.; Schauer-Vukasinovic, V.; Almstetter, M.; Furer, P.; Herdtweck, E.; Dömling, A., *Biorg. Med. Chem. Lett.*, **2004**, 14, 3121.
25.	Constable, F.; Ugi, I., *Tetrahedron*, **2001**, 57, 5785-5789.
26.	Stroker, A. M.; Keating, T. A.; Tempest, P. A.; Armstrong, R. W., *Tetrahedron Lett.*, **1996**, 37, 8, 1149-1152.
27.	Ugi, I.; Werner, B.; Domling, A., *Molecules*, **2003**, 8, 53-66.
28.	Lin, X. P.; Wenwei, L.; Xiaomin, Z., *Synthesis*, **2002**, 1017.
29.	R. A. Weinberg, *The biology of cancer*, **2007**, Published by Garland Science, Taylor & Francis Group, LLC, New York. b) Schollkopf, U., *Angew. Chem. Int. Ed. Engl.*,

	1977, 16, 6, 339-348; Kitano, Y.; Chiba, K.; Tada, M., *Tetrahedron Lett.*, **1998**, 39, 1911-1912.
30.	Bergers, G.; Benjamin, L. E., *Nature Rev. Cancer*, **2003**, 3, 401-410.
31.	Carmeliet, P., *Nature Medicine*, **2000**, 6, 3, 389-395.
32.	Hlatky, L.; Hahnfeldt, P.; Folkman, J., J. *Nat. Cancer Inst.*, **2002**, 94, 12, 883-893.
33.	Jin, H.; Varner, J., *British Journal of Cancer*, **2004**, 90, 561-565.
34.	www.e22.physik.tu-muenchen.de/bausch/Oli_ECM.html.
35.	Moschos, S. J.; Drogowski, L. M.; Reppert, S. I.; Kirkwood, J. M., *Oncology*, **2007**, 21, 13-20.
36.	Haubner, R.; Kuhnast, B.; Mang, C.; Weber, W. A.; Kessler, H.; Wester, H. J.; Schwaiger, M., *Bioconj. Chem.*, **2004**, 15, 61-69.
37.	Varner, J. A.; Cheresh,D. A., *Curr. Opin. Cell Biol.*, **1996**, 8, 5, 724-730.
38.	Liu, Z.; Wang, F.; Chen, X., *Drug Dev. Res.*, **2008**, 69, 329-339.
39.	Schottelius, M.; Laufer, B.; Kessler, H.; Wester, H-J., *Accounts of Chemical Research*, **2009**, 42, 7, 969-980.
40.	Milbradt, A. G.; Loweneck, M.; Krupka, S. S.; Reif, M.; Sinner, E-K.; Moroder, L.; Renner, C., *Biopolymers*, **2005**, 77, 5, 304-313.
41.	Xiong, J-P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A., *Science*, **2002**, 296, 151-155.
42.	Craig, W. S.; Cheng, S.; Mullen, D. G.; Blevitt, J.; Pierschbacher, M. D., *Biopolymers*, **1995**, 37, 157-175.
43.	Haubner, R.; Gratias, R.; Diefenbach, B.; Gooman, S. L.; Jonczyk, A.; Kessler, H., *J. Am. Chem. Soc.*, **1996**, 118, 7461-7472.
44.	Haubner, R.; Brucherseifer, F.; Bock, M.; Kessler, H.; Schwaiger, M.; Wester, H. J., *Nuclear Medicine*, **2004**, 43, 26-32.
45.	Thumshirn, G.; Hersel, U.; Goodman, S. L.; Kessler, H., *Chem. Eur. J.*, **2003**, 9, 2717-2725. also *Chemistry*, **2003**, 9, 2717-2725.
46.	Chen, X.; Sievers, E.; Hou, Y.; Park, R.; Tohme, M.; Bart, R.; Bremner, R.; Bading, J. R.; Conti, P. S., *Neoplasia*, **2005**, 7, 3, 271-279.
47.	Chen, X.; Hou, Y.; Tohme, M.; Park, R.; Khankaldyyan, V.; Gonzales-Gomez, I.; Bading, J. R.; Laug, W. E.; Conti, P. S., *J. Nucl. Med.*, **2004**, 45, 1776-1783.
48.	Chen, X.; Park, R.; Shahinian, A. H.; Bading, J. R.; Conti, P. S., *Nucl. Med. Biol.* **2004**, 31, 11-19.
49.	Chen, X.; Hou, Y.; Tohme, M.; Park, R.; Khankaldyyan, V.; Gonzales-Gomez, I.; Bading, J. R.; Laug, W. E.; Conti, P. S., *J. Nucl. Med.*, **2004**, 45, 1776-1783.
50.	Ntziachristos, V.; Ripoli, J.; Weissleder, R., *Opp. Lett.* **2002**, 27, 1652.
51.	Von Wallbrunn, Holtke, C.; Zuhlsdorf, M.; Heindel, W.; Schafers, M.; Bremer, C., *Eur. J. Nucl. Med. Mol. Imaging*, **2005**, 4, 439-447.

52.	Chen, X.; Conti, P. S.; Moats, R. A., *Cancer Res.*, **2004**, 64, 8009-8014.
53.	Wang, W.; Ke, S.; Wu, Q.; Chamsangavej, C.; Gurfinkel, M.; Gelovani, J. G.; Abruzzese, J. L.; Sevick-Muraca, E. M.; Li, C., *Mol. Imaging*, **2004**, 3, 343-351.
54.	Cheng, Z.; Wu, Y.; Xiong, Z.; Gambhir, S. S.; Chen, X. *Bioconj. Chem.*, **2005**, 16, 1433-1441.
55.	Achilefu, S.; Bloch, S.; Markiewicz, M. A.; Zhong, T.; Ye, Y.; Dorshow, R. B.; Chance, B.; Liang, K., *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, 102, 7976-7681.
56.	Cai, W.; Gambhir, S. S.; Chen, X., *Molecular Imaging*, **2005**, 14-25.
57.	Shokeen, M.; Anderson, C. J., *Accounts of Chemical Research*, **2009**, 42, 7, 832-841.
58.	Mulder, W. J. M.; Strijkers, G. J.; Van Tilborg, G. A. F.; Cormode, D. P.; Fayad, Z. A.; Nicolay, K., *Accounts of Chemical Research*, **2009**, 42, 7, 904-914.
59.	Mulder, W. J.; Strijkers, G. J.; Habets, J. W.; Bleeker, E. J.; van der Schaft, D. W.; Storm, G.; Koning, G. A.; Griffioen, A. W.; Nicolay, K., *FASEB J.*, **2005**, 19, 2008-2010.
60.	a) Lindhorst, T.; Bock, H.; Ugi, I., *Tetrahedron*, **1999**, 55, 7411-7420. b) Ugi, I.; Demharter, A.; Horl, W.; Schmid, T., *Tetrahedron*, **1996**, 52, 35, 11657-11664.
61.	a) Merrifield R. B., *J. Am. Chem. Soc.*, **1963**, 85, 2149. b) Mitchell A. R., *Peptide Sci.*, **2008**, 90, 3, 175-184. c) Kent, S.; Alewood, P., Int. J. Pep. Res. & Ther., **2007**, 13, 29.
62.	Kim, S. W.; Bauer, S. M.; Armstrong, R. W., *Tetrahedron Lett.*, **1998**, 39, 6993-6996.
63.	Anelli, P. L.; Lattuada, L.; Gabellini, M.; Recanati, P., *Bioconjugate Chem.* **2001**, 12, 1081-1084.
64.	Yoo, J.; Reichert, D. E.; Welch, M. J., *Chem. Commun.* **2003**, 766-767.
65.	Mishra, A. K.; Draillard, Faivre-Chauvet, A.; Gestin, J. F.; Curtet, C.; Chatal, J-F., *Tetrahedron Lett.*, **1996**, 37, 42, 7515-7518.
66.	Chaux, F.; Denat, F.; Espinosa, E.; Guilard, R., *Chem. Commun.* **2006**, 5054-5056.
67.	Li, C.; Wong, W-T., *Tetrahedron*, **2004**, 60, 5595-5601.
68.	Li, C.; Wong, W-T., *Tetrahedron Lett.*, **2002**, 43, 3217-3220.
69.	Oliver, M.; Jorgensen, M. R.; Miller, A. D., *Synlett.* **2004**, 3, 453-456.
70.	Luis, M.; Leon-Rodriguez, D.; Kovacs, Z., *Bioconjugate Chem.*, **2008**, 19, 2, 391-402.
71.	Main, M.; Snaith, J. S.; Meloni, M. M.; Jauregui, M.; Sykes, D.; Faulkner, S.; Kenwright, A. M., *Chem. Commun.* **2008**, 5212-5214.
72.	Parker, D.; Dickins, R. S.; Puschmann, H.; Crossland, C.; Howard, J. A. K., *Chem. Rev.*, **2002**, 102, 1977-2010.
73.	Parker, D., *Chem. Soc. Rev.*, **2004**, 33, 156-165.

74.	Bonzide, A.; Sauve, G., Tetrahedron Lett., **1997**, 38, 5945-5948.
75.	Henkel, B.; Sax, M.; Domling, A., *Tet. Lett.*, **2003**, 44, 7015-7018.
76.	Boturyn, D.; Dumy, P., *Tetrahedron Lett.*, **2001**, 42, 2787-2790.
77.	Valldosera, M.; Monso, M.; Xavier, C.; Raposinho, P.; Correia, J. D. G.; Santos, I.; Gomes, P., Int. J. Pep. & Ther., **2008**, 273-281.
78.	Diaz-Mochon, J. J.; Bialy, L.; Bradley, M., *Org. Lett.*, **2004**, 6, 7, 1127-1129.
79.	Decostaire, I. P.; Lelievre, D.; Aucagne, V.; Delmas, A. F., *Tetrahedron Lett.*, **2007**, 48, 6523-6526.
80.	Guibe, F., *Tetrahedron*, **1998**, 54, 2967-3042.
81.	Trzeciak, A.; Bannwarth, W., *Tetrahedron Lett.*, **1992**, 33, 32, 4557-4560.
82.	Thieriet, N.; Alsina, J.; Giralt, E.; Guibe, F.; Albericio, F., *Tetrahedron Lett.*, **1997**, 38, 41, 7275-7278.
83.	Farrera-Sinfreu, J.; Royo, M.; Albericio, F., *Tetrahedron Lett.*, **2002**, 43, 7813-7815.
84.	Freund, E.; Robinson, J. A., *Chem. Commun.*, **1999**, 2509-2510.
85.	Vutukuri, D, R.; Bharathi, P.; Yu, Z.; Rajasekaran, K.; Tran, M-H.; Thayumanavan, S., *J. Org. Chem.*, **2003**, 68, 1146-1149.
86.	Bouteiller, C.; Clave, G.; Bernardin, A.; Chipon, B.; Massonneau, M.; Renard, P-Y.; Romieu, A., *Bioconj. Chem.*, **2007**, 18, 1303-1317.
87.	Shealy, D. B.; Lipowska, M.; Lipowska, J.; Narayanan, N.; Sutter, S.; Strekowski, L.; Patonay, G., *Anal. Chem.*, **1995**, 67, 247-251.
88.	Kircher, M. F.; Mahmood, U.; King, R. S.; Weissleder, R.; Josephson, L., *Cancer Res.*, **2003**, 63, 8122-8125.
89.	Reynolds, G. A.; Drexhage, K. H., *J. Org. Chem.*, **1977**, 42, 5, 885-888.
90.	Lipowska, M.; Patonay, G.; Strekowski, L., *Synth. Commun.*, **1993**, 23, 21, 3087-3094.
91.	Klonis, N.; Sawyer, W. H., *Photochemistry & Photobiology*, **2003**, 77, 5, 502-509.
92.	Thumshirn, G.; Hersel, U.; Goodman, S. L.; Kessler, H., *Chem. Eur. J.*, **2003**, 9, 2717-2725. also *Chemistry*, **2003**, 9, 2717-2725.
93.	Peek, B. M.; Ross, G. T.; Edwards, S. W.; Meyer, G. J.; Meyer, T. J.; Erickson, B. W., *Int. J. Peptide Protein Res.*, **1991**, 38, 114-123.
94.	Lay, P. A.; Sasse, W. H. F., *Inorg. Chem.*, **1984**, 23, 4123-4125.
95.	Hoertz, P. G.; Staniszewski, A.; Marton, A.; Higgins, G. T.; Incarvito, C. D.; Rheingold, A. L.; Meyer, G. J., *J. Am. Chem. Soc.*, **2006**, 128, 8234-8245.
96.	Lin, R-J.; Chang, I-J., *J. Chin. Chem. Soc.*, **2002**, 49, 2, 161-164.
97.	Kendric Gino, M. J.; Dawson, J. H., *Inorganica Chimica Acta*, **1985**, 97, L41.
98.	Kober, E. M., Casper, J. V.; Sullivan, B. P.; Meyer, T. J., *Inorganic Chemistry*, **1988**, 27, 25, 4587-4598.

99.	Pope, S. J. A.; Coe, B. J.; Faulkner, S.; Bichenkova, E. V.; Yu, X.; Douglas, K. T., *J. Am. Chem. Soc.*, **2004**, 126, 9490-9491.
100.	Kerkhoff, E.; Fedorov, L. M.; Siefken, R.; Walter, A. O.; Papadopoulos, T.; Rapp, U. R. *Cell Growth & Differentiation*, **2000**, 11, 185-190.
101.	Ehrhardt, A.; Bartels, T.; Geick, A.; Klocke, R.; Paul, D.; Halter, R. *British Journal of Cancer*, **2001**, 84, 6, 813-818.
102.	Dalemans, W.; Perraud, F.; Le Meur, M.; Gerlinger, P.; Courtney, M.; Pavirani, A. *Biologicals*, **1990**, 18, 191-198.
103.	Tonjes, R. R.; Lohler, J.; O'Sullivan, J. F.; Kay, G. F.; Schmidt, G. H.; Dalemans, W.; Pavirani, A.; Paul, D. *Oncogene*, **1995**, 10, 765-768.
104.	Keren, S.; Gheysens, O.; Levin, C. S.; Gambhir, S. S. *IEEE Transactions on Medical Imaging*, **2008**, 27, 1, 58-63.
105.	Yao, N.; Xiao, W.; Wang, X.; Marik, J.; Park, S. H.; Takada, Y.; Lam, K. S. *J. Med. Chem.*, **2009**, 52, 126-133.
106.	Ye, Y.; Bloch, S.; Xu, B.; Achilefu, S. *J. Med. Chem.*, **2006**, 49, 2268-2275.
107.	Morales, A. R.; Schafer-Hales, K. J.; Marcus, A. I.; Belfield, K. D. *Bioconjugate Chem.*, **2008**, 19, 2559-2567.
108.	Bloch, S.; Xu, B.; Ye, Y.; Liang, K.; Nikiforovich, G. V.; Achilefu, S. *Molecular Pharmaceutics*, **2006**, 3, 5, 539-549.
109.	Haubner, R.; Brucherseifer, F.; Bock, M.; Kessler, H.; Schwaiger, M.; Wester, H. J., *Nuclear Medicine*, **2004**, 43, 26-32.
110.	Cheng, Z.; Wu, Y.; Xiong, Z.; Gambhir, S. S.; Chen, X. *Bioconj. Chem.*, **2005**, 16, 1433-1441.
111.	Segars, W. P.; Tsui, B. M. W.; Frey, E. C.; Johnson, G. A.; Berr, S. S. *Molecular Imaging and Biology*, **2004**, 6, 3, 149-159.
112.	Cook, M. J. *The Anatomy of the Laboratory Mouse*, **1965**, Academic Press Inc. (London) Ltd.
113.	Mincu, N. *Advanced research technologies Inc.* **2006**, lecture notes.
114.	Cavanaugh, D.; Johnson, E.; Price, R. E.; Kurie, J.; Travis, E. L.; Cody, D. D. *Molecular Imaging*, **2004**, 3, 1, 55-62
115.	Garbow, J. R.; Zhang, Z.; You, M. *Cancer Res.*, **2004**, 64, 2740-2742.
116.	Namati, E.; Chon, D.; Thiesse, J.; Hoffman, E. A.; de Ryk, J.; Ross, A.; McLennan, G. *Phys. Med. Biol.*, **2006**, 51, 6061-6075.
117.	Coll, J-L.; Josserand, V. *Drug Discovery Today: Disease Models*, **2006**, 3, 3, 219-224.
118.	Xiong, L.; Yu, M.; Cheng, M.; Zhang, M.; Zhang, X.; Xu, C.; Li, F. *Mol. Biosyst.*, **2009**, 5, 241-243.
119.	Davis, B. G.; Fairbanks, A. J., *Carbohydrate Chemistry*, **2002**, OUP Oxford. b)

	Hales, C. A.; Cowman, M. K.; Hari, G. G., *Carbohydrate Chemistry, Biology and Medical applications*, **2008**, Elsevier Ltd.
120.	Ernst, B., *Carbohydrates in Chemistry & Biology: A comprehensive handbook*, **2000**, Wiley VCH. b) Kannagi, R.; Izawa, M.; Kioke, T.; Miyazaki, K.; Kimura, N., *Cancer Science*, **2004**, 95, 5, 377-384.
121.	Handerson, T.; Camp, R.; Harigopal, M.; Rimm, D.; Pawalek, J., *Clinical Cancer Res.*, **2005**, 11, 2969-2973. b) Osborn, H.; Khan, T. H., *Oligosaccharides: Their synthesis & biological role.*, **2000**, OUP Oxford.
122.	a) Ben, R. N.; Eniade, A. A.; Hauer, L., *Org. Lett.* **1999**, 1, 11, 1759-1762. b) Nangia-Makker, P.; Honjo, Y.; Sarvis, R.; Akahani, S.; Hogan, V.; Piata, K. J.; Raz, A., *Am. J. Pathol.*, **2000**, 156, 899-909. c) Van den Brule, F.; Califine, S.; Castronovo, V., *Glycoconj. J.*, **2004**, 19, 537-542. d) Wada, J.; Makino, H., *Acta Med.Okayama*, **2001**, 55, 1, 11-17.
123.	Leffler, H.; Carlsson, S.; Hedlund, M.; Qian, Y.; Poirier, F., *Glycoconj. Journal*, **2004**, 19, 433-440.
124.	Oka, N.; Takenaka, Y.; Raz, A., *J. Cell Biochem.*, **2004**, 91, 118-124.
125.	Barondes, S. H.; Cooper, D. N. W.; Gitt, M. A.; Leffler, H., *J. Bio. Chem.*, **1994**, 269, 33, 20807-20810.
126.	Takenaka, Y.; Fukomori, T.; Raz, A., *Glycoconj. Journal*, **2004**, 19, 543-549.
127.	Liu, F-T.; Rabinovich, G. A., *Nature Rev. Cancer*, **2005**, 5, 29-41.
128.	Nakahara, S.; Oka, N.; Raz, A., *Apoptosis*, **2005**, 10, 2, 267.
129.	Krzeslak, A.; Lipinska, A., *Cellular & Molecular Bio. Lett.*, **2004**, 9, 305-328.
130.	Honjo, Y.; Nangia-Makker, P.; Inohara, H.; Raz, A., *Clin. Cancer. Res.*, **2001**, 7, 661-668.
131.	Yoshii, T.; Iohara, H.; Takenaka, Y.; Honjo, Y.; Akahari, S.; Nomura, T.; Raz, A.; Kubo, T., *Int. J. Oncol.*, **2001**, 18, 787-792.
132.	Takenaka, Y.; Iohara, H.; Yoshii, T.; Kazuo, O.; Susumu, N.; Shiro, A.; Yuichiro, H.; Yoshifumi, Y.; Avraham, R.; Takeshi, K., *Cancer Lett.*, **2003**, 195, 111-119.
133.	a) Sorme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. S.; Nilsson, U. J., *J. Am. Chem. Soc.*, **2005**, 127, 1737-1743. b) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids. Res.*, **2000**, 28, 235-242.
134.	Sorme, P.; Qian, Y.; Nyholm, P-G.; Leffler, H.; Nilsson, U. J., *ChemBioChem*, **2002**, 3, 183-189.
135.	Pieters, R. J., *ChemBioChem*, **2006**, 7, 721-728.
136.	Sorme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J., *J. Am. Chem. Soc.*, **2006**, 127, 1737-1743. b) Nilsson, U. J.; Fournier, J-L.; Hindsgaul, O.*, Bioorrg. Med. Chem.*, **1998**, 6, 1563-1575. c) Ritter, T. K.; Wong, C-H.; *Tetrahedron Lett.*, **2001**, 42, 615-618. d) zu Reckendorf, W. M., *Methods*

Carbohydr. Chem, **1972**, 6, 129-131. e) Kretzschmar, G.; Stahl, W., *Tetrahedron*, **1998**, 54, 6341-6358.

137. Cumpstey, I.; Sundin, A.; Leffler, H.; Nilsson, U. J., *Angew. Chem.*, **2005**, 117, 5240-5242.

138. Cumptsey, I.; Salomonsson, E.; Sundin, A.; Leffler, H.; Nilsson, U. J., *Chem. Eur. J.*, **2008**, 14, 4233-4245.

139. Cumptsey, I.; Salomonsson, E.; Sundin, A.; Leffler, H.; Nilsson, U. J., *ChemBioChem*, **2007**, 8, 1389-1398.

140. Valerio, S.; Iadonisi, A.; Adinolfi, M.; Ravida, A., *J. Org. Chem.*, **2007**, 72, 6097-6106.

141. Adinolfi, M.; Iadonisi, A.; Ravida, A.; Schiattarella, M., *Tetrahedron Lett.*, **2003**, 44, 7863-7866.

142. Debenham, J. S.; Debenham, S. D.; Fraser-Reid, B., *Bioorg.& Med. Chem.*, **1996**, 4, 11, 1909-1918.

143. a) Castro-Palomino, J. C.; Schmidt, R. R., *Tetrahedron Lett.*, **1995**, 36, 30, 5343-5346; b) *Liebigs Ann.*, **1996**, 1623-1626.

144. Ellervik, U.; Magnusson, G., *J. Org. Chem.*, **1998**, 63, 9314-9322.

145. Kochetkov, N. K.; Byramova, N. E.; Tsvetkov, Y. E.; Backinowsky, L. V., *Tetrahedron*, **1985**, 41, 16, 3363-3375.

146. a) Silva, D. J.; Wang, H.; Allanson, N. M.; Jain, R. K.; Sofia, M. J., *J. Org. Chem.*, **1999**, 64, 5926-5929.

147. Mukhopadhyay, B.; Kartha, K. P. R.; Russell, D. A.; Field, R. A., *J. Org. Chem.*, **2004**, 69, 7758-7760.

148. Melcher, R.; Hillebrand, A.; Bahr, U.; Schroder, B.; Karas, M.; Hasilik, A., *BioChem. J.*, **2000**, 348, 3, 507-515.

149. Kerns, R. J.; Linhardt, R. J., *J. Chromatography A*, **1995**, 705, 369-373.

150. Tariq, M. A.; Hayashi, K.; Tokuyasu, K.; Nagata, T., *Carbohydrate Res.* **1995**, 275, 67-72.

151. Boulineau, F, P.; Wei, A., *Org. Lett.*, **2004**, 6, 1, 119-121.

152. Dahmen, J.; Gnosspelius, G.; Larsson, A-C.; Lave, T.; Noori, G.; Palsson, K., *Carbohydr. Res.* **1985**, 138, 17-28.

153. Ogawa, S.; Matsunaga, N.; Li, H.; Palcic, M. M., *Eur. J. Org. Chem.*, **1999**, 631-642.

154. O' Brien, A. O.; Lynch, C.; O' Boyle, K. M.; Murphy, P. V., *Carbohydr. Res.*, **2004**, 339, 2343-2354.

155. Gottlieb, H. E.; Kotlyar, V.; Nudelman, A., *J. Org. Chem.*, **1997**, 62, 7512-7515.

156. Kuklinski, S.; Pesheva, P.; Heimann, C.; Urschel, S.; Gloor, S.; Graeber, S.; Herzog, V.; Pietsch, T.; Wiestler, O. D.; Probstmeier, R., *J. Neuro. Res.*, **2001**, 60, 1, 45-57.

157.	Perillo, N. L.; Marcus, M. E.; Baum, L. G., *J. Mol. Med.*, **1998**, 76, 402-412.
158.	Lichtenthaler, F. W.; Kaji, E.; Weprek, S., *J. Org. Chem.*, **1985**, 50, 3505-3515.
159.	Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyo, D.; chait, B. T.; Muir, T. W., *J. Am. Chem. Soc.*, **2003**, 125, 2416-2425. b) Butterworth, J. F.; Moran, J. R.; Whitesides, G. M.; Strichartz, G. R., *J. Med. Chem.*, **1987**, 30, 1295-1302. c) Roy, B. C.; Mallik, S., *J. Org. Chem.*, **1999**, 64, 2969-2974. d) Notta, J., *PhD Thesis*, **2004**, University of Birmingham.

5.1 Appendix 1

Crystal structures

Appendix 1a: Benzyl Ugi product M05M. Crystal structure

Table 1. Crystal data and structure refinement for m150m

.

Identification code	m150m
Empirical formula	C54 H74 N6 Na O12
Formula weight	1022.18
Temperature	296(2) K
Wavelength	1.54178 Å
Crystal system	Monoclinic
Space group	P21/n

Unit cell dimensions
$a = 12.8283(17)$ Å $= 90°.$
$b = 21.445(3)$ Å $= 93.842(6)°.$
$c = 21.426(3)$ Å $= 90°.$

Volume	5881.3(13) Å^3
Z	4
Density (calculated)	1.154 Mg/m^3
Absorption coefficient	0.730 mm^{-1}
F(000)	2188
Crystal size	0.30 x 0.20 x 0.20 mm^3
Theta range for data collection	3.90 to 56.45°.
Index ranges	-12<=h<=13, -22<=k<=22, -23<=l<=22
Reflections collected	36735
Independent reflections	7471 [R(int) = 0.0790]
Completeness to theta = 56.45°	96.2 %
Max. and min. transmission	0.8678 and 0.8108
Refinement method	Full-matrix least-squares on F^2
Goodness-of-fit on F^2	1.148
Final R indices [I>2sigma(I)]	R1 = 0.1185, wR2 = 0.3078
R indices (all data)	R1 = 0.1967, wR2 = 0.3668
Extinction coefficient	0.0011(2)
Largest diff. peak and hole	0.732 and -0.326 e.Å^{-3}

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3)

for m150mm. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
C(1)	8631(7)	2770(4)	8274(3)	106(2)
C(2)	9676(7)	2404(4)	8381(3)	110(3)
C(3)	10965(7)	3217(4)	8399(4)	118(3)
C(4)	11590(8)	3632(5)	7992(4)	137(3)
C(5)	10278(8)	4395(4)	7630(4)	121(3)
C(6)	9418(7)	4553(3)	7136(4)	115(3)
C(7)	7957(6)	3943(4)	7467(4)	106(2)
C(8)	7505(6)	3290(4)	7486(3)	101(2)
C(9)	11289(7)	2199(4)	7959(4)	127(3)
C(10)	10986(8)	1805(4)	7399(5)	130(3)
C(11)	11490(12)	926(7)	6802(7)	281(11)
C(12)	12120(30)	1164(19)	6279(9)	600(30)
C(13)	11637(8)	4057(5)	6943(5)	148(3)
C(14)	11923(10)	3515(6)	6566(6)	178(6)
C(15)	12702(14)	3190(17)	5590(15)	560(30)
C(16)	11722(18)	2888(12)	5310(14)	500(20)
C(17)	8204(7)	4088(4)	6372(3)	112(3)
C(18)	8984(10)	3977(4)	5879(4)	112(3)
C(19)	9311(12)	4149(6)	4817(5)	204(6)
C(20)	8860(15)	4511(6)	4290(5)	262(8)
C(21)	7839(5)	2200(3)	7387(3)	87(2)
C(22)	7827(5)	2148(3)	6683(3)	77(2)
C(23)	7080(3)	1673(1)	5713(1)	77(2)
C(24)	7897(3)	1337(1)	5377(1)	80(2)
C(25)	7575(3)	942(1)	4886(1)	104(2)
C(26)	8311(3)	617(1)	4568(1)	133(3)
C(27)	9369(3)	686(1)	4739(1)	126(3)
C(28)	9691(3)	1081(1)	5229(1)	131(3)
C(29)	8956(3)	1406(1)	5548(1)	104(2)
C(30)	6791(6)	2322(3)	5447(4)	89(2)
C(31)	7284(7)	2583(4)	4966(4)	111(3)
C(32)	7018(12)	3168(5)	4722(5)	147(4)
C(33)	6233(14)	3487(6)	4963(7)	169(6)
C(34)	5708(10)	3240(6)	5433(8)	171(5)
C(35)	5975(7)	2641(4)	5680(5)	137(3)
C(36)	6813(2)	1175(1)	6758(1)	80(2)
C(37)	6192(2)	227(1)	6111(1)	88(2)
C(38)	6991(2)	530(1)	6460(1)	76(2)
C(39)	7985(2)	270(1)	6520(1)	89(2)
C(40)	8180(2)	-294(1)	6230(1)	96(2)
C(41)	7381(2)	-598(1)	5881(1)	83(2)
C(42)	6387(2)	-337(1)	5821(1)	79(2)
C(43)	5588(2)	-641(1)	5471(1)	113(2)
C(44)	5783(2)	-1205(1)	5181(1)	137(3)
C(45)	6777(2)	-1465(1)	5241(1)	133(3)
C(46)	7576(2)	-1162(1)	5591(1)	116(3)
C(47)	5667(7)	1322(3)	6856(4)	92(2)
C(48)	4386(3)	1460(3)	7658(3)	166(5)

243

C(49)	4400(3)	2149(3)	7866(3)	139(4)
C(50)	4473(3)	2303(3)	8498(3)	187(5)
C(51)	4518(3)	2926(3)	8679(3)	215(8)
C(52)	4491(3)	3393(3)	8229(3)	210(8)
C(53)	4418(3)	3239(3)	7598(3)	175(4)
C(54)	4373(3)	2617(3)	7416(3)	161(4)
N(1)	8275(4)	2795(2)	7608(2)	79(1)
N(2)	10503(5)	2673(3)	8060(3)	98(2)
N(3)	10964(6)	3875(3)	7449(3)	111(2)
N(4)	8717(5)	4012(2)	6993(2)	93(2)
N(5)	7265(4)	1683(2)	6396(2)	70(1)
N(6)	5438(5)	1277(3)	7466(3)	111(2)
O(1)	11658(6)	1333(3)	7377(4)	187(3)
O(2)	10266(5)	1880(3)	7032(3)	145(2)
O(3)	12503(9)	3646(6)	6108(6)	270(5)
O(4)	11637(7)	2984(5)	6649(4)	178(3)
O(5)	8660(6)	4263(3)	5344(3)	150(2)
O(6)	9758(5)	3663(3)	5950(3)	124(2)
O(7)	8303(3)	2532(2)	6379(2)	82(1)
O(8)	5040(4)	1468(3)	6439(3)	121(2)
Na(1)	9767(2)	3011(1)	6942(1)	85(1)
O(9)	7407(18)	904(8)	8473(8)	128(5)
O(10)	6705(15)	1060(7)	8547(7)	195(6)
O(11)	7941(16)	739(7)	8223(8)	176(5)
O(12)	8680(20)	776(11)	8031(11)	208(9)
O(13)	8812(17)	979(12)	9298(10)	230(8)
O(14)	8578(13)	342(9)	9272(8)	194(6)
O(15)	7574(14)	675(8)	9610(8)	209(6)
O(16)	8364(16)	1447(10)	9401(9)	223(7)

Table 3. Bond lengths [Å] and angles [°] for m150mm.

C(1)-N(1)	1.470(8)	C(47)-N(6)	1.364(9)
C(1)-C(2)	1.557(11)	C(48)-N(6)	1.489(7)
C(2)-N(2)	1.426(9)	C(48)-C(49)	1.5424
C(3)-N(2)	1.478(9)	C(49)-C(54)	1.3900
C(3)-C(4)	1.514(11)	C(49)-C(50)	1.3901
C(4)-N(3)	1.465(10)	C(50)-C(51)	1.3900
C(5)-N(3)	1.487(10)	C(51)-C(52)	1.3900
C(5)-C(6)	1.517(11)	C(52)-C(53)	1.3900
C(6)-N(4)	1.487(9)	C(53)-C(54)	1.3900
C(7)-N(4)	1.461(9)	N(1)-Na(1)	2.505(5)
C(7)-C(8)	1.517(9)	N(2)-Na(1)	2.617(6)
C(8)-N(1)	1.462(8)	N(3)-Na(1)	2.598(6)
C(9)-N(2)	1.457(10)	N(4)-Na(1)	2.540(6)
C(9)-C(10)	1.498(12)	O(2)-Na(1)	2.514(7)
C(10)-O(2)	1.184(10)	O(4)-Na(1)	2.521(9)
C(10)-O(1)	1.331(9)	O(6)-Na(1)	2.544(7)
C(11)-O(1)	1.514(8)	O(7)-Na(1)	2.394(5)
C(11)-C(12)	1.515(10)	O(9)-O(11)	0.965(19)
C(13)-C(14)	1.476(14)	O(9)-O(10)	0.98(2)
C(13)-N(3)	1.482(11)	O(11)-O(12)	1.06(2)
C(14)-O(4)	1.215(5)	O(13)-O(16)	1.19(2)
C(14)-O(3)	1.301(14)	O(13)-O(14)	1.40(2)
C(14)-Na(1)	3.126(13)	O(14)-O(15)	1.68(2)
C(15)-C(16)	1.502(10)		
C(15)-O(3)	1.513(9)	N(1)-C(1)-C(2)	111.6(6)
C(17)-N(4)	1.455(8)	N(2)-C(2)-C(1)	112.7(6)
C(17)-C(18)	1.520(12)	N(2)-C(3)-C(4)	113.0(7)
C(18)-O(6)	1.201(10)	N(3)-C(4)-C(3)	112.5(7)
C(18)-O(5)	1.342(10)	N(3)-C(5)-C(6)	113.6(6)
C(19)-C(20)	1.457(16)	N(4)-C(6)-C(5)	111.7(6)
C(19)-O(5)	1.468(12)	N(4)-C(7)-C(8)	112.7(6)
C(21)-N(1)	1.459(7)	N(1)-C(8)-C(7)	114.8(6)
C(21)-C(22)	1.512(9)	N(2)-C(9)-C(10)	111.6(7)
C(22)-O(7)	1.236(7)	O(2)-C(10)-O(1)	123.9(9)
C(22)-N(5)	1.353(7)	O(2)-C(10)-C(9)	127.2(8)
C(22)-Na(1)	3.121(8)	O(1)-C(10)-C(9)	108.9(9)
C(23)-N(5)	1.468(5)	O(1)-C(11)-C(12)	110.7(9)
C(23)-C(24)	1.4965	C(14)-C(13)-N(3)	111.9(8)
C(23)-C(30)	1.539(8)	O(4)-C(14)-O(3)	120.4(14)
C(24)-C(29)	1.3900	O(4)-C(14)-C(13)	124.7(12)
C(24)-C(25)	1.3901	O(3)-C(14)-C(13)	114.9(11)
C(25)-C(26)	1.3900	O(4)-C(14)-Na(1)	49.9(7)
C(26)-C(27)	1.3900	O(3)-C(14)-Na(1)	146.1(10)
C(27)-C(28)	1.3900	C(13)-C(14)-Na(1)	82.7(6)
C(28)-C(29)	1.3900	C(16)-C(15)-O(3)	113.2(9)
C(30)-C(31)	1.367(10)	N(4)-C(17)-C(18)	109.9(7)
C(30)-C(35)	1.371(11)	O(6)-C(18)-O(5)	124.6(10)
C(31)-C(32)	1.392(12)	O(6)-C(18)-C(17)	125.3(8)
C(32)-C(33)	1.349(17)	O(5)-C(18)-C(17)	110.0(10)
C(33)-C(34)	1.355(17)	C(20)-C(19)-O(5)	106.9(12)
C(34)-C(35)	1.423(14)	N(1)-C(21)-C(22)	111.6(5)
C(36)-N(5)	1.479(5)	O(7)-C(22)-N(5)	121.3(6)
C(36)-C(47)	1.531(9)	O(7)-C(22)-C(21)	120.2(6)
C(36)-C(38)	1.5464	N(5)-C(22)-C(21)	118.5(6)
C(37)-C(38)	1.3900	O(7)-C(22)-Na(1)	44.1(3)
C(37)-C(42)	1.3900	N(5)-C(22)-Na(1)	155.8(4)
C(38)-C(39)	1.3900	C(21)-C(22)-Na(1)	79.8(4)
C(39)-C(40)	1.3900	N(5)-C(23)-C(24)	114.5(2)
C(40)-C(41)	1.3900	N(5)-C(23)-C(30)	112.1(4)
C(41)-C(46)	1.3900	C(24)-C(23)-C(30)	114.7(4)
C(41)-C(42)	1.3900	C(29)-C(24)-C(25)	120.0
C(42)-C(43)	1.3900	C(29)-C(24)-C(23)	121.6
C(43)-C(44)	1.3900	C(25)-C(24)-C(23)	118.4
C(44)-C(45)	1.3900	C(26)-C(25)-C(24)	120.0
C(45)-C(46)	1.3900	C(25)-C(26)-C(27)	120.0
C(47)-O(8)	1.203(8)	C(28)-C(27)-C(26)	120.0
		C(27)-C(28)-C(29)	120.0

C(24)-C(29)-C(28)	120.0
C(31)-C(30)-C(35)	118.3(8)
C(31)-C(30)-C(23)	122.5(7)
C(35)-C(30)-C(23)	119.2(8)
C(30)-C(31)-C(32)	122.6(10)
C(33)-C(32)-C(31)	118.9(12)
C(32)-C(33)-C(34)	120.6(13)
C(33)-C(34)-C(35)	120.6(13)
C(30)-C(35)-C(34)	119.1(11)
N(5)-C(36)-C(47)	109.5(4)
N(5)-C(36)-C(38)	111.6(2)
C(47)-C(36)-C(38)	114.2(3)
C(38)-C(37)-C(42)	120.0
C(37)-C(38)-C(39)	120.0
C(37)-C(38)-C(36)	121.1
C(39)-C(38)-C(36)	118.8
C(40)-C(39)-C(38)	120.0
C(41)-C(40)-C(39)	120.0
C(40)-C(41)-C(46)	120.0
C(40)-C(41)-C(42)	120.0
C(46)-C(41)-C(42)	120.0
C(43)-C(42)-C(41)	120.0
C(43)-C(42)-C(37)	120.0
C(41)-C(42)-C(37)	120.0
C(42)-C(43)-C(44)	120.0
C(45)-C(44)-C(43)	120.0
C(44)-C(45)-C(46)	120.0
C(41)-C(46)-C(45)	120.0
O(8)-C(47)-N(6)	123.7(7)
O(8)-C(47)-C(36)	123.5(6)
N(6)-C(47)-C(36)	112.8(7)
N(6)-C(48)-C(49)	109.8(3)
C(54)-C(49)-C(50)	120.0
C(54)-C(49)-C(48)	119.4
C(50)-C(49)-C(48)	120.6
C(51)-C(50)-C(49)	120.0
C(50)-C(51)-C(52)	120.0
C(53)-C(52)-C(51)	120.0
C(52)-C(53)-C(54)	120.0
C(49)-C(54)-C(53)	120.0
C(21)-N(1)-C(8)	109.7(5)
C(21)-N(1)-C(1)	111.8(5)
C(8)-N(1)-C(1)	111.3(6)
C(21)-N(1)-Na(1)	105.5(4)
C(8)-N(1)-Na(1)	107.2(4)
C(1)-N(1)-Na(1)	111.1(4)
C(2)-N(2)-C(9)	109.4(7)
C(2)-N(2)-C(3)	111.7(6)
C(9)-N(2)-C(3)	111.4(6)
C(2)-N(2)-Na(1)	108.6(4)
C(9)-N(2)-Na(1)	105.6(4)
C(3)-N(2)-Na(1)	109.9(5)
C(4)-N(3)-C(13)	111.0(8)
C(4)-N(3)-C(5)	111.2(7)
C(13)-N(3)-C(5)	112.2(8)
C(4)-N(3)-Na(1)	110.8(5)
C(13)-N(3)-Na(1)	103.8(5)
C(5)-N(3)-Na(1)	107.5(5)
C(17)-N(4)-C(7)	111.4(6)
C(17)-N(4)-C(6)	109.3(6)
C(7)-N(4)-C(6)	111.1(6)
C(17)-N(4)-Na(1)	105.3(4)
C(7)-N(4)-Na(1)	108.9(4)
C(6)-N(4)-Na(1)	110.7(4)
C(22)-N(5)-C(23)	120.6(5)
C(22)-N(5)-C(36)	121.4(5)
C(23)-N(5)-C(36)	118.1(4)
C(47)-N(6)-C(48)	120.2(7)
C(10)-O(1)-C(11)	114.2(8)
C(10)-O(2)-Na(1)	111.4(6)
C(14)-O(3)-C(15)	123.2(17)
C(14)-O(4)-Na(1)	108.4(9)
C(18)-O(5)-C(19)	115.0(9)
C(18)-O(6)-Na(1)	111.7(6)
C(22)-O(7)-Na(1)	114.8(4)
O(7)-Na(1)-N(1)	66.72(16)
O(7)-Na(1)-O(2)	79.11(19)
N(1)-Na(1)-O(2)	88.8(2)
O(7)-Na(1)-O(4)	126.2(3)
N(1)-Na(1)-O(4)	156.0(3)
O(2)-Na(1)-O(4)	75.8(3)
O(7)-Na(1)-N(4)	89.13(18)
N(1)-Na(1)-N(4)	72.8(2)
O(2)-Na(1)-N(4)	161.0(3)
O(4)-Na(1)-N(4)	123.1(3)
O(7)-Na(1)-O(6)	81.69(19)
N(1)-Na(1)-O(6)	128.1(2)
O(2)-Na(1)-O(6)	125.5(2)
O(4)-Na(1)-O(6)	75.8(3)
N(4)-Na(1)-O(6)	66.4(2)
O(7)-Na(1)-N(3)	159.8(2)
N(1)-Na(1)-N(3)	110.3(2)
O(2)-Na(1)-N(3)	121.1(3)
O(4)-Na(1)-N(3)	64.9(3)
N(4)-Na(1)-N(3)	71.3(2)
O(6)-Na(1)-N(3)	85.8(2)
O(7)-Na(1)-N(2)	124.4(2)
N(1)-Na(1)-N(2)	70.8(2)
O(2)-Na(1)-N(2)	65.7(2)
O(4)-Na(1)-N(2)	86.0(3)
N(4)-Na(1)-N(2)	110.8(2)
O(6)-Na(1)-N(2)	153.9(2)
N(3)-Na(1)-N(2)	69.4(2)
O(7)-Na(1)-C(22)	21.07(14)
N(1)-Na(1)-C(22)	50.33(17)
O(2)-Na(1)-C(22)	68.7(2)
O(4)-Na(1)-C(22)	134.6(3)
N(4)-Na(1)-C(22)	95.20(19)
O(6)-Na(1)-C(22)	102.6(2)
N(3)-Na(1)-C(22)	160.0(2)
N(2)-Na(1)-C(22)	103.5(2)
O(7)-Na(1)-C(14)	134.1(3)
N(1)-Na(1)-C(14)	159.0(3)
O(2)-Na(1)-C(14)	97.4(3)
O(4)-Na(1)-C(14)	21.6(2)
N(4)-Na(1)-C(14)	101.5(3)
O(6)-Na(1)-C(14)	63.2(3)
N(3)-Na(1)-C(14)	49.6(3)
N(2)-Na(1)-C(14)	93.5(3)
C(22)-Na(1)-C(14)	150.2(3)
O(11)-O(9)-O(10)	155(2)
O(9)-O(11)-O(12)	151(3)
O(16)-O(13)-O(14)	137(2)
O(13)-O(14)-O(15)	74.7(14)

Symmetry transformations used to generate equivalent atoms

245

:

Table 4.   Anisotropic displacement parameters (Å^2x 10^3)for m150mm. The anisotropic displacement factor exponent takes the form: -2 ^{2}[$h^2a^{*2}U^{11}$ + ... + 2 h k a* b* U^{12}]

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
C(1)	119(7)	122(6)	78(5)	-19(4)	18(4)	-41(5)
C(2)	121(7)	131(6)	72(5)	-8(4)	-26(5)	1(6)
C(3)	111(6)	146(7)	92(5)	-11(5)	-31(5)	-12(5)
C(4)	132(7)	172(8)	100(6)	-8(6)	-27(6)	-47(7)
C(5)	145(7)	111(6)	106(6)	-22(5)	-7(6)	-57(6)
C(6)	164(8)	69(5)	110(6)	-17(4)	4(6)	-15(5)
C(7)	100(6)	108(6)	112(6)	-28(5)	9(5)	12(5)
C(8)	83(5)	118(6)	104(5)	-32(5)	13(4)	-6(5)
C(9)	118(7)	138(7)	120(6)	-24(6)	-43(5)	15(6)
C(10)	118(7)	120(7)	149(8)	-12(6)	-14(7)	44(6)
C(11)	271(18)	228(15)	320(20)	-131(16)	-137(17)	115(14)
C(12)	730(70)	760(70)	300(30)	-30(40)	30(40)	-220(60)
C(13)	144(8)	166(9)	137(8)	-13(7)	27(7)	-56(7)
C(14)	124(10)	280(20)	133(9)	7(13)	50(7)	-36(12)
C(15)	260(20)	840(60)	580(40)	-580(40)	70(30)	0(30)
C(16)	460(40)	420(30)	580(40)	-340(30)	-160(30)	190(30)
C(17)	137(7)	99(6)	100(6)	-3(4)	-2(6)	32(5)
C(18)	157(9)	87(6)	91(6)	-5(5)	-5(7)	-16(6)
C(19)	315(17)	224(12)	74(6)	12(7)	27(8)	-49(11)
C(20)	460(30)	204(12)	112(8)	33(8)	-36(11)	-82(14)
C(21)	88(5)	98(5)	78(4)	-8(4)	9(3)	-23(4)
C(22)	65(4)	88(5)	76(4)	-5(4)	-4(4)	5(4)
C(23)	72(4)	83(5)	74(4)	-4(3)	-10(3)	2(3)
C(24)	90(5)	88(5)	62(4)	4(3)	7(4)	5(4)
C(25)	138(7)	87(5)	87(5)	-12(4)	16(5)	7(5)
C(26)	212(10)	94(6)	98(6)	-8(4)	48(7)	9(7)
C(27)	147(8)	117(7)	119(7)	37(5)	56(6)	52(6)
C(28)	127(7)	156(8)	115(6)	21(6)	42(6)	24(6)
C(29)	96(6)	130(6)	89(5)	-2(4)	27(4)	9(5)
C(30)	95(6)	80(5)	88(5)	-11(4)	-26(4)	-6(4)
C(31)	146(7)	96(6)	86(5)	1(5)	-30(5)	-6(5)
C(32)	224(13)	94(8)	116(7)	18(6)	-48(8)	-8(7)
C(33)	207(15)	107(9)	179(13)	21(9)	-85(11)	20(9)
C(34)	148(10)	125(10)	234(15)	5(9)	-26(10)	45(7)
C(35)	112(7)	125(7)	171(8)	11(6)	-13(6)	47(6)
C(36)	77(5)	93(5)	71(4)	-16(4)	12(3)	-8(3)
C(37)	84(5)	88(5)	93(5)	-2(4)	4(4)	1(4)
C(38)	88(5)	80(4)	61(4)	0(3)	16(3)	-13(4)
C(39)	67(4)	112(5)	90(4)	-1(4)	10(4)	-3(4)
C(40)	93(5)	100(5)	99(5)	3(4)	32(4)	8(4)
C(41)	113(6)	65(4)	74(4)	-6(3)	26(4)	-11(4)
C(42)	86(5)	78(4)	74(4)	-1(3)	10(4)	-1(4)
C(43)	130(7)	88(5)	115(6)	-19(4)	-19(5)	1(5)
C(44)	180(9)	99(6)	127(7)	-2(5)	-29(7)	-5(6)
C(45)	205(11)	92(6)	101(6)	-23(5)	11(7)	-5(6)
C(46)	134(7)	111(6)	106(6)	9(5)	34(5)	10(5)
C(47)	96(6)	94(5)	89(5)	-29(4)	30(5)	-17(4)
C(48)	131(8)	190(10)	190(9)	-86(8)	110(7)	-75(7)
C(49)	92(6)	172(10)	160(9)	-69(8)	57(6)	-26(6)
C(50)	178(11)	238(13)	148(9)	-57(9)	36(8)	-23(10)
C(51)	167(11)	229(15)	254(16)	-152(14)	57(10)	-41(10)
C(52)	111(8)	239(16)	284(18)	-136(15)	39(10)	-18(8)
C(53)	118(8)	164(10)	243(14)	-13(9)	17(9)	46(7)
C(54)	137(8)	163(10)	188(11)	-61(9)	41(8)	2(7)
N(1)	78(4)	89(4)	71(3)	-16(3)	2(3)	-9(3)
N(2)	92(4)	112(5)	88(4)	-13(4)	-5(4)	-8(4)
N(3)	115(5)	116(5)	104(5)	-22(4)	20(4)	-34(4)
N(4)	113(5)	83(4)	82(4)	-12(3)	-4(4)	5(3)
N(5)	69(3)	72(3)	67(3)	-5(3)	3(3)	-9(3)
N(6)	118(5)	103(5)	114(5)	-18(4)	23(4)	-30(4)
O(1)	173(6)	149(5)	226(8)	-54(5)	-83(6)	61(5)
O(2)	141(5)	154(5)	131(5)	-34(4)	-57(4)	52(4)
O(3)	228(10)	333(14)	259(11)	12(10)	97(9)	-68(10)
O(4)	131(6)	229(9)	177(7)	-23(7)	23(5)	2(6)
O(5)	228(7)	117(4)	100(4)	10(4)	-19(5)	-13(4)
O(6)	136(5)	133(5)	104(4)	-5(4)	20(4)	-4(4)
O(7)	86(3)	87(3)	73(3)	1(2)	-2(2)	-13(2)
O(8)	83(4)	166(5)	113(4)	-23(4)	6(3)	5(3)
Na(1)	78(2)	87(2)	90(2)	-10(1)	3(1)	-3(1)

Table 5.   Hydrogen coordinates (x 10^4) and isotropic displacement parameters (Å^2x 10^3) for m150mm.

	x	y	z	U(eq)
H(1A)	8099	2570	8506	127
H(1B)	8727	3191	8433	127
H(2A)	9877	2392	8825	132
H(2B)	9568	1977	8239	132
H(3A)	11416	3071	8751	142
H(3B)	10409	3461	8565	142
H(4A)	11872	3979	8240	164
H(4B)	12174	3397	7848	164
H(5A)	10704	4763	7715	145
H(5B)	9963	4282	8014	145
H(6A)	9010	4899	7281	138
H(6B)	9728	4684	6756	138
H(7A)	7393	4238	7382	128
H(7B)	8292	4042	7874	128
H(8A)	7008	3276	7807	122

H(8B)	7124	3207	7088	122
H(9A)	11377	1936	8327	153
H(9B)	11953	2401	7901	153
H(11A)	10755	925	6663	337
H(11B)	11696	502	6905	337
H(12A)	11996	905	5916	894
H(12B)	12853	1152	6411	894
H(12C)	11922	1585	6178	894
H(13A)	11271	4360	6673	178
H(13B)	12267	4253	7126	178
H(15A)	13042	3408	5263	670
H(15B)	13175	2868	5754	670
H(16A)	11902	2540	5057	745
H(16B)	11329	3186	5056	745
H(16C)	11308	2746	5639	745
H(17A)	7921	4506	6327	135
H(17B)	7631	3794	6314	135
H(19A)	10025	4280	4924	245
H(19B)	9313	3709	4713	245
H(20A)	9270	4453	3936	392
H(20B)	8857	4945	4401	392
H(20C)	8158	4374	4187	392
H(21A)	7132	2160	7516	105
H(21B)	8251	1862	7576	105
H(23)	6446	1423	5635	92
H(25)	6867	896	4772	124
H(26)	8096	352	4240	160
H(27)	9862	468	4525	151
H(28)	10399	1127	5343	157
H(29)	9171	1671	5875	125
H(31)	7819	2362	4794	133
H(32)	7377	3336	4399	177
H(33)	6049	3878	4806	202
H(34)	5168	3465	5596	205
H(35)	5601	2468	5994	164
H(36)	7186	1173	7173	96
H(37)	5527	401	6071	106
H(39)	8519	473	6754	107
H(40)	8845	-469	6270	115
H(43)	4923	-466	5431	135
H(44)	5249	-1408	4947	164
H(45)	6908	-1843	5047	159
H(46)	8241	-1336	5631	139
H(48A)	4192	1197	8000	199
H(48B)	3871	1403	7310	199
H(50)	4491	1991	8799	224
H(51)	4567	3029	9101	258
H(52)	4522	3810	8350	252
H(53)	4400	3552	7296	210
H(54)	4324	2513	6994	194
H(6)	5902	1143	7742	133

Appendix 1b: Benzyl Europium Ugi product M12M. Crystal structure
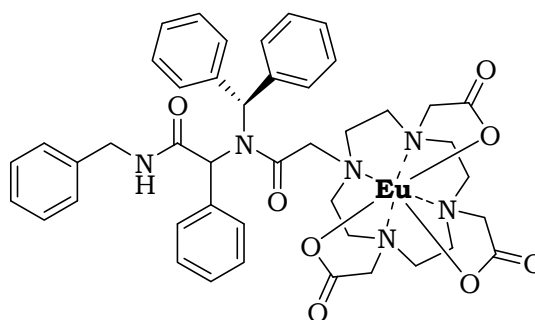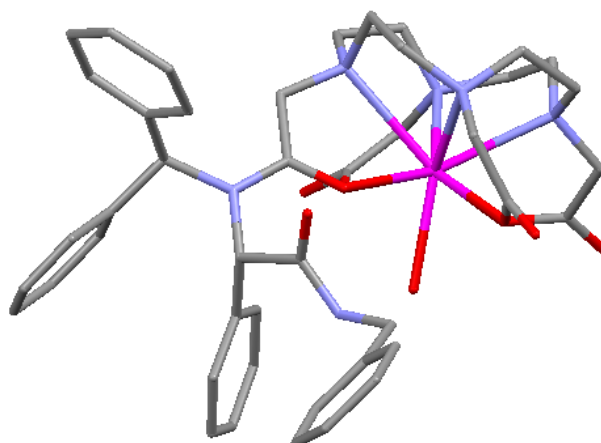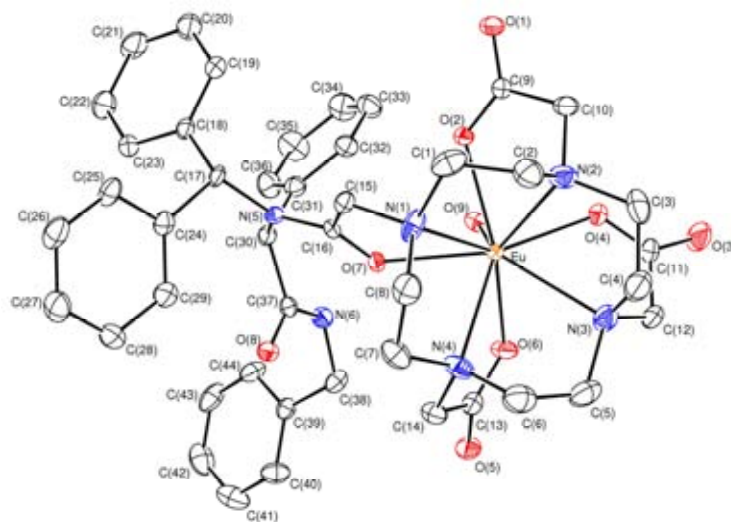
Table 1. Crystal data and structure refinement for 2009 src0262.

Identification code	2009src0262, B_2Eu
Empirical formula	$C_{44} H_{49} D_2 Eu N_6 O_9$, 9.84($D_2O$)
Formula weight	1158.97
Temperature	120(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	P 2(1)/c

Unit cell dimensions	a = 11.2327(2) Å	= 90°.
	b = 43.0382(9) Å	= 112.1750(10)°.
	c = 11.6011(2) Å	= 90°.

Volume	5193.56(17) $Å^3$
Z	4
Density (calculated)	1.482 Mg/m^3
Absorption coefficient	1.285 mm^{-1}
F(000)	2361
Crystal size	0.42 x 0.08 x 0.04 mm^3
Theta range for data collection	3.03 to 27.47°.
Index ranges	-14<=h<=14, -55<=k<=52, -15<=l<=14
Reflections collected	43192
Independent reflections	11813 [R(int) = 0.0652]
Completeness to theta = 27.47°	99.2 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.9504 and 0.6145
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	11813 / 27 / 685
Goodness-of-fit on F^2	1.041
Final R indices [I>2sigma(I)]	R1 = 0.0519, wR2 = 0.1363
R indices (all data)	R1 = 0.0672, wR2 = 0.1459
Largest diff. peak and hole	1.796 and -1.912 e.$Å^{-3}$

Notes:
The occupancy of the water molecule oxygen atom O(001) was initially refined and then fixed at the final value, 84 %.
The deuterium atoms belonging to the water molecules O(901) and O(001) could not be successfully located in the electron density map. Thus, the hydrogen bonds involving these molecules could not be tabulated (see Table 6 for hydrogen bonds).
The hydrogen atoms in the europium complex were fixed as riding models while the deuterium atoms (belonging to both bound and free water molecules) were located in the electron density map. The positions of these deuterium atoms were refined subject to restraints while the isotropic thermal parameters were fixed at a value of $1.5U_{eq}$ of the parent oxygen atom.

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3)

for 2009src0262. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
C(1)	1972(5)	1057(1)	-139(4)	30(1)
C(2)	1878(5)	718(1)	-350(5)	32(1)
C(3)	1604(6)	203(1)	382(5)	33(1)
C(4)	2941(5)	124(1)	1239(5)	35(1)
C(5)	4658(5)	215(1)	3244(5)	34(1)
C(6)	5226(5)	505(1)	2992(5)	35(1)
C(7)	5041(5)	1078(1)	2733(5)	30(1)
C(8)	4180(5)	1103(1)	1394(5)	31(1)
C(9)	-456(4)	894(1)	473(4)	17(1)
C(10)	-77(4)	604(1)	-62(4)	24(1)
C(11)	1235(5)	31(1)	2833(5)	24(1)
C(12)	2647(5)	-34(1)	3099(5)	31(1)
C(13)	4611(4)	660(1)	5418(4)	22(1)
C(14)	5303(4)	836(1)	4712(4)	23(1)
C(15)	2549(4)	1457(1)	1527(4)	17(1)
C(16)	2571(4)	1489(1)	2846(4)	15(1)
C(17)	2189(4)	2056(1)	2468(4)	16(1)
C(18)	1174(4)	2267(1)	2616(4)	17(1)
C(19)	-127(5)	2203(1)	1941(5)	27(1)
C(20)	-1070(5)	2389(1)	2091(6)	35(1)
C(21)	-742(5)	2639(1)	2903(6)	34(1)
C(22)	559(5)	2706(1)	3591(5)	28(1)
C(23)	1498(4)	2521(1)	3428(4)	20(1)
C(24)	3462(4)	2224(1)	2675(4)	18(1)
C(25)	3432(5)	2470(1)	1876(5)	24(1)
C(26)	4543(5)	2638(1)	2031(5)	31(1)
C(27)	5693(5)	2560(1)	2985(5)	33(1)
C(28)	5740(5)	2315(1)	3759(5)	29(1)
C(29)	4620(5)	2143(1)	3608(5)	26(1)
C(30)	2299(4)	1792(1)	4495(4)	17(1)
C(31)	951(4)	1727(1)	4476(4)	20(1)

C(32)	78(4)	1531(1)	3615(4)	21(1)
C(33)	-1120(5)	1473(1)	3666(5)	28(1)
C(34)	-1460(5)	1606(1)	4576(5)	34(1)
C(35)	-591(5)	1804(2)	5450(5)	37(1)
C(36)	597(5)	1867(1)	5389(5)	29(1)
C(37)	3411(4)	1613(1)	5504(4)	18(1)
C(38)	4116(5)	1306(1)	7402(4)	22(1)
C(39)	4619(4)	1520(1)	8532(4)	19(1)
C(40)	5897(5)	1495(1)	9348(4)	27(1)
C(41)	6373(5)	1677(2)	10422(5)	39(1)
C(42)	5583(6)	1889(1)	10684(5)	38(1)
C(43)	4317(6)	1916(1)	9875(5)	34(1)
C(44)	3841(5)	1734(1)	8807(4)	26(1)
N(1)	2834(4)	1139(1)	1242(4)	26(1)
N(2)	1329(4)	550(1)	413(4)	29(1)
N(3)	3214(4)	199(1)	2529(4)	30(1)
N(4)	4724(4)	787(1)	3352(4)	26(1)
N(5)	2375(3)	1767(1)	3253(3)	15(1)
N(6)	3117(4)	1447(1)	6322(3)	20(1)
O(1)	-1565(3)	1002(1)	-98(3)	23(1)
O(2)	338(3)	1006(1)	1467(3)	17(1)
O(3)	560(4)	-193(1)	2915(4)	40(1)
O(4)	823(3)	306(1)	2555(3)	23(1)
O(5)	5192(3)	618(1)	6565(3)	25(1)
O(6)	3463(3)	572(1)	4811(3)	24(1)
O(7)	2742(3)	1256(1)	3529(3)	16(1)
O(8)	4534(3)	1651(1)	5568(3)	21(1)
O(9)	976(3)	830(1)	4001(3)	23(1)
D(9A)	1300(50)	877(14)	4850(20)	34
D(9B)	80(20)	803(13)	3720(50)	34
Eu	2213(1)	738(1)	2749(1)	15(1)
O(101)	1120(4)	-770(1)	3868(4)	28(1)
D(1WA)	1080(60)	-561(6)	3600(50)	42
D(1WB)	1580(50)	-879(11)	3460(50)	42
O(201)	2618(3)	5977(1)	7603(3)	26(1)
D(2WA)	2200(50)	5937(13)	6740(20)	40
D(2WB)	3390(30)	5860(12)	7860(50)	40
O(301)	1364(4)	970(1)	6473(3)	33(1)
D(3WA)	490(20)	960(13)	6360(60)	50
D(3WB)	1750(50)	774(8)	6740(60)	50
O(401)	-649(4)	1636(1)	-396(4)	35(1)
D(4WA)	-1270(50)	1668(14)	-1210(30)	53
D(4WB)	-950(60)	1457(10)	-100(50)	53
O(501)	1497(4)	5640(1)	1391(4)	39(1)
D(5WA)	1260(70)	5426(6)	1250(60)	58
D(5WB)	1600(70)	5712(13)	660(40)	58
O(601)	7306(4)	1642(1)	7069(4)	35(1)
D(6WA)	6420(20)	1694(13)	6770(60)	52
D(6WB)	7310(60)	1426(5)	6920(60)	52
O(701)	3145(5)	462(1)	7368(6)	63(2)
D(7WA)	3600(60)	490(20)	6830(60)	94
D(7WB)	3760(60)	380(20)	8150(40)	94
O(801)	3154(5)	5741(2)	3917(6)	81(2)
D(8WA)	3260(110)	5538(12)	4350(90)	121
D(8WB)	2600(90)	5849(19)	4240(90)	121
O(901)	4881(5)	327(1)	-235(6)	69(2)
O(001)	1200(7)	4994(1)	1209(7)	71(2)

Table 3.   Bond lengths [Å] and angles [°] for 2009src0262.

C(1)-C(2)	1.480(7)
C(1)-N(1)	1.567(7)
C(1)-H(1A)	0.9900
C(1)-H(1B)	0.9900
C(2)-N(2)	1.449(7)
C(2)-H(2A)	0.9900
C(2)-H(2B)	0.9900
C(3)-C(4)	1.493(8)
C(3)-N(2)	1.528(7)
C(3)-H(3A)	0.9900
C(3)-H(3B)	0.9900
C(4)-N(3)	1.446(7)
C(4)-H(4A)	0.9900
C(4)-H(4B)	0.9900
C(5)-C(6)	1.479(8)
C(5)-N(3)	1.519(7)
C(5)-H(5A)	0.9900
C(5)-H(5B)	0.9900
C(6)-N(4)	1.465(7)
C(6)-H(6A)	0.9900
C(6)-H(6B)	0.9900
C(7)-C(8)	1.492(7)
C(7)-N(4)	1.548(7)
C(7)-H(7A)	0.9900
C(7)-H(7B)	0.9900
C(8)-N(1)	1.462(7)
C(8)-H(8A)	0.9900
C(8)-H(8B)	0.9900
C(9)-O(2)	1.257(5)
C(9)-O(1)	1.259(5)
C(9)-C(10)	1.524(6)
C(9)-Eu	3.230(4)
C(10)-N(2)	1.481(6)
C(10)-H(10A)	0.9900
C(10)-H(10B)	0.9900
C(11)-O(3)	1.253(6)
C(11)-O(4)	1.267(5)
C(11)-C(12)	1.523(7)
C(11)-Eu	3.249(4)
C(12)-N(3)	1.471(7)
C(12)-H(12A)	0.9900
C(12)-H(12B)	0.9900
C(13)-O(5)	1.255(6)
C(13)-O(6)	1.272(5)
C(13)-C(14)	1.526(7)
C(13)-Eu	3.264(5)
C(14)-N(4)	1.478(6)
C(14)-H(14A)	0.9900
C(14)-H(14B)	0.9900
C(15)-N(1)	1.473(5)
C(15)-C(16)	1.528(6)
C(15)-H(15A)	0.9900
C(15)-H(15B)	0.9900
C(16)-O(7)	1.249(5)
C(16)-N(5)	1.335(5)
C(16)-Eu	3.255(4)
C(17)-N(5)	1.505(5)
C(17)-C(18)	1.519(6)
C(17)-C(24)	1.540(6)
C(17)-H(17)	1.0000
C(18)-C(23)	1.397(6)
C(18)-C(19)	1.401(6)
C(19)-C(20)	1.388(7)
C(19)-H(19)	0.9500
C(20)-C(21)	1.387(8)
C(20)-H(20)	0.9500
C(21)-C(22)	1.405(7)
C(21)-H(21)	0.9500
C(22)-C(23)	1.390(7)
C(22)-H(22)	0.9500
C(23)-H(23)	0.9500
C(24)-C(29)	1.387(6)
C(24)-C(25)	1.399(6)
C(25)-C(26)	1.394(7)
C(25)-H(25)	0.9500
C(26)-C(27)	1.388(8)
C(26)-H(26)	0.9500
C(27)-C(28)	1.374(7)
C(27)-H(27)	0.9500
C(28)-C(29)	1.412(7)
C(28)-H(28)	0.9500
C(29)-H(29)	0.9500
C(30)-N(5)	1.479(5)
C(30)-C(31)	1.532(6)
C(30)-C(37)	1.556(6)
C(30)-H(30)	1.0000
C(31)-C(32)	1.390(6)
C(31)-C(36)	1.401(7)
C(32)-C(33)	1.392(6)
C(32)-H(32)	0.9500
C(33)-C(34)	1.376(7)
C(33)-H(33)	0.9500
C(34)-C(35)	1.400(8)
C(34)-H(34)	0.9500
C(35)-C(36)	1.389(7)
C(35)-H(35)	0.9500
C(36)-H(36)	0.9500
C(37)-O(8)	1.247(5)
C(37)-N(6)	1.326(6)
C(38)-N(6)	1.462(6)
C(38)-C(39)	1.525(6)
C(38)-H(38A)	0.9900
C(38)-H(38B)	0.9900
C(39)-C(44)	1.387(6)
C(39)-C(40)	1.393(6)
C(40)-C(41)	1.398(7)
C(40)-H(40)	0.9500
C(41)-C(42)	1.383(9)
C(41)-H(41)	0.9500
C(42)-C(43)	1.382(8)
C(42)-H(42)	0.9500
C(43)-C(44)	1.391(7)
C(43)-H(43)	0.9500
C(44)-H(44)	0.9500
N(1)-Eu	2.728(4)
N(2)-Eu	2.638(4)
N(3)-Eu	2.633(4)
N(4)-Eu	2.646(4)
N(6)-H(6)	0.8800
O(2)-Eu	2.369(3)
O(4)-Eu	2.383(3)
O(6)-Eu	2.380(3)
O(7)-Eu	2.393(3)
O(9)-Eu	2.391(3)
O(9)-D(9A)	0.932(19)
O(9)-D(9B)	0.941(19)
O(101)-D(1WA)	0.949(19)
O(101)-D(1WB)	0.946(19)
O(201)-D(2WA)	0.952(19)
O(201)-D(2WB)	0.948(19)
O(301)-D(3WA)	0.942(19)
O(301)-D(3WB)	0.942(19)
O(401)-D(4WA)	0.950(19)
O(401)-D(4WB)	0.955(19)
O(501)-D(5WA)	0.957(19)
O(501)-D(5WB)	0.95(2)
O(601)-D(6WA)	0.950(19)
O(601)-D(6WB)	0.946(19)
O(701)-D(7WA)	0.95(2)
O(701)-D(7WB)	0.97(2)
O(801)-D(8WA)	0.99(2)
O(801)-D(8WB)	0.96(2)
C(2)-C(1)-N(1)	
C(2)-C(1)-H(1A)	109.3
N(1)-C(1)-H(1A)	109.3
C(2)-C(1)-H(1B)	109.3
N(1)-C(1)-H(1B)	109.3
H(1A)-C(1)-H(1B)	108.0
N(2)-C(2)-C(1)	114.3(4)
N(2)-C(2)-H(2A)	108.7
C(1)-C(2)-H(2A)	108.7
N(2)-C(2)-H(2B)	108.7
C(1)-C(2)-H(2B)	108.7
H(2A)-C(2)-H(2B)	107.6
C(4)-C(3)-N(2)	111.6(4)
C(4)-C(3)-H(3A)	109.3
N(2)-C(3)-H(3A)	109.3
C(4)-C(3)-H(3B)	109.3
N(2)-C(3)-H(3B)	109.3
H(3A)-C(3)-H(3B)	108.0
N(3)-C(4)-C(3)	113.9(5)
N(3)-C(4)-H(4A)	108.8
C(3)-C(4)-H(4A)	108.8
N(3)-C(4)-H(4B)	108.8
C(3)-C(4)-H(4B)	108.8
H(4A)-C(4)-H(4B)	107.7
C(6)-C(5)-N(3)	112.5(4)
C(6)-C(5)-H(5A)	109.1
N(3)-C(5)-H(5A)	109.1
C(6)-C(5)-H(5B)	109.1
N(3)-C(5)-H(5B)	109.1
H(5A)-C(5)-H(5B)	107.8
N(4)-C(6)-C(5)	113.7(4)
N(4)-C(6)-H(6A)	108.8
C(5)-C(6)-H(6A)	108.8
N(4)-C(6)-H(6B)	108.8
C(5)-C(6)-H(6B)	108.8
H(6A)-C(6)-H(6B)	107.7
C(8)-C(7)-N(4)	111.6(4)
C(8)-C(7)-H(7A)	109.3
N(4)-C(7)-H(7A)	109.3
C(8)-C(7)-H(7B)	109.3
N(4)-C(7)-H(7B)	109.3
H(7A)-C(7)-H(7B)	108.0
N(1)-C(8)-C(7)	111.6(4)
N(1)-C(8)-H(8A)	109.3
C(7)-C(8)-H(8A)	109.3
N(1)-C(8)-H(8B)	109.3
C(7)-C(8)-H(8B)	109.3
H(8A)-C(8)-H(8B)	108.0
O(2)-C(9)-O(1)	124.0(4)
O(2)-C(9)-C(10)	118.5(4)
O(1)-C(9)-C(10)	117.4(4)
O(2)-C(9)-Eu	38.0(2)
O(1)-C(9)-Eu	159.4(3)
C(10)-C(9)-Eu	81.2(2)
N(2)-C(10)-C(9)	113.5(4)

N(2)-C(10)-H(10A)	108.9	C(33)-C(34)-H(34)	120.3
C(9)-C(10)-H(10A)	108.9	C(35)-C(34)-H(34)	120.3
N(2)-C(10)-H(10B)	108.9	C(36)-C(35)-C(34)	120.1(5)
C(9)-C(10)-H(10B)	108.9	C(36)-C(35)-H(35)	119.9
H(10A)-C(10)-H(10B)	107.7	C(34)-C(35)-H(35)	119.9
O(3)-C(11)-O(4)	124.5(5)	C(35)-C(36)-C(31)	120.4(5)
O(3)-C(11)-C(12)	117.2(4)	C(35)-C(36)-H(36)	119.8
O(4)-C(11)-C(12)	118.3(4)	C(31)-C(36)-H(36)	119.8
O(3)-C(11)-Eu	160.9(4)	O(8)-C(37)-N(6)	123.2(4)
O(4)-C(11)-Eu	38.1(2)	O(8)-C(37)-C(30)	119.1(4)
C(12)-C(11)-Eu	80.8(3)	N(6)-C(37)-C(30)	117.4(4)
N(3)-C(12)-C(11)	112.3(4)	N(6)-C(38)-C(39)	113.7(4)
N(3)-C(12)-H(12A)	109.2	N(6)-C(38)-H(38A)	108.8
C(11)-C(12)-H(12A)	109.2	C(39)-C(38)-H(38A)	108.8
N(3)-C(12)-H(12B)	109.2	N(6)-C(38)-H(38B)	108.8
C(11)-C(12)-H(12B)	109.2	C(39)-C(38)-H(38B)	108.8
H(12A)-C(12)-H(12B)	107.9	H(38A)-C(38)-H(38B)	107.7
O(5)-C(13)-O(6)	123.9(4)	C(44)-C(39)-C(40)	118.2(4)
O(5)-C(13)-C(14)	118.2(4)	C(44)-C(39)-C(38)	122.5(4)
O(6)-C(13)-C(14)	117.8(4)	C(40)-C(39)-C(38)	119.3(4)
O(5)-C(13)-Eu	159.0(3)	C(39)-C(40)-C(41)	120.8(5)
O(6)-C(13)-Eu	37.3(2)	C(39)-C(40)-H(40)	119.6
C(14)-C(13)-Eu	81.1(2)	C(41)-C(40)-H(40)	119.6
N(4)-C(14)-C(13)	113.2(4)	C(42)-C(41)-C(40)	120.3(5)
N(4)-C(14)-H(14A)	108.9	C(42)-C(41)-H(41)	119.8
C(13)-C(14)-H(14A)	108.9	C(40)-C(41)-H(41)	119.8
N(4)-C(14)-H(14B)	108.9	C(43)-C(42)-C(41)	119.1(5)
C(13)-C(14)-H(14B)	108.9	C(43)-C(42)-H(42)	120.4
H(14A)-C(14)-H(14B)	107.7	C(41)-C(42)-H(42)	120.4
N(1)-C(15)-C(16)	112.8(3)	C(42)-C(43)-C(44)	120.6(5)
N(1)-C(15)-H(15A)	109.0	C(42)-C(43)-H(43)	119.7
C(16)-C(15)-H(15A)	109.0	C(44)-C(43)-H(43)	119.7
N(1)-C(15)-H(15B)	109.0	C(39)-C(44)-C(43)	121.0(5)
C(16)-C(15)-H(15B)	109.0	C(39)-C(44)-H(44)	119.5
H(15A)-C(15)-H(15B)	107.8	C(43)-C(44)-H(44)	119.5
O(7)-C(16)-N(5)	120.3(4)	C(8)-N(1)-C(15)	112.0(4)
O(7)-C(16)-C(15)	120.2(4)	C(8)-N(1)-C(1)	108.1(4)
N(5)-C(16)-C(15)	119.5(4)	C(15)-N(1)-C(1)	109.6(4)
O(7)-C(16)-Eu	37.81(19)	C(8)-N(1)-Eu	111.0(3)
N(5)-C(16)-Eu	150.3(3)	C(15)-N(1)-Eu	108.3(3)
C(15)-C(16)-Eu	85.4(2)	C(1)-N(1)-Eu	107.8(3)
N(5)-C(17)-C(18)	111.4(3)	C(2)-N(2)-C(10)	109.5(4)
N(5)-C(17)-C(24)	112.6(3)	C(2)-N(2)-C(3)	109.6(4)
C(18)-C(17)-C(24)	113.1(4)	C(10)-N(2)-C(3)	110.0(4)
N(5)-C(17)-H(17)	106.4	C(2)-N(2)-Eu	113.1(3)
C(18)-C(17)-H(17)	106.4	C(10)-N(2)-Eu	105.6(3)
C(24)-C(17)-H(17)	106.4	C(3)-N(2)-Eu	108.9(3)
C(23)-C(18)-C(19)	118.8(4)	C(4)-N(3)-C(12)	110.9(4)
C(23)-C(18)-C(17)	121.9(4)	C(4)-N(3)-C(5)	109.9(4)
C(19)-C(18)-C(17)	119.3(4)	C(12)-N(3)-C(5)	109.9(4)
C(20)-C(19)-C(18)	120.1(5)	C(4)-N(3)-Eu	111.4(3)
C(20)-C(19)-H(19)	120.0	C(12)-N(3)-Eu	106.4(3)
C(18)-C(19)-H(19)	120.0	C(5)-N(3)-Eu	108.2(3)
C(21)-C(20)-C(19)	120.8(5)	C(6)-N(4)-C(14)	111.2(4)
C(21)-C(20)-H(20)	119.6	C(6)-N(4)-C(7)	111.0(4)
C(19)-C(20)-H(20)	119.6	C(14)-N(4)-C(7)	107.2(4)
C(20)-C(21)-C(22)	119.8(5)	C(6)-N(4)-Eu	110.2(3)
C(20)-C(21)-H(21)	120.1	C(14)-N(4)-Eu	106.8(3)
C(22)-C(21)-H(21)	120.1	C(7)-N(4)-Eu	110.5(3)
C(23)-C(22)-C(21)	119.1(5)	C(16)-N(5)-C(30)	119.1(3)
C(23)-C(22)-H(22)	120.4	C(16)-N(5)-C(17)	121.9(3)
C(21)-C(22)-H(22)	120.4	C(30)-N(5)-C(17)	118.9(3)
C(22)-C(23)-C(18)	121.4(4)	C(37)-N(6)-C(38)	121.3(4)
C(22)-C(23)-H(23)	119.3	C(37)-N(6)-H(6)	119.3
C(18)-C(23)-H(23)	119.3	C(38)-N(6)-H(6)	119.3
C(29)-C(24)-C(25)	119.3(4)	C(9)-O(2)-Eu	122.9(3)
C(29)-C(24)-C(17)	123.1(4)	C(11)-O(4)-Eu	122.8(3)
C(25)-C(24)-C(17)	117.7(4)	C(13)-O(6)-Eu	123.8(3)
C(26)-C(25)-C(24)	120.7(5)	C(16)-O(7)-Eu	123.5(3)
C(26)-C(25)-H(25)	119.7	Eu-O(9)-D(9A)	126(3)
C(24)-C(25)-H(25)	119.7	Eu-O(9)-D(9B)	124(3)
C(27)-C(26)-C(25)	119.7(5)	D(9A)-O(9)-D(9B)	109(4)
C(27)-C(26)-H(26)	120.1	O(2)-Eu-O(6)	145.54(11)
C(25)-C(26)-H(26)	120.1	O(2)-Eu-O(4)	85.68(10)
C(28)-C(27)-C(26)	120.1(5)	O(6)-Eu-O(4)	88.69(11)
C(28)-C(27)-H(27)	120.0	O(2)-Eu-O(9)	73.62(11)
C(26)-C(27)-H(27)	120.0	O(6)-Eu-O(9)	72.39(11)
C(27)-C(28)-C(29)	120.6(5)	O(4)-Eu-O(9)	71.14(11)
C(27)-C(28)-H(28)	119.7	O(2)-Eu-O(7)	79.61(10)
C(29)-C(28)-H(28)	119.7	O(6)-Eu-O(7)	86.11(11)
C(24)-C(29)-C(28)	119.6(5)	O(4)-Eu-O(7)	145.31(11)
C(24)-C(29)-H(29)	120.2	O(9)-Eu-O(7)	74.61(11)
C(28)-C(29)-H(29)	120.2	O(2)-Eu-N(3)	132.74(12)
N(5)-C(30)-C(31)	113.2(3)	O(6)-Eu-N(3)	73.81(13)
N(5)-C(30)-C(37)	111.8(3)	O(4)-Eu-N(3)	65.89(12)
C(31)-C(30)-C(37)	115.4(4)	O(9)-Eu-N(3)	124.92(13)
N(5)-C(30)-H(30)	105.1	O(7)-Eu-N(3)	143.42(12)
C(31)-C(30)-H(30)	105.1	O(2)-Eu-N(2)	66.64(11)
C(37)-C(30)-H(30)	105.1	O(6)-Eu-N(2)	142.21(13)
C(32)-C(31)-C(36)	118.7(4)	O(4)-Eu-N(2)	71.57(13)
C(32)-C(31)-C(30)	123.2(4)	O(9)-Eu-N(2)	126.51(12)
C(36)-C(31)-C(30)	118.1(4)	O(7)-Eu-N(2)	128.17(12)
C(31)-C(32)-C(33)	120.7(4)	N(3)-Eu-N(2)	68.78(14)
C(31)-C(32)-H(32)	119.7	O(2)-Eu-N(4)	137.18(12)
C(33)-C(32)-H(32)	119.7	O(6)-Eu-N(4)	66.00(12)
C(34)-C(33)-C(32)	120.6(5)	O(4)-Eu-N(4)	133.02(12)
C(34)-C(33)-H(33)	119.7	O(9)-Eu-N(4)	129.22(11)
C(32)-C(33)-H(33)	119.7	O(7)-Eu-N(4)	74.81(12)
C(33)-C(34)-C(35)	119.4(5)		

N(3)-Eu-N(4)	69.17(13)	N(3)-Eu-C(11)	48.62(13)
N(2)-Eu-N(4)	104.26(13)	N(2)-Eu-C(11)	75.06(13)
O(2)-Eu-N(1)	70.84(12)	N(4)-Eu-C(11)	113.94(13)
O(6)-Eu-N(1)	129.62(12)	N(1)-Eu-C(11)	141.47(12)
O(4)-Eu-N(1)	138.55(12)	C(9)-Eu-C(11)	89.92(11)
O(9)-Eu-N(1)	129.23(12)	O(2)-Eu-C(16)	66.88(10)
O(7)-Eu-N(1)	64.34(11)	O(6)-Eu-C(16)	104.22(11)
N(3)-Eu-N(1)	105.83(13)	O(4)-Eu-C(16)	147.80(10)
N(2)-Eu-N(1)	67.90(13)	O(9)-Eu-C(16)	84.65(11)
N(4)-Eu-N(1)	67.27(13)	O(7)-Eu-C(16)	18.65(10)
O(2)-Eu-C(9)	19.06(10)	N(3)-Eu-C(16)	145.82(12)
O(6)-Eu-C(9)	153.77(11)	N(2)-Eu-C(16)	109.55(12)
O(4)-Eu-C(9)	73.25(11)	N(4)-Eu-C(16)	78.82(12)
O(9)-Eu-C(9)	83.64(11)	N(1)-Eu-C(16)	48.41(11)
O(7)-Eu-C(9)	97.90(10)	C(9)-Eu-C(16)	83.49(10)
N(3)-Eu-C(9)	113.69(13)	C(11)-Eu-C(16)	165.48(12)
N(2)-Eu-C(9)	49.48(12)	D(1WA)-O(101)-D(1WB)	106(3)
N(4)-Eu-C(9)	140.10(12)	D(2WA)-O(201)-D(2WB)	105(3)
N(1)-Eu-C(9)	74.20(12)	D(3WA)-O(301)-D(3WB)	109(4)
O(2)-Eu-C(11)	104.11(11)	D(4WA)-O(401)-D(4WB)	104(4)
O(6)-Eu-C(11)	76.48(12)	D(5WA)-O(501)-D(5WB)	106(4)
O(4)-Eu-C(11)	19.14(11)	D(6WA)-O(601)-D(6WB)	104(4)
O(9)-Eu-C(11)	81.76(12)	D(7WA)-O(701)-D(7WB)	107(4)
O(7)-Eu-C(11)	154.03(11)	D(8WA)-O(801)-D(8WB)	101(4)

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($\text{Å}^2\text{x }10^3$) for 2009src0262. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2 a^{*2} U^{11} + ... + 2 h k a^* b^* U^{12}]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
C(1)	32(3)	38(3)	23(2)	12(2)	15(2)	12(2)
C(2)	32(3)	34(3)	26(3)	-3(2)	7(2)	0(2)
C(3)	46(3)	25(3)	27(3)	-9(2)	13(2)	-10(2)
C(4)	34(3)	24(3)	47(3)	4(2)	15(3)	4(2)
C(5)	31(3)	36(3)	34(3)	2(2)	10(2)	15(2)
C(6)	27(3)	43(3)	35(3)	-3(2)	10(2)	5(2)
C(7)	24(2)	37(3)	30(3)	-4(2)	13(2)	-9(2)
C(8)	31(3)	29(3)	38(3)	2(2)	19(2)	-1(2)
C(9)	17(2)	15(2)	18(2)	1(2)	5(2)	-1(2)
C(10)	19(2)	26(2)	21(2)	-4(2)	-1(2)	6(2)
C(11)	26(2)	13(2)	29(2)	2(2)	7(2)	-2(2)
C(12)	23(2)	15(2)	48(3)	5(2)	4(2)	2(2)
C(13)	21(2)	17(2)	25(2)	1(2)	8(2)	1(2)
C(14)	18(2)	28(2)	19(2)	1(2)	3(2)	-2(2)
C(15)	18(2)	14(2)	20(2)	2(2)	9(2)	1(2)
C(16)	12(2)	13(2)	18(2)	-1(2)	3(2)	-3(2)
C(17)	23(2)	9(2)	15(2)	3(2)	5(2)	2(2)
C(18)	18(2)	13(2)	20(2)	3(2)	7(2)	0(2)
C(19)	22(2)	17(2)	35(3)	2(2)	2(2)	1(2)
C(20)	19(2)	26(3)	55(4)	4(2)	10(2)	1(2)
C(21)	21(2)	29(3)	57(4)	8(3)	19(2)	8(2)
C(22)	34(3)	23(2)	32(3)	-3(2)	16(2)	3(2)
C(23)	19(2)	17(2)	23(2)	-1(2)	6(2)	-1(2)
C(24)	20(2)	16(2)	21(2)	-2(2)	11(2)	2(2)
C(25)	29(2)	16(2)	33(3)	2(2)	17(2)	6(2)
C(26)	32(3)	22(2)	49(3)	10(2)	25(2)	5(2)
C(27)	27(3)	28(3)	49(3)	3(2)	21(2)	-3(2)
C(28)	20(2)	29(3)	34(3)	3(2)	6(2)	-4(2)
C(29)	24(2)	23(2)	31(3)	5(2)	8(2)	0(2)
C(30)	20(2)	14(2)	16(2)	1(2)	6(2)	0(2)
C(31)	21(2)	20(2)	18(2)	1(2)	6(2)	1(2)
C(32)	18(2)	24(2)	22(2)	-1(2)	9(2)	0(2)
C(33)	21(2)	30(3)	32(3)	-2(2)	9(2)	-3(2)
C(34)	26(3)	40(3)	43(3)	0(3)	20(2)	-1(2)
C(35)	34(3)	49(4)	36(3)	-10(3)	23(3)	0(3)
C(36)	32(3)	29(3)	28(3)	-7(2)	14(2)	-2(2)
C(37)	20(2)	14(2)	16(2)	-6(2)	4(2)	0(2)
C(38)	27(2)	20(2)	19(2)	3(2)	9(2)	6(2)
C(39)	23(2)	17(2)	19(2)	4(2)	8(2)	1(2)
C(40)	21(2)	39(3)	22(2)	0(2)	9(2)	6(2)
C(41)	25(3)	60(4)	27(3)	2(3)	4(2)	-8(3)
C(42)	52(3)	38(3)	23(3)	-6(2)	12(2)	-14(3)
C(43)	51(3)	26(3)	26(3)	10(2)	16(2)	12(2)
C(44)	27(2)	28(3)	22(2)	4(2)	6(2)	11(2)
N(1)	41(2)	17(2)	29(2)	1(2)	23(2)	8(2)
N(2)	20(2)	34(2)	28(2)	-8(2)	4(2)	7(2)
N(3)	22(2)	21(2)	49(3)	1(2)	15(2)	2(2)
N(4)	21(2)	38(2)	14(2)	0(2)	1(2)	-6(2)
N(5)	16(2)	14(2)	12(2)	2(1)	1(1)	0(1)
N(6)	21(2)	21(2)	17(2)	2(2)	6(2)	-1(2)
O(1)	18(2)	25(2)	20(2)	-2(1)	0(1)	5(1)
O(2)	15(1)	16(2)	16(1)	-2(1)	1(1)	2(1)
O(3)	30(2)	20(2)	66(3)	9(2)	14(2)	-2(2)
O(4)	20(2)	14(2)	29(2)	1(1)	5(1)	0(1)
O(5)	27(2)	27(2)	16(2)	3(1)	2(1)	3(1)
O(6)	15(2)	31(2)	21(2)	6(1)	2(1)	0(1)
O(7)	18(1)	13(1)	15(1)	-1(1)	4(1)	-2(1)
O(8)	16(2)	20(2)	23(2)	4(1)	4(1)	-1(1)
O(9)	16(2)	30(2)	19(2)	1(1)	4(1)	-4(1)
Eu	13(1)	12(1)	16(1)	1(1)	2(1)	0(1)
O(101)	30(2)	20(2)	38(2)	5(2)	19(2)	2(1)
O(201)	23(2)	33(2)	19(2)	2(1)	2(1)	0(2)
O(301)	28(2)	43(2)	30(2)	1(2)	13(2)	-9(2)
O(401)	30(2)	26(2)	42(2)	5(2)	5(2)	-6(2)

O(501)	25(2)	51(2)	40(2)	-1(2)	14(2)	2(2)
O(601)	26(2)	37(2)	41(2)	1(2)	10(2)	-4(2)
O(701)	80(4)	41(3)	102(4)	17(3)	73(4)	9(3)
O(801)	42(3)	126(6)	77(4)	-49(4)	27(3)	-2(3)
O(901)	67(3)	64(3)	101(4)	30(3)	60(3)	23(3)
O(001)	93(5)	47(3)	100(5)	1(3)	68(5)	-4(3)

Table 5.   Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for 2009src0262.

	x	y	z	U(eq)
H(1A)	1100	1144	-340	36
H(1B)	2340	1155	-704	36
H(2A)	1348	678	-1237	38
H(2B)	2750	635	-185	38
H(3A)	985	83	627	40
H(3B)	1484	142	-478	40
H(4A)	3084	-101	1171	42
H(4B)	3555	238	964	42
H(5A)	4840	201	4146	41
H(5B)	5069	35	3014	41
H(6A)	6169	498	3448	42
H(6B)	5058	515	2091	42
H(7A)	4945	1266	3179	36
H(7B)	5946	1066	2802	36
H(8A)	4443	1283	1017	37
H(8B)	4271	913	948	37
H(10A)	-497	421	139	29
H(10B)	-403	623	-980	29
H(12A)	3131	-35	4011	38
H(12B)	2729	-242	2777	38
H(14A)	5290	1061	4888	27
H(14B)	6213	769	5023	27
H(15A)	1690	1519	924	20
H(15B)	3189	1601	1424	20
H(17)	1843	1985	1582	20
H(19)	-366	2033	1378	33
H(20)	-1951	2343	1632	42
H(21)	-1395	2765	2994	41
H(22)	794	2875	4159	34
H(23)	2379	2568	3878	24
H(25)	2646	2523	1222	29
H(26)	4513	2805	1486	38
H(27)	6448	2677	3104	40
H(28)	6533	2260	4401	35
H(29)	4660	1973	4143	32
H(30)	2463	2016	4728	20
H(32)	301	1436	2985	25
H(33)	-1709	1339	3067	34
H(34)	-2276	1564	4610	41
H(35)	-814	1895	6087	44
H(36)	1173	2006	5970	35
H(38A)	4843	1244	7168	26
H(38B)	3767	1115	7635	26
H(40)	6451	1352	9171	33
H(41)	7243	1656	10976	47
H(42)	5907	2014	11411	45
H(43)	3767	2060	10050	40
H(44)	2970	1756	8257	32
H(6)	2304	1421	6212	24

Table 6. Hydrogen bonds for 2009src0262 [Å and °].

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
N(6)-H(6)...O(301)	0.88	2.28	2.894(5)	126.5
O(9)-D(9A)...O(301)	0.932(19)	1.90(3)	2.798(5)	161(5)
O(9)-D(9B)...O(501)#1	0.941(19)	1.87(3)	2.764(5)	159(5)
O(101)-D(1WA)...O(3)	0.949(19)	1.77(3)	2.694(5)	164(6)
O(101)-D(1WB)...O(201)#2	0.946(19)	1.84(3)	2.764(5)	163(5)
O(201)-D(2WA)...O(1)#3	0.952(19)	1.78(3)	2.695(4)	160(5)
O(201)-D(2WB)...O(5)#4	0.948(19)	1.81(2)	2.751(5)	175(6)
O(301)-D(3WA)...O(101)#5	0.942(19)	1.91(3)	2.802(5)	157(5)
O(301)-D(3WB)...O(701)	0.942(19)	1.99(3)	2.879(7)	158(5)
O(401)-D(4WA)...O(601)#6	0.950(19)	2.03(2)	2.971(5)	168(5)
O(401)-D(4WB)...O(1)	0.955(19)	2.08(3)	2.983(5)	157(6)
O(501)-D(5WA)...O(001)	0.957(19)	1.86(3)	2.800(8)	167(6)
O(501)-D(5WB)...O(101)#2	0.95(2)	1.95(4)	2.850(6)	156(6)
O(601)-D(6WA)...O(8)	0.950(19)	2.06(3)	2.938(5)	154(5)
O(601)-D(6WB)...O(201)#7	0.946(19)	2.01(3)	2.885(5)	154(6)
O(701)-D(7WA)...O(5)	0.95(2)	2.01(3)	2.868(6)	150(7)
O(701)-D(7WA)...O(6)	0.95(2)	2.32(4)	3.155(6)	146(6)
O(701)-D(7WB)...O(901)#8	0.97(2)	1.84(3)	2.789(9)	166(8)
O(801)-D(8WA)...O(901)#9	0.99(2)	2.15(9)	2.799(9)	121(8)
O(801)-D(8WB)...O(1)#3	0.96(2)	1.91(3)	2.859(6)	169(8)

Symmetry transformations used to generate equivalent atoms:
#1 -x,y-1/2,-z+1/2   #2 x,-y+1/2,z-1/2   #3 -x,y+1/2,-z+1/2
#4 -x+1,y+1/2,-z+3/2   #5 -x,-y,-z+1   #6 x-1,y,z-1
#7 -x+1,y-1/2,-z+3/2   #8 x,y,z+1   #9 -x+1,y+1/2,-z+1/2

Appendix 1c: Benzyl Terbium Ugi complex M16M. Crystal structure

Table 1. Crystal data and structure refinement for 2008JS2.

Identification code	2008JS2, B_2Tb
Empirical formula	$C_{44} H_{49} N_6 O_9 Tb, 10(D_2O)$
Formula weight	1169.13
Temperature	120(2) K
Wavelength	1.54178 Å
Crystal system	Monoclinic
Space group	P 2(1)/c
Unit cell dimensions	a = 11.20490(10) Å = 90°.
	b = 42.9538(5) Å = 112.1950(10)°.
	c = 11.5570(2) Å = 90°.
Volume	5150.16(12) $Å^3$
Z	4
Density (calculated)	1.508 Mg/m^3
Absorption coefficient	7.418 mm^{-1}
F(000)	2376
Crystal size	0.30 x 0.12 x 0.12 mm^3
Theta range for data collection	6.61 to 66.59°.
Index ranges	-13<=h<=13, -50<=k<=51, -12<=l<=13
Reflections collected	52752
Independent reflections	8969 [R(int) = 0.0267]
Completeness to theta = 66.59°	98.6 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.4697 and 0.2144
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	8969 / 31 / 704
Goodness-of-fit on F^2	1.041
Final R indices [I>2sigma(I)]	R1 = 0.0263, wR2 = 0.0653
R indices (all data)	R1 = 0.0267, wR2 = 0.0656
Largest diff. peak and hole	0.999 and -0.560 e.$Å^{-3}$

Notes:
The water molecule O(001)/O(01') is disordered over two sites at a ratio of 0.835 (8): 0.165 (8). Only one deuterium atom could be successfully located in the electron density map for the major site, O(001) while no deuterium atoms could be located for O(01').
The hydrogen atoms in the terbium complex were fixed as riding models while the deuterium atoms (belonging to both bound and free water molecules) were located in the electron density map. The positions of these deuterium atoms were refined subject to restraints while the isotropic thermal parameters were fixed at a value of $1.5U_{eq}$ of the parent oxygen atom.

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3)

for 2008js2. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
C(1)	1993(2)	1061(1)	-141(2)	15(1)
C(2)	1898(2)	713(1)	-368(2)	17(1)
C(3)	1606(2)	204(1)	382(2)	19(1)
C(4)	2964(2)	123(1)	1248(2)	21(1)
C(5)	4669(2)	219(1)	3255(2)	19(1)
C(6)	5240(2)	517(1)	2984(2)	19(1)
C(7)	5047(2)	1079(1)	2734(2)	17(1)
C(8)	4185(2)	1104(1)	1373(2)	17(1)
C(9)	-437(2)	893(1)	470(2)	12(1)
C(10)	-70(2)	598(1)	-42(2)	16(1)
C(11)	1251(2)	41(1)	2860(2)	20(1)
C(12)	2665(2)	-23(1)	3152(2)	20(1)
C(13)	4600(2)	665(1)	5405(2)	15(1)
C(14)	5292(2)	846(1)	4709(2)	17(1)
C(15)	2531(2)	1457(1)	1487(2)	12(1)
C(16)	2570(2)	1488(1)	2810(2)	11(1)
C(17)	2202(2)	2055(1)	2444(2)	13(1)
C(18)	3468(2)	2224(1)	2655(2)	14(1)
C(19)	4636(2)	2145(1)	3590(2)	21(1)
C(20)	5746(2)	2316(1)	3753(3)	27(1)
C(21)	5699(3)	2565(1)	2981(3)	27(1)
C(22)	4544(3)	2640(1)	2024(3)	26(1)
C(23)	3436(2)	2472(1)	1866(2)	20(1)
C(24)	1176(2)	2266(1)	2597(2)	15(1)
C(25)	1504(2)	2518(1)	3410(2)	19(1)
C(26)	560(3)	2702(1)	3567(3)	25(1)
C(27)	-729(3)	2636(1)	2898(3)	32(1)
C(28)	-1068(3)	2388(1)	2074(3)	33(1)
C(29)	-118(2)	2203(1)	1918(2)	24(1)
C(30)	2311(2)	1791(1)	4471(2)	13(1)
C(31)	968(2)	1728(1)	4463(2)	15(1)
C(32)	81(2)	1536(1)	3584(2)	18(1)

C(33)	-1118(2)	1481(1)	3642(2)	24(1)
C(34)	-1443(3)	1615(1)	4570(3)	32(1)
C(35)	-568(3)	1808(1)	5441(3)	32(1)
C(36)	623(2)	1866(1)	5379(2)	24(1)
C(37)	3430(2)	1610(1)	5469(2)	14(1)
C(38)	4135(2)	1307(1)	7385(2)	18(1)
C(39)	4622(2)	1521(1)	8512(2)	15(1)
C(40)	5902(2)	1504(1)	9319(2)	23(1)
C(41)	6366(3)	1689(1)	10388(3)	33(1)
C(42)	5552(3)	1892(1)	10651(2)	31(1)
C(43)	4278(3)	1911(1)	9850(2)	26(1)
C(44)	3809(2)	1728(1)	8783(2)	20(1)
N(1)	2807(2)	1135(1)	1195(2)	13(1)
N(2)	1335(2)	544(1)	425(2)	15(1)
N(3)	3246(2)	208(1)	2569(2)	16(1)
N(4)	4715(2)	803(1)	3340(2)	15(1)
N(5)	2387(2)	1768(1)	3224(2)	11(1)
N(6)	3134(2)	1445(1)	6301(2)	17(1)
O(1)	-1531(1)	1008(1)	-112(2)	18(1)
O(2)	362(1)	1003(1)	1472(1)	13(1)
O(3)	564(2)	-183(1)	2925(2)	38(1)
O(4)	850(1)	315(1)	2584(2)	16(1)
O(5)	5191(2)	619(1)	6544(1)	19(1)
O(6)	3450(1)	583(1)	4792(1)	16(1)
O(7)	2734(1)	1254(1)	3490(1)	11(1)
O(8)	4543(2)	1646(1)	5537(2)	18(1)
O(9)	996(2)	842(1)	3983(2)	17(1)
D(9A)	1280(30)	902(7)	4785(17)	25
D(9B)	179(19)	766(6)	3700(30)	25
Tb	2222(1)	742(1)	2743(1)	10(1)
O(101)	2605(2)	5972(1)	7611(2)	22(1)
D(1WA)	3330(20)	5850(7)	7870(30)	33
D(1WB)	2210(30)	5960(7)	6765(17)	33
O(201)	1102(2)	-763(1)	3872(2)	24(1)
D(2WA)	1610(30)	-860(6)	3510(30)	35
D(2WB)	1020(30)	-559(4)	3570(30)	35
O(301)	8522(2)	647(1)	3605(2)	30(1)
D(3WA)	8510(40)	719(7)	4360(20)	45
D(3WB)	8640(30)	435(4)	3670(30)	45
O(401)	1390(2)	965(1)	6478(2)	30(1)
D(4WA)	1880(30)	793(6)	6830(30)	45
D(4WB)	556(19)	936(8)	6430(30)	45
O(501)	-652(2)	1638(1)	-424(2)	31(1)
D(5WA)	-950(30)	1466(6)	-140(30)	47
D(5WB)	-1240(30)	1670(8)	-1220(20)	47
O(601)	2682(2)	6636(1)	7964(2)	31(1)
D(6WA)	3542(19)	6685(7)	8360(30)	46
D(6WB)	2670(30)	6423(4)	8000(30)	46
O(701)	3182(3)	459(1)	7384(3)	55(1)
D(7WA)	3830(30)	514(11)	7100(40)	82
D(7WB)	2500(30)	398(11)	6620(30)	82
O(801)	4889(3)	329(1)	-205(3)	68(1)
D(8WA)	5700(30)	398(12)	-70(40)	102
D(8WB)	4400(40)	244(12)	-990(30)	102
O(901)	6855(3)	753(1)	1081(3)	76(1)
D(9WA)	7440(40)	748(13)	1940(20)	113
D(9WB)	7330(40)	813(13)	590(40)	113
O(001)	1201(4)	5002(1)	1181(4)	61(1)
D(0WA)	920(60)	4958(14)	1820(40)	92
O(01')	1926(11)	5085(3)	2262(14)	32(4)

Table 3. Bond lengths [Å] and angles [°] for 2008js2.

C(1)-N(1)	1.500(3)
C(1)-C(2)	1.516(3)
C(1)-H(1A)	0.9900
C(1)-H(1B)	0.9900
C(2)-N(2)	1.483(3)
C(2)-H(2A)	0.9900
C(2)-H(2B)	0.9900
C(3)-N(2)	1.494(3)
C(3)-C(4)	1.513(3)
C(3)-H(3A)	0.9900
C(3)-H(3B)	0.9900
C(4)-N(3)	1.482(3)
C(4)-H(4A)	0.9900
C(4)-H(4B)	0.9900
C(5)-N(3)	1.490(3)
C(5)-C(6)	1.517(3)
C(5)-H(5A)	0.9900
C(5)-H(5B)	0.9900
C(6)-N(4)	1.483(3)
C(6)-H(6A)	0.9900
C(6)-H(6B)	0.9900
C(7)-N(4)	1.495(3)
C(7)-C(8)	1.507(3)
C(7)-H(7A)	0.9900
C(7)-H(7B)	0.9900
C(8)-N(1)	1.485(3)
C(8)-H(8A)	0.9900
C(8)-H(8B)	0.9900
C(9)-O(1)	1.254(3)
C(9)-O(2)	1.257(3)
C(9)-C(10)	1.521(3)
C(10)-N(2)	1.476(3)
C(10)-H(10A)	0.9900
C(10)-H(10B)	0.9900
C(11)-O(3)	1.250(3)
C(11)-O(4)	1.261(3)
C(11)-C(12)	1.515(3)
C(12)-N(3)	1.482(3)
C(12)-H(12A)	0.9900
C(12)-H(12B)	0.9900
C(13)-O(5)	1.246(3)
C(13)-O(6)	1.264(3)
C(13)-C(14)	1.523(3)
C(14)-N(4)	1.478(3)
C(14)-H(14A)	0.9900
C(14)-H(14B)	0.9900
C(15)-N(1)	1.484(3)
C(15)-C(16)	1.519(3)
C(15)-H(15A)	0.9900
C(15)-H(15B)	0.9900
C(16)-O(7)	1.246(3)
C(16)-N(5)	1.338(3)
C(17)-N(5)	1.497(3)
C(17)-C(24)	1.525(3)
C(17)-C(18)	1.529(3)
C(17)-H(17)	1.0000
C(18)-C(19)	1.389(3)
C(18)-C(23)	1.393(3)
C(19)-C(20)	1.397(3)
C(19)-H(19)	0.9500
C(20)-C(21)	1.381(4)
C(20)-H(20)	0.9500
C(21)-C(22)	1.385(4)
C(21)-H(21)	0.9500
C(22)-C(23)	1.387(4)
C(22)-H(22)	0.9500
C(23)-H(23)	0.9500
C(24)-C(25)	1.390(3)
C(24)-C(29)	1.390(3)

257

C(25)-C(26)	1.386(3)	N(3)-C(5)-H(5B)	109.3
C(25)-H(25)	0.9500	C(6)-C(5)-H(5B)	109.3
C(26)-C(27)	1.387(4)	H(5A)-C(5)-H(5B)	108.0
C(26)-H(26)	0.9500	N(4)-C(6)-C(5)	113.53(19)
C(27)-C(28)	1.383(4)	N(4)-C(6)-H(6A)	108.9
C(27)-H(27)	0.9500	C(5)-C(6)-H(6A)	108.9
C(28)-C(29)	1.393(4)	N(4)-C(6)-H(6B)	108.9
C(28)-H(28)	0.9500	C(5)-C(6)-H(6B)	108.9
C(29)-H(29)	0.9500	H(6A)-C(6)-H(6B)	107.7
C(30)-N(5)	1.478(3)	N(4)-C(7)-C(8)	111.69(18)
C(30)-C(31)	1.525(3)	N(4)-C(7)-H(7A)	109.3
C(30)-C(37)	1.554(3)	C(8)-C(7)-H(7A)	109.3
C(30)-H(30)	1.0000	N(4)-C(7)-H(7B)	109.3
C(31)-C(36)	1.391(3)	C(8)-C(7)-H(7B)	109.3
C(31)-C(32)	1.392(3)	H(7A)-C(7)-H(7B)	107.9
C(32)-C(33)	1.390(3)	N(1)-C(8)-C(7)	111.93(18)
C(32)-H(32)	0.9500	N(1)-C(8)-H(8A)	109.2
C(33)-C(34)	1.382(4)	C(7)-C(8)-H(8A)	109.2
C(33)-H(33)	0.9500	N(1)-C(8)-H(8B)	109.2
C(34)-C(35)	1.384(4)	C(7)-C(8)-H(8B)	109.2
C(34)-H(34)	0.9500	H(8A)-C(8)-H(8B)	107.9
C(35)-C(36)	1.386(4)	O(1)-C(9)-O(2)	124.0(2)
C(35)-H(35)	0.9500	O(1)-C(9)-C(10)	118.11(19)
C(36)-H(36)	0.9500	O(2)-C(9)-C(10)	117.83(18)
C(37)-O(8)	1.230(3)	N(2)-C(10)-C(9)	113.32(17)
C(37)-N(6)	1.331(3)	N(2)-C(10)-H(10A)	108.9
C(38)-N(6)	1.456(3)	C(9)-C(10)-H(10A)	108.9
C(38)-C(39)	1.518(3)	N(2)-C(10)-H(10B)	108.9
C(38)-H(38A)	0.9900	C(9)-C(10)-H(10B)	108.9
C(38)-H(38B)	0.9900	H(10A)-C(10)-H(10B)	107.7
C(39)-C(40)	1.386(3)	O(3)-C(11)-O(4)	124.4(2)
C(39)-C(44)	1.392(3)	O(3)-C(11)-C(12)	117.7(2)
C(40)-C(41)	1.394(4)	O(4)-C(11)-C(12)	117.88(19)
C(40)-H(40)	0.9500	N(3)-C(12)-C(11)	112.34(18)
C(41)-C(42)	1.376(4)	N(3)-C(12)-H(12A)	109.1
C(41)-H(41)	0.9500	C(11)-C(12)-H(12A)	109.1
C(42)-C(43)	1.380(4)	N(3)-C(12)-H(12B)	109.1
C(42)-H(42)	0.9500	C(11)-C(12)-H(12B)	109.1
C(43)-C(44)	1.388(4)	H(12A)-C(12)-H(12B)	107.9
C(43)-H(43)	0.9500	O(5)-C(13)-O(6)	124.8(2)
C(44)-H(44)	0.9500	O(5)-C(13)-C(14)	117.7(2)
N(1)-Tb	2.7150(18)	O(6)-C(13)-C(14)	117.45(19)
N(2)-Tb	2.6216(18)	N(4)-C(14)-C(13)	113.43(18)
N(3)-Tb	2.6037(18)	N(4)-C(14)-H(14A)	108.9
N(4)-Tb	2.6250(18)	C(13)-C(14)-H(14A)	108.9
N(6)-H(6)	0.8800	N(4)-C(14)-H(14B)	108.9
O(2)-Tb	2.3329(14)	C(13)-C(14)-H(14B)	108.9
O(4)-Tb	2.3527(14)	H(14A)-C(14)-H(14B)	107.7
O(6)-Tb	2.3467(15)	N(1)-C(15)-C(16)	112.41(17)
O(7)-Tb	2.3545(14)	N(1)-C(15)-H(15A)	109.1
O(9)-Tb	2.3698(15)	C(16)-C(15)-H(15A)	109.1
O(9)-D(9A)	0.897(17)	N(1)-C(15)-H(15B)	109.1
O(9)-D(9B)	0.909(17)	C(16)-C(15)-H(15B)	109.1
O(101)-D(1WA)	0.917(17)	H(15A)-C(15)-H(15B)	107.9
O(101)-D(1WB)	0.910(17)	O(7)-C(16)-N(5)	120.34(19)
O(201)-D(2WA)	0.920(17)	O(7)-C(16)-C(15)	120.34(18)
O(201)-D(2WB)	0.937(17)	N(5)-C(16)-C(15)	119.29(18)
O(301)-D(3WA)	0.933(18)	N(5)-C(17)-C(24)	111.04(18)
O(301)-D(3WB)	0.922(18)	N(5)-C(17)-C(18)	112.77(17)
O(401)-D(4WA)	0.921(18)	C(24)-C(17)-C(18)	113.19(18)
O(401)-D(4WB)	0.923(18)	N(5)-C(17)-H(17)	106.4
O(501)-D(5WA)	0.924(18)	C(24)-C(17)-H(17)	106.4
O(501)-D(5WB)	0.915(18)	C(18)-C(17)-H(17)	106.4
O(601)-D(6WA)	0.921(18)	C(19)-C(18)-C(23)	118.8(2)
O(601)-D(6WB)	0.917(18)	C(19)-C(18)-C(17)	123.4(2)
O(701)-D(7WA)	0.941(19)	C(23)-C(18)-C(17)	117.8(2)
O(701)-D(7WB)	0.959(19)	C(18)-C(19)-C(20)	120.2(2)
O(801)-D(8WA)	0.915(19)	C(18)-C(19)-H(19)	119.9
O(801)-D(8WB)	0.935(19)	C(20)-C(19)-H(19)	119.9
O(901)-D(9WA)	0.960(19)	C(21)-C(20)-C(19)	120.5(2)
O(901)-D(9WB)	0.946(19)	C(21)-C(20)-H(20)	119.8
O(001)-D(0WA)	0.93(2)	C(19)-C(20)-H(20)	119.8
O(01')-D(0WA)	1.18(6)N(1)-C(1)-C(2)	C(20)-C(21)-C(22)	119.5(2)
N(1)-C(1)-H(1A)	109.4	C(20)-C(21)-H(21)	120.3
C(2)-C(1)-H(1A)	109.4	C(22)-C(21)-H(21)	120.3
N(1)-C(1)-H(1B)	109.4	C(21)-C(22)-C(23)	120.2(2)
C(2)-C(1)-H(1B)	109.4	C(21)-C(22)-H(22)	119.9
H(1A)-C(1)-H(1B)	108.0	C(23)-C(22)-H(22)	119.9
N(2)-C(2)-C(1)	113.23(18)	C(22)-C(23)-C(18)	120.7(2)
N(2)-C(2)-H(2A)	108.9	C(22)-C(23)-H(23)	119.6
C(1)-C(2)-H(2A)	108.9	C(18)-C(23)-H(23)	119.6
N(2)-C(2)-H(2B)	108.9	C(25)-C(24)-C(29)	119.2(2)
C(1)-C(2)-H(2B)	108.9	C(25)-C(24)-C(17)	121.5(2)
H(2A)-C(2)-H(2B)	107.7	C(29)-C(24)-C(17)	119.3(2)
N(2)-C(3)-C(4)	111.71(18)	C(26)-C(25)-C(24)	120.9(2)
N(2)-C(3)-H(3A)	109.3	C(26)-C(25)-H(25)	119.6
C(4)-C(3)-H(3A)	109.3	C(24)-C(25)-H(25)	119.6
N(2)-C(3)-H(3B)	109.3	C(25)-C(26)-C(27)	119.7(2)
C(4)-C(3)-H(3B)	109.3	C(25)-C(26)-H(26)	120.2
H(3A)-C(3)-H(3B)	107.9	C(27)-C(26)-H(26)	120.2
N(3)-C(4)-C(3)	113.21(19)	C(28)-C(27)-C(26)	120.0(2)
N(3)-C(4)-H(4A)	108.9	C(28)-C(27)-H(27)	120.0
C(3)-C(4)-H(4A)	108.9	C(26)-C(27)-H(27)	120.0
N(3)-C(4)-H(4B)	108.9	C(27)-C(28)-C(29)	120.3(2)
C(3)-C(4)-H(4B)	108.9	C(27)-C(28)-H(28)	119.9
H(4A)-C(4)-H(4B)	107.7	C(29)-C(28)-H(28)	119.9
N(3)-C(5)-C(6)	111.44(18)	C(24)-C(29)-C(28)	120.0(2)
N(3)-C(5)-H(5A)	109.3	C(24)-C(29)-H(29)	120.0
C(6)-C(5)-H(5A)	109.3	C(28)-C(29)-H(29)	120.0

N(5)-C(30)-C(31)	113.72(17)	C(14)-N(4)-C(6)	109.98(18)
N(5)-C(30)-C(37)	111.28(17)	C(14)-N(4)-C(7)	108.41(17)
C(31)-C(30)-C(37)	115.69(18)	C(6)-N(4)-C(7)	109.38(18)
N(5)-C(30)-H(30)	105.0	C(14)-N(4)-Tb	106.60(13)
C(31)-C(30)-H(30)	105.0	C(6)-N(4)-Tb	109.82(13)
C(37)-C(30)-H(30)	105.0	C(7)-N(4)-Tb	112.60(13)
C(36)-C(31)-C(32)	118.8(2)	C(16)-N(5)-C(30)	118.78(17)
C(36)-C(31)-C(30)	118.5(2)	C(16)-N(5)-C(17)	122.04(18)
C(32)-C(31)-C(30)	122.7(2)	C(30)-N(5)-C(17)	119.13(17)
C(33)-C(32)-C(31)	120.1(2)	C(37)-N(6)-C(38)	121.20(19)
C(33)-C(32)-H(32)	119.9	C(37)-N(6)-H(6)	119.4
C(31)-C(32)-H(32)	119.9	C(38)-N(6)-H(6)	119.4
C(34)-C(33)-C(32)	120.6(2)	C(9)-O(2)-Tb	123.82(13)
C(34)-C(33)-H(33)	119.7	C(11)-O(4)-Tb	123.31(14)
C(32)-C(33)-H(33)	119.7	C(13)-O(6)-Tb	124.70(14)
C(33)-C(34)-C(35)	119.5(2)	C(16)-O(7)-Tb	124.45(13)
C(33)-C(34)-H(34)	120.3	Tb-O(9)-D(9A)	128.1(18)
C(35)-C(34)-H(34)	120.3	Tb-O(9)-D(9B)	117.0(18)
C(34)-C(35)-C(36)	120.1(2)	D(9A)-O(9)-D(9B)	113(2)
C(34)-C(35)-H(35)	119.9	O(2)-Tb-O(6)	144.98(5)
C(36)-C(35)-H(35)	119.9	O(2)-Tb-O(4)	85.57(5)
C(35)-C(36)-C(31)	120.8(2)	O(6)-Tb-O(4)	88.25(5)
C(35)-C(36)-H(36)	119.6	O(2)-Tb-O(7)	79.51(5)
C(31)-C(36)-H(36)	119.6	O(6)-Tb-O(7)	86.16(5)
O(8)-C(37)-N(6)	123.0(2)	O(4)-Tb-O(7)	144.85(5)
O(8)-C(37)-C(30)	119.84(19)	O(2)-Tb-O(9)	73.10(5)
N(6)-C(37)-C(30)	116.78(19)	O(6)-Tb-O(9)	72.27(5)
N(6)-C(38)-C(39)	113.66(19)	O(4)-Tb-O(9)	71.22(5)
N(6)-C(38)-H(38A)	108.8	O(7)-Tb-O(9)	74.02(5)
C(39)-C(38)-H(38A)	108.8	O(2)-Tb-N(3)	133.54(6)
N(6)-C(38)-H(38B)	108.8	O(6)-Tb-N(3)	73.42(6)
C(39)-C(38)-H(38B)	108.8	O(4)-Tb-N(3)	66.63(6)
H(38A)-C(38)-H(38B)	107.7	O(7)-Tb-N(3)	142.86(5)
C(40)-C(39)-C(44)	118.7(2)	O(9)-Tb-N(3)	125.46(6)
C(40)-C(39)-C(38)	119.5(2)	O(2)-Tb-N(2)	66.96(5)
C(44)-C(39)-C(38)	121.8(2)	O(6)-Tb-N(2)	142.04(6)
C(39)-C(40)-C(41)	120.8(2)	O(4)-Tb-N(2)	71.56(6)
C(39)-C(40)-H(40)	119.6	O(7)-Tb-N(2)	128.55(5)
C(41)-C(40)-H(40)	119.6	O(9)-Tb-N(2)	126.47(6)
C(42)-C(41)-C(40)	120.0(2)	N(3)-Tb-N(2)	69.14(6)
C(42)-C(41)-H(41)	120.0	O(2)-Tb-N(4)	136.58(5)
C(40)-C(41)-H(41)	120.0	O(6)-Tb-N(4)	66.64(6)
C(41)-C(42)-C(43)	119.6(2)	O(4)-Tb-N(4)	134.19(6)
C(41)-C(42)-H(42)	120.2	O(7)-Tb-N(4)	73.90(5)
C(43)-C(42)-H(42)	120.2	O(9)-Tb-N(4)	128.71(6)
C(42)-C(43)-C(44)	120.7(2)	N(3)-Tb-N(4)	69.61(6)
C(42)-C(43)-H(43)	119.7	N(2)-Tb-N(4)	104.83(6)
C(44)-C(43)-H(43)	119.7	O(2)-Tb-N(1)	70.54(5)
C(43)-C(44)-C(39)	120.2(2)	O(6)-Tb-N(1)	130.58(5)
C(43)-C(44)-H(44)	119.9	O(4)-Tb-N(1)	138.25(5)
C(39)-C(44)-H(44)	119.9	O(7)-Tb-N(1)	64.76(5)
C(15)-N(1)-C(8)	110.40(17)	O(9)-Tb-N(1)	128.65(6)
C(15)-N(1)-C(1)	109.47(16)	N(3)-Tb-N(1)	105.87(6)
C(8)-N(1)-C(1)	108.62(17)	N(2)-Tb-N(1)	67.74(6)
C(15)-N(1)-Tb	107.91(12)	N(4)-Tb-N(1)	67.26(6)
C(8)-N(1)-Tb	110.45(13)	D(1WA)-O(101)-D(1WB)	109(2)
C(1)-N(1)-Tb	109.98(12)	D(2WA)-O(201)-D(2WB)	104(2)
C(10)-N(2)-C(2)	109.92(17)	D(3WA)-O(301)-D(3WB)	107(3)
C(10)-N(2)-C(3)	110.04(17)	D(4WA)-O(401)-D(4WB)	110(3)
C(2)-N(2)-C(3)	108.54(17)	D(5WA)-O(501)-D(5WB)	105(3)
C(10)-N(2)-Tb	105.26(13)	D(6WA)-O(601)-D(6WB)	104(3)
C(2)-N(2)-Tb	112.77(13)	D(7WA)-O(701)-D(7WB)	101(3)
C(3)-N(2)-Tb	110.28(13)	D(8WA)-O(801)-D(8WB)	119(3)
C(12)-N(3)-C(4)	110.03(18)	D(9WA)-O(901)-D(9WB)	107(3)
C(12)-N(3)-C(5)	110.11(17)		
C(4)-N(3)-C(5)	108.94(18)		
C(12)-N(3)-Tb	105.83(13)	Symmetry transformations used to generate equivalent atoms:	
C(4)-N(3)-Tb	111.61(13)		
C(5)-N(3)-Tb	110.29(13)		

Table 4. Anisotropic displacement parameters (Å^2x 10^3)for 2008js2. The anisotropic displacement factor exponent takes the form: -2 π^2[$h^2a^{*2}U^{11}$ + ... + 2 h k a* b* U^{12}]

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
C(1)	18(1)	16(1)	12(1)	2(1)	6(1)	1(1)
C(2)	20(1)	15(1)	16(1)	-3(1)	6(1)	0(1)
C(3)	23(1)	11(1)	21(1)	-5(1)	6(1)	-1(1)
C(4)	22(1)	13(1)	26(1)	-1(1)	8(1)	4(1)
C(5)	13(1)	18(1)	23(1)	2(1)	4(1)	5(1)
C(6)	12(1)	22(1)	22(1)	1(1)	6(1)	2(1)
C(7)	13(1)	18(1)	21(1)	0(1)	7(1)	-3(1)
C(8)	16(1)	17(1)	21(1)	2(1)	10(1)	0(1)
C(9)	12(1)	13(1)	12(1)	2(1)	4(1)	-1(1)
C(10)	12(1)	15(1)	15(1)	-2(1)	0(1)	0(1)
C(11)	17(1)	14(1)	24(1)	4(1)	3(1)	-4(1)
C(12)	18(1)	10(1)	29(1)	5(1)	6(1)	2(1)
C(13)	16(1)	11(1)	17(1)	1(1)	6(1)	3(1)
C(14)	10(1)	21(1)	14(1)	1(1)	-1(1)	-2(1)
C(15)	14(1)	9(1)	14(1)	0(1)	5(1)	0(1)
C(16)	6(1)	11(1)	13(1)	-1(1)	1(1)	-2(1)
C(17)	17(1)	9(1)	11(1)	1(1)	3(1)	0(1)
C(18)	18(1)	10(1)	16(1)	-2(1)	9(1)	1(1)
C(19)	21(1)	18(1)	24(1)	5(1)	7(1)	-1(1)

C(20)	19(1)	30(1)	30(1)	4(1)	6(1)	-3(1)
C(21)	24(1)	22(1)	41(2)	0(1)	20(1)	-5(1)
C(22)	31(1)	17(1)	38(2)	9(1)	22(1)	3(1)
C(23)	21(1)	16(1)	26(1)	5(1)	12(1)	5(1)
C(24)	17(1)	11(1)	18(1)	4(1)	7(1)	2(1)
C(25)	19(1)	16(1)	20(1)	-1(1)	6(1)	0(1)
C(26)	30(1)	17(1)	33(1)	-2(1)	16(1)	3(1)
C(27)	23(1)	23(1)	55(2)	6(1)	22(1)	7(1)
C(28)	16(1)	25(1)	52(2)	4(1)	9(1)	1(1)
C(29)	19(1)	13(1)	32(1)	1(1)	3(1)	-1(1)
C(30)	15(1)	11(1)	12(1)	-2(1)	3(1)	0(1)
C(31)	16(1)	15(1)	13(1)	3(1)	4(1)	3(1)
C(32)	18(1)	18(1)	18(1)	0(1)	7(1)	2(1)
C(33)	16(1)	27(1)	28(1)	-2(1)	6(1)	-1(1)
C(34)	21(1)	41(2)	39(2)	1(1)	17(1)	1(1)
C(35)	33(2)	40(2)	32(2)	-5(1)	22(1)	5(1)
C(36)	26(1)	25(1)	21(1)	-5(1)	10(1)	1(1)
C(37)	18(1)	10(1)	11(1)	-4(1)	4(1)	-1(1)
C(38)	21(1)	16(1)	16(1)	3(1)	6(1)	3(1)
C(39)	18(1)	14(1)	13(1)	4(1)	7(1)	-2(1)
C(40)	18(1)	31(1)	22(1)	1(1)	8(1)	2(1)
C(41)	23(1)	47(2)	22(1)	-3(1)	2(1)	-10(1)
C(42)	47(2)	25(1)	20(1)	-5(1)	12(1)	-12(1)
C(43)	45(2)	16(1)	22(1)	3(1)	18(1)	7(1)
C(44)	23(1)	19(1)	18(1)	5(1)	7(1)	7(1)
N(1)	15(1)	11(1)	14(1)	-1(1)	7(1)	1(1)
N(2)	14(1)	12(1)	17(1)	-3(1)	3(1)	1(1)
N(3)	12(1)	12(1)	24(1)	2(1)	5(1)	2(1)
N(4)	12(1)	18(1)	13(1)	1(1)	3(1)	-1(1)
N(5)	13(1)	10(1)	11(1)	1(1)	3(1)	0(1)
N(6)	17(1)	18(1)	14(1)	2(1)	4(1)	0(1)
O(1)	11(1)	23(1)	16(1)	-1(1)	0(1)	4(1)
O(2)	10(1)	13(1)	13(1)	0(1)	2(1)	2(1)
O(3)	24(1)	19(1)	66(1)	15(1)	12(1)	-4(1)
O(4)	14(1)	9(1)	23(1)	2(1)	5(1)	-3(1)
O(5)	19(1)	21(1)	14(1)	3(1)	2(1)	0(1)
O(6)	12(1)	19(1)	16(1)	5(1)	3(1)	0(1)
O(7)	13(1)	9(1)	12(1)	-1(1)	4(1)	-2(1)
O(8)	14(1)	18(1)	18(1)	2(1)	4(1)	-1(1)
O(9)	12(1)	24(1)	15(1)	-2(1)	6(1)	-3(1)
Tb	8(1)	8(1)	11(1)	1(1)	2(1)	-1(1)
O(101)	18(1)	28(1)	16(1)	3(1)	2(1)	1(1)
O(201)	26(1)	18(1)	32(1)	6(1)	17(1)	3(1)
O(301)	20(1)	41(1)	31(1)	2(1)	12(1)	-2(1)
O(401)	26(1)	41(1)	23(1)	1(1)	10(1)	-10(1)
O(501)	26(1)	24(1)	36(1)	3(1)	2(1)	-5(1)
O(601)	20(1)	33(1)	37(1)	-5(1)	7(1)	2(1)
O(701)	71(2)	35(1)	90(2)	8(1)	67(2)	2(1)
O(801)	66(2)	75(2)	88(2)	29(2)	58(2)	27(2)
O(901)	37(1)	132(3)	62(2)	51(2)	25(1)	13(2)
O(001)	88(3)	34(2)	89(3)	-5(2)	66(3)	-6(2)
O(01')	23(7)	24(6)	58(10)	4(6)	26(6)	1(5)

Table 5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters (Å^{2}x 10^3) for 2008js2.

	x	y	z	U(eq)
H(1A)	2373	1160	-695	18
H(1B)	1118	1148	-349	18
H(2A)	1361	673	-1257	21
H(2B)	2770	629	-204	21
H(3A)	1484	146	-483	22
H(3B)	985	83	624	22
H(4A)	3096	-104	1199	25
H(4B)	3581	232	961	25
H(5A)	5068	38	3006	22
H(5B)	4870	203	4164	22
H(6A)	6186	513	3443	23
H(6B)	5072	526	2080	23
H(7A)	5955	1063	2811	20
H(7B)	4962	1270	3174	20
H(8A)	4446	1286	1002	20
H(8B)	4295	916	927	20
H(10A)	-485	418	188	19
H(10B)	-408	610	-966	19
H(12A)	3139	-20	4069	24
H(12B)	2759	-234	2848	24
H(14A)	5278	1070	4900	20
H(14B)	6204	778	5018	20
H(15A)	1667	1520	889	15
H(15B)	3172	1601	1377	15
H(17)	1856	1985	1553	16
H(19)	4679	1972	4121	25
H(20)	6540	2261	4400	32
H(21)	6452	2684	3105	32
H(22)	4511	2807	1475	31
H(23)	2647	2527	1213	24
H(25)	2387	2565	3865	22
H(26)	795	2873	4131	31
H(27)	-1380	2760	3006	38
H(28)	-1951	2343	1612	39
H(29)	-354	2033	1348	28
H(30)	2483	2015	4711	16
H(32)	295	1442	2943	21

H(33)	-1719	1350	3038	29
H(34)	-2260	1575	4610	38
H(35)	-785	1901	6082	39
H(36)	1211	2002	5970	28
H(38A)	4869	1246	7155	22
H(38B)	3791	1115	7618	22
H(40)	6470	1365	9141	28
H(41)	7245	1675	10935	39
H(42)	5865	2018	11381	37
H(43)	3716	2052	10030	32
H(44)	2931	1744	8237	24
H(6)	2320	1420	6193	20

Table 6. Hydrogen bonds for 2008js2 [Å and °].

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
N(6)-H(6)...O(401)	0.88	2.29	2.900(3)	126.0
O(9)-D(9A)...O(401)	0.897(17)	1.932(19)	2.797(2)	162(3)
O(9)-D(9B)...O(301)#1	0.909(17)	1.886(19)	2.767(2)	163(3)
O(101)-D(1WA)...O(5)#2	0.917(17)	1.827(18)	2.743(2)	177(3)
O(101)-D(1WB)...O(1)#3	0.910(17)	1.781(18)	2.681(2)	170(3)
O(201)-D(2WA)...O(101)#4	0.920(17)	1.856(19)	2.761(2)	167(3)
O(201)-D(2WB)...O(3)	0.937(17)	1.771(18)	2.698(2)	169(3)
O(301)-D(3WA)...O(201)#5	0.933(18)	1.93(2)	2.828(3)	161(3)
O(301)-D(3WB)...O(001)#6	0.922(18)	1.869(18)	2.790(4)	177(3)
O(301)-D(3WB)...O(01')#6	0.922(18)	1.82(3)	2.592(11)	140(3)
O(401)-D(4WA)...O(701)	0.921(18)	1.970(19)	2.874(3)	167(3)
O(401)-D(4WB)...O(201)#7	0.923(18)	1.91(2)	2.803(3)	164(3)
O(501)-D(5WA)...O(1)	0.924(18)	2.08(2)	2.950(2)	157(3)
O(501)-D(5WB)...O(601)#8	0.915(18)	2.06(2)	2.957(3)	167(3)
O(601)-D(6WA)...O(8)#2	0.921(18)	2.05(2)	2.933(2)	160(3)
O(601)-D(6WB)...O(101)	0.917(18)	1.98(2)	2.877(3)	165(3)
O(701)-D(7WA)...O(5)	0.941(19)	1.911(19)	2.851(3)	179(5)
O(701)-D(7WA)...O(6)	0.941(19)	2.55(4)	3.166(3)	123(3)
O(701)-D(7WB)...O(001)#9	0.959(19)	2.19(4)	2.908(5)	131(4)
O(801)-D(8WA)...O(901)	0.915(19)	2.11(4)	2.811(5)	132(4)
O(801)-D(8WB)...O(701)#10	0.935(19)	2.08(5)	2.776(4)	130(5)
O(901)-D(9WA)...O(301)	0.960(19)	1.90(2)	2.844(4)	168(5)
O(901)-D(9WB)...O(1)#11	0.946(19)	1.94(2)	2.874(3)	167(5)
O(001)-D(0WA)...O(3)#3	0.93(2)	1.89(4)	2.673(4)	140(5)

Symmetry transformations used to generate equivalent atoms:
#1 x-1,y,z   #2 -x+1,y+1/2,-z+3/2   #3 -x,y+1/2,-z+1/2
#4 x,-y+1/2,z-1/2   #5 -x+1,-y,-z+1   #6 -x+1,y-1/2,-z+1/2
#7 -x,-y,-z+1   #8 -x,y-1/2,-z+1/2   #9 x,-y+1/2,z+1/2
#10 x,y,z-1   #11 x+1,y,z

Appendix 1d: Benzyl Ytterbium Ugi complex M20M. Crystal structure
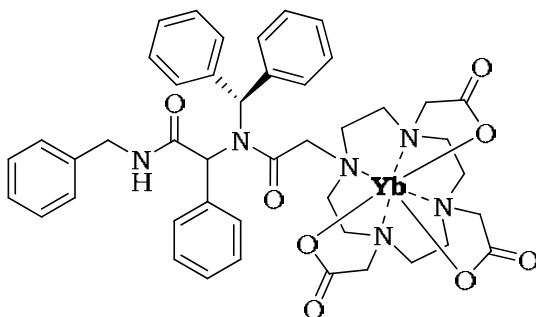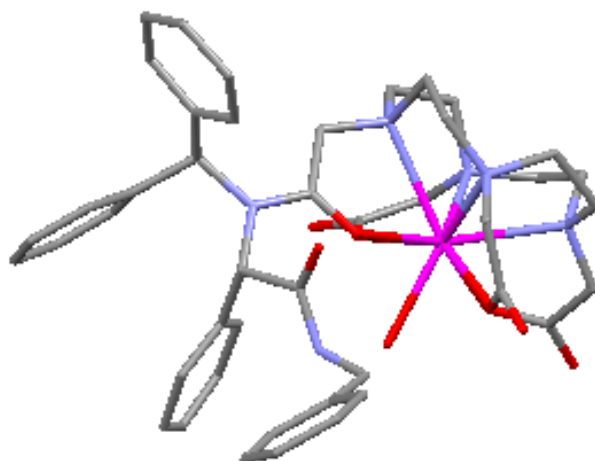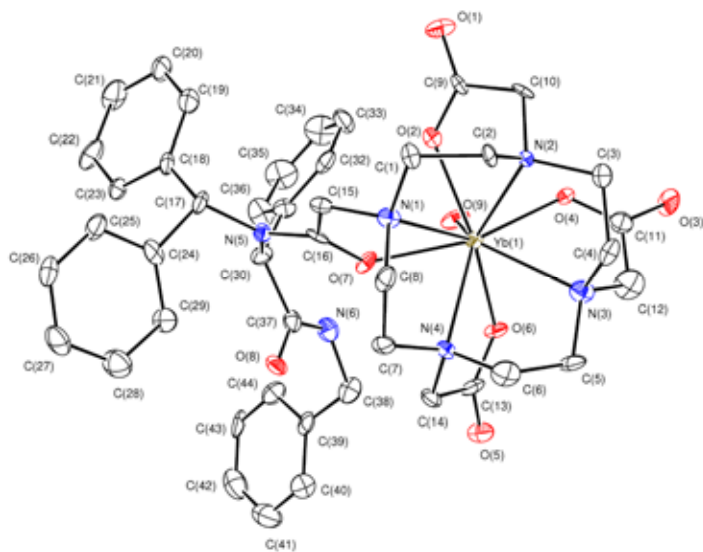
Table 1. Crystal data and structure refinement for B$_2$Yb.

Identification code	B$_2$Yb, 2009src0812
Empirical formula	C$_{44}$ H$_{49}$ D$_2$ N$_6$ O$_9$ Yb, 10(D$_2$ O)
Formula weight	1183.25
Temperature	120(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	P 2(1)/c
Unit cell dimensions	a = 11.1950(4) Å a= 90°.
	b = 42.9194(14) Å b= 112.2910(10)°.
	c = 11.5718(4) Å g = 90°.
Volume	5144.5(3) Å^3
Z	4
Density (calculated)	1.528 Mg/m^3
Absorption coefficient	1.896 mm^{-1}
F(000)	2396
Crystal size	0.68 x 0.20 x 0.06 mm^3
Theta range for data collection	3.04 to 25.68°.
Index ranges	-13<=h<=13, -52<=k<=48, -14<=l<=14
Reflections collected	27619
Independent reflections	9207 [R(int) = 0.0566]
Completeness to theta = 25.68°	94.2 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.8947 and 0.3588
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	9207 / 104 / 693
Goodness-of-fit on F^2	1.085
Final R indices [I>2sigma(I)]	R1 = 0.0629, wR2 = 0.1889
R indices (all data)	R1 = 0.0863, wR2 = 0.2337
Largest diff. peak and hole	1.720 and -2.443 e.Å^{-3}

Notes:

The water molecule O(701)/O(71') is disordered over two sites at a ratio of 0.51 (7): 0.49 (7) and is refined isotropically while O(001)/O(01') is disordered over two sites at a ratio of 0.73 (3): 0.27 (3) and O(01') is refined isotropically. No deuterium atoms could be located in the electron density map for O(701)/O(71') or for O(01').

The hydrogen atoms in the ytterbium complex were fixed as riding models while the deuterium atoms (belonging to both bound and free water molecules) were located in the electron density map. The positions of these deuterium atoms were refined subject to restraints while the isotropic thermal parameters were fixed at a value of 1.5Ueq of the parent oxygen atom.

Table 2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^2 x 10^3) for B$_2$Yb. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
C(1)	1986(8)	1057(2)	-161(8)	19(2)
C(2)	1898(9)	709(2)	-367(8)	16(2)
C(3)	1624(9)	197(2)	399(9)	22(2)
C(4)	2986(9)	129(2)	1279(9)	22(2)
C(5)	4694(8)	235(2)	3286(8)	22(2)
C(6)	5251(9)	534(2)	2993(9)	26(2)
C(7)	5026(8)	1095(2)	2720(8)	19(2)
C(8)	4168(8)	1109(2)	1338(9)	21(2)
C(9)	-425(8)	892(2)	473(8)	14(2)
C(10)	-59(8)	591(2)	-39(8)	19(2)
C(11)	1262(9)	54(2)	2871(8)	24(2)
C(12)	2681(10)	-4(3)	3173(11)	34(2)
C(13)	4562(9)	681(2)	5377(9)	22(2)
C(14)	5245(8)	866(2)	4690(8)	17(2)
C(15)	2475(8)	1456(2)	1443(7)	16(2)
C(16)	2513(7)	1483(2)	2773(7)	15(2)
C(17)	2166(8)	2055(2)	2415(8)	18(2)
C(18)	1136(8)	2267(2)	2572(8)	18(2)
C(19)	-168(9)	2202(2)	1873(10)	28(2)
C(20)	-1120(10)	2386(2)	2035(11)	32(2)
C(21)	-782(10)	2633(3)	2864(12)	38(3)
C(22)	528(10)	2704(2)	3537(10)	31(2)
C(23)	1474(9)	2517(2)	3380(9)	20(2)
C(24)	3433(8)	2224(2)	2640(9)	20(2)
C(25)	3394(9)	2473(2)	1826(9)	25(2)
C(26)	4506(10)	2637(2)	2003(9)	27(2)
C(27)	5654(10)	2566(3)	2943(10)	31(2)
C(28)	5693(10)	2315(3)	3712(10)	34(3)

C(29)	4598(9)	2140(2)	3575(9)	25(2)
C(30)	2255(8)	1786(2)	4425(8)	19(2)
C(31)	923(8)	1723(2)	4420(8)	19(2)
C(32)	28(9)	1536(2)	3548(9)	22(2)
C(33)	-1191(9)	1484(2)	3587(9)	27(2)
C(34)	-1489(10)	1618(3)	4528(11)	42(3)
C(35)	-623(10)	1808(3)	5398(11)	41(3)
C(36)	581(10)	1870(3)	5335(10)	32(2)
C(37)	3380(9)	1604(2)	5423(8)	19(2)
C(38)	4092(10)	1304(2)	7369(8)	26(2)
C(39)	4580(9)	1523(2)	8477(8)	20(2)
C(40)	5861(9)	1503(2)	9283(9)	27(2)
C(41)	6331(11)	1688(3)	10331(11)	41(3)
C(42)	5515(12)	1895(3)	10605(10)	39(3)
C(43)	4247(10)	1916(2)	9807(9)	27(2)
C(44)	3778(9)	1728(2)	8753(9)	25(2)
N(1)	2777(6)	1135(2)	1162(6)	16(2)
N(2)	1360(6)	545(2)	446(6)	14(2)
N(3)	3287(7)	223(2)	2608(7)	21(2)
N(4)	4695(7)	816(2)	3315(7)	15(2)
N(5)	2343(7)	1762(2)	3191(6)	15(2)
N(6)	3087(8)	1443(2)	6263(7)	23(2)
O(1)	-1515(6)	1010(2)	-106(6)	24(2)
O(2)	392(5)	996(1)	1491(5)	16(1)
O(3)	561(6)	-169(2)	2911(7)	35(2)
O(4)	884(5)	334(1)	2586(5)	17(1)
O(5)	5138(6)	618(2)	6494(6)	25(2)
O(6)	3391(5)	597(2)	4719(5)	18(1)
O(7)	2683(6)	1249(1)	3439(5)	17(1)
O(8)	4495(6)	1643(2)	5509(6)	22(1)
O(9)	1011(6)	856(2)	3942(6)	23(1)
Yb(1)	2222(1)	746(1)	2716(1)	13(1)
D(9A)	210(30)	753(16)	3600(50)	20
D(9B)	1490(40)	794(18)	4770(50)	20
O(101)	1120(7)	-751(1)	3874(7)	25(2)
D(1WA)	1710(80)	-820(20)	3510(90)	38
D(1WB)	770(90)	-559(13)	3490(90)	38
O(201)	7381(6)	962(2)	7391(6)	24(1)
D(2WA)	6650(70)	830(20)	7220(80)	35
D(2WB)	7810(80)	970(20)	8280(20)	35
O(301)	9307(7)	1635(2)	9548(7)	36(2)
D(3WA)	8880(90)	1463(19)	9730(100)	54
D(3WB)	8730(80)	1710(20)	8770(60)	54
O(401)	8558(7)	647(2)	3615(7)	38(2)
D(4WA)	8560(120)	690(30)	4420(40)	57
D(4WB)	7890(90)	760(30)	3010(70)	57
O(501)	7286(7)	1632(2)	6998(7)	37(2)
D(5WA)	6420(40)	1700(20)	6670(110)	55
D(5WB)	7290(100)	1412(5)	7020(120)	55
O(601)	1386(7)	959(2)	6456(7)	42(2)
D(6WA)	680(80)	1070(30)	6550(120)	63
O(701)	6973(19)	837(8)	1300(20)	46(7)
O(71')	6814(19)	715(8)	990(20)	38(7)
O(801)	3194(12)	453(2)	7374(11)	71(3)
D(8WA)	3870(80)	480(40)	7080(120)	106
D(8WB)	2400(70)	500(50)	6680(90)	106
O(901)	4928(12)	337(3)	9824(13)	90(4)
D(9WA)	4520(150)	340(50)	8940(30)	135
D(9WB)	5780(80)	420(50)	10070(150)	135
O(001)	8787(16)	-1(3)	3769(17)	73(7)
D(0WA)	8700(200)	217(11)	3850(160)	91
D(0WB)	9340(180)	-80(40)	4560(90)	91
O(01')	8080(40)	80(10)	2680(40)	61(14)

Table 3. Bond lengths [Å] and angles [°] for B₂Yb.

C(1)-N(1)	1.486(11)	C(10)-N(2)	1.484(10)
C(1)-C(2)	1.510(11)	C(10)-H(10A)	0.9900
C(1)-H(1A)	0.9900	C(10)-H(10B)	0.9900
C(1)-H(1B)	0.9900	C(11)-O(3)	1.249(11)
C(2)-N(2)	1.474(11)	C(11)-O(4)	1.276(11)
C(2)-H(2A)	0.9900	C(11)-C(12)	1.512(13)
C(2)-H(2B)	0.9900	C(11)-Yb(1)	3.189(10)
C(3)-C(4)	1.507(12)	C(12)-N(3)	1.474(13)
C(3)-N(2)	1.524(11)	C(12)-H(12A)	0.9900
C(3)-H(3A)	0.9900	C(12)-H(12B)	0.9900
C(3)-H(3B)	0.9900	C(13)-O(5)	1.235(11)
C(4)-N(3)	1.498(12)	C(13)-O(6)	1.292(11)
C(4)-H(4A)	0.9900	C(13)-C(14)	1.521(13)
C(4)-H(4B)	0.9900	C(13)-Yb(1)	3.210(9)
C(5)-N(3)	1.471(11)	C(14)-N(4)	1.488(11)
C(5)-C(6)	1.520(13)	C(14)-H(14A)	0.9900
C(5)-H(5A)	0.9900	C(14)-H(14B)	0.9900
C(5)-H(5B)	0.9900	C(15)-N(1)	1.485(11)
C(6)-N(4)	1.472(12)	C(15)-C(16)	1.528(12)
C(6)-H(6A)	0.9900	C(15)-H(15A)	0.9900
C(6)-H(6B)	0.9900	C(15)-H(15B)	0.9900
C(7)-N(4)	1.495(11)	C(16)-O(7)	1.236(11)
C(7)-C(8)	1.522(12)	C(16)-N(5)	1.334(11)
C(7)-H(7A)	0.9900	C(16)-Yb(1)	3.175(9)
C(7)-H(7B)	0.9900	C(17)-N(5)	1.515(11)
C(8)-N(1)	1.496(11)	C(17)-C(24)	1.525(12)
C(8)-H(8A)	0.9900	C(17)-C(18)	1.533(12)
C(8)-H(8B)	0.9900	C(17)-H(17)	1.0000
C(9)-O(1)	1.256(10)	C(18)-C(23)	1.376(12)
C(9)-O(2)	1.267(10)	C(18)-C(19)	1.403(13)
C(9)-C(10)	1.538(12)	C(19)-C(20)	1.393(14)
C(9)-Yb(1)	3.173(8)	C(19)-H(19)	0.9500

C(20)-C(21)	1.385(16)	C(39)-C(40)	1.386(13)
C(20)-H(20)	0.9500	C(40)-C(41)	1.377(15)
C(21)-C(22)	1.407(15)	C(40)-H(40)	0.9500
C(21)-H(21)	0.9500	C(41)-C(42)	1.394(17)
C(22)-C(23)	1.395(12)	C(41)-H(41)	0.9500
C(22)-H(22)	0.9500	C(42)-C(43)	1.372(16)
C(23)-H(23)	0.9500	C(42)-H(42)	0.9500
C(24)-C(29)	1.390(13)	C(43)-C(44)	1.389(14)
C(24)-C(25)	1.414(13)	C(43)-H(43)	0.9500
C(25)-C(26)	1.377(13)	C(44)-H(44)	0.9500
C(25)-H(25)	0.9500	N(1)-Yb(1)	2.692(7)
C(26)-C(27)	1.367(14)	N(2)-Yb(1)	2.581(7)
C(26)-H(26)	0.9500	N(3)-Yb(1)	2.570(8)
C(27)-C(28)	1.387(15)	N(4)-Yb(1)	2.605(7)
C(27)-H(27)	0.9500	N(6)-H(6)	0.8800
C(28)-C(29)	1.396(14)	O(2)-Yb(1)	2.270(6)
C(28)-H(28)	0.9500	O(4)-Yb(1)	2.286(6)
C(29)-H(29)	0.9500	O(6)-Yb(1)	2.278(6)
C(30)-N(5)	1.471(11)	O(7)-Yb(1)	2.299(6)
C(30)-C(31)	1.514(12)	O(9)-Yb(1)	2.351(6)
C(30)-C(37)	1.558(12)	O(9)-D(9A)	0.94(2)
C(30)-H(30)	1.0000	O(9)-D(9B)	0.94(2)
C(31)-C(32)	1.377(13)	O(101)-D(1WA)	0.95(2)
C(31)-C(36)	1.406(13)	O(101)-D(1WB)	0.95(2)
C(32)-C(33)	1.400(13)	O(201)-D(2WA)	0.95(2)
C(32)-H(32)	0.9500	O(201)-D(2WB)	0.95(2)
C(33)-C(34)	1.381(15)	O(301)-D(3WA)	0.95(2)
C(33)-H(33)	0.9500	O(301)-D(3WB)	0.95(2)
C(34)-C(35)	1.370(17)	O(401)-D(4WA)	0.95(2)
C(34)-H(34)	0.9500	O(401)-D(4WB)	0.94(2)
C(35)-C(36)	1.403(14)	O(501)-D(5WA)	0.95(2)
C(35)-H(35)	0.9500	O(501)-D(5WB)	0.95(2)
C(36)-H(36)	0.9500	O(601)-D(6WA)	0.96(2)
C(37)-O(8)	1.225(11)	O(801)-D(8WA)	0.95(2)
C(37)-N(6)	1.332(12)	O(801)-D(8WB)	0.96(2)
C(38)-N(6)	1.472(12)	O(901)-D(9WA)	0.95(2)
C(38)-C(39)	1.514(13)	O(901)-D(9WB)	0.95(2)
C(38)-H(38A)	0.9900	O(001)-D(0WA)	0.95(2)
C(38)-H(38B)	0.9900	O(001)-D(0WB)	0.95(2)
C(39)-C(44)	1.379(13)		
N(1)-C(1)-C(2)	111.2(7)	O(3)-C(11)-O(4)	125.5(9)
N(1)-C(1)-H(1A)	109.4	O(3)-C(11)-C(12)	119.0(9)
C(2)-C(1)-H(1A)	109.4	O(4)-C(11)-C(12)	115.5(9)
N(1)-C(1)-H(1B)	109.4	O(3)-C(11)-Yb(1)	161.0(7)
C(2)-C(1)-H(1B)	109.4	C(12)-C(11)-Yb(1)	79.6(6)
H(1A)-C(1)-H(1B)	108.0	N(3)-C(12)-C(11)	114.1(9)
N(2)-C(2)-C(1)	112.9(7)	N(3)-C(12)-H(12A)	108.7
N(2)-C(2)-H(2A)	109.0	C(11)-C(12)-H(12A)	108.7
C(1)-C(2)-H(2A)	109.0	N(3)-C(12)-H(12B)	108.7
N(2)-C(2)-H(2B)	109.0	C(11)-C(12)-H(12B)	108.7
C(1)-C(2)-H(2B)	109.0	H(12A)-C(12)-H(12B)	107.6
H(2A)-C(2)-H(2B)	107.8	O(5)-C(13)-O(6)	124.6(9)
C(4)-C(3)-N(2)	109.1(7)	O(5)-C(13)-C(14)	119.6(8)
C(4)-C(3)-H(3A)	109.9	O(6)-C(13)-C(14)	115.9(8)
N(2)-C(3)-H(3A)	109.9	O(5)-C(13)-Yb(1)	159.2(7)
C(4)-C(3)-H(3B)	109.9	C(14)-C(13)-Yb(1)	80.9(5)
N(2)-C(3)-H(3B)	109.9	N(4)-C(14)-C(13)	113.4(7)
H(3A)-C(3)-H(3B)	108.3	N(4)-C(14)-H(14A)	108.9
N(3)-C(4)-C(3)	114.7(7)	C(13)-C(14)-H(14A)	108.9
N(3)-C(4)-H(4A)	108.6	N(4)-C(14)-H(14B)	108.9
C(3)-C(4)-H(4A)	108.6	C(13)-C(14)-H(14B)	108.9
N(3)-C(4)-H(4B)	108.6	H(14A)-C(14)-H(14B)	107.7
C(3)-C(4)-H(4B)	108.6	N(1)-C(15)-C(16)	111.7(7)
H(4A)-C(4)-H(4B)	107.6	N(1)-C(15)-H(15A)	109.3
N(3)-C(5)-C(6)	111.0(7)	C(16)-C(15)-H(15A)	109.3
N(3)-C(5)-H(5A)	109.4	N(1)-C(15)-H(15B)	109.3
C(6)-C(5)-H(5A)	109.4	C(16)-C(15)-H(15B)	109.3
N(3)-C(5)-H(5B)	109.4	H(15A)-C(15)-H(15B)	107.9
C(6)-C(5)-H(5B)	109.4	O(7)-C(16)-N(5)	120.9(7)
H(5A)-C(5)-H(5B)	108.0	O(7)-C(16)-C(15)	120.4(8)
N(4)-C(6)-C(5)	113.0(7)	N(5)-C(16)-C(15)	118.7(8)
N(4)-C(6)-H(6A)	109.0	N(5)-C(16)-Yb(1)	151.0(6)
C(5)-C(6)-H(6A)	109.0	C(15)-C(16)-Yb(1)	86.6(5)
N(4)-C(6)-H(6B)	109.0	N(5)-C(17)-C(24)	112.9(7)
C(5)-C(6)-H(6B)	109.0	N(5)-C(17)-C(18)	111.0(7)
H(6A)-C(6)-H(6B)	107.8	C(24)-C(17)-C(18)	113.0(7)
N(4)-C(7)-C(8)	110.3(7)	N(5)-C(17)-H(17)	106.5
N(4)-C(7)-H(7A)	109.6	C(24)-C(17)-H(17)	106.5
C(8)-C(7)-H(7A)	109.6	C(18)-C(17)-H(17)	106.5
N(4)-C(7)-H(7B)	109.6	C(23)-C(18)-C(19)	120.3(9)
C(8)-C(7)-H(7B)	109.6	C(23)-C(18)-C(17)	121.1(8)
H(7A)-C(7)-H(7B)	108.1	C(19)-C(18)-C(17)	118.6(8)
N(1)-C(8)-C(7)	110.8(7)	C(20)-C(19)-C(18)	119.5(9)
N(1)-C(8)-H(8A)	109.5	C(20)-C(19)-H(19)	120.3
C(7)-C(8)-H(8A)	109.5	C(18)-C(19)-H(19)	120.3
N(1)-C(8)-H(8B)	109.5	C(21)-C(20)-C(19)	120.3(10)
C(7)-C(8)-H(8B)	109.5	C(21)-C(20)-H(20)	119.8
H(8A)-C(8)-H(8B)	108.1	C(19)-C(20)-H(20)	119.8
O(1)-C(9)-O(2)	124.4(8)	C(20)-C(21)-C(22)	120.1(9)
O(1)-C(9)-C(10)	118.8(7)	C(20)-C(21)-H(21)	120.0
O(2)-C(9)-C(10)	116.8(7)	C(22)-C(21)-H(21)	120.0
O(1)-C(9)-Yb(1)	158.9(6)	C(23)-C(22)-C(21)	119.2(10)
C(10)-C(9)-Yb(1)	81.3(4)	C(23)-C(22)-H(22)	120.4
N(2)-C(10)-C(9)	111.8(6)	C(21)-C(22)-H(22)	120.4
N(2)-C(10)-H(10A)	109.3	C(18)-C(23)-C(22)	120.6(9)
C(9)-C(10)-H(10A)	109.3	C(18)-C(23)-H(23)	119.7
N(2)-C(10)-H(10B)	109.3	C(22)-C(23)-H(23)	119.7
C(9)-C(10)-H(10B)	109.3	C(29)-C(24)-C(25)	119.8(8)
H(10A)-C(10)-H(10B)	107.9	C(29)-C(24)-C(17)	123.2(9)

C(25)-C(24)-C(17)	117.0(8)
C(26)-C(25)-C(24)	119.4(9)
C(26)-C(25)-H(25)	120.3
C(24)-C(25)-H(25)	120.3
C(27)-C(26)-C(25)	121.9(9)
C(27)-C(26)-H(26)	119.1
C(25)-C(26)-H(26)	119.1
C(26)-C(27)-C(28)	118.4(9)
C(26)-C(27)-H(27)	120.8
C(28)-C(27)-H(27)	120.8
C(27)-C(28)-C(29)	122.2(10)
C(27)-C(28)-H(28)	118.9
C(29)-C(28)-H(28)	118.9
C(24)-C(29)-C(28)	118.3(9)
C(24)-C(29)-H(29)	120.9
C(28)-C(29)-H(29)	120.9
N(5)-C(30)-C(31)	114.3(7)
N(5)-C(30)-C(37)	110.6(7)
C(31)-C(30)-C(37)	115.5(7)
N(5)-C(30)-H(30)	105.1
C(31)-C(30)-H(30)	105.1
C(37)-C(30)-H(30)	105.1
C(32)-C(31)-C(36)	118.9(9)
C(32)-C(31)-C(30)	123.4(8)
C(36)-C(31)-C(30)	117.7(8)
C(31)-C(32)-C(33)	121.2(9)
C(31)-C(32)-H(32)	119.4
C(33)-C(32)-H(32)	119.4
C(34)-C(33)-C(32)	119.2(9)
C(34)-C(33)-H(33)	120.4
C(32)-C(33)-H(33)	120.4
C(35)-C(34)-C(33)	120.8(10)
C(35)-C(34)-H(34)	119.6
C(33)-C(34)-H(34)	119.6
C(34)-C(35)-C(36)	120.0(10)
C(34)-C(35)-H(35)	120.0
C(36)-C(35)-H(35)	120.0
C(35)-C(36)-C(31)	119.7(10)
C(35)-C(36)-H(36)	120.1
C(31)-C(36)-H(36)	120.1
O(8)-C(37)-N(6)	122.4(8)
O(8)-C(37)-C(30)	120.3(8)
N(6)-C(37)-C(30)	116.6(8)
N(6)-C(38)-C(39)	113.4(8)
N(6)-C(38)-H(38A)	108.9
C(39)-C(38)-H(38A)	108.9
N(6)-C(38)-H(38B)	108.9
C(39)-C(38)-H(38B)	108.9
H(38A)-C(38)-H(38B)	107.7
C(44)-C(39)-C(40)	118.8(9)
C(44)-C(39)-C(38)	122.3(8)
C(40)-C(39)-C(38)	118.7(8)
C(41)-C(40)-C(39)	120.6(10)
C(41)-C(40)-H(40)	119.7
C(39)-C(40)-H(40)	119.7
C(40)-C(41)-C(42)	120.2(10)
C(40)-C(41)-H(41)	119.9
C(42)-C(41)-H(41)	119.9
C(43)-C(42)-C(41)	119.3(10)
C(43)-C(42)-H(42)	120.3
C(41)-C(42)-H(42)	120.3
C(42)-C(43)-C(44)	120.2(10)
C(42)-C(43)-H(43)	119.9
C(44)-C(43)-H(43)	119.9
C(39)-C(44)-C(43)	120.8(9)
C(39)-C(44)-H(44)	119.6
C(43)-C(44)-H(44)	119.6
C(15)-N(1)-C(1)	109.7(6)
C(15)-N(1)-C(8)	110.4(7)
C(1)-N(1)-C(8)	107.9(6)
C(15)-N(1)-Yb(1)	107.4(5)
C(1)-N(1)-Yb(1)	110.5(5)
C(8)-N(1)-Yb(1)	110.9(5)
C(2)-N(2)-C(10)	109.4(7)
C(2)-N(2)-C(3)	108.2(7)
C(10)-N(2)-C(3)	108.3(6)
C(2)-N(2)-Yb(1)	113.8(5)
C(10)-N(2)-Yb(1)	105.8(5)
C(3)-N(2)-Yb(1)	111.2(5)
C(5)-N(3)-C(12)	111.5(8)
C(5)-N(3)-C(4)	109.6(7)
C(12)-N(3)-C(4)	108.4(8)
C(5)-N(3)-Yb(1)	111.4(6)
C(12)-N(3)-Yb(1)	104.8(6)
C(4)-N(3)-Yb(1)	110.9(5)
C(6)-N(4)-C(14)	110.2(7)
C(6)-N(4)-C(7)	109.9(7)
C(14)-N(4)-C(7)	107.6(7)
C(6)-N(4)-Yb(1)	110.3(5)
C(14)-N(4)-Yb(1)	105.5(5)
C(7)-N(4)-Yb(1)	113.2(5)
C(16)-N(5)-C(30)	119.1(7)
C(16)-N(5)-C(17)	122.4(7)
C(30)-N(5)-C(17)	118.4(7)
C(37)-N(6)-C(38)	121.7(8)
C(37)-N(6)-H(6)	119.1
C(38)-N(6)-H(6)	119.1
C(9)-O(2)-Yb(1)	125.2(5)
C(11)-O(4)-Yb(1)	124.6(6)

C(13)-O(6)-Yb(1)	125.7(6)
C(16)-O(7)-Yb(1)	125.1(5)
Yb(1)-O(9)-D(9A)	109(3)
Yb(1)-O(9)-D(9B)	109(3)
D(9A)-O(9)-D(9B)	112(4)
O(2)-Yb(1)-O(6)	143.3(2)
O(2)-Yb(1)-O(4)	84.9(2)
O(6)-Yb(1)-O(4)	87.6(2)
O(2)-Yb(1)-O(7)	79.0(2)
O(6)-Yb(1)-O(7)	86.0(2)
O(4)-Yb(1)-O(7)	143.2(2)
O(2)-Yb(1)-O(9)	72.1(2)
O(6)-Yb(1)-O(9)	71.5(2)
O(4)-Yb(1)-O(9)	71.5(2)
O(7)-Yb(1)-O(9)	72.2(2)
O(2)-Yb(1)-N(3)	135.1(2)
O(6)-Yb(1)-N(3)	72.9(2)
O(4)-Yb(1)-N(3)	67.9(2)
O(7)-Yb(1)-N(3)	142.5(2)
O(9)-Yb(1)-N(3)	126.1(2)
O(2)-Yb(1)-N(2)	67.8(2)
O(6)-Yb(1)-N(2)	142.1(2)
O(4)-Yb(1)-N(2)	72.0(2)
O(7)-Yb(1)-N(2)	129.1(2)
O(9)-Yb(1)-N(2)	127.1(2)
N(3)-Yb(1)-N(2)	70.0(2)
O(2)-Yb(1)-N(4)	136.8(2)
O(6)-Yb(1)-N(4)	67.6(2)
O(4)-Yb(1)-N(4)	135.3(2)
O(7)-Yb(1)-N(4)	74.1(2)
O(9)-Yb(1)-N(4)	127.9(2)
N(3)-Yb(1)-N(4)	69.4(2)
N(2)-Yb(1)-N(4)	105.0(2)
O(2)-Yb(1)-N(1)	70.9(2)
O(6)-Yb(1)-N(1)	131.6(2)
O(4)-Yb(1)-N(1)	138.3(2)
O(7)-Yb(1)-N(1)	65.5(2)
O(9)-Yb(1)-N(1)	127.7(2)
N(3)-Yb(1)-N(1)	106.2(2)
N(2)-Yb(1)-N(1)	67.6(2)
N(4)-Yb(1)-N(1)	67.3(2)
O(2)-Yb(1)-C(9)	19.0(2)
O(6)-Yb(1)-C(9)	152.3(2)
O(4)-Yb(1)-C(9)	73.3(2)
O(7)-Yb(1)-C(9)	97.1(2)
O(9)-Yb(1)-C(9)	83.2(2)
N(3)-Yb(1)-C(9)	116.0(2)
N(2)-Yb(1)-C(9)	50.3(2)
N(4)-Yb(1)-C(9)	139.8(2)
N(1)-Yb(1)-C(9)	73.3(2)
O(2)-Yb(1)-C(16)	66.7(2)
O(6)-Yb(1)-C(16)	104.1(2)
O(4)-Yb(1)-C(16)	146.3(2)
O(7)-Yb(1)-C(16)	18.57(19)
O(9)-Yb(1)-C(16)	82.3(2)
N(3)-Yb(1)-C(16)	145.7(2)
N(2)-Yb(1)-C(16)	110.5(2)
N(4)-Yb(1)-C(16)	77.8(2)
N(1)-Yb(1)-C(16)	49.5(2)
C(9)-Yb(1)-C(16)	83.0(2)
O(2)-Yb(1)-C(11)	103.4(2)
O(6)-Yb(1)-C(11)	75.9(2)
O(4)-Yb(1)-C(11)	19.2(2)
O(7)-Yb(1)-C(11)	152.6(2)
O(9)-Yb(1)-C(11)	82.5(2)
N(3)-Yb(1)-C(11)	50.2(2)
N(2)-Yb(1)-C(11)	75.0(2)
N(4)-Yb(1)-C(11)	116.1(2)
N(1)-Yb(1)-C(11)	141.5(2)
C(9)-Yb(1)-C(11)	90.0(2)
C(16)-Yb(1)-C(11)	163.9(2)
D(1WA)-O(101)-D(1WB)	107(4)
D(2WA)-O(201)-D(2WB)	107(4)
D(3WA)-O(301)-D(3WB)	106(4)
D(4WA)-O(401)-D(4WB)	110(5)
D(5WA)-O(501)-D(5WB)	110(5)
D(8WA)-O(801)-D(8WB)	106(5)
D(9WA)-O(901)-D(9WB)	110(5)
D(0WA)-O(001)-D(0WB)	109(5)

266

Symmetry transformations used to generate equivalent atoms:

Table 4. Anisotropic displacement parameters ($\text{Å}^2 \times 10^3$) for B_2Yb. The anisotropic displacement factor exponent takes the form: $-2p^2[h^2a^{*2}U^{11} + ... + 2hka^*b^*U^{12}]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
C(1)	23(4)	14(5)	21(5)	-4(4)	10(4)	-4(4)
C(2)	22(5)	9(4)	17(5)	-7(3)	5(4)	-3(3)
C(3)	24(5)	15(5)	25(5)	-4(4)	6(4)	1(4)
C(4)	27(5)	14(5)	27(5)	4(4)	14(4)	7(4)
C(5)	13(4)	29(5)	18(5)	10(4)	0(4)	5(4)
C(6)	19(4)	34(6)	30(5)	-2(5)	15(4)	2(4)
C(7)	10(4)	22(5)	23(5)	8(4)	5(4)	-3(4)
C(8)	24(5)	15(5)	27(5)	7(4)	13(4)	2(4)
C(9)	15(4)	13(4)	15(4)	-5(4)	6(4)	-8(3)
C(10)	14(4)	21(5)	13(4)	-7(4)	-5(3)	-6(4)
C(11)	27(5)	23(5)	17(5)	-2(4)	2(4)	0(4)
C(12)	30(5)	38(7)	41(6)	3(5)	22(5)	3(5)
C(13)	15(4)	23(5)	23(5)	6(4)	2(4)	10(3)
C(14)	15(4)	18(4)	15(4)	-5(3)	2(3)	-4(3)
C(15)	13(3)	21(4)	11(3)	-1(3)	1(3)	-1(3)
C(16)	6(3)	24(5)	13(4)	-4(3)	2(3)	-8(3)
C(17)	26(4)	9(4)	19(4)	4(3)	8(3)	4(3)
C(18)	23(4)	9(4)	23(4)	3(3)	9(4)	-1(3)
C(19)	26(5)	18(5)	38(6)	1(4)	9(4)	1(4)
C(20)	24(5)	15(5)	52(7)	0(5)	9(5)	5(4)
C(21)	25(5)	28(6)	71(8)	11(6)	28(6)	11(5)
C(22)	36(6)	22(5)	43(6)	8(5)	23(5)	18(4)
C(23)	21(4)	11(4)	27(5)	-2(4)	8(4)	5(4)
C(24)	21(4)	14(5)	29(5)	-15(4)	13(4)	-2(4)
C(25)	27(5)	18(5)	31(5)	9(4)	13(4)	4(4)
C(26)	38(6)	13(5)	32(5)	5(4)	15(5)	0(4)
C(27)	26(5)	30(6)	41(6)	-1(5)	16(5)	-12(4)
C(28)	28(5)	38(7)	36(6)	2(5)	13(5)	-6(5)
C(29)	22(5)	29(6)	25(5)	11(4)	11(4)	-5(4)
C(30)	22(4)	21(5)	11(4)	3(4)	4(4)	3(4)
C(31)	20(4)	25(5)	13(4)	0(4)	6(4)	2(4)
C(32)	24(5)	16(5)	28(5)	11(4)	13(4)	6(4)
C(33)	23(5)	22(5)	34(6)	-11(4)	9(4)	-6(4)
C(34)	28(5)	59(8)	48(7)	-1(6)	26(5)	2(6)
C(35)	32(6)	56(8)	49(7)	2(6)	28(6)	2(6)
C(36)	36(6)	33(6)	34(6)	-11(5)	24(5)	-3(5)
C(37)	28(5)	15(5)	13(4)	-1(4)	5(4)	-4(4)
C(38)	37(5)	24(5)	19(5)	4(4)	14(4)	5(4)
C(39)	26(4)	14(4)	23(4)	10(3)	11(3)	2(3)
C(40)	29(4)	30(5)	26(5)	-2(4)	14(4)	2(4)
C(41)	30(6)	51(8)	35(6)	1(6)	4(5)	-9(5)
C(42)	52(7)	36(7)	30(6)	-6(5)	15(5)	-15(6)
C(43)	45(6)	7(5)	30(5)	1(4)	14(5)	5(4)
C(44)	28(4)	27(5)	22(4)	5(4)	12(4)	7(4)
N(1)	14(3)	20(4)	15(3)	-6(3)	7(3)	3(3)
N(2)	13(3)	9(4)	15(4)	-2(3)	-1(3)	0(3)
N(3)	17(4)	25(4)	22(4)	-3(3)	9(3)	1(3)
N(4)	13(3)	15(4)	15(4)	-2(3)	3(3)	-3(3)
N(5)	18(4)	13(4)	11(3)	-1(3)	4(3)	-1(3)
N(6)	27(4)	22(4)	23(4)	-5(3)	11(3)	-3(3)
O(1)	17(3)	36(4)	17(3)	3(3)	4(3)	10(3)
O(2)	16(3)	15(3)	16(3)	-2(3)	5(2)	-2(2)
O(3)	26(3)	17(4)	58(5)	17(3)	12(3)	-5(3)
O(4)	15(3)	13(3)	17(3)	3(3)	0(2)	2(2)
O(5)	19(3)	31(4)	23(4)	4(3)	5(3)	2(3)
O(6)	12(3)	23(4)	15(3)	8(3)	3(2)	4(3)
O(7)	21(3)	17(3)	17(3)	11(3)	11(3)	6(3)
O(8)	17(3)	20(3)	23(3)	-8(3)	2(3)	-4(3)
O(9)	16(3)	35(4)	22(3)	11(3)	10(3)	9(3)
Yb(1)	12(1)	13(1)	14(1)	0(1)	4(1)	-1(1)
O(101)	28(4)	16(4)	35(4)	7(3)	16(3)	2(3)
O(201)	23(3)	27(4)	15(3)	2(3)	0(3)	1(3)
O(301)	33(4)	25(4)	46(5)	0(4)	10(4)	-9(3)
O(401)	22(4)	60(5)	35(4)	3(4)	15(3)	-7(4)
O(501)	30(4)	42(5)	37(4)	3(4)	10(3)	-5(4)
O(601)	38(4)	62(6)	28(4)	-1(4)	14(3)	-8(4)
O(801)	92(7)	49(6)	108(9)	9(6)	79(7)	-6(6)
O(901)	93(8)	101(10)	113(10)	39(9)	81(8)	37(8)
O(001)	98(13)	32(7)	124(16)	34(8)	82(12)	24(7)

Table 5. Hydrogen coordinates ($x\,10^4$) and isotropic displacement parameters ($\text{Å}^2 \times 10^3$) for B_2Yb.

	x	y	z	U(eq)
H(1A)	2376	1154	-711	22
H(1B)	1107	1143	-386	22
H(2A)	1350	666	-1250	20
H(2B)	2772	626	-210	20
H(3A)	1008	76	645	27
H(3B)	1511	137	-462	27
H(4A)	3146	-97	1255	26

H(4B)	3589	239	976	26
H(5A)	5098	54	3049	26
H(5B)	4897	223	4196	26
H(6A)	6197	535	3461	31
H(6B)	5089	538	2091	31
H(7A)	5942	1084	2818	22
H(7B)	4907	1287	3141	22
H(8A)	4413	1291	952	25
H(8B)	4299	919	916	25
H(10A)	-467	411	198	22
H(10B)	-396	602	-963	22
H(12A)	3144	0	4091	41
H(12B)	2785	-216	2881	41
H(14A)	5192	1091	4862	21
H(14B)	6169	807	5020	21
H(15A)	1606	1515	843	19
H(15B)	3108	1602	1335	19
H(17)	1821	1987	1522	22
H(19)	-401	2035	1294	34
H(20)	-2005	2340	1575	39
H(21)	-1435	2756	2977	46
H(22)	764	2876	4092	37
H(23)	2360	2562	3835	24
H(25)	2608	2526	1164	30
H(26)	4474	2805	1455	32
H(27)	6407	2686	3067	38
H(28)	6492	2261	4355	41
H(29)	4648	1968	4106	30
H(30)	2427	2010	4665	22
H(32)	241	1441	2910	27
H(33)	-1806	1357	2972	32
H(34)	-2305	1578	4574	50
H(35)	-838	1897	6045	50
H(36)	1163	2011	5908	38
H(38A)	4828	1240	7147	31
H(38B)	3741	1115	7612	31
H(40)	6421	1360	9112	33
H(41)	7215	1675	10869	49
H(42)	5834	2020	11337	47
H(43)	3689	2060	9976	33
H(44)	2894	1741	8215	30
H(6)	2270	1419	6154	28

Table 6. Hydrogen bonds for B₂Yb [Å and °].

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
N(6)-H(6)...O(601)	0.88	2.29	2.883(11)	124.4
O(9)-D(9A)...O(401)#1	0.94(2)	1.92(3)	2.776(10)	151(4)
O(9)-D(9B)...O(601)	0.94(2)	2.12(4)	2.812(10)	129(5)
O(101)-D(1WA)...O(201)#2	0.95(2)	1.82(3)	2.765(10)	171(10)
O(101)-D(1WB)...O(3)	0.95(2)	1.78(4)	2.708(9)	164(9)
O(201)-D(2WA)...O(5)	0.95(2)	1.82(3)	2.752(9)	166(8)
O(201)-D(2WB)...O(1)#3	0.95(2)	1.74(2)	2.692(8)	175(11)
O(301)-D(3WA)...O(1)#3	0.95(2)	2.02(5)	2.911(10)	156(11)
O(301)-D(3WB)...O(501)	0.95(2)	2.10(6)	2.960(11)	149(9)
O(401)-D(4WA)...O(101)#2	0.95(2)	1.89(3)	2.824(11)	170(10)
O(401)-D(4WB)...O(701)	0.94(2)	1.88(6)	2.72(2)	147(9)
O(401)-D(4WB)...O(71')	0.94(2)	2.20(8)	2.94(3)	135(9)
O(501)-D(5WA)...O(8)	0.95(2)	2.08(6)	2.942(9)	150(9)
O(501)-D(5WB)...O(201)	0.95(2)	1.97(3)	2.908(11)	169(12)
O(601)-D(6WA)...O(101)#4	0.96(2)	2.32(13)	2.829(10)	113(10)
O(801)-D(8WA)...O(5)	0.95(2)	1.88(4)	2.821(11)	170(16)
O(801)-D(8WB)...O(601)	0.96(2)	2.26(17)	2.882(14)	122(15)
O(901)-D(9WA)...O(801)	0.95(2)	1.92(9)	2.807(19)	154(19)
O(901)-D(9WB)...O(71')#5	0.95(2)	1.77(13)	2.60(4)	142(19)
O(901)-D(9WB)...O(701)#5	0.95(2)	2.37(15)	3.13(4)	137(17)
O(001)-D(0WA)...O(401)	0.95(2)	1.86(6)	2.789(16)	165(18)
O(001)-D(0WB)...O(001)#6	0.95(2)	2.28(12)	3.10(4)	144(16)

Symmetry transformations used to generate equivalent atoms:
#1 x-1,y,z   #2 -x+1,-y,-z+1   #3 x+1,y,z+1   #4 -x,-y,-z+1
#5 x,y,z+1   #6 -x+2,-y,-z

5.2 Appendix 2

Variable temperature (VT) nmr

Compound **04**: Benzyl Ugi
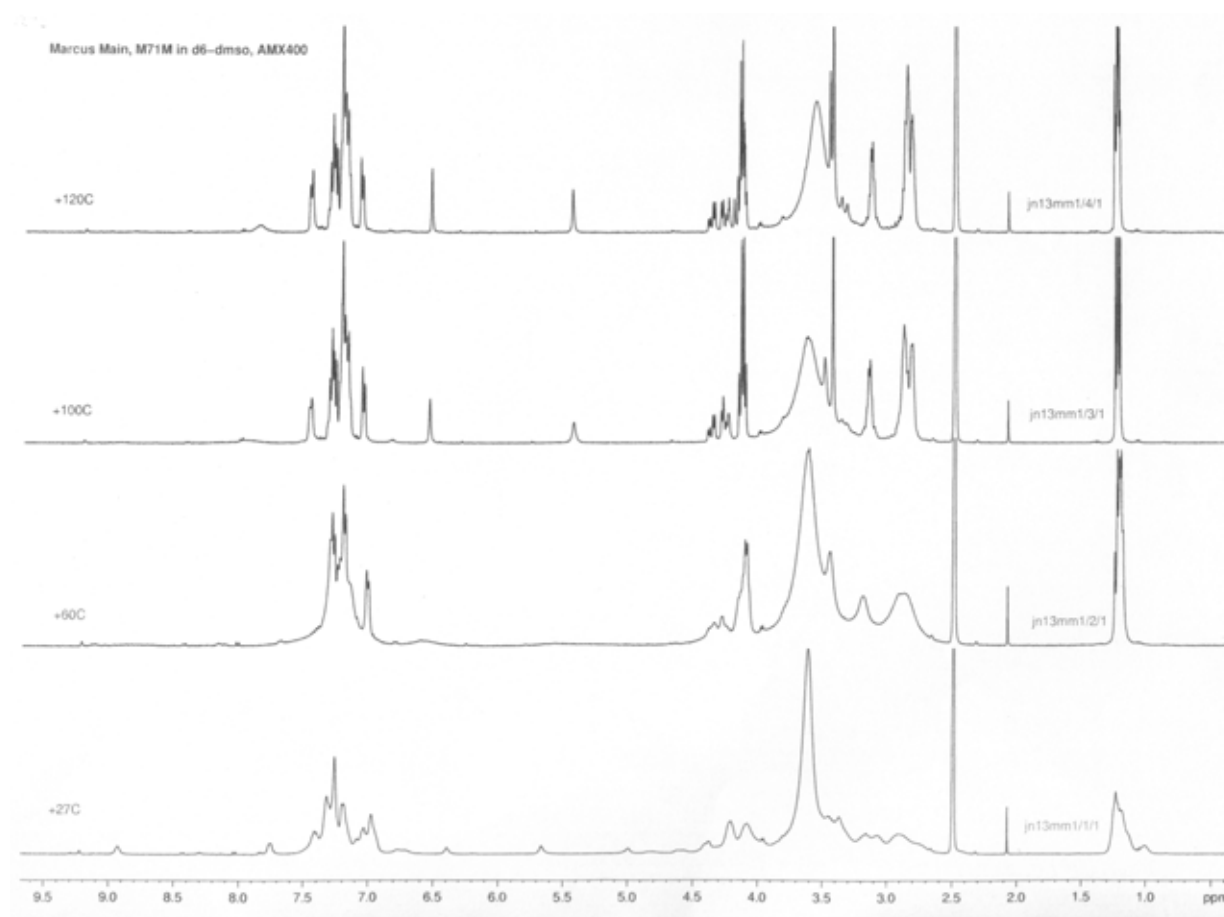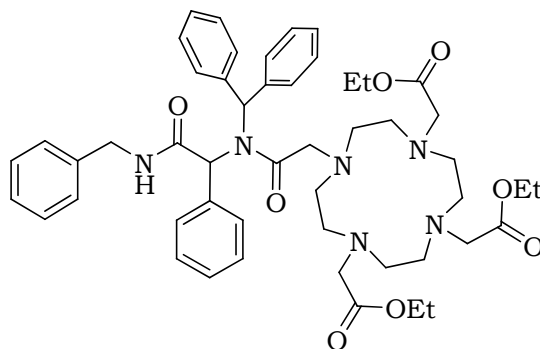
Compound **05**: Naphthyl Ugi

Marcus Main, M73M in d6–dmso, AMX400

+120C   jn14mm1/4/1

+100C   jn14mm1/3/1

+60C   jn14mm1/2/1

+27C   jn14mm1/1/1
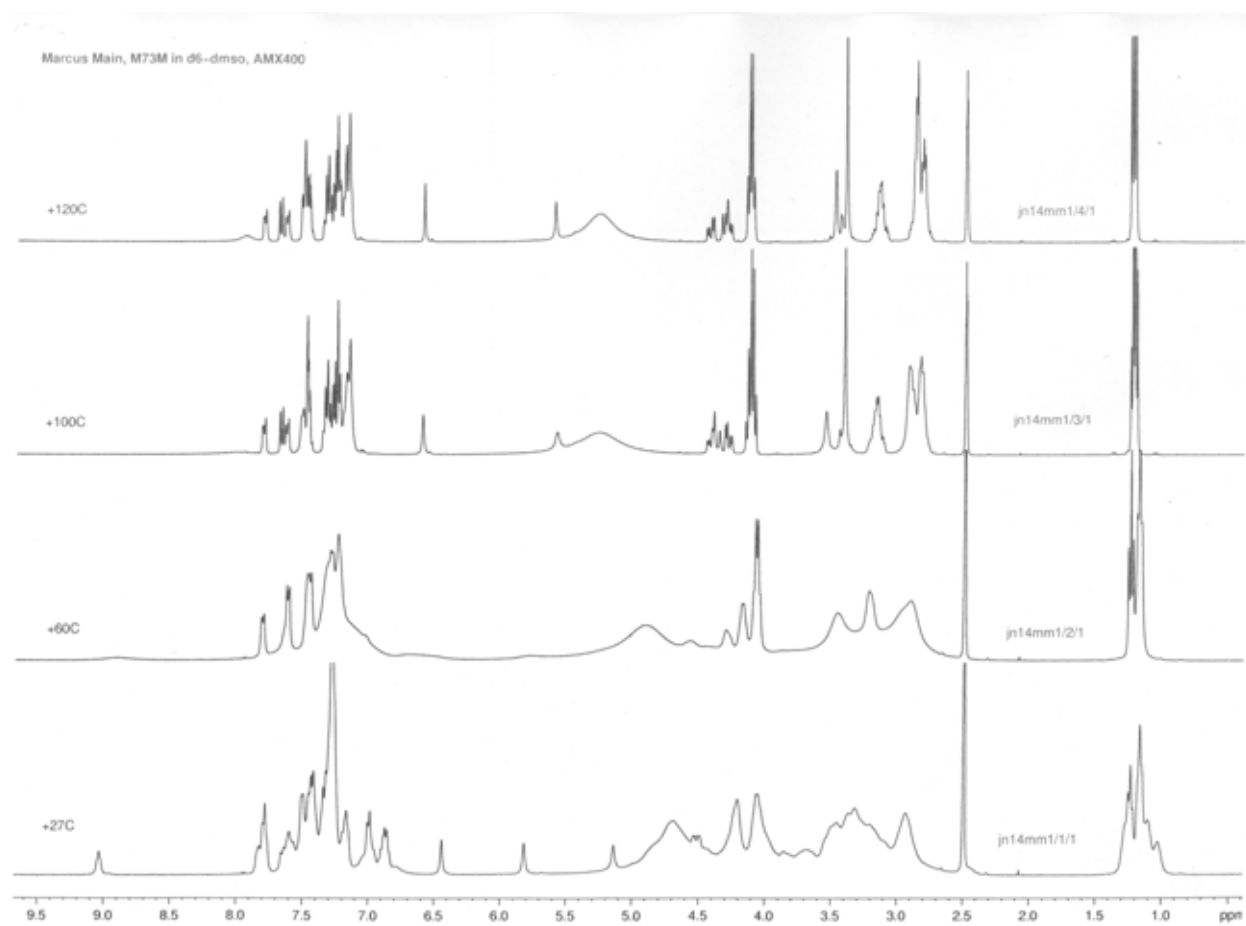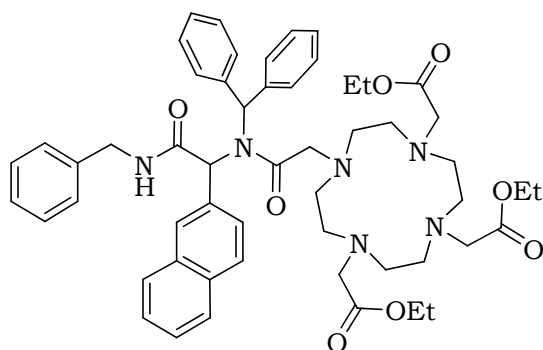
Compound **06**: Pyrenyl Ugi
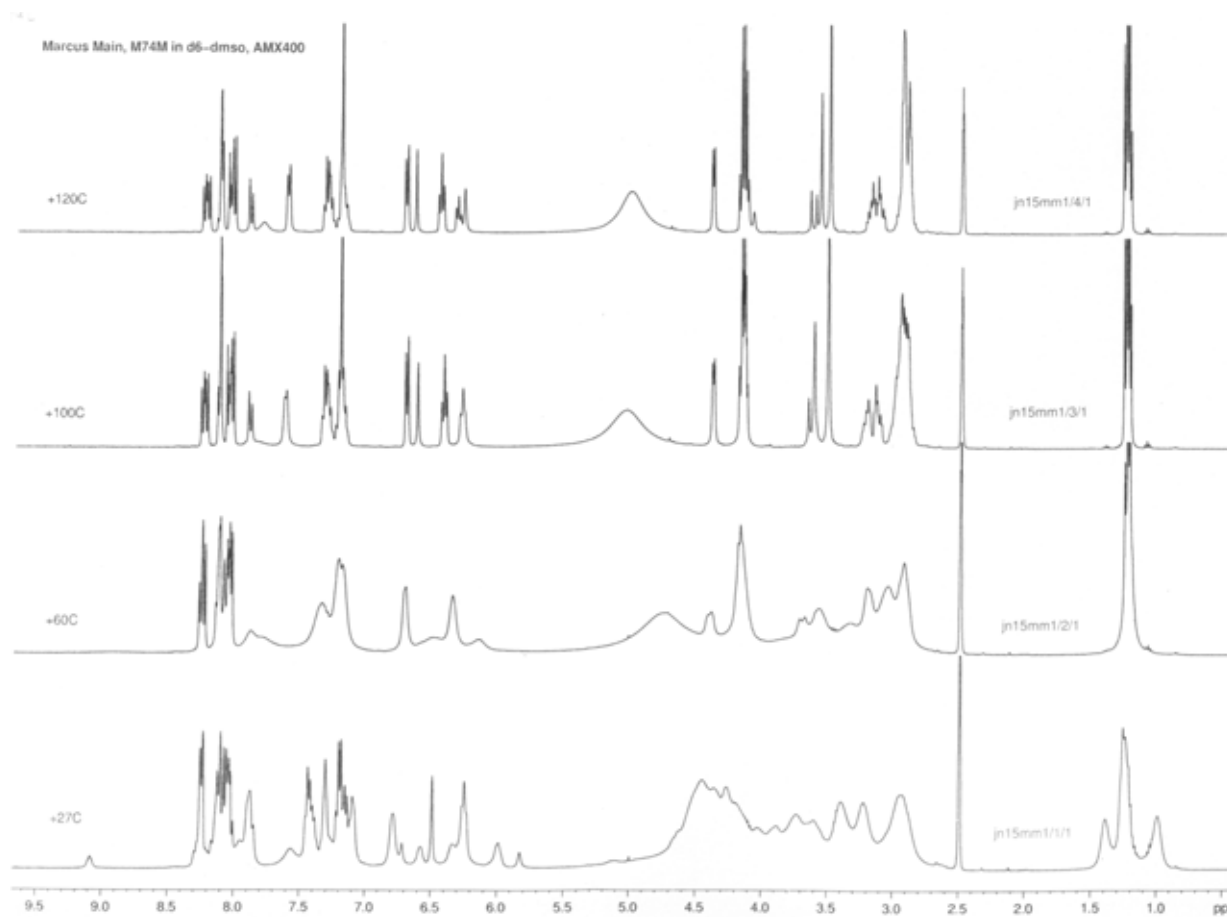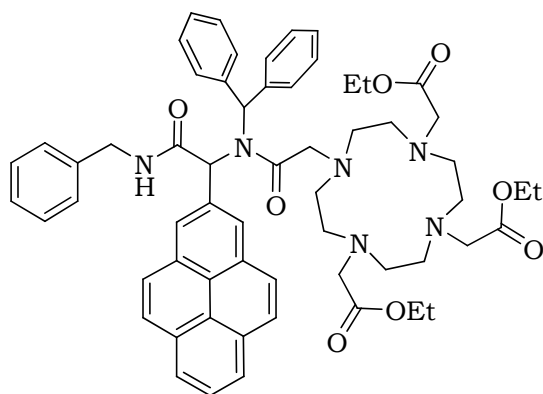
5.3 Appendix 3

Lanthanide luminescence spectra

Eu Benzyl D$_2$O

Excitation of metal centre
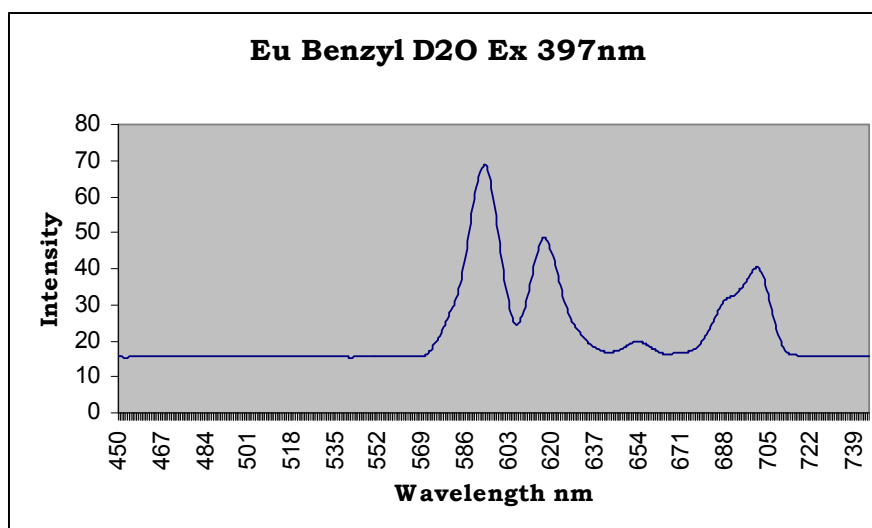
Eu Benzyl D2O Ex 397nm

Excitation of chromophore

Eu Benzyl D2O Ex 258nm

Combination: excitation of metal centre vs energy-transfer from chromophore

Eu Benzyl D2O

Lifetime

LIFETIME Eu Benzyl D2O Ex 258nm; Em 617nm

Eu Benzyl H₂O

Excitation of metal centre

Eu Benzyl H2O Ex 397nm

Excitation of chromophore

Eu Benzyl H2O Ex 258nm

Combination: Excitation of metal centre vs energy-transfer from chromophore

Eu benzyl H2O comparison

Lifetime

LIFETIME Eu Benzyl H2O Ex 258nm; Em 617nm
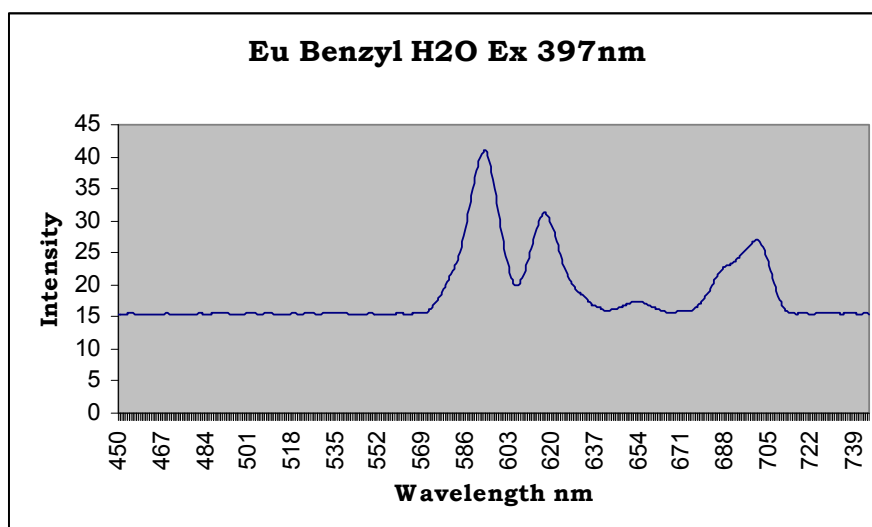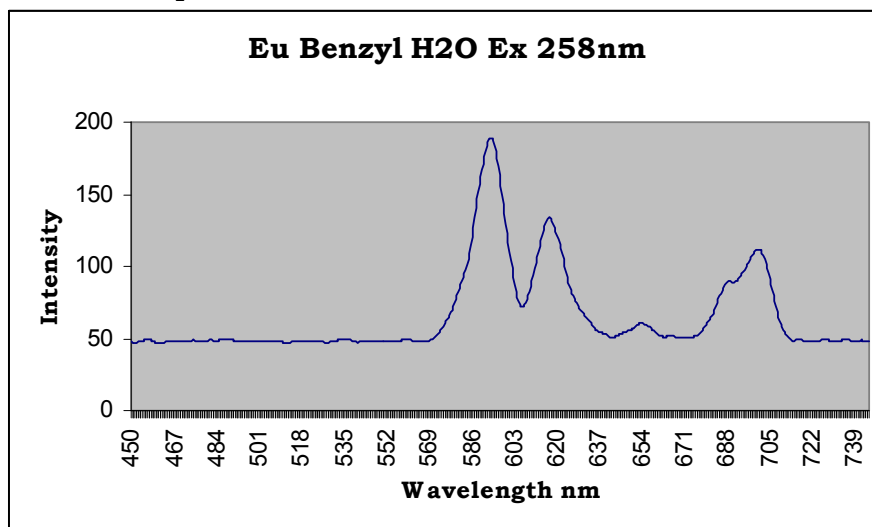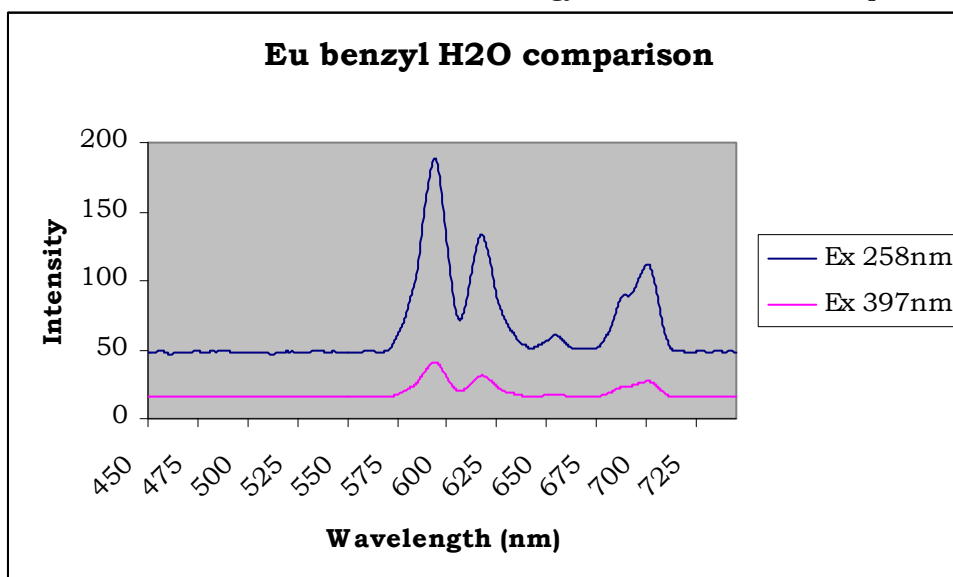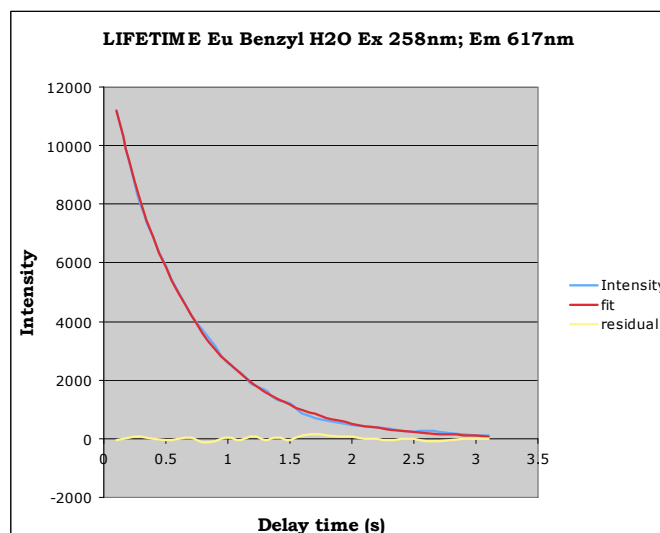
Eu Naphthyl H$_2$O

Absorption spectrum

Eu Naphthyl Abs

Excitation of metal

Eu Napthyl H2O Ex 259nm

Excitation of chromophore

Eu Naphthyl H2O Ex 290nm

Comparison: Excitation of metal centre vs excitation of chromophore

Eu Naphthyl H2O comparison

Eu Indolyl D$_2$O

Excitation of metal centre

Eu Indole D2O Ex 595nm

Excitation of chromophore

Eu Indole Ex 397nm

Eu Indole D2O Ex 275nm
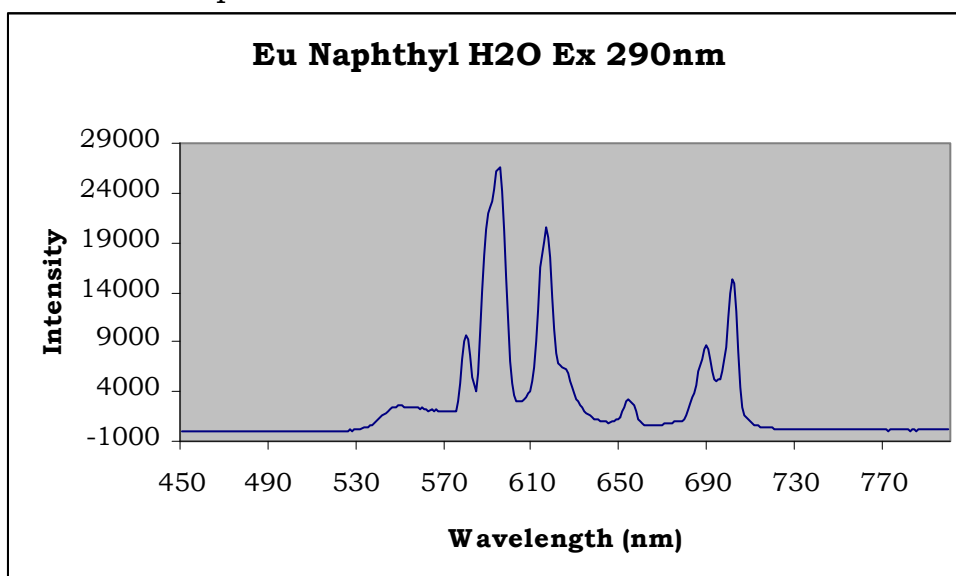
Comparison: Excitation of metal centre vs excitation of chromophore
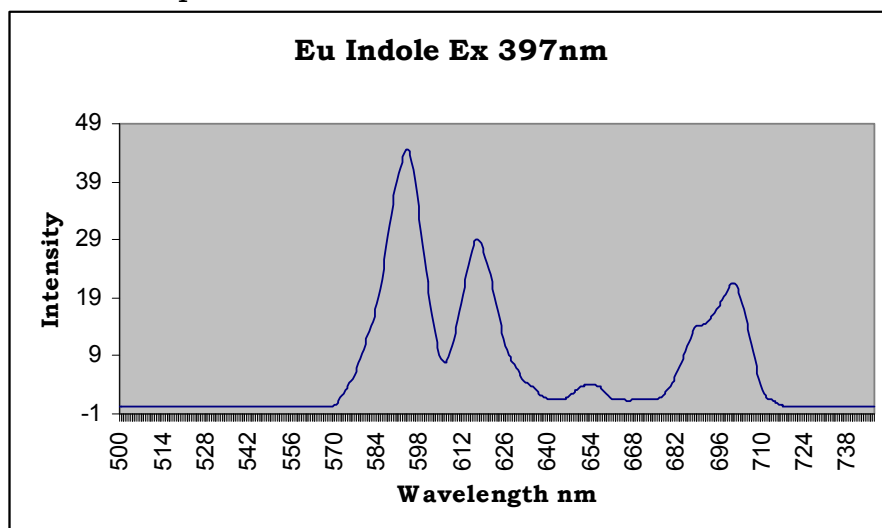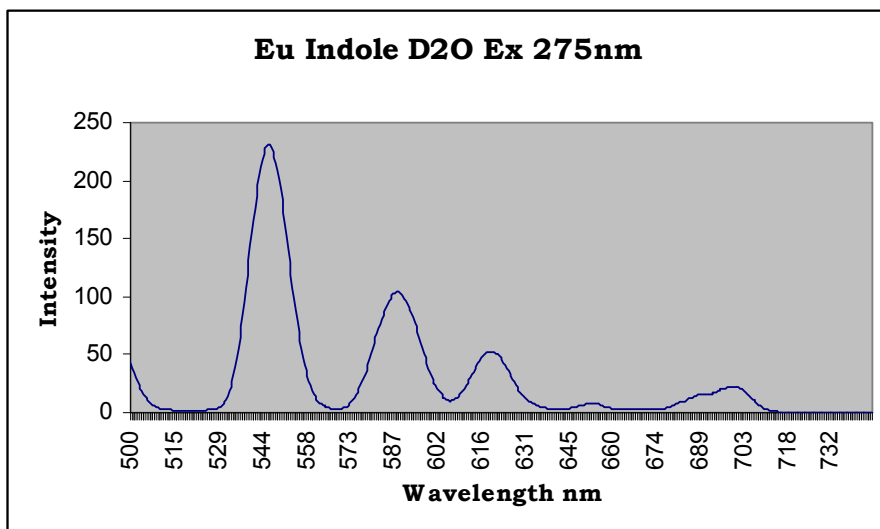
Eu Indolyl D2O comparison

Eu Indole H$_2$O
Absorption

Eu Indole H2O Abs

Excitation of metal centre

Eu Indole H2O Ex 545nm

Excitation of chromophore

Eu Indole H2O Ex 397nm

Eu Indole H2O Ex 275nm

Comparison: Excitation of metal centre vs excitation of chromophore

Eu Indolyl H2O comparison

Tb Benzyl D$_2$O

Excitation of metal centre

Tb Benzyl D2O Ex 366nm

Excitation of chromophore

Tb Benzyl D2O Ex 258nm

Tb Benzyl D2O Ex 258nm

Comparison: Excitation of metal centre vs excitation of chromophore

Tb benzyl D2O comparison

Lifetime

LIFETIME Tb Benzyl D2O Ex 258nm; Em 545nm

Tb Benzyl H$_2$O

Excitation of metal centre

Tb Benzyl H2O Ex 366nm

Excitation of chromophore

Tb Benzyl H2O Ex 258nm

Tb Benzyl H2O Ex 366nm

Comparison: Excitation of metal centre vs excitation of chromophore

Tb benzyl H2O comparison

Lifetime

LIFETIME Tb Benzyl Ex 258nm; Em 545nm

Tb Naphthyl H$_2$O (Ex 267; Em 550nm)

Excitation of metal centre

Tb Naphthyl H2O Ex 259nm
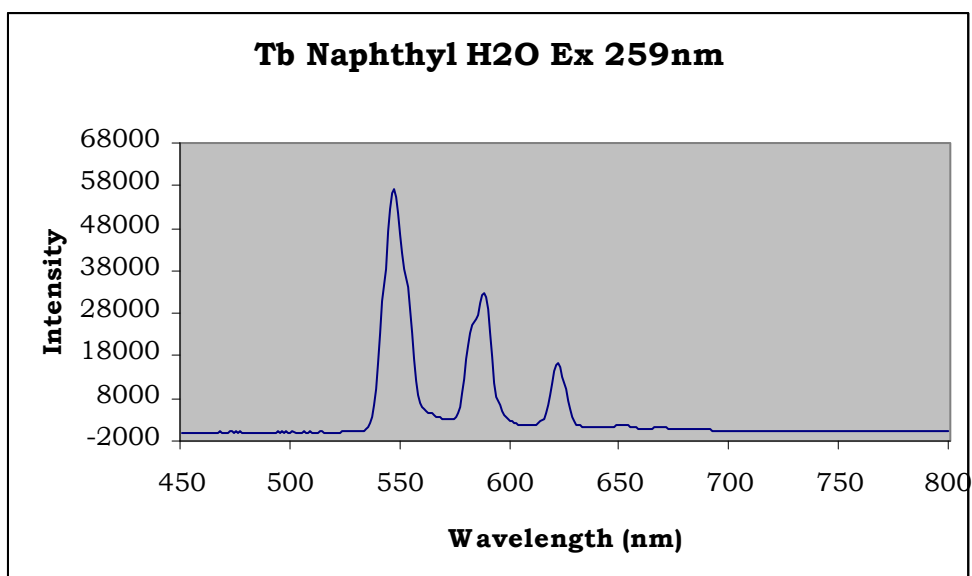
Excitation of chromophore

Comparison: Excitation of metal centre vs excitation of chromophore
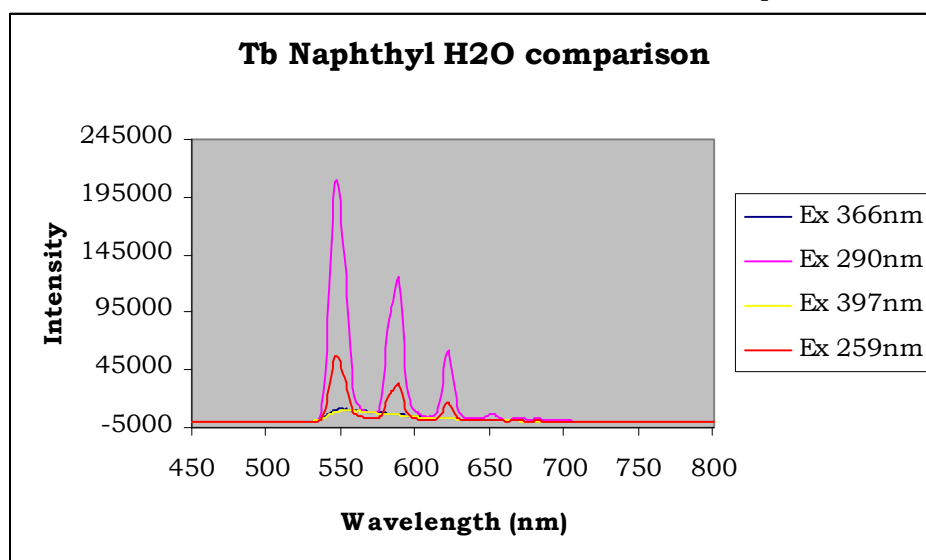
Tb Indole D$_2$O

Absorption

Excitation of metal centre

Tb Indolyl D2O Ex 366nm

Excitation of chromophore

Tb Indolyl D2O Ex 275nm

Combination: excitation of metal centre vs energy-transfer from chromophore

Tb Indolyl D2O comparison
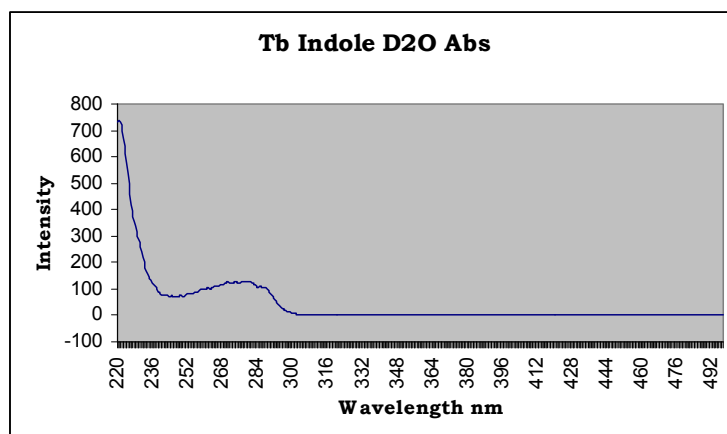
Tb Indolyl H$_2$O

Excitation of metal centre
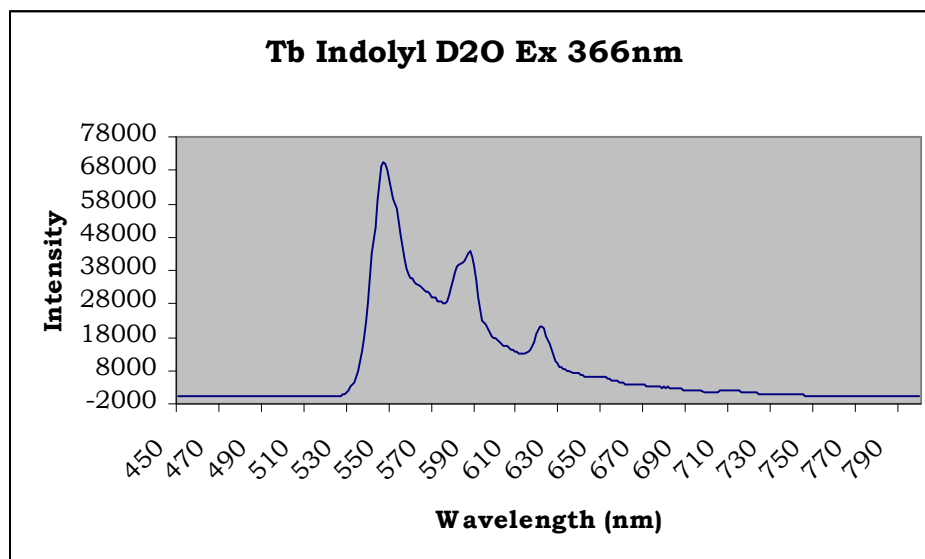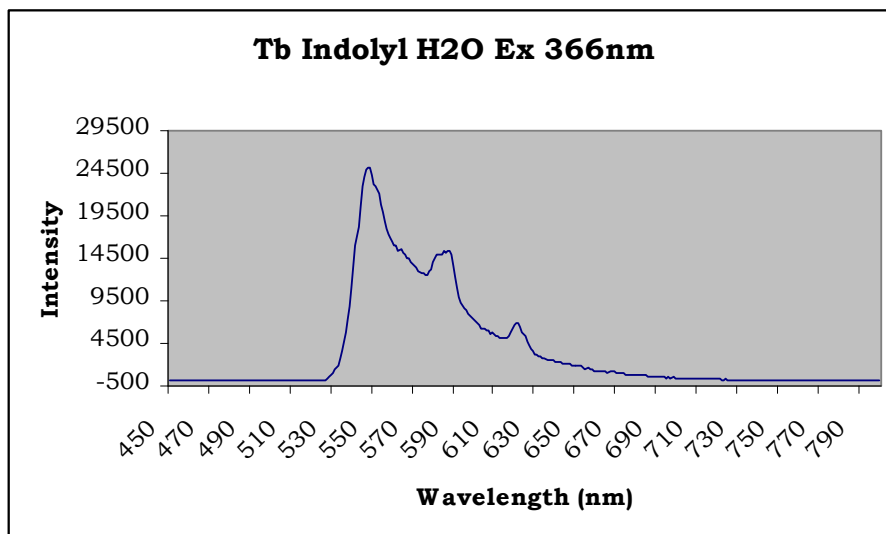
Excitation of chromophore
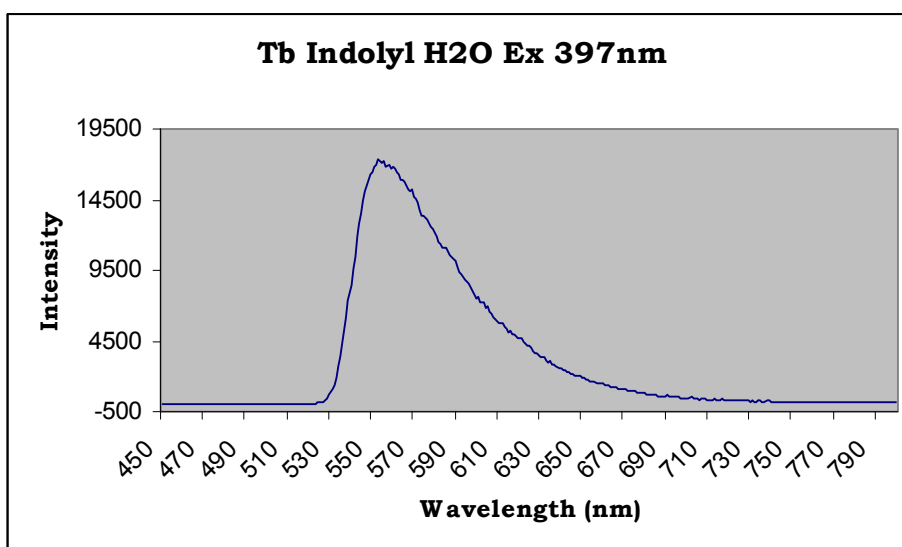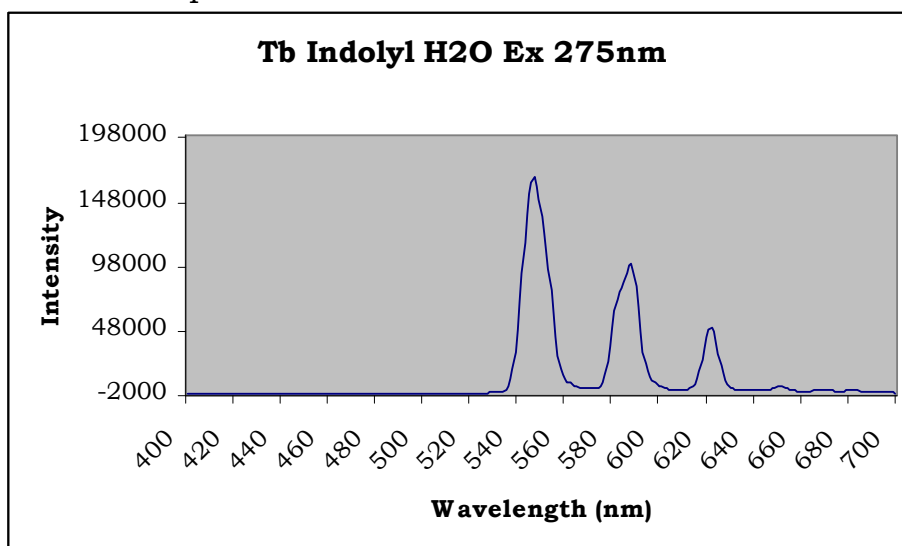
Combination: excitation of metal centre vs energy-transfer from chromophore

Tb Indolyl H2O comparison

Ex 275 nm
Ex 397 nm
Ex 366 nm

Intensity

Wavelength (nm)
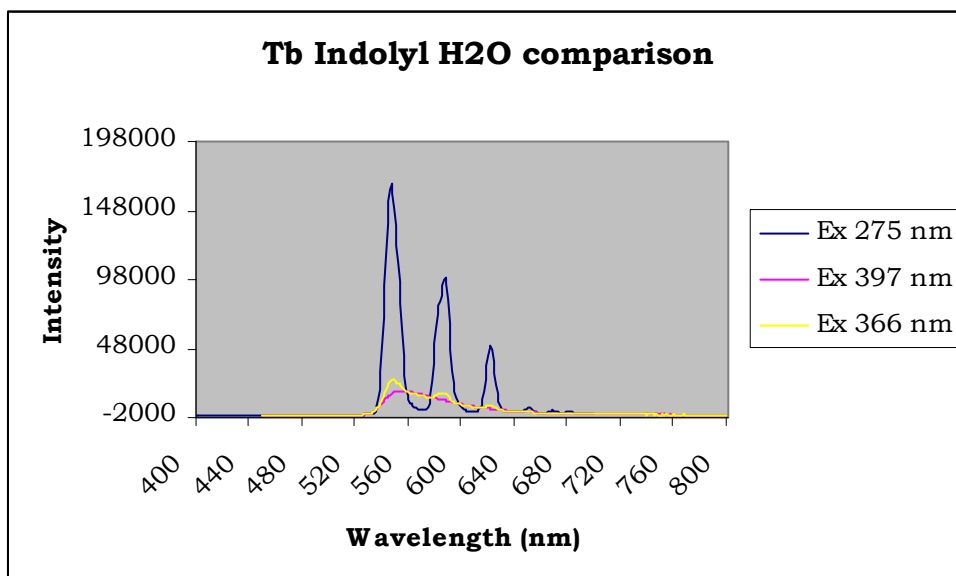
Yb Indolyl D$_2$O (Ex 337nm, Em 980)

decay
fit
difference
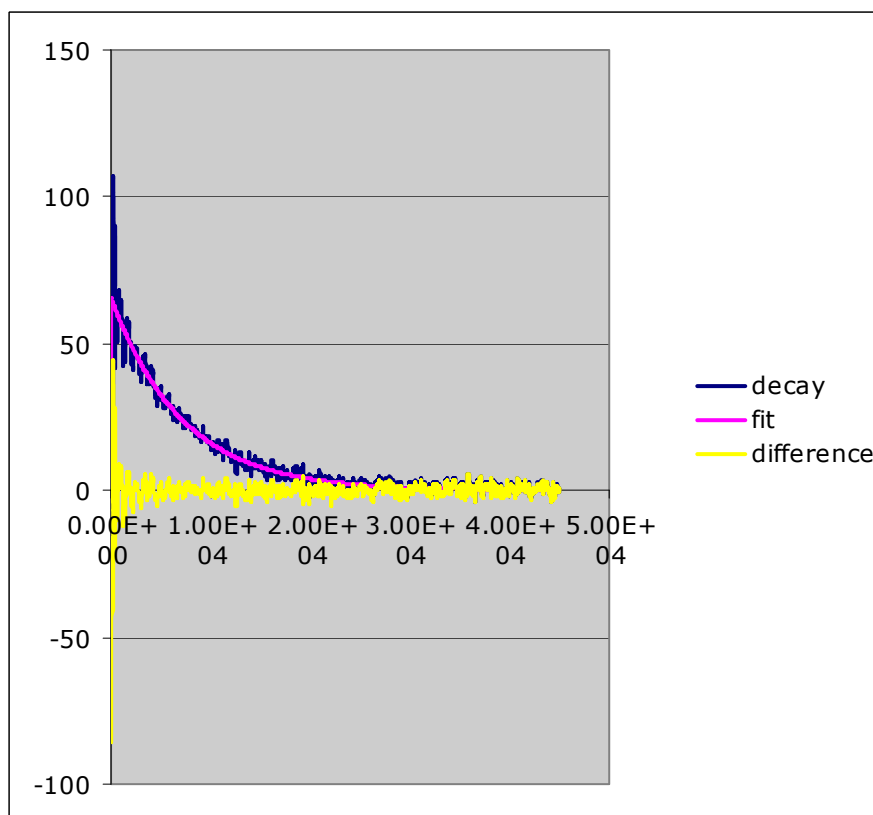
Yb Pyrenyl MeOH
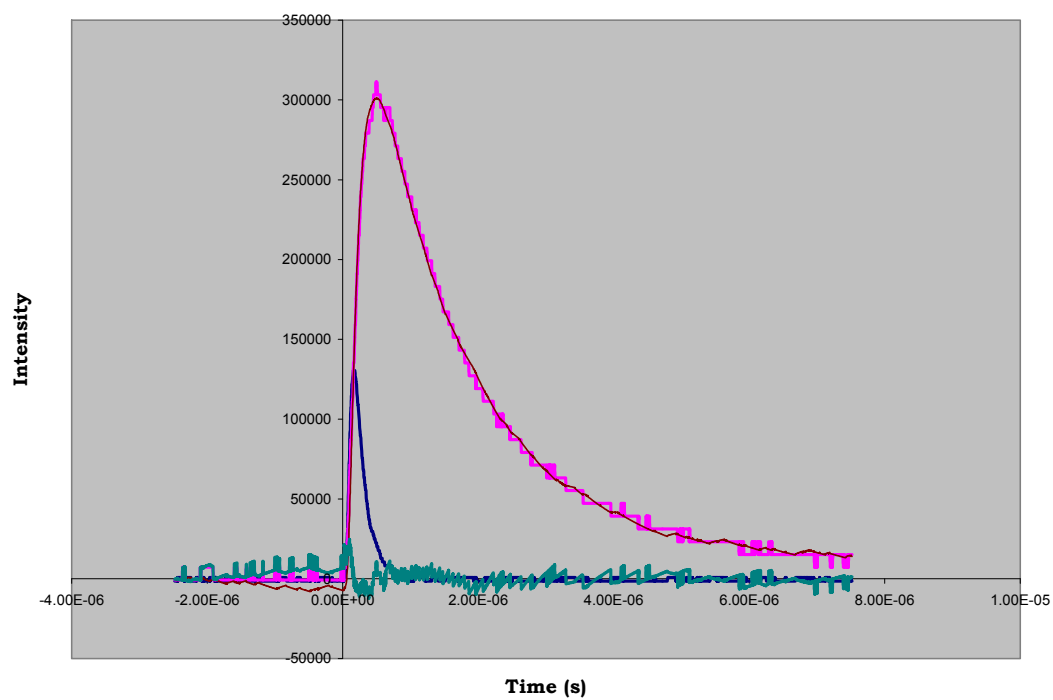
Yb Pyrene d4-MeOH
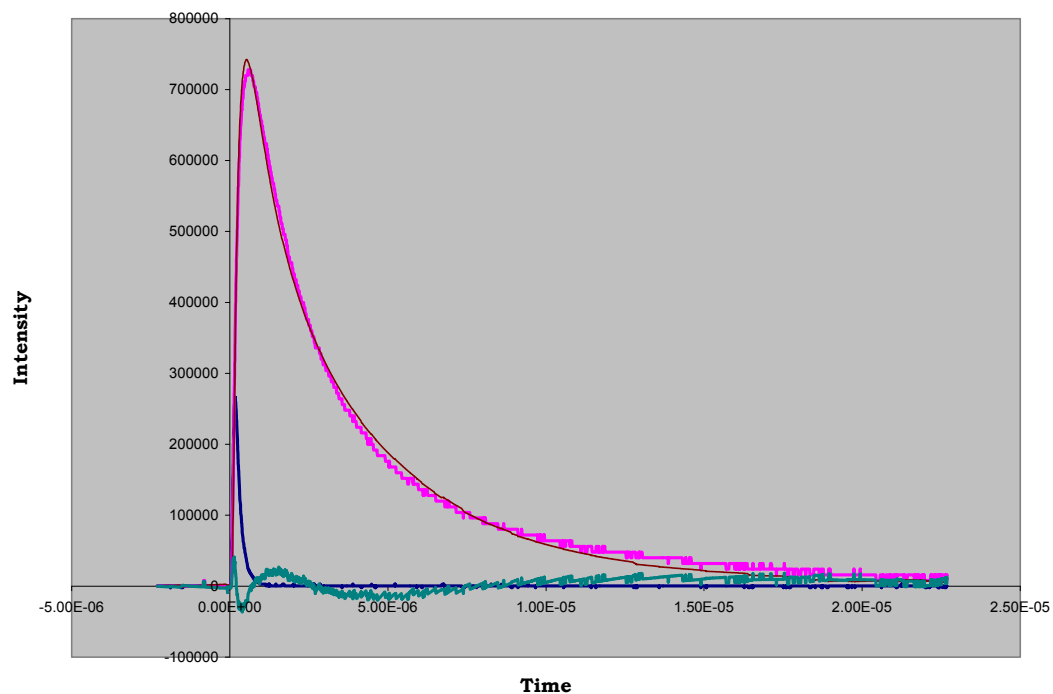
Yb Naphthyl d4-MeOH
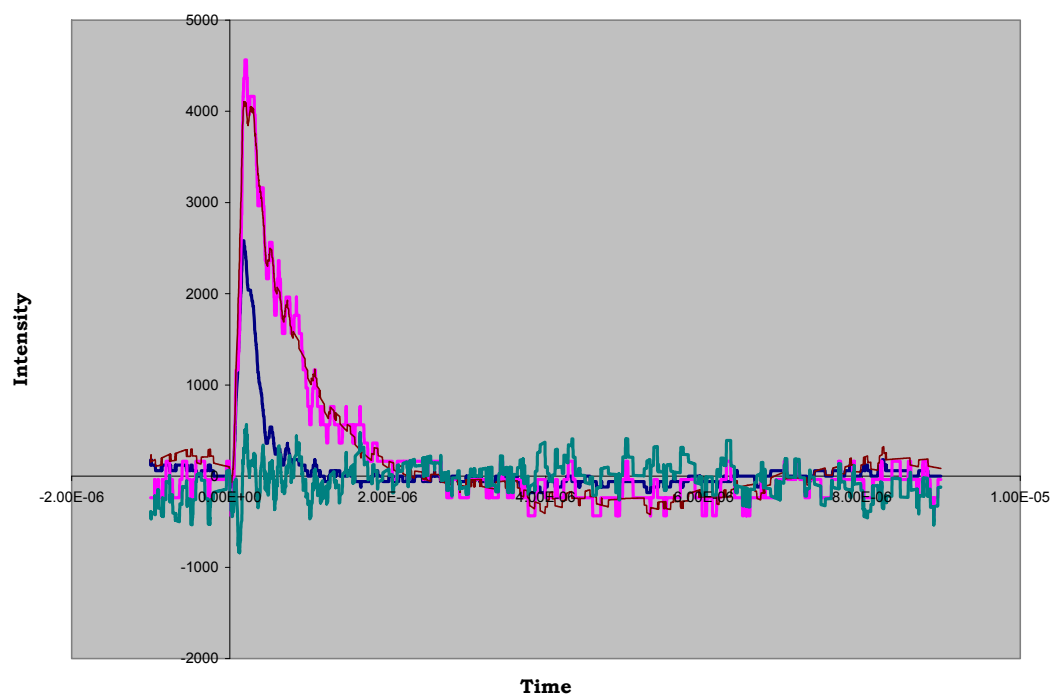
Yb Naphthyl d4-MeOH
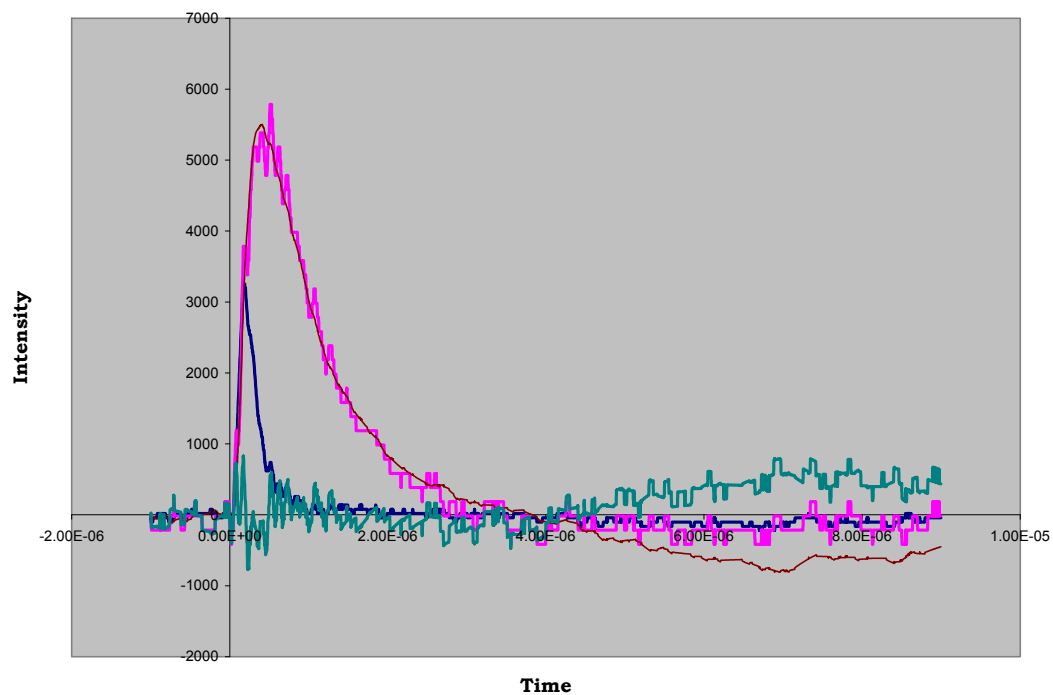
Compound 69: Tb Benzyl peptide
Excitation of metal
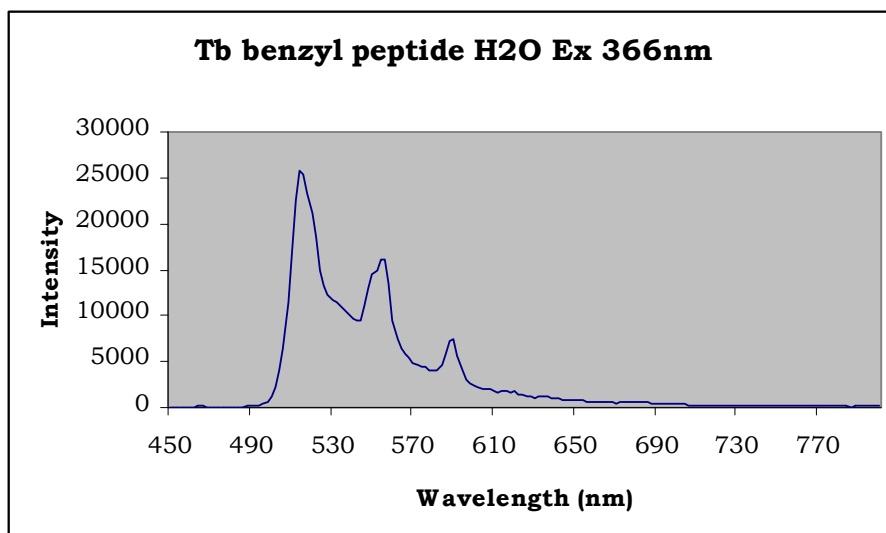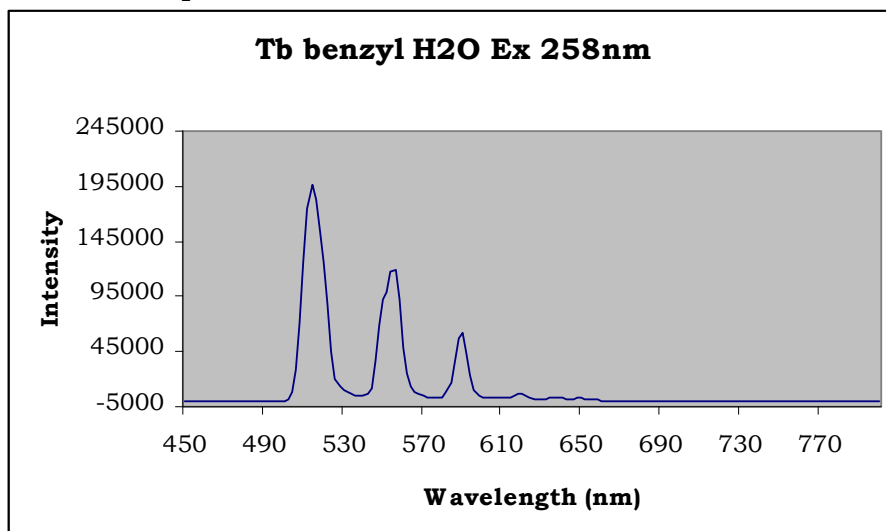
Tb benzyl peptide H2O Ex 366nm

Excitation of chromophore

Tb benzyl H2O Ex 258nm

Combination: Excitation of metal centre vs energy-transfer from chromophore

Tb Benzyl peptide H2O comparison
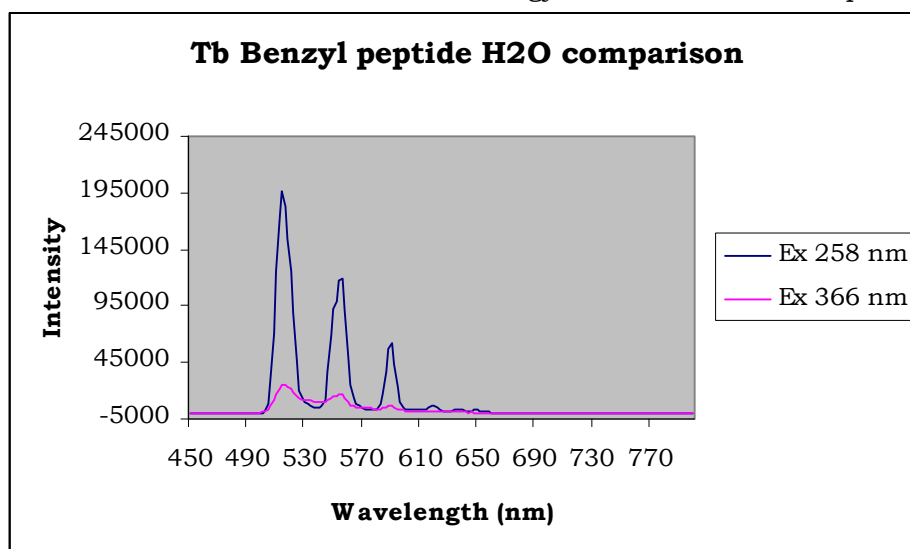
Compound **34:** Tb Indolyl caproic linker
Excitation of metal
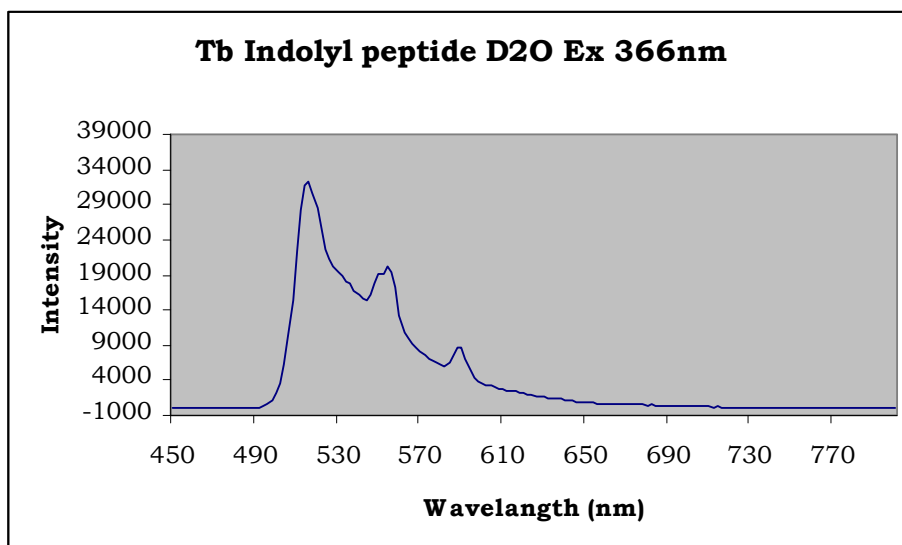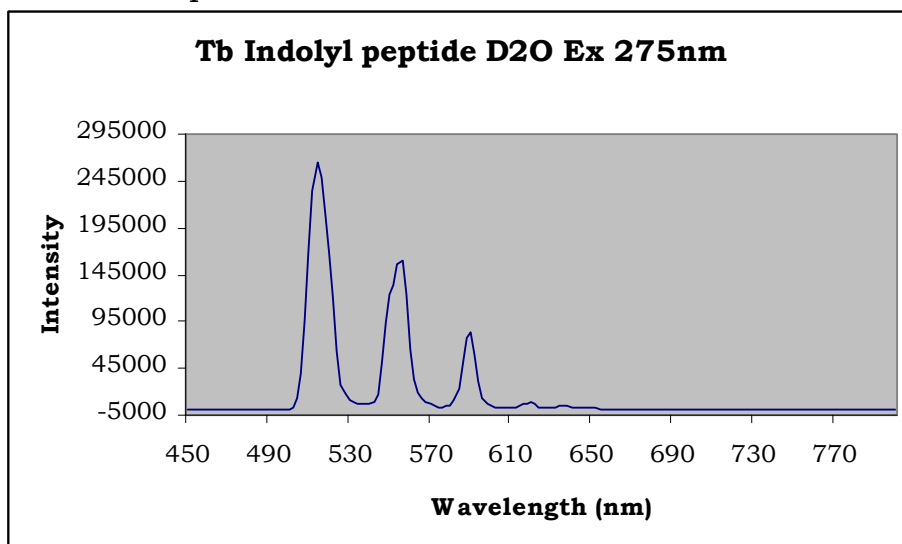
Tb Indolyl peptide D2O Ex 366nm

Excitation of chromophore

Tb Indolyl peptide D2O Ex 275nm

Combination: Excitation of metal centre vs energy-transfer from chromophore

Tb Indolyl peptide comparison
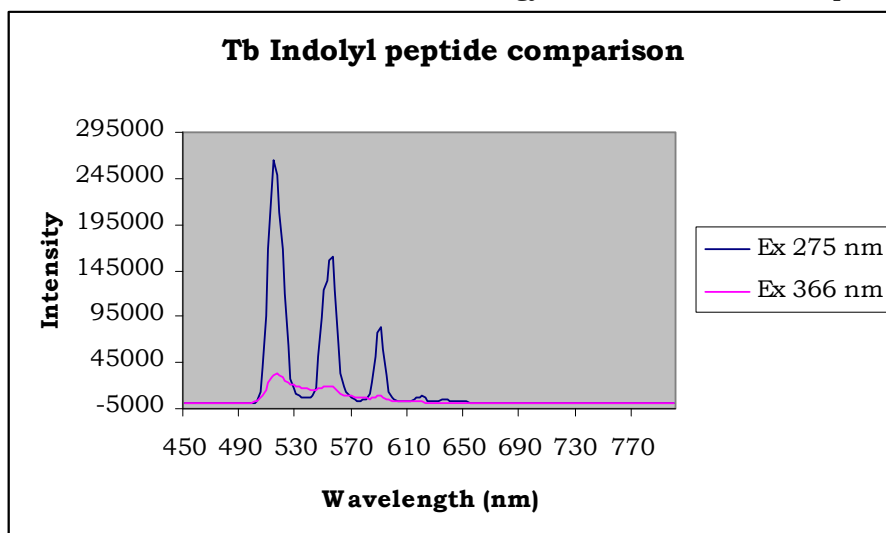
5.4 Appendix 4

Final product absorption & emission spectra

Compound **61**, cPep-NIR797

UV absorption Spectrum

cPepNIR797 UV Absorption Spectrum (0.1 mg/mL)

Emission spectrum
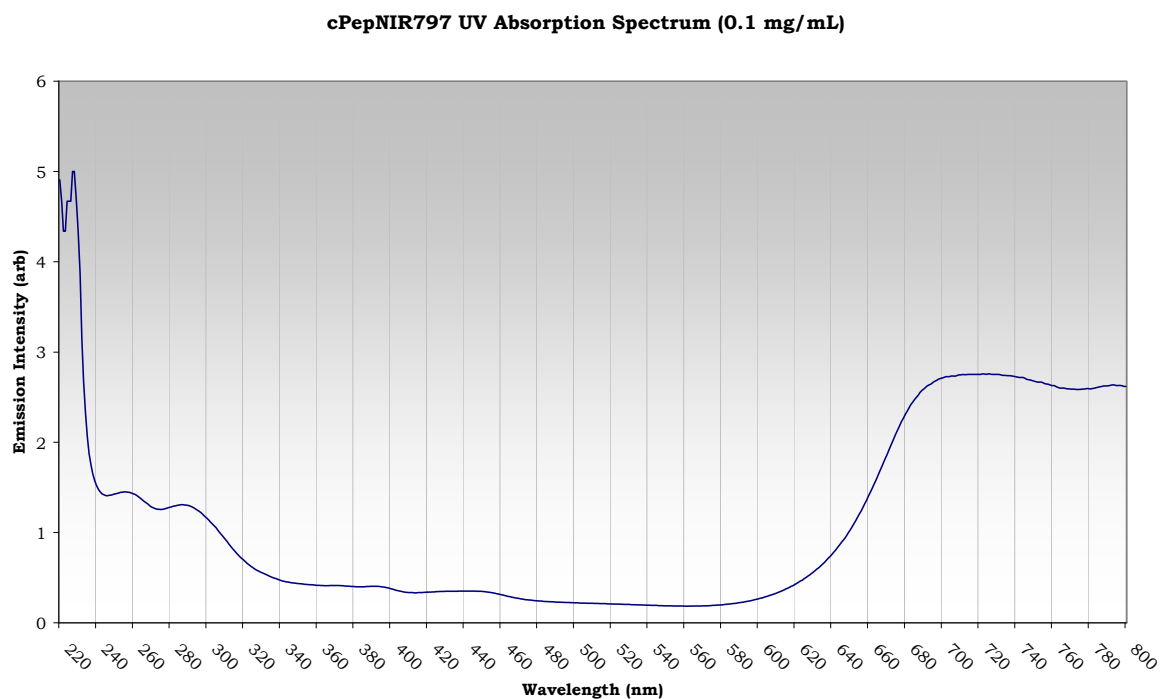
cPep-NIR797 emission spectrum (0.1 mg/mL). Ex 710nm
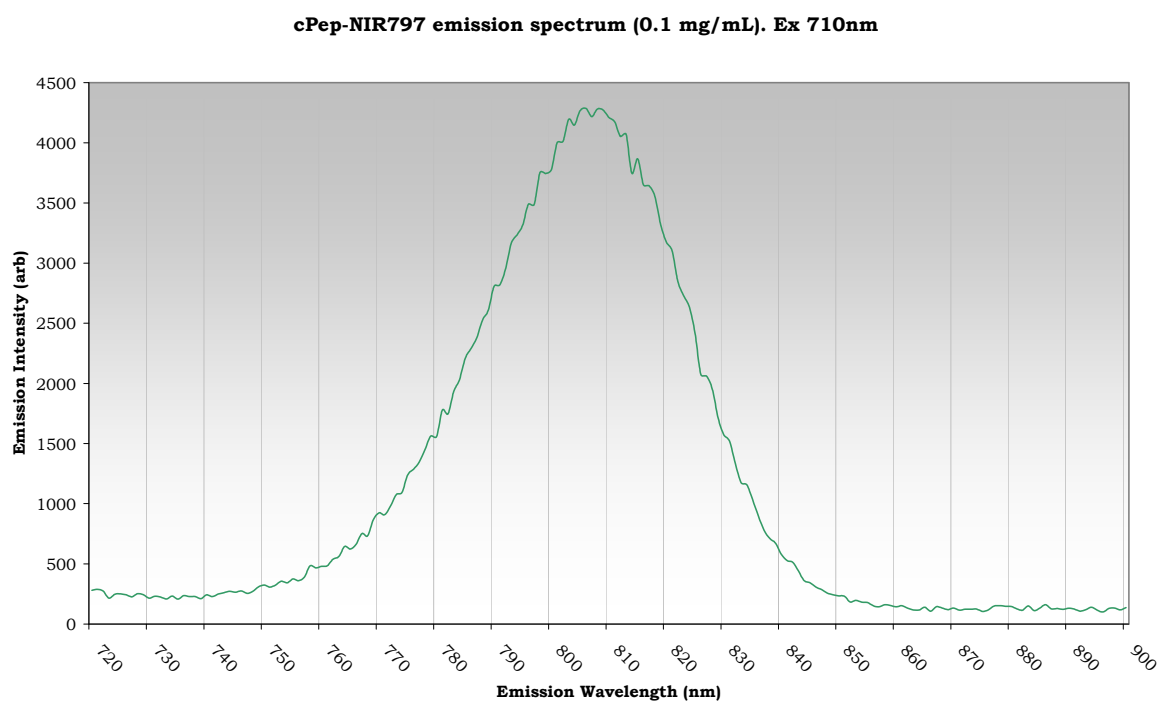
Emission at different excitation wavelengths

cPepNIR797 Emission Spectra (0.1 mg/mL)
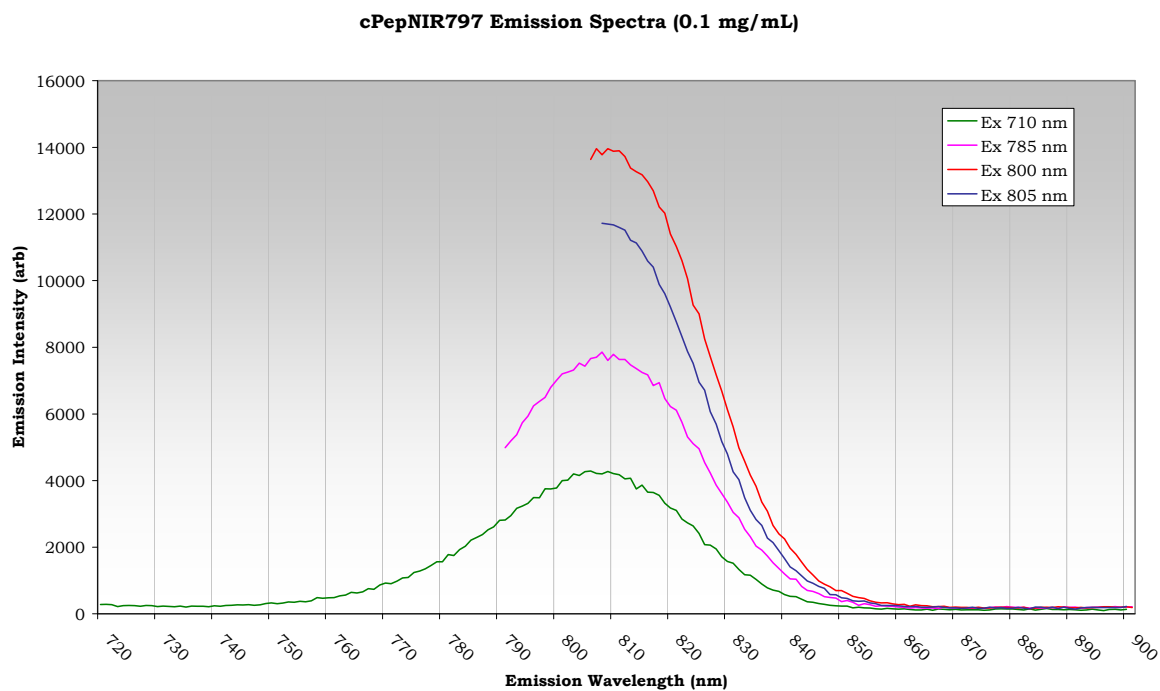
Excitation Spectrum (with filter)

cPepNIR797 Excitation Spectrum enlargement (Em 810 nm)
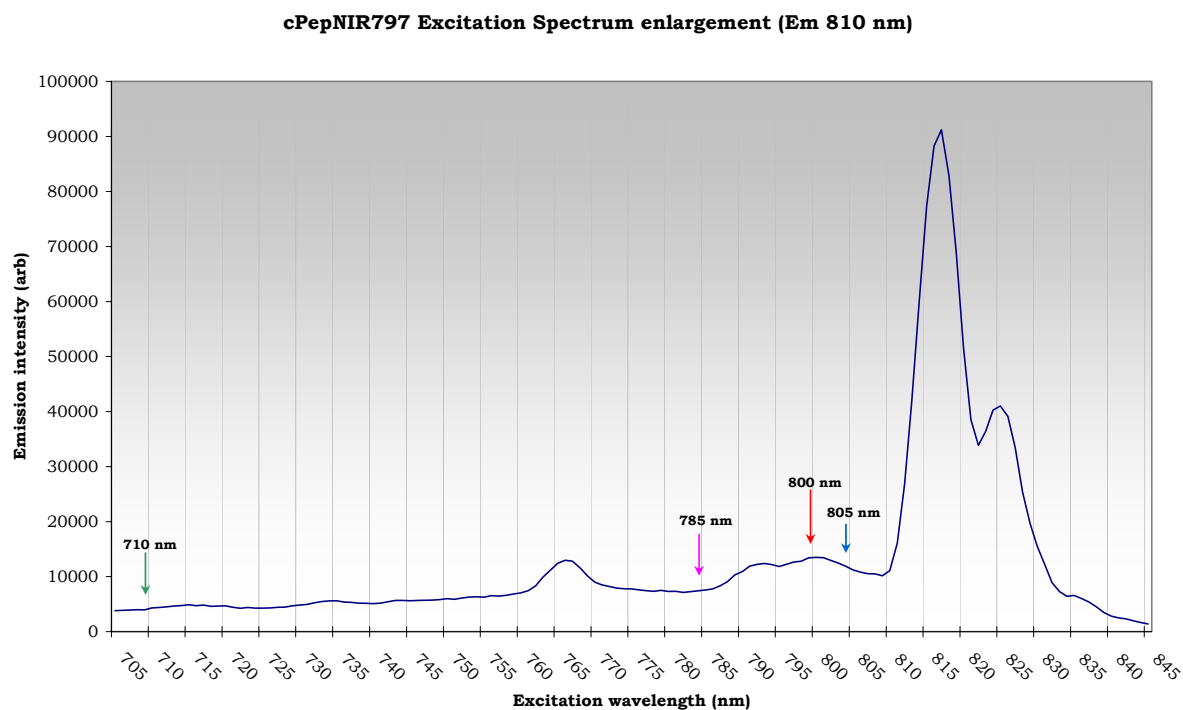
Excitation spectrum (without filter)

cPepNIR797 excitation spectrum (0.1 mg/mL). Em 810nm
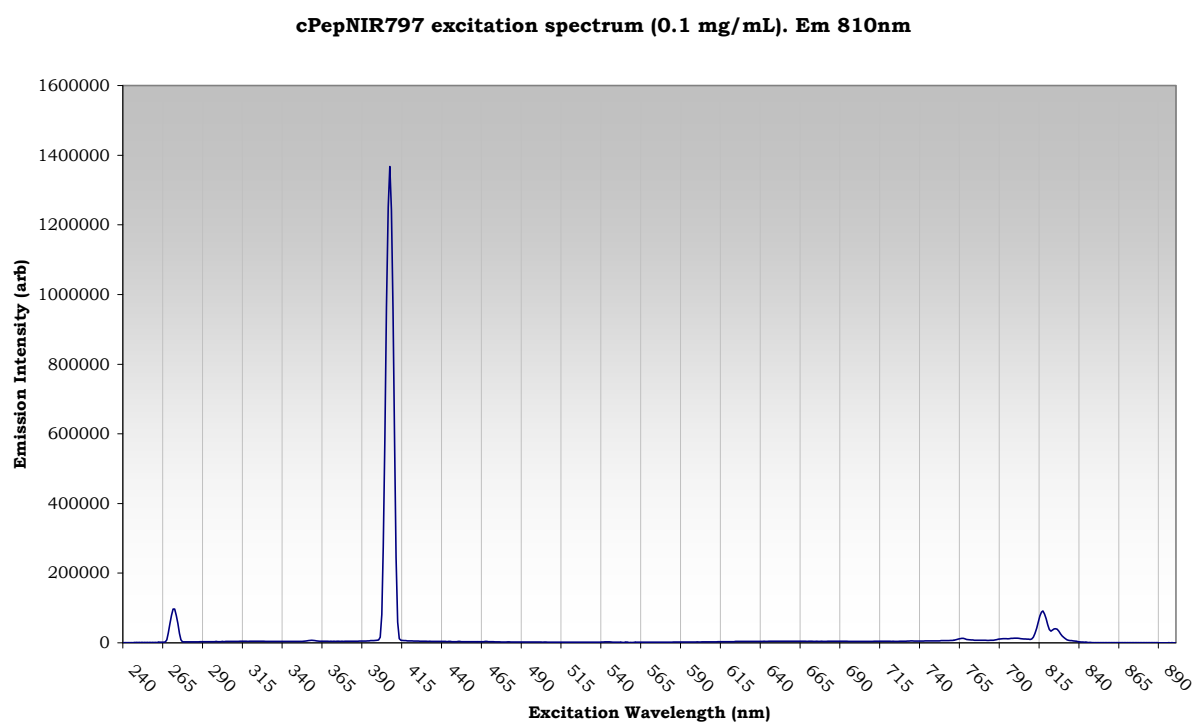
Full excitation spectrum with 715nm filter

cPepNIR797 Excitation Spectrum (0.1 mg/mL). Em 810nm
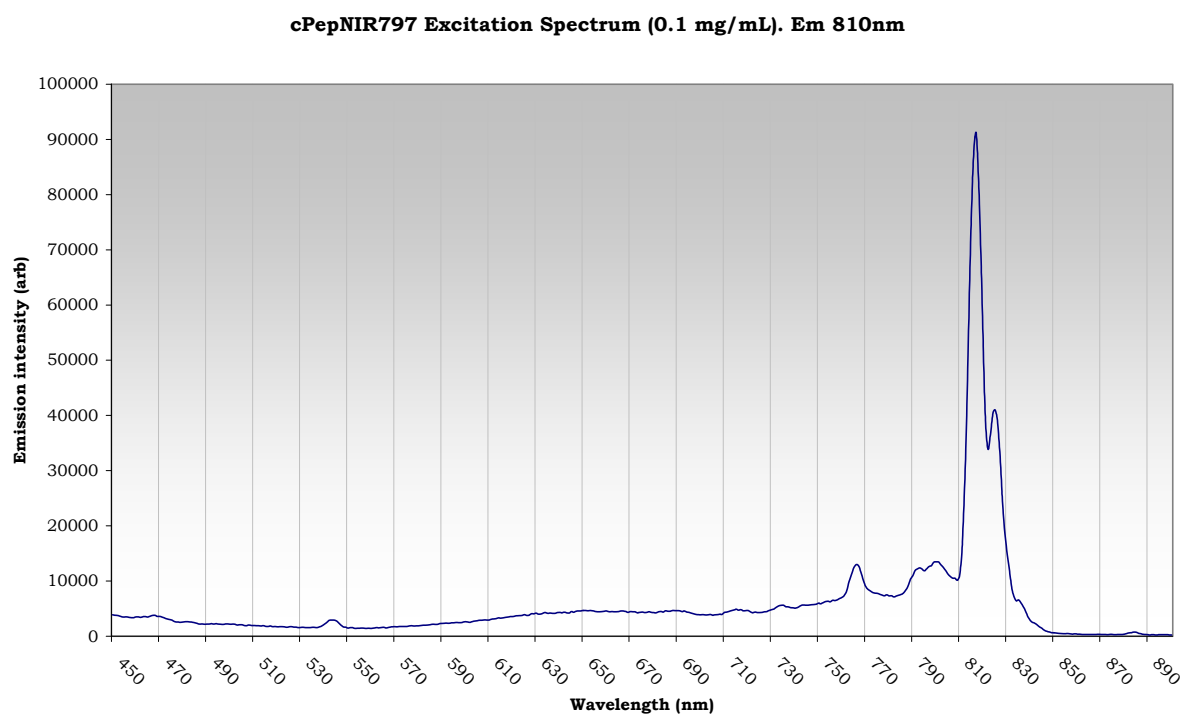
Compound **68**, cPep-cap-Os

UV absorption spectrum

cPep-Cap-Os UV absorption spectrum
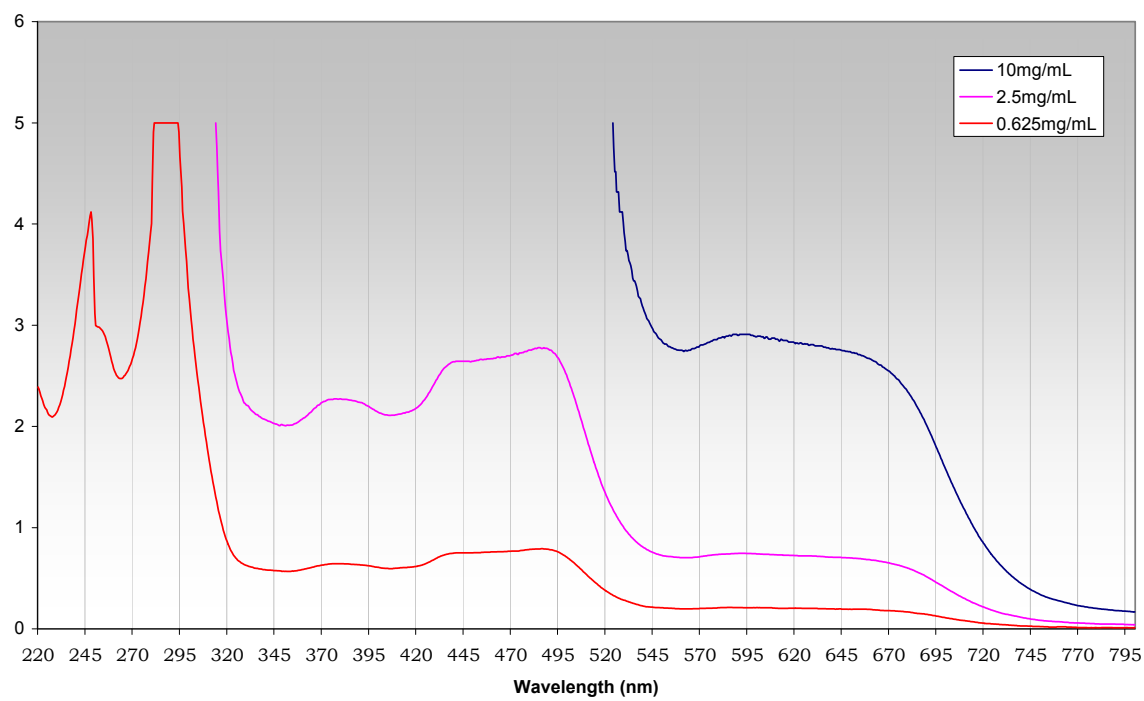
Fluorescence emission spectrum

cPep-Cap-Os emission spectrum
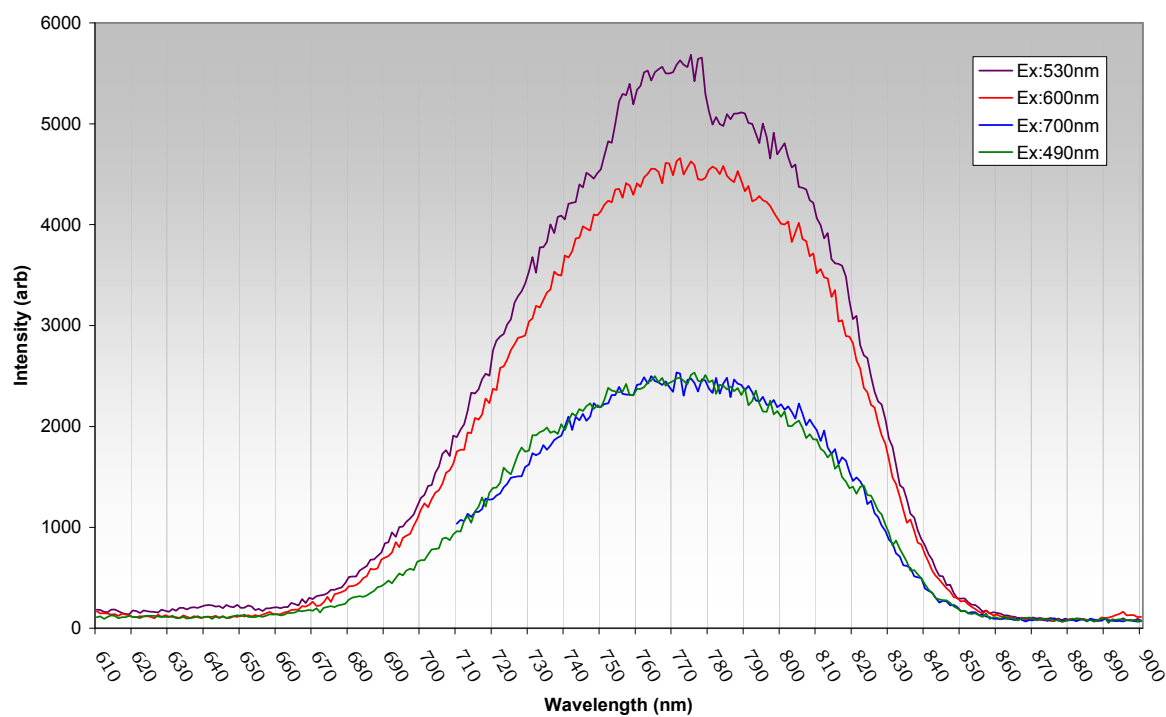
Excitation spectrum (scanning emission at specifically 770nm (maximum above) across a range of excitation wavelengths).

cPep-Cap-Os excitation spectrum Em:770nm
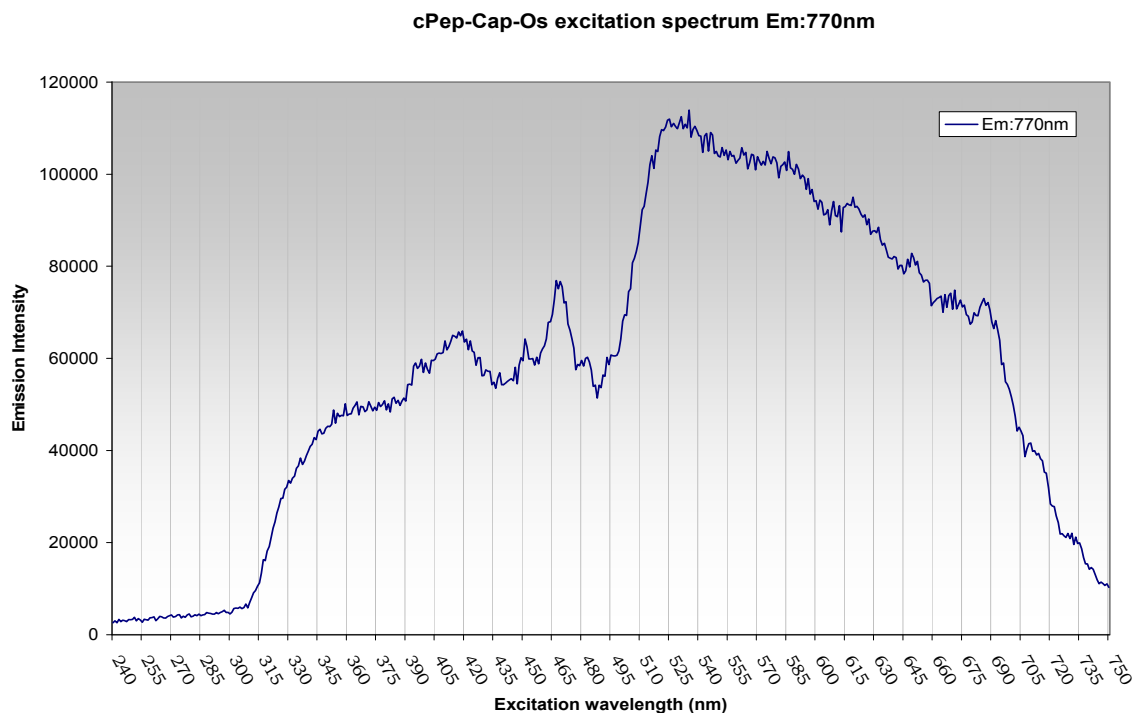
Excitation spectrum (as above) (with and without filter)

cPep-Cap-Os excitation spectrum Em:770nm
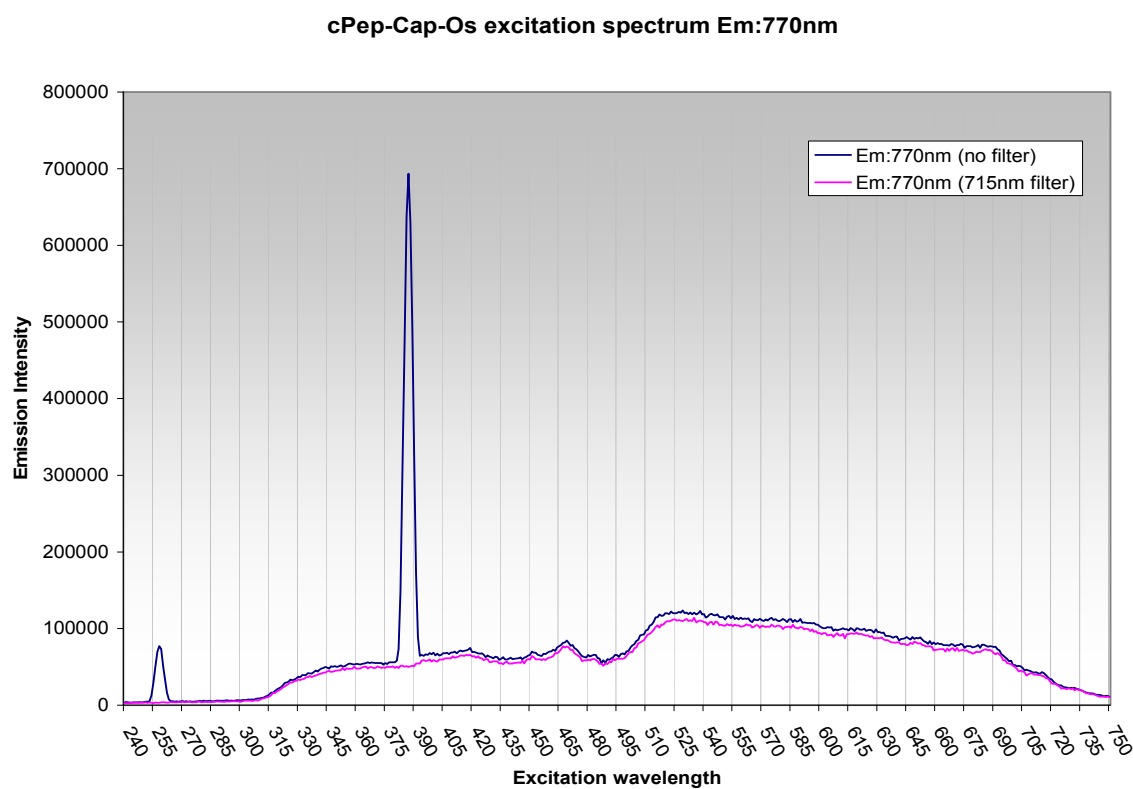
Compound **60**, cPep-Fluorescein

Absorption spectrum

UV absorption spectrum Fluorescein

Emission spectrum
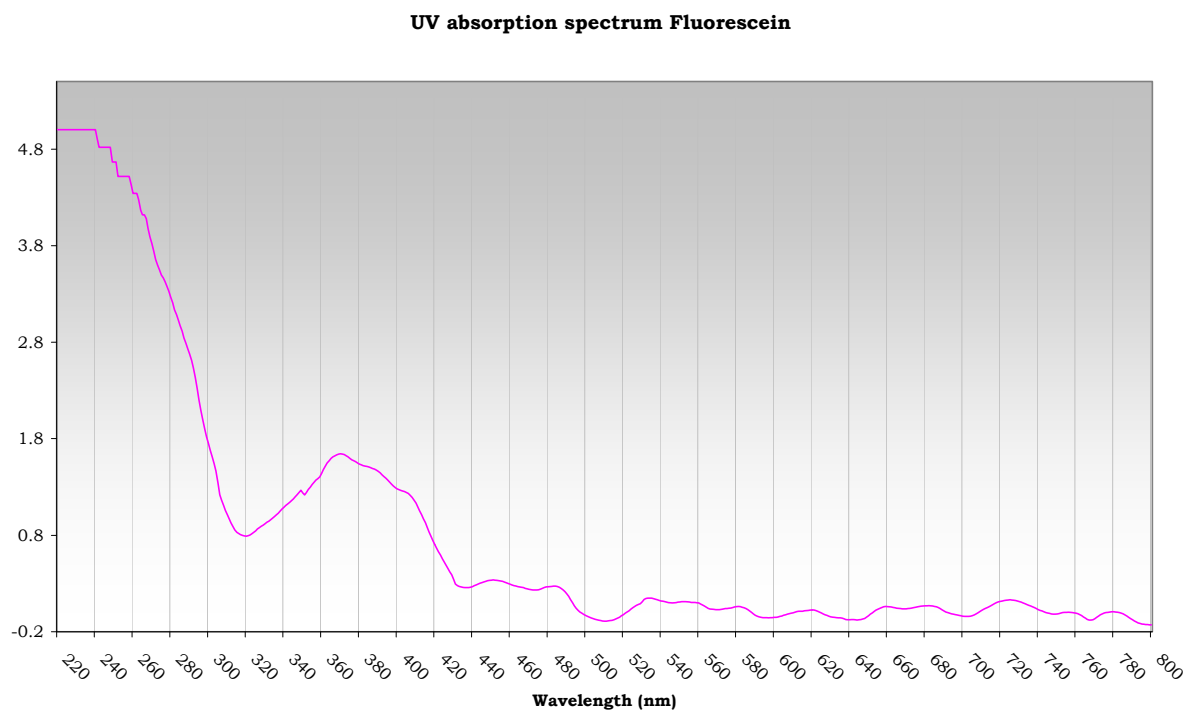
Appended Fluorescein Emission Spectrum (1.0 mg/mL, Ex 400 nm)

Compound **88**, Disaccharide-NIR797

UV absorption spectrum

Disaccharide-NIR797 (H2O)

Emission spectrum

Disaccharide-NIR797 H2O Emission spectra
0.0001mg/mL
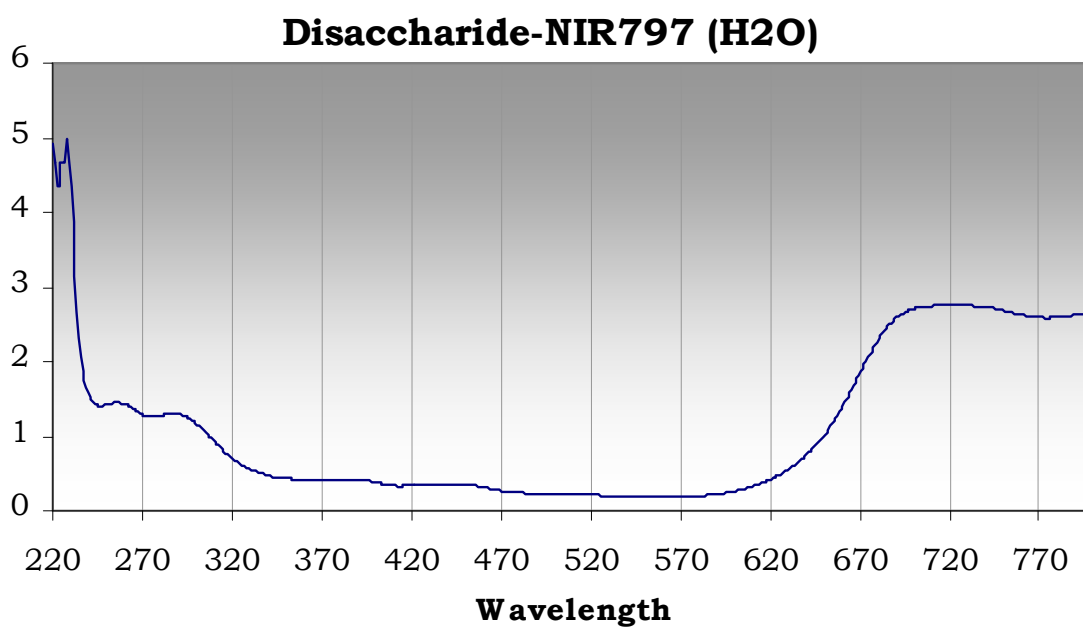
Excitation spectrum

Excitation spectrum Em 815nm

Compound **89**, Disaccharide-Osmium

Osmium complex absorption spectrum

Appended Osmium complex absorption spectrum

Osmium complex emission spectrum
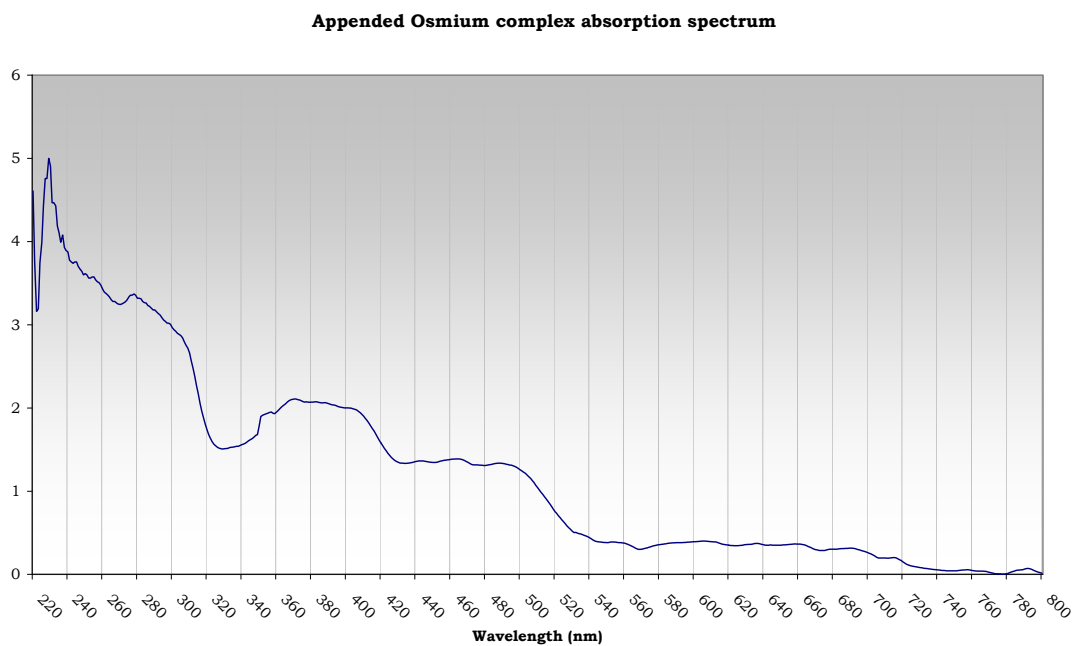
Appended Osmium Complex Emission Spectrum (1.0 mg/mL, Ex 600nm)
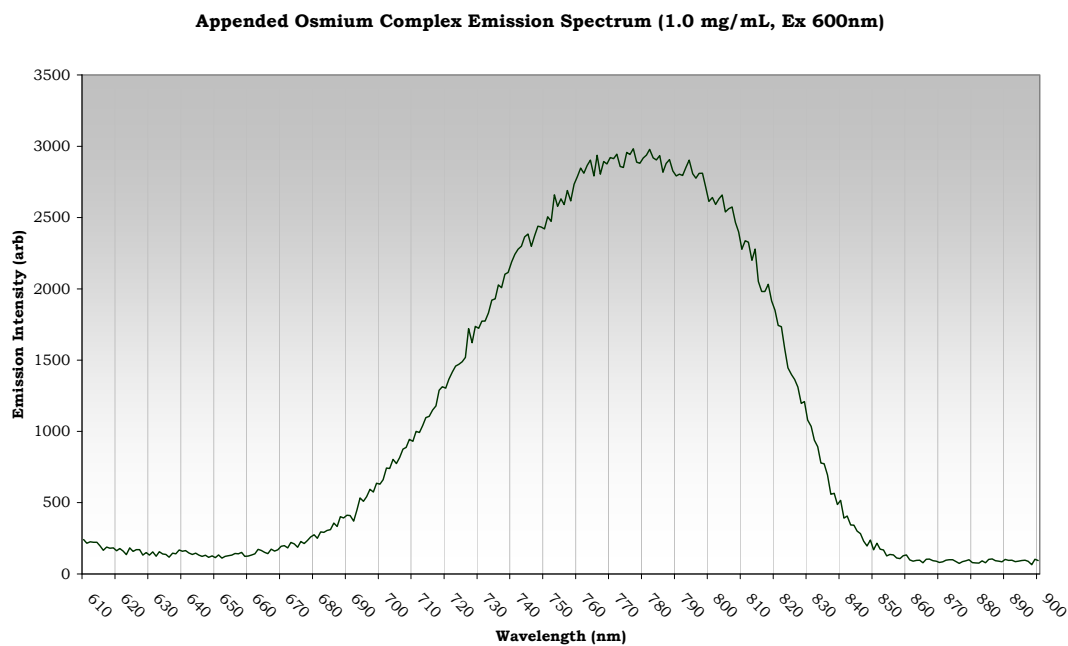
Compound **87**, Disaccharide-Fluorescein

UV absorption spectrum

Disaccharide-Fluorescein (H2O)
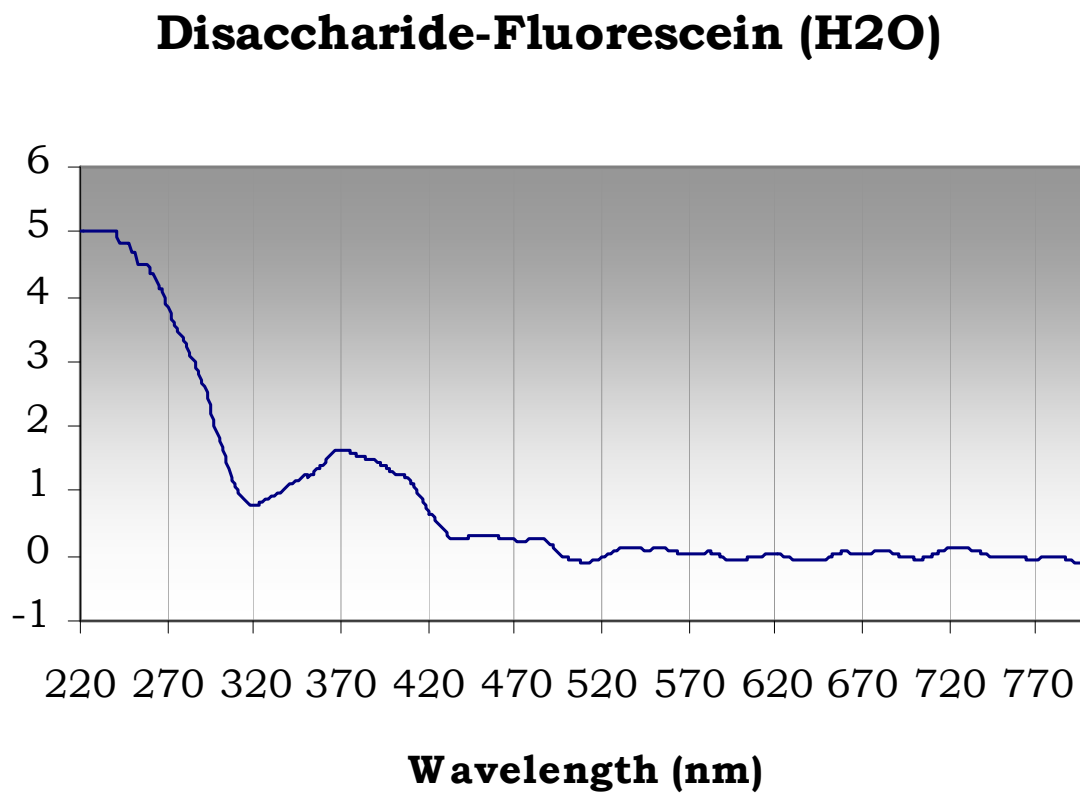
Emission spectra

Disaccharide-Fluorescein (H2O) Ex 400nm
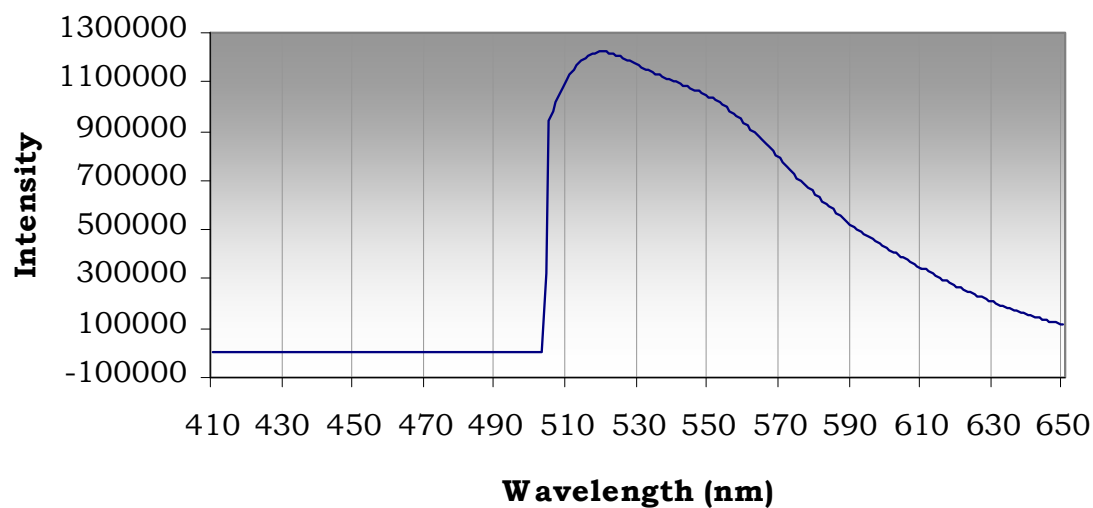
Disaccharide-Fluorescein (H2O) Ex 280nm
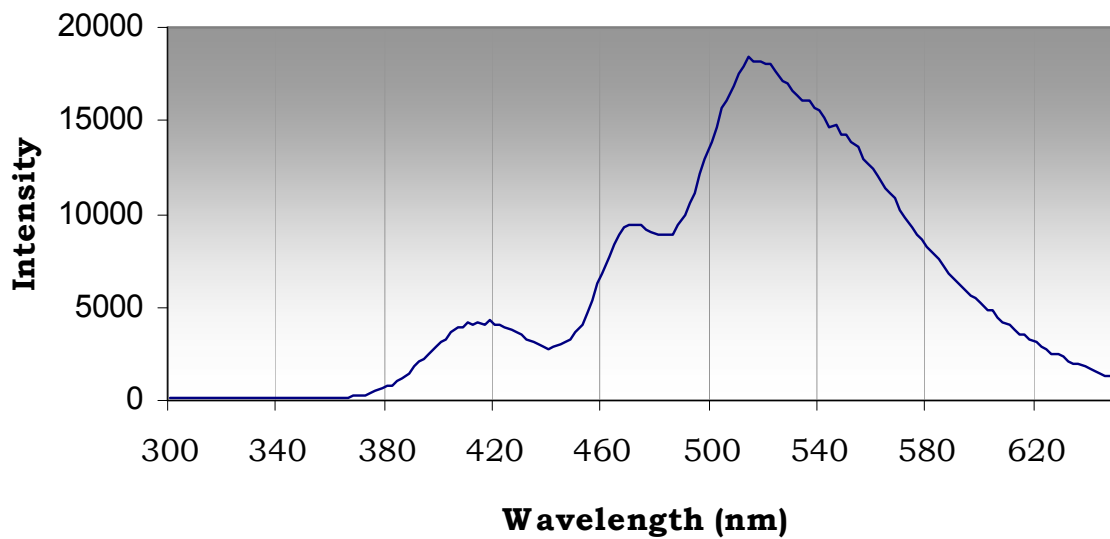
Excitation spectrum. Emission set at 520nm.

Disaccharide-Fluorescein (H2O) Em 520nm
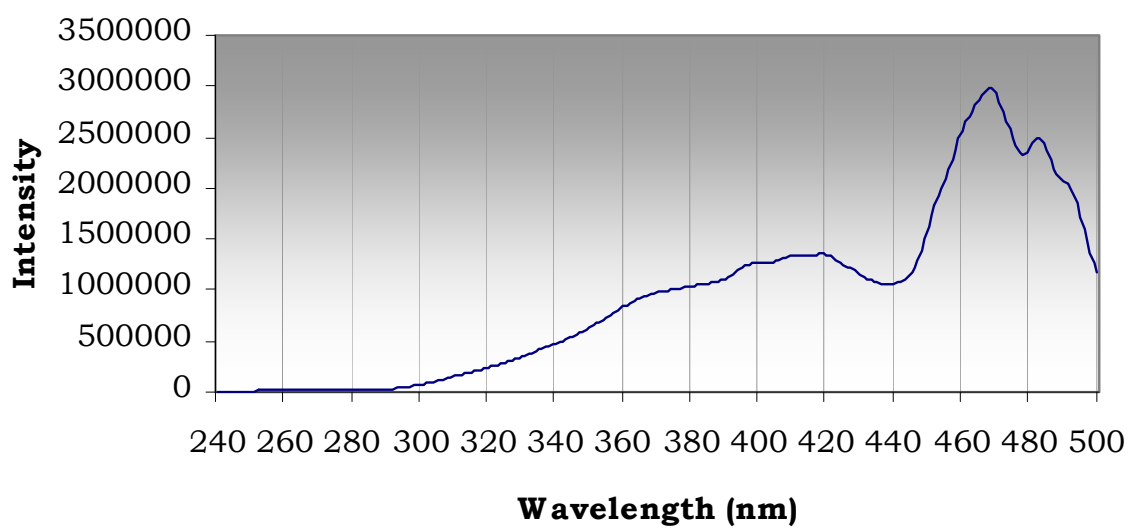
Fluorescein vs Osmium vs NIR797 emission spectra
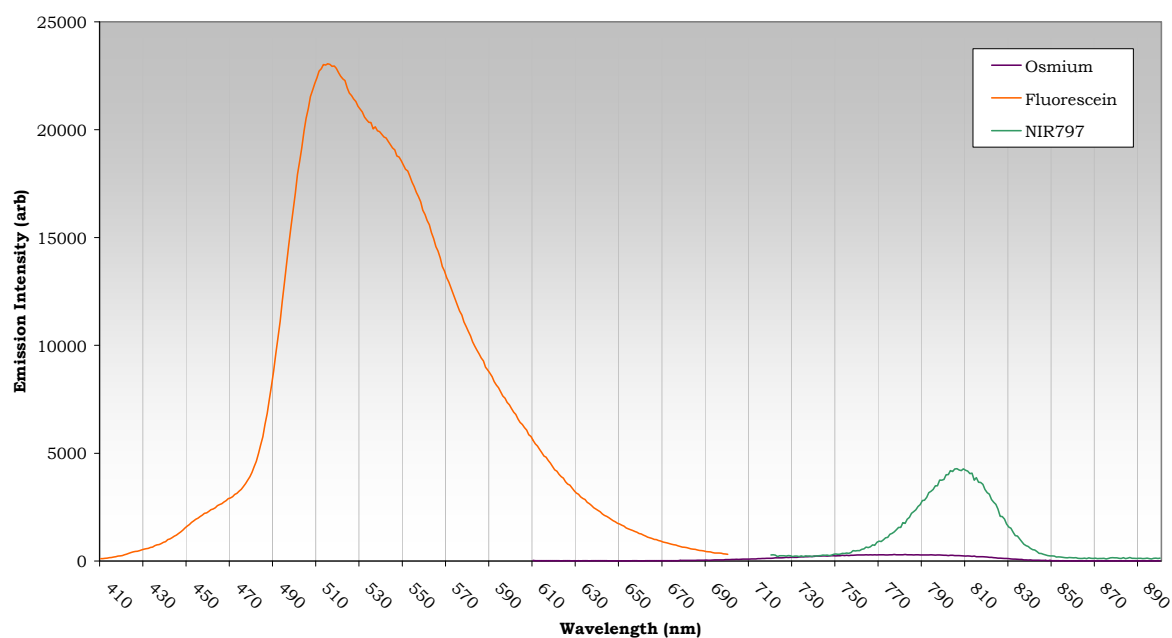
5.5 Appendix 5: HPLC & m/z data

Ugi-lanthanide complexes

Compound **12**: Eu Benzyl
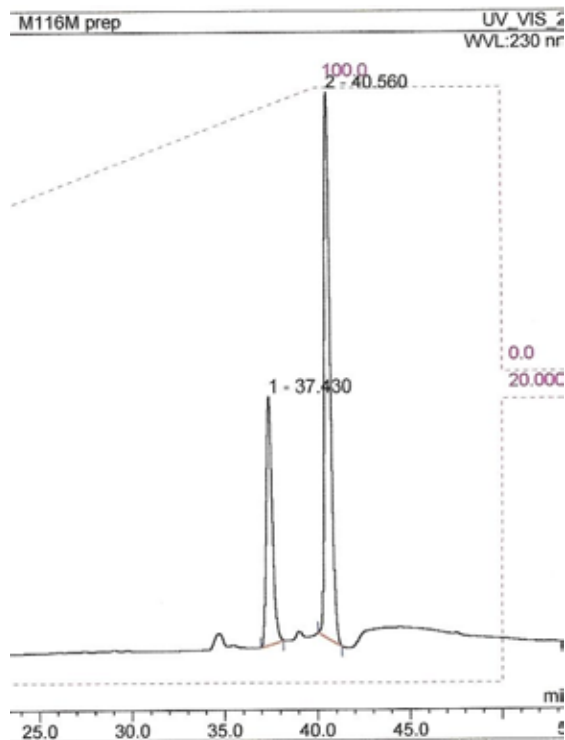
HPLC   m/z
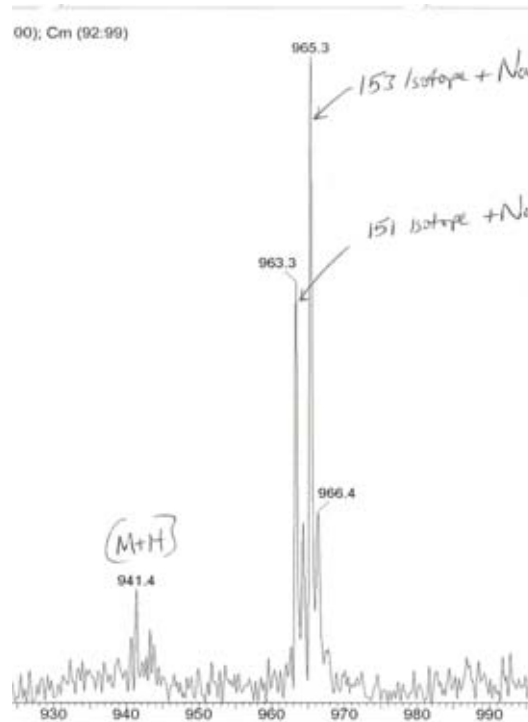
Compound **13**: Eu Naphthyl
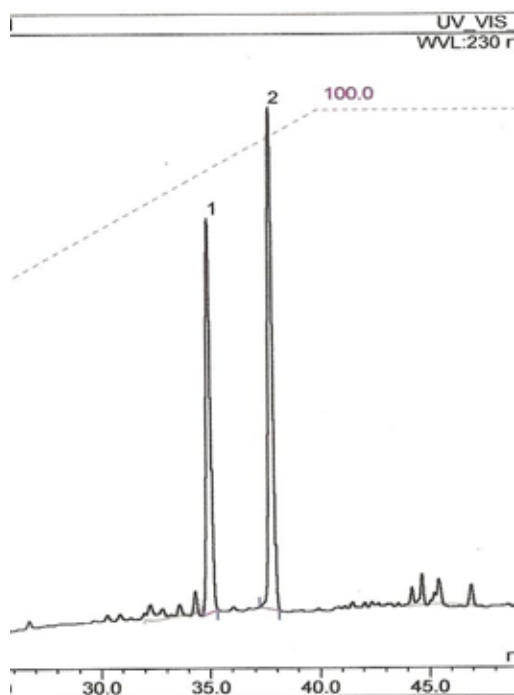
HPLC   m/z
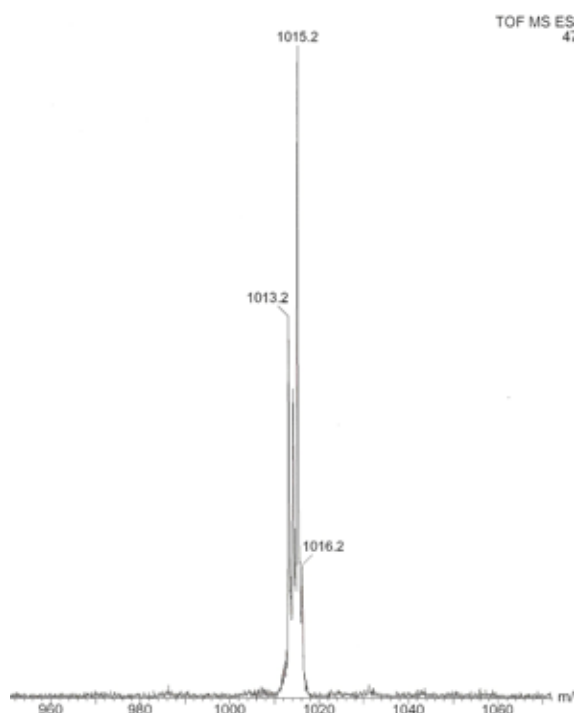
Compound **14**: Eu Pyrenyl
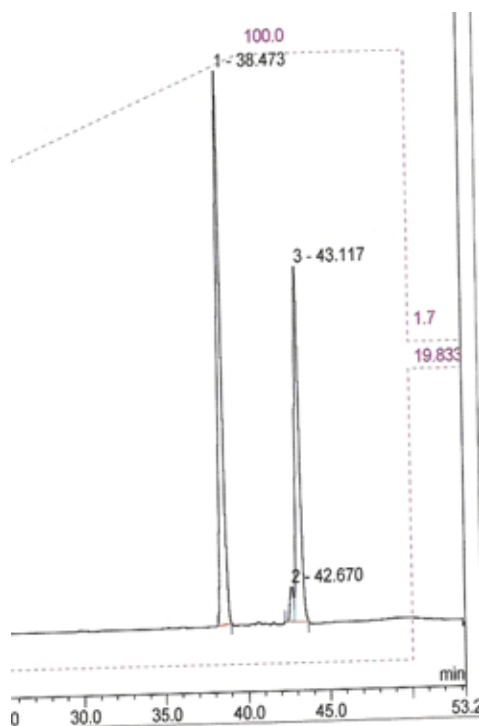
HPLC   m/z
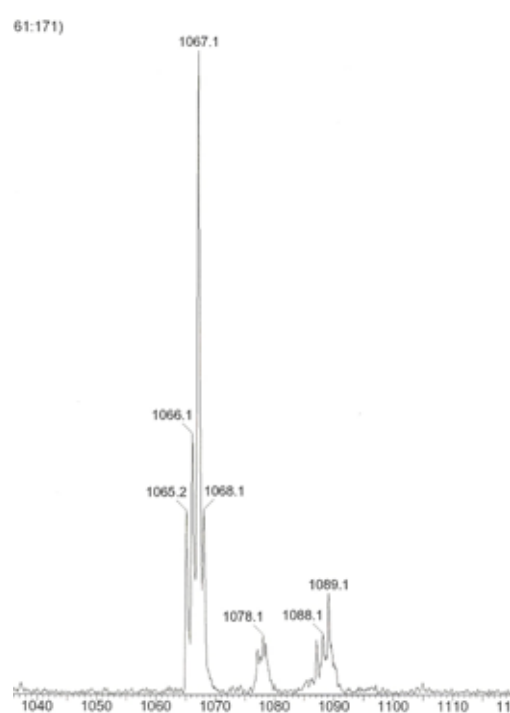
Compound **15**: Eu Indolyl
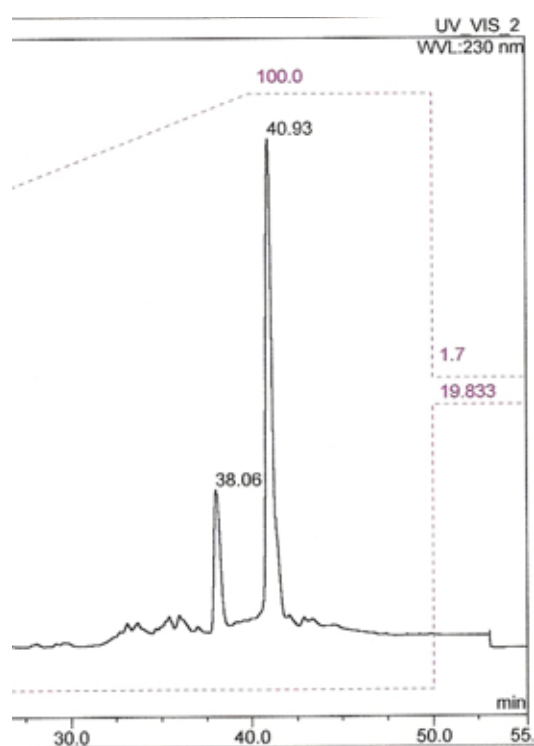
HPLC   m/z
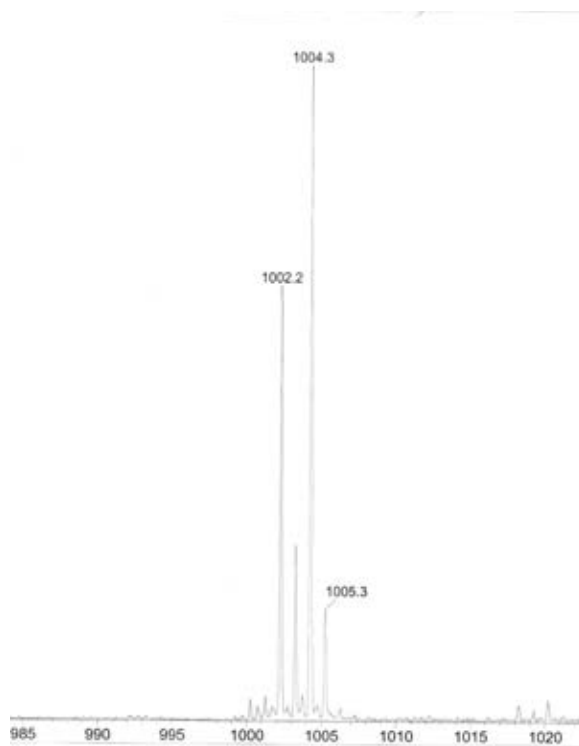
Compound **16**: Tb Benzyl

HPLC   m/z
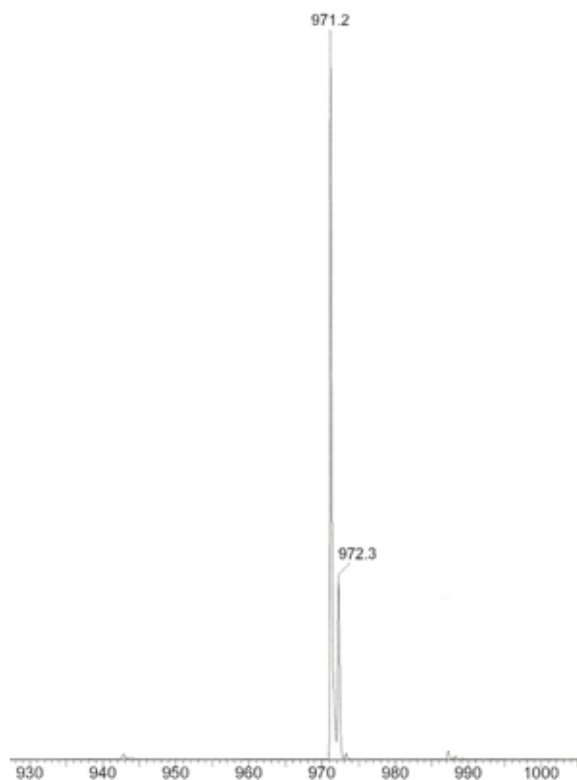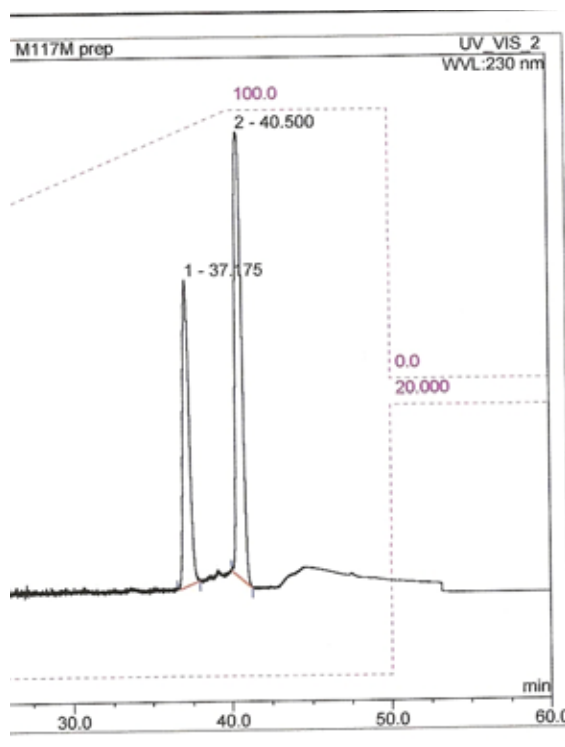
Compound **17**: Tb Naphthyl

HPLC   m/z
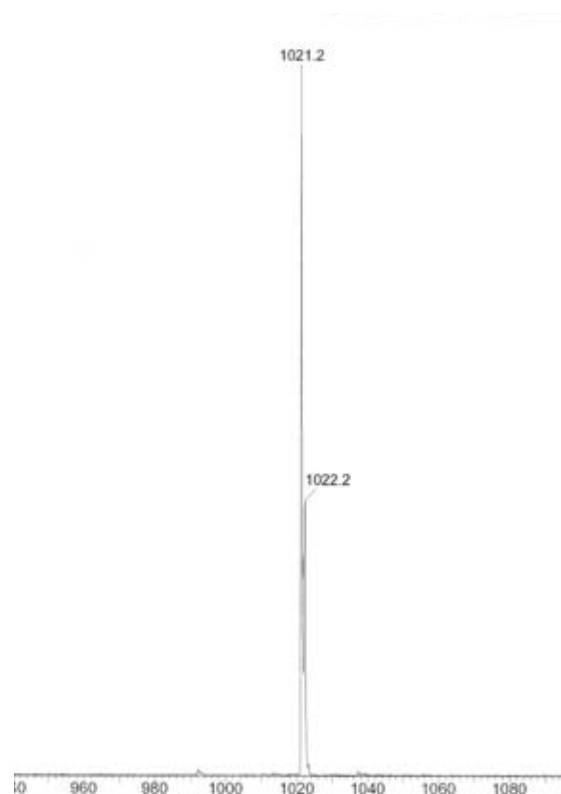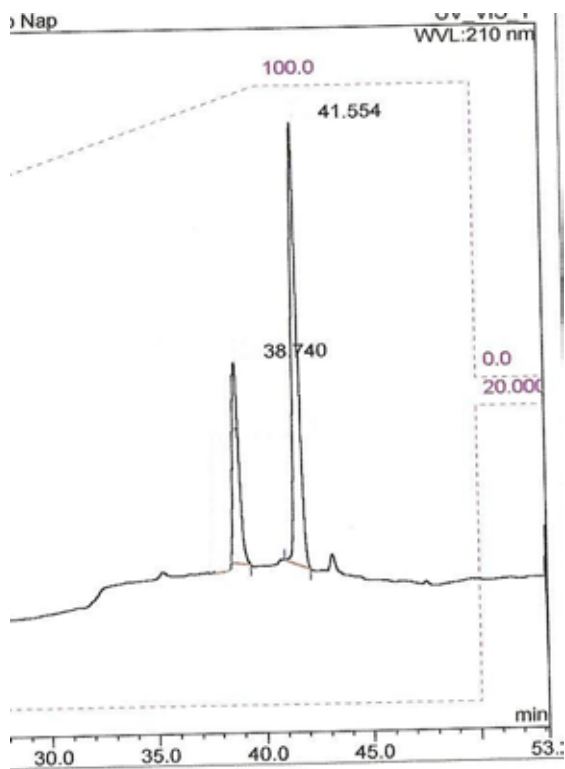
Compound **18**: Tb Pyrenyl

HPLC   m/z
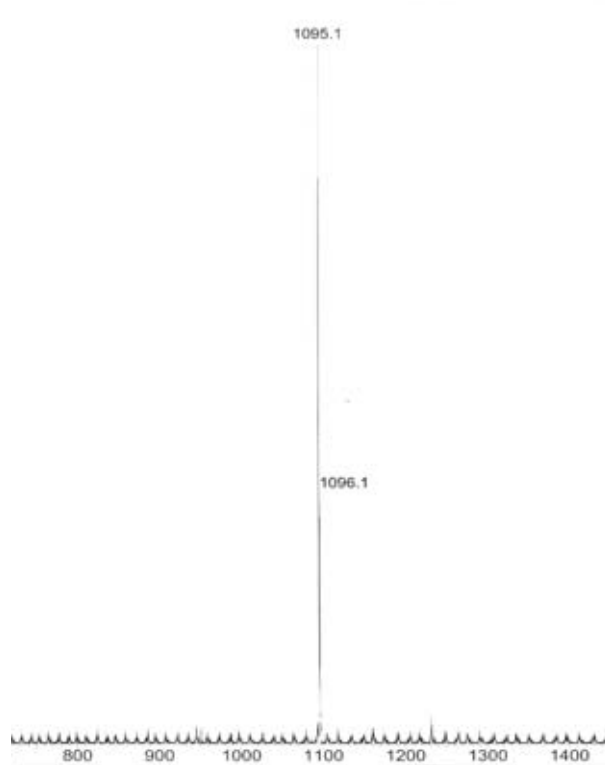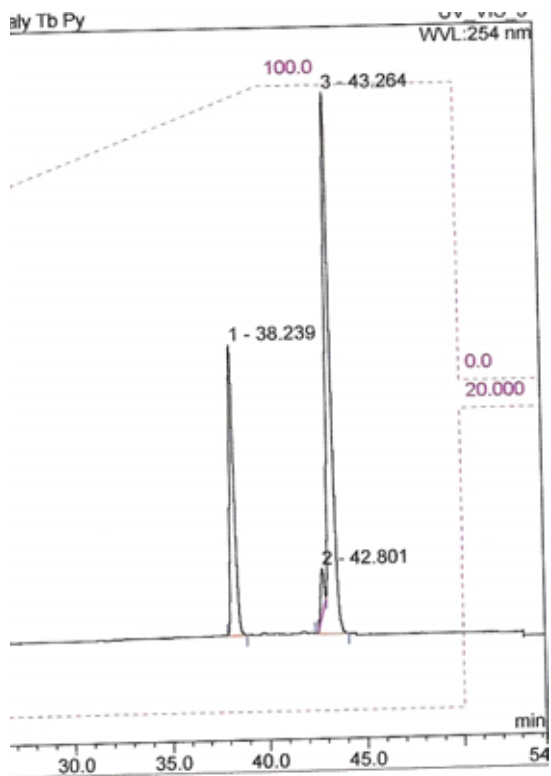
Compound **19**: Tb Indolyl

HPLC   m/z

305
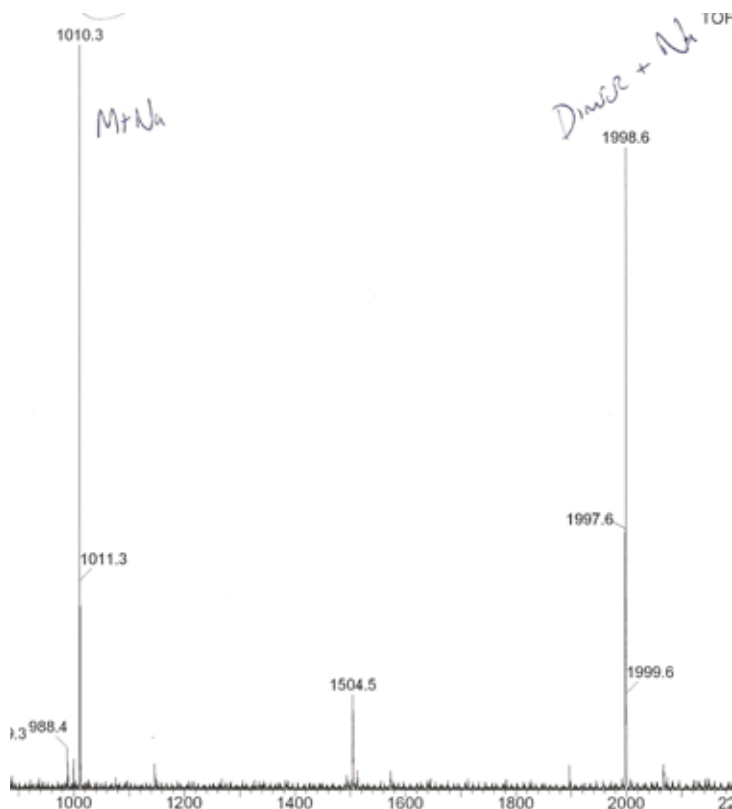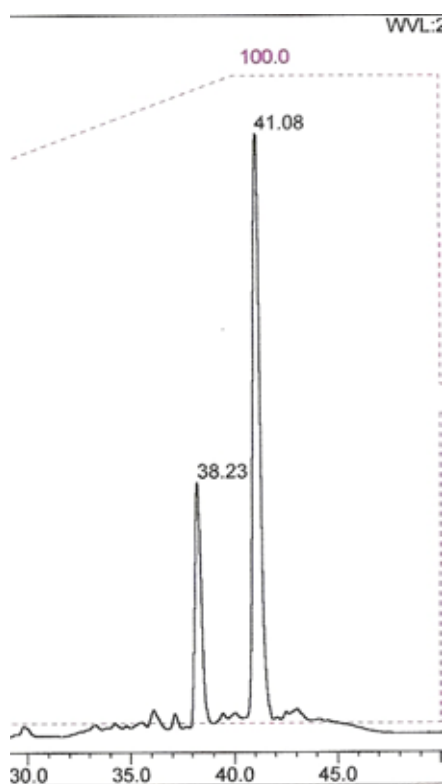
Compound **20**: Yb Benzyl

HPLC   m/z
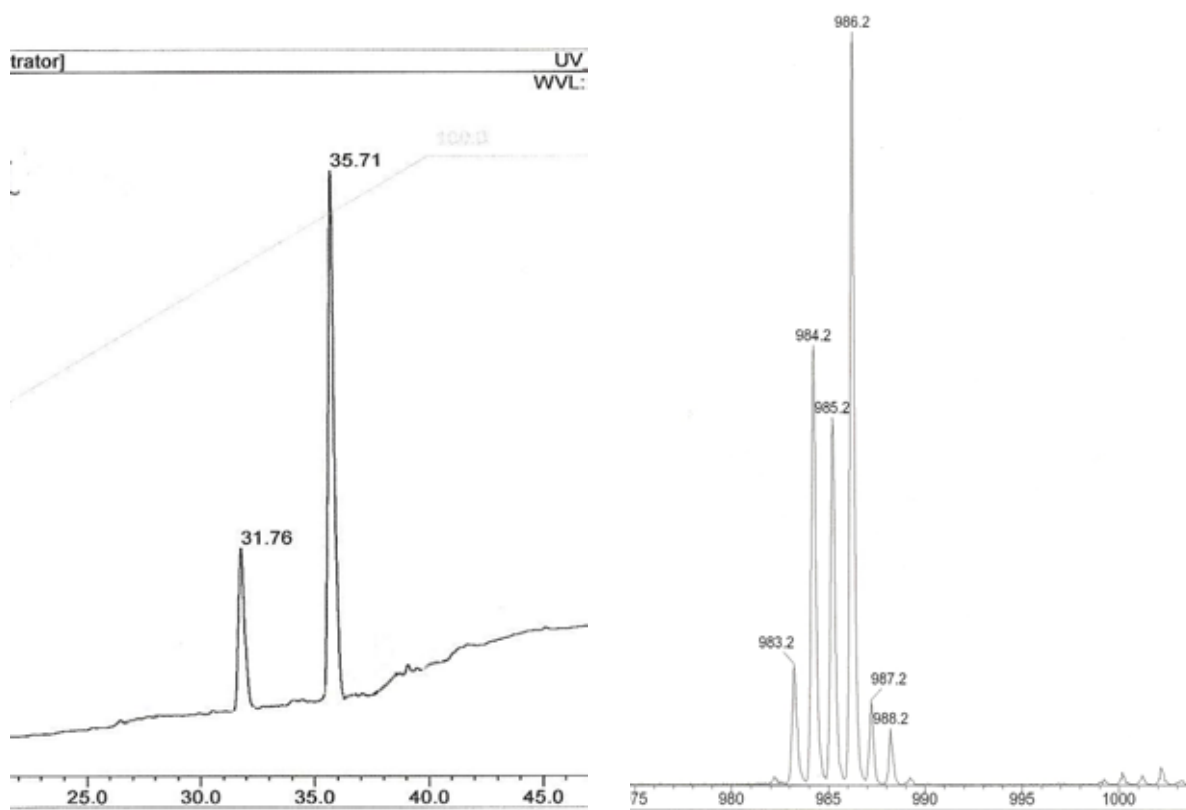
Compound **21**: Yb Naphthyl

HPLC   m/z
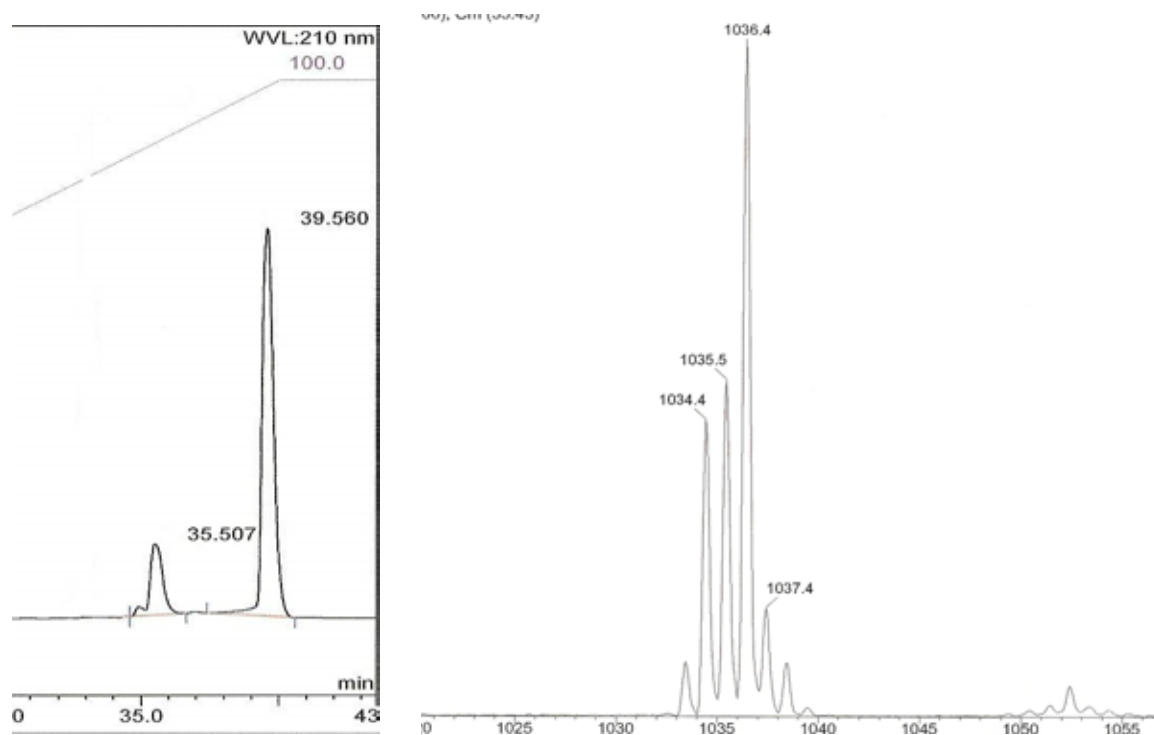
Compound **22**: Yb Pyrenyl
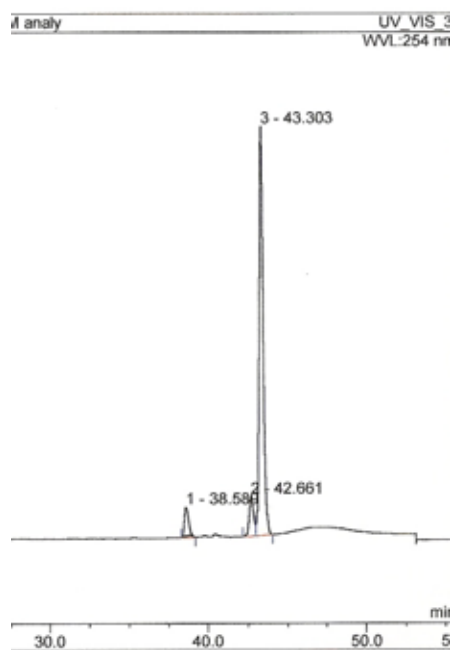
HPLC   m/z
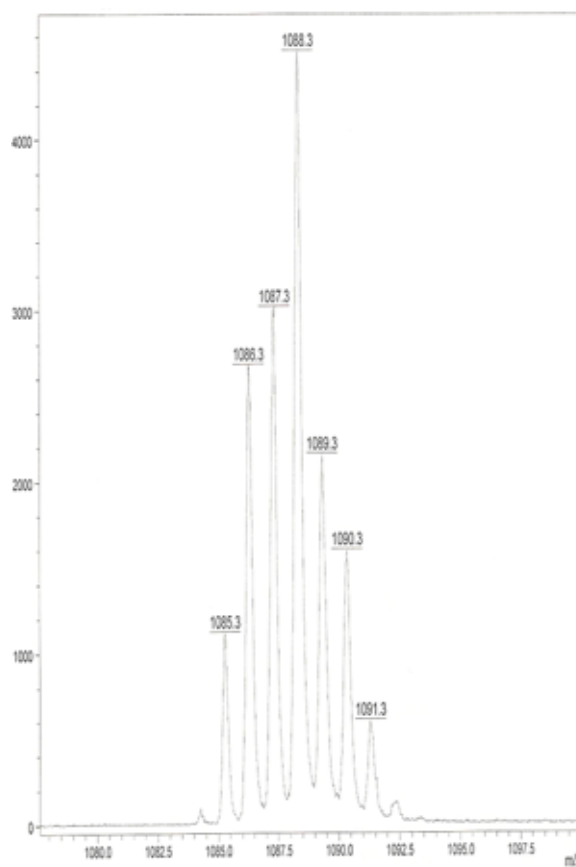
Compound **23**: Yb Indolyl
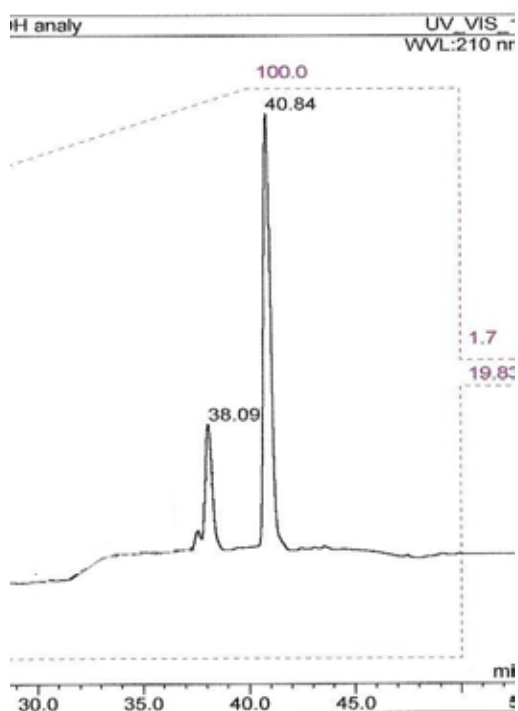
HPLC   m/z
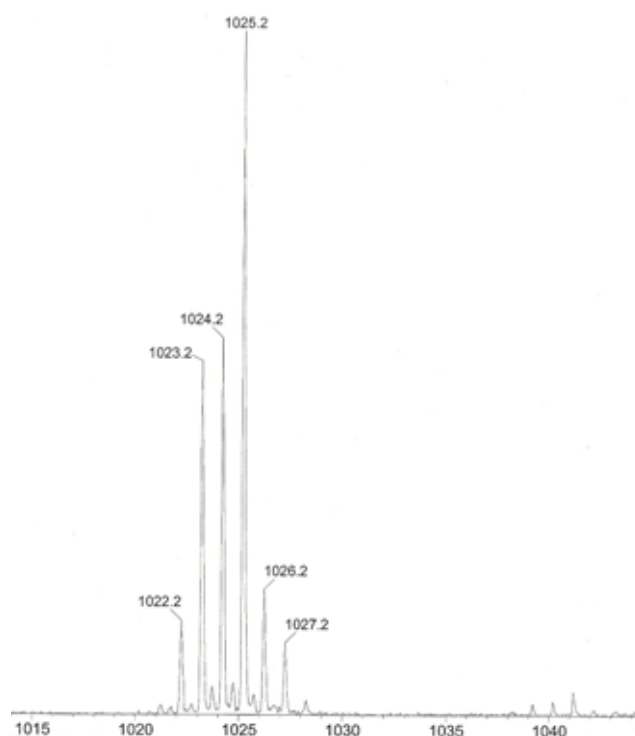
Peptides

Compound **54 Linear** Asp(tBu)-DPhe-Lys(Alloc)-Arg(pmc)-Gly

m/z
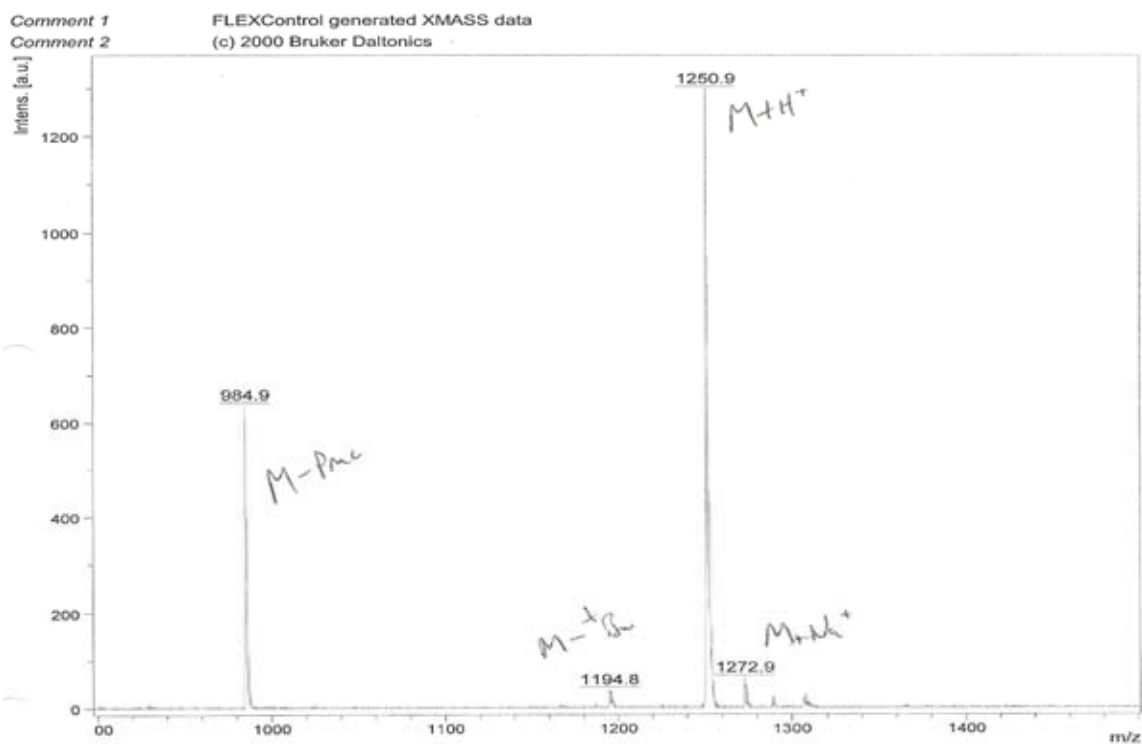
HPLC
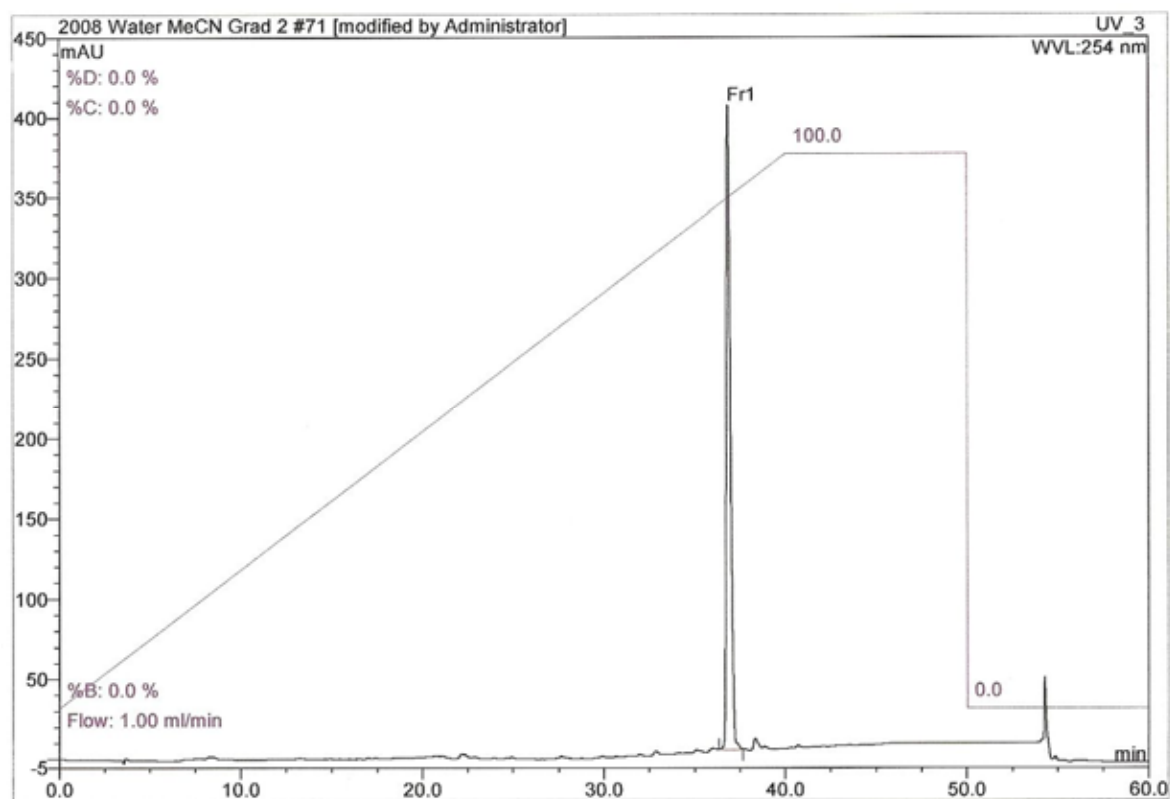
Cyc Asp(tBu)-DPhe-Lys(Alloc)-Arg(pmc)-Gly **56**

m/z

Comment 1   FLEXControl generated XMASS data
Comment 2   (c) 2000 Bruker Daltonics
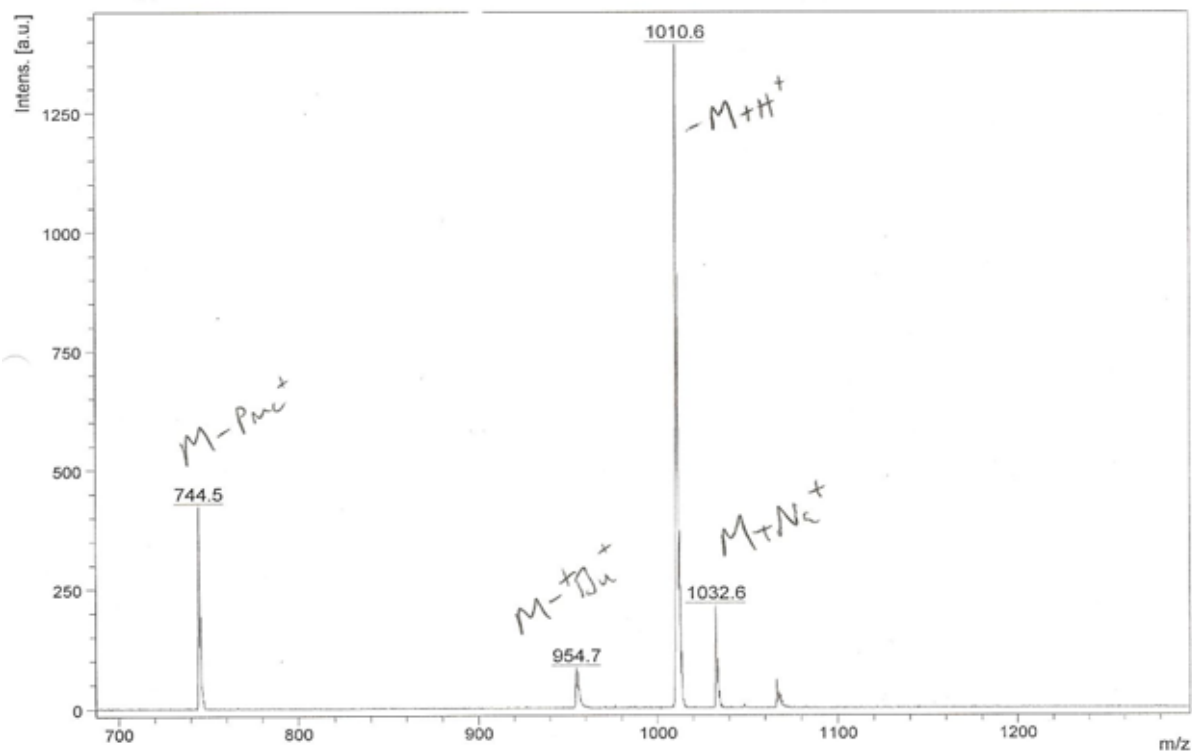
HPLC
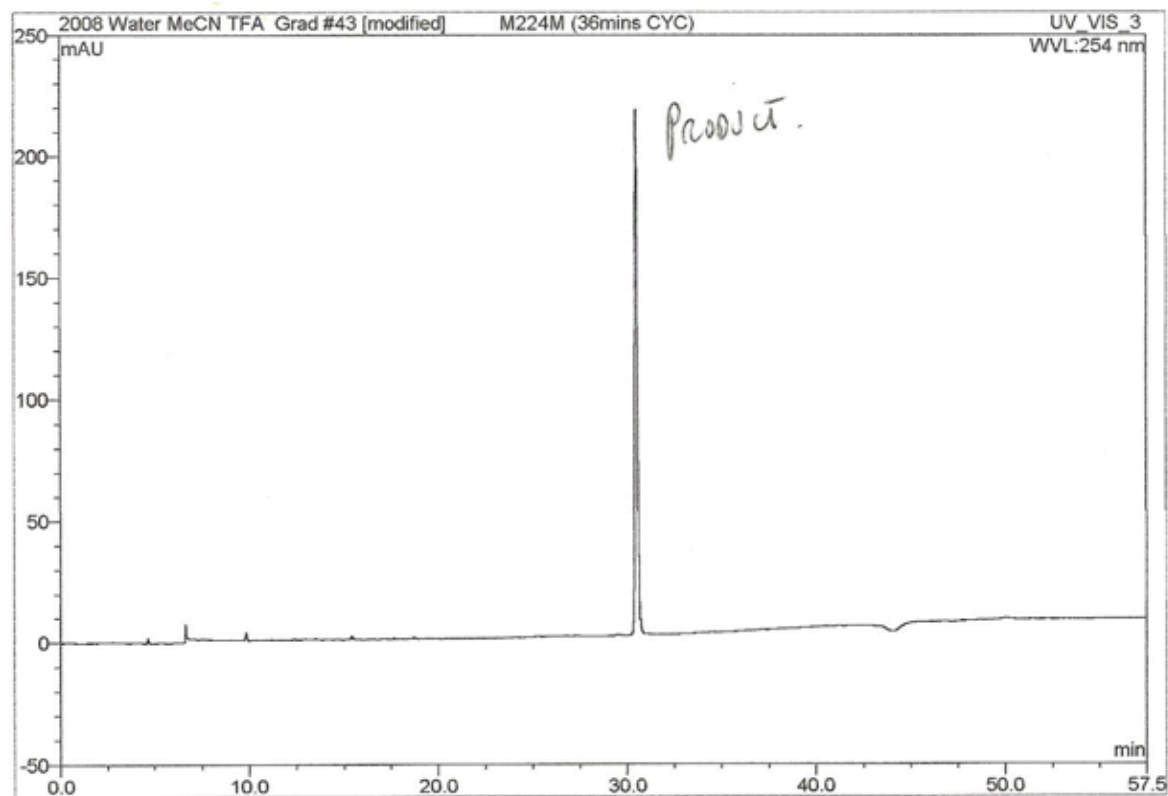
Compound **58 Cyc** Asp(^{t}Bu)-DPhe-Lys(NH$_2$)-Arg(pmc)-Gly

m/z
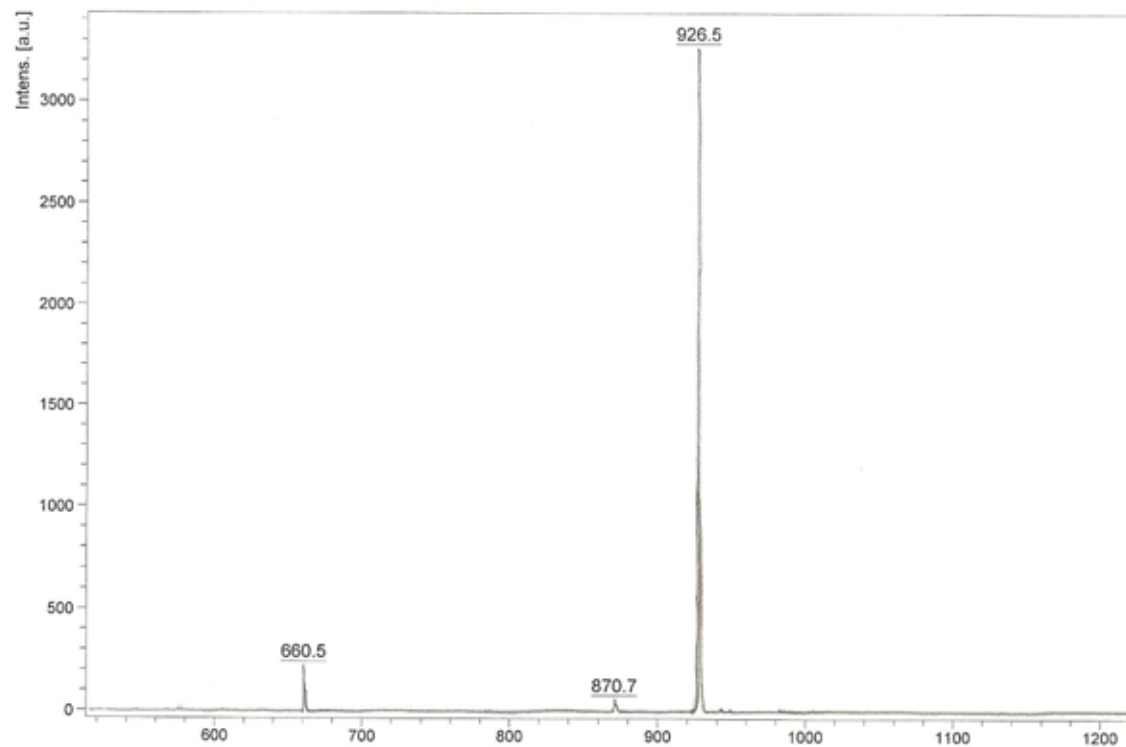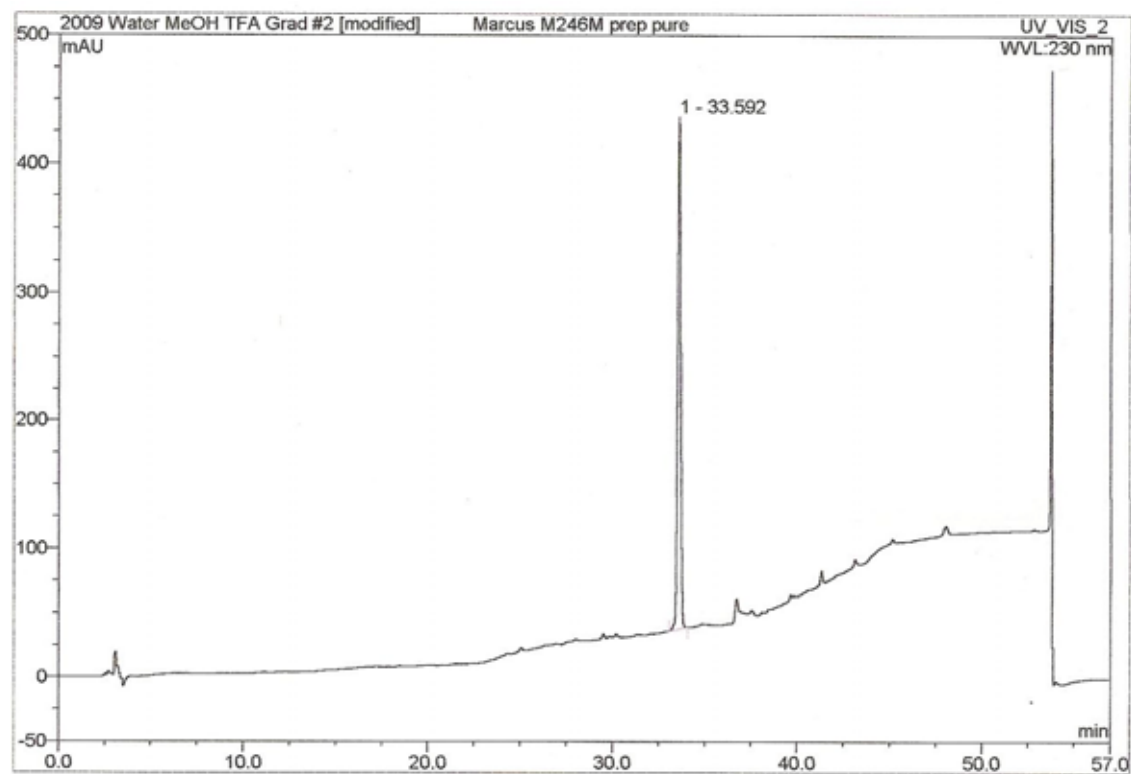
HPLC

Compound **59 Cyc** Asp-DPhe-Lys(NH$_2$)-Arg-Gly

m/z
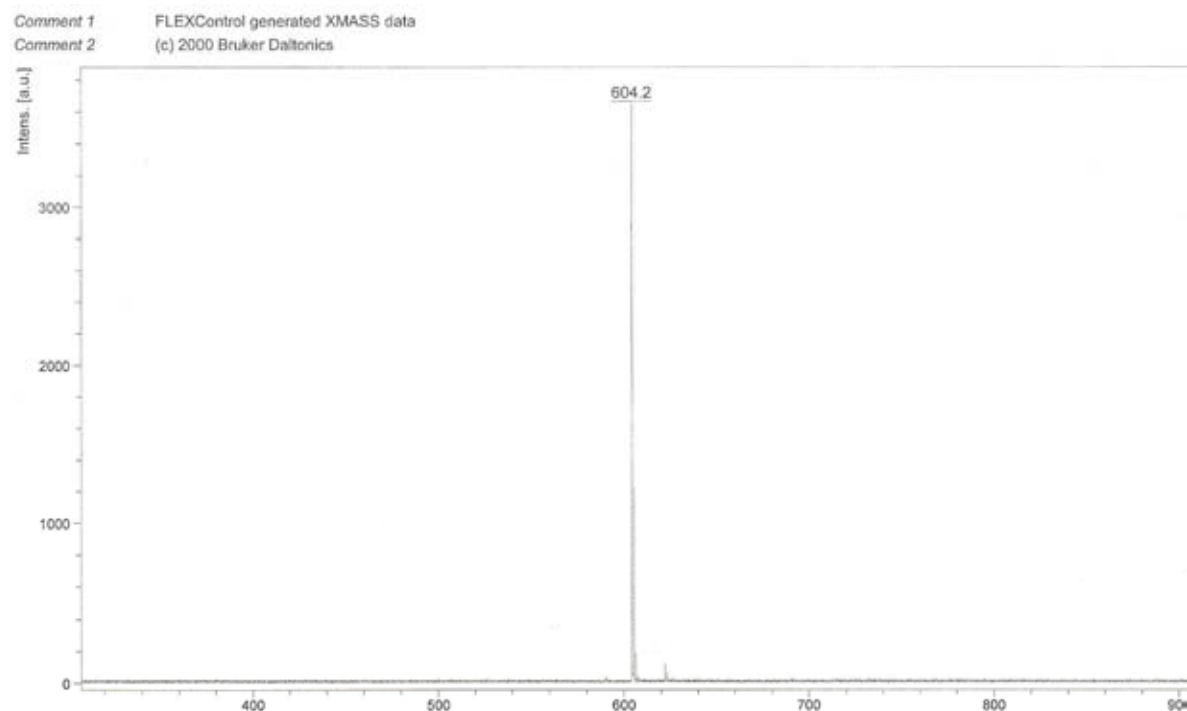
Compound **60**: cPep-Fluorescein

HPLC
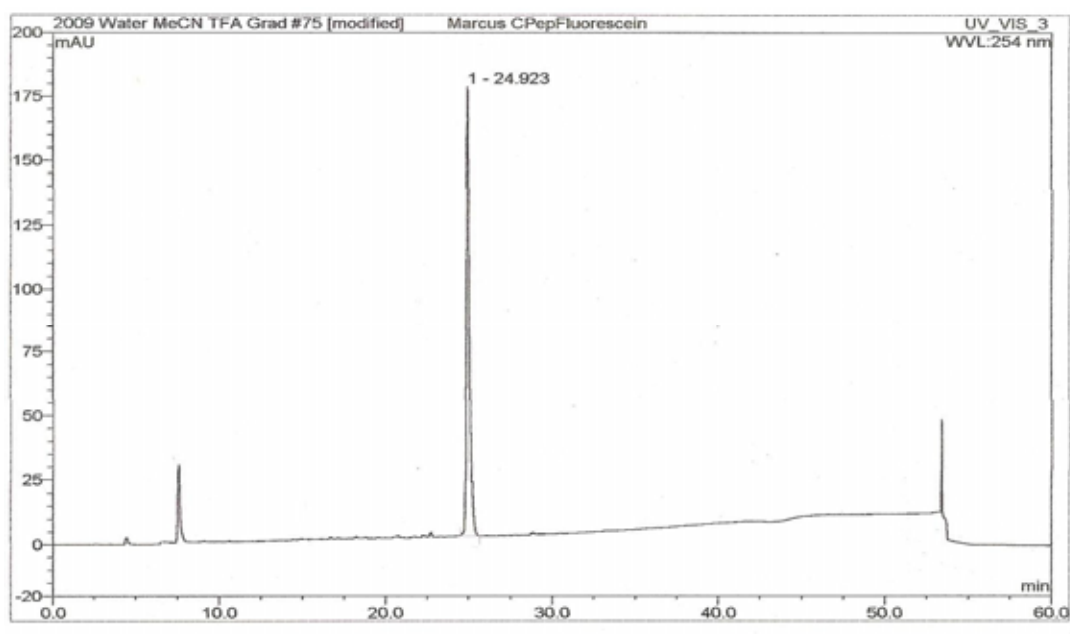
m/z
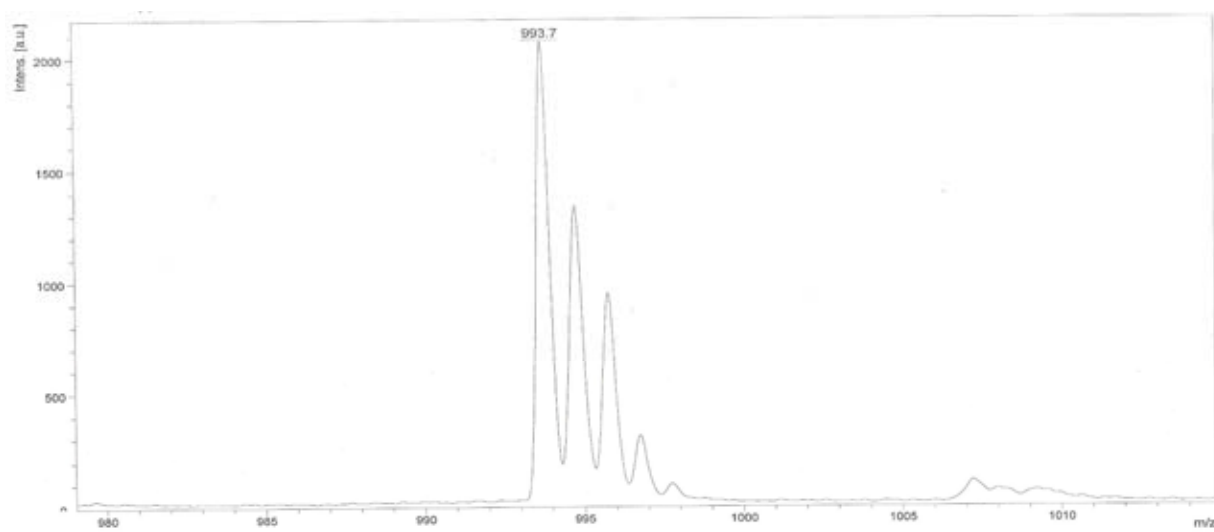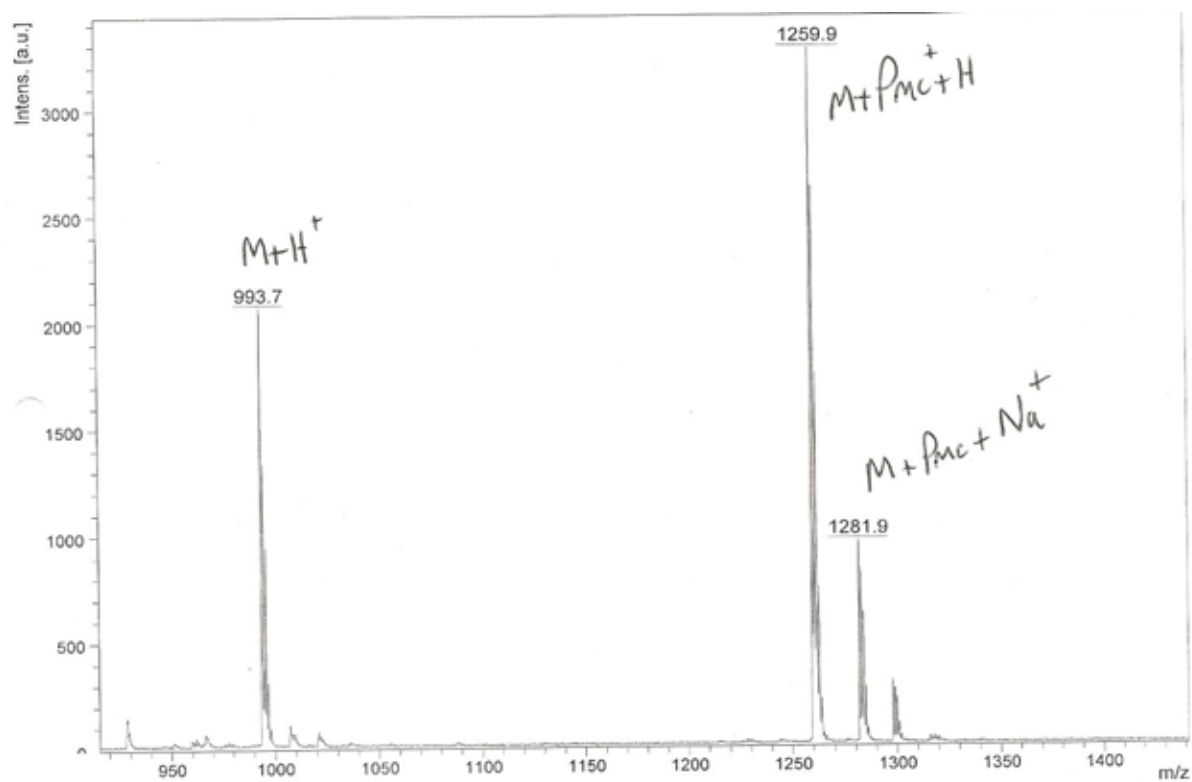
Compound **61**: cPep-NIR797

m/z
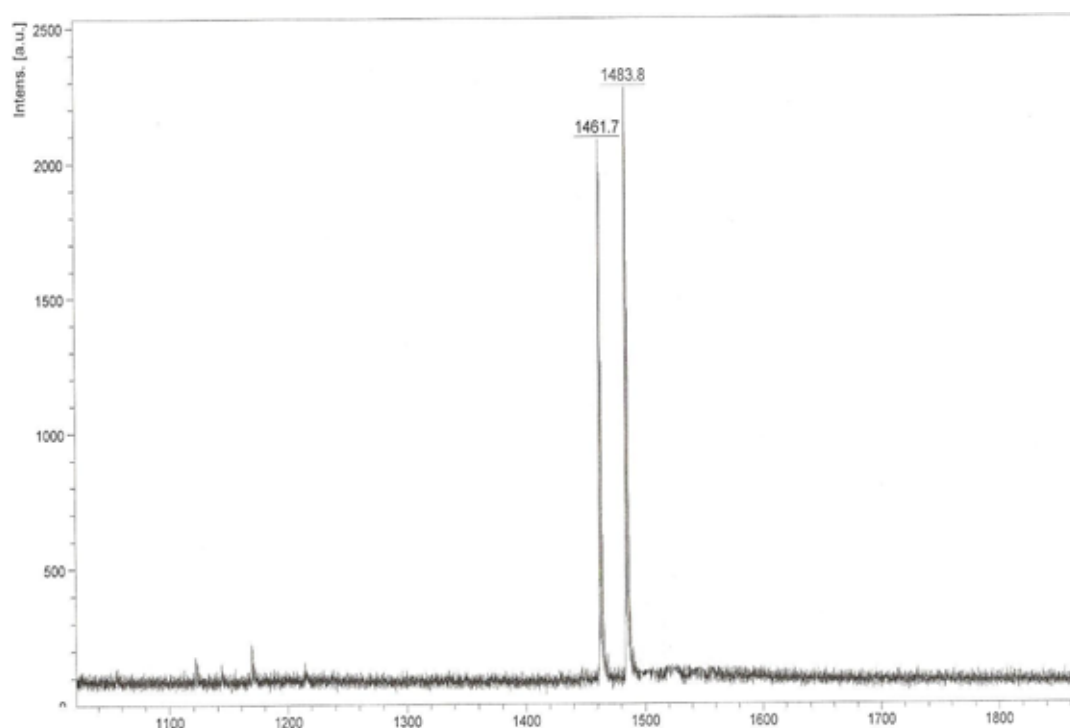
HPLC

Compound **68**: cPep-cap-Os

m/z
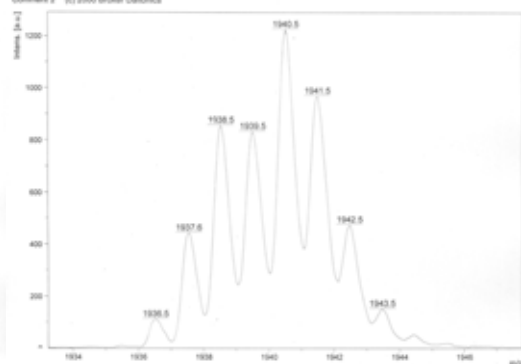
Compound **69**: cPep-cap-benzyl Tb Ugi complex

m/z
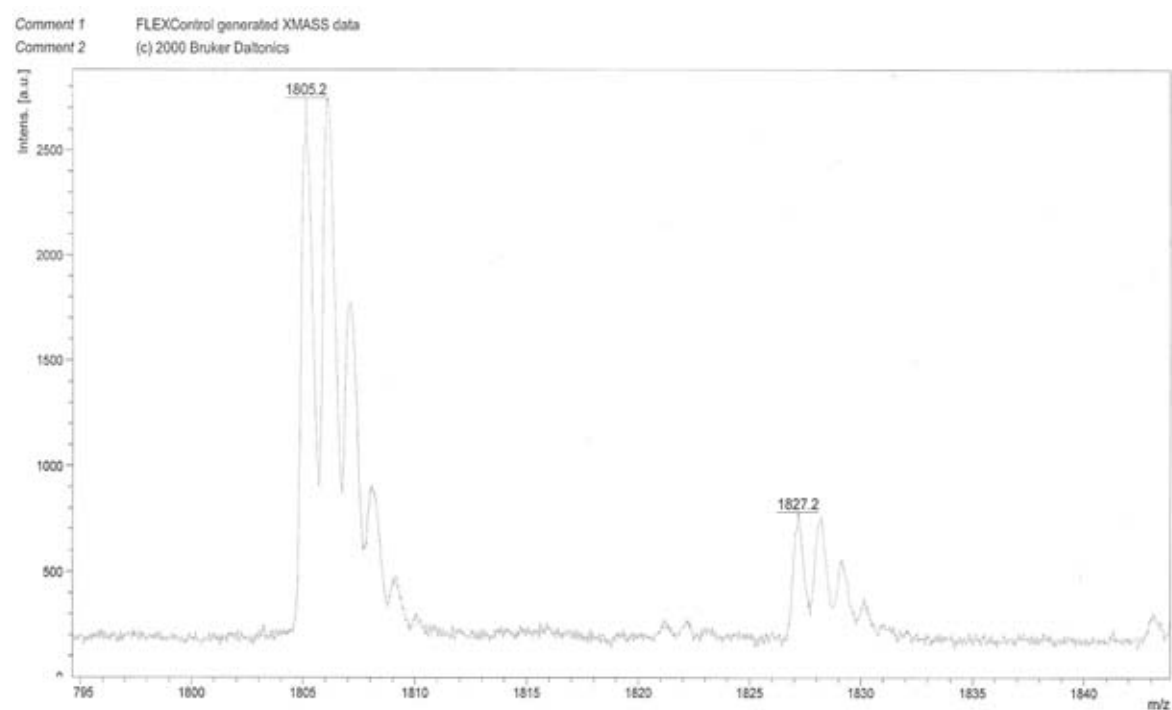
Compound **87** Disaccharide-Fluorescein

m/z

(acetylated)
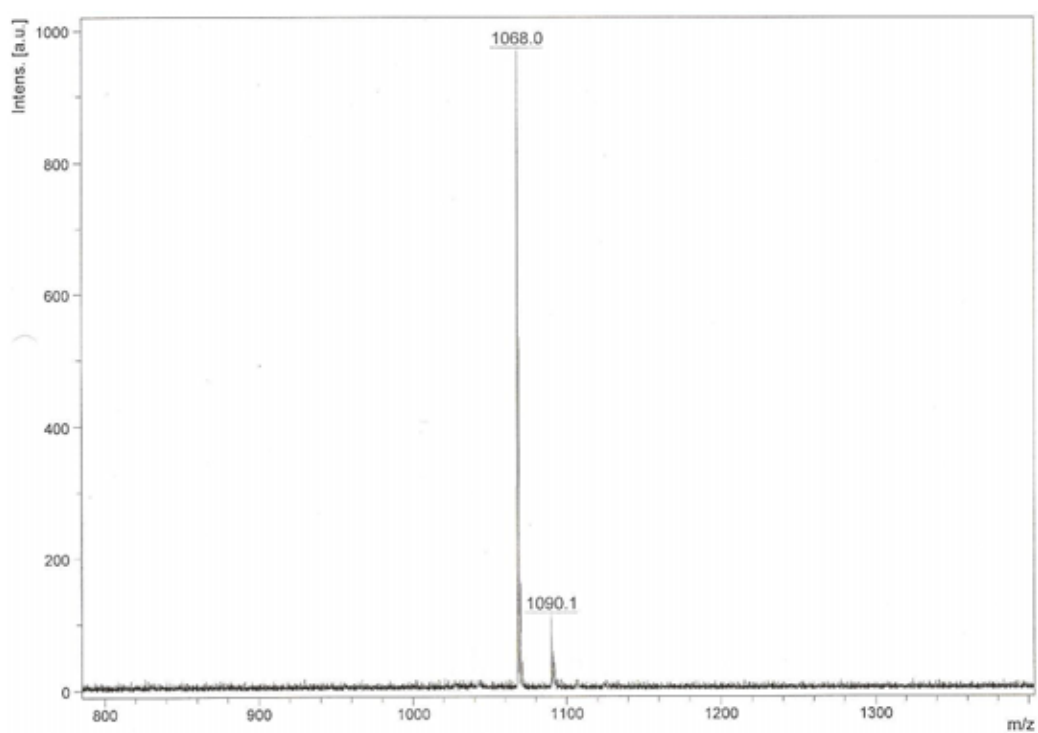
(deacetylated)
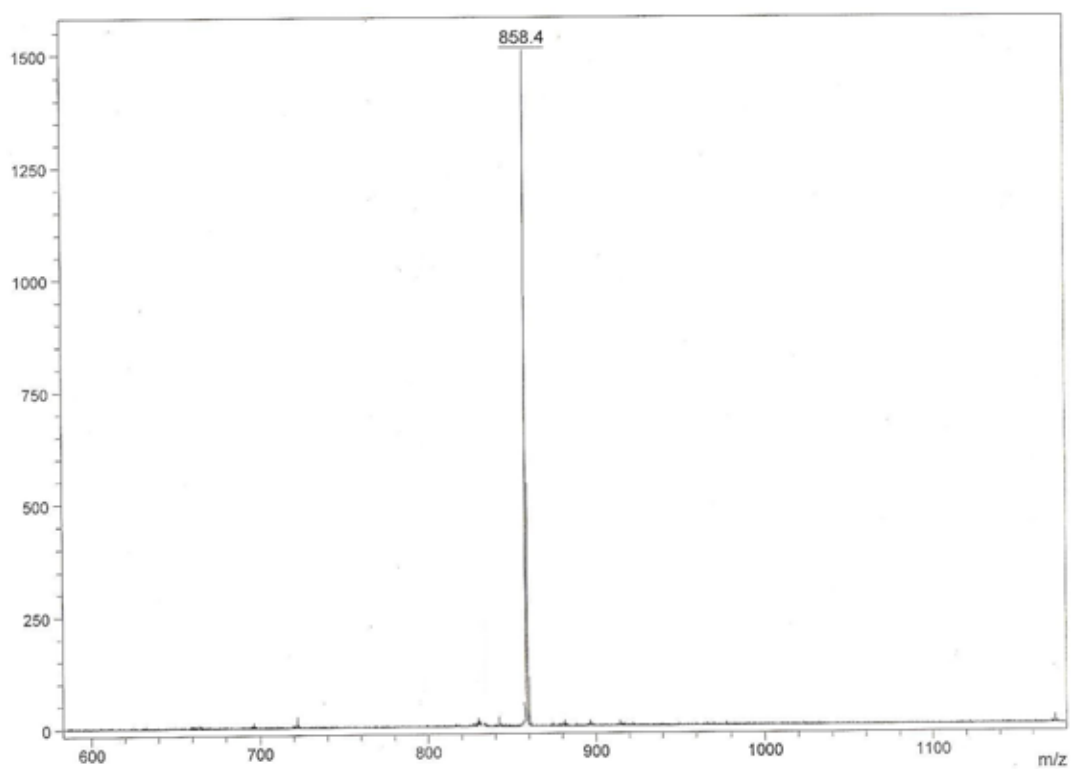
Cut-out of product peaks
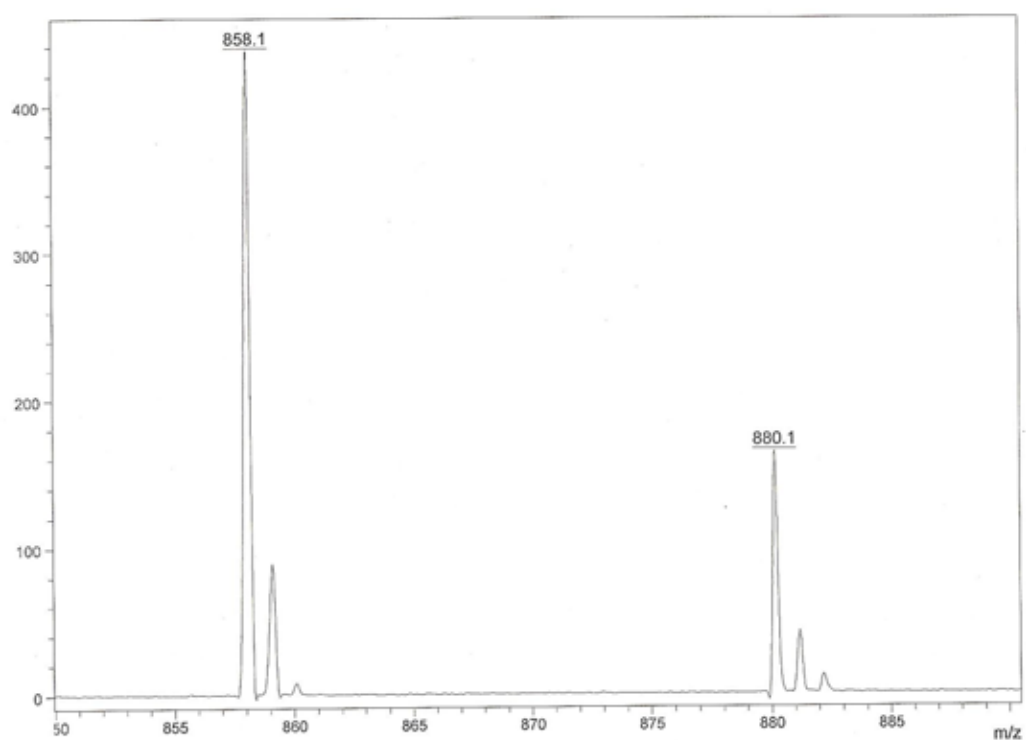
HPLC
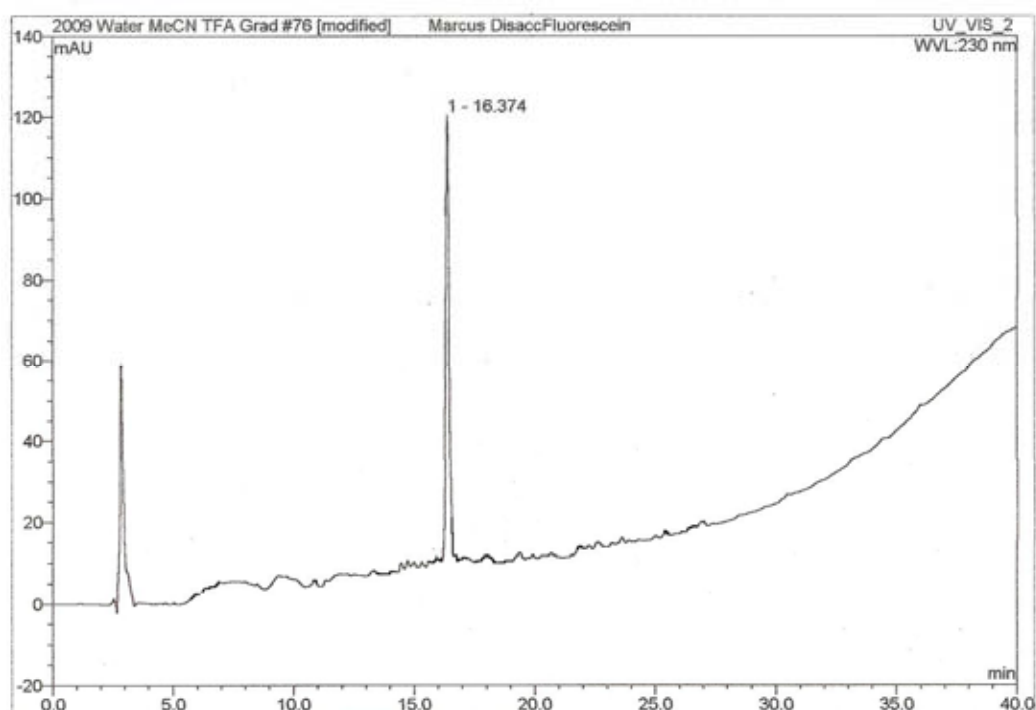
Compound **88**: Disaccharide-NIR797

m/z
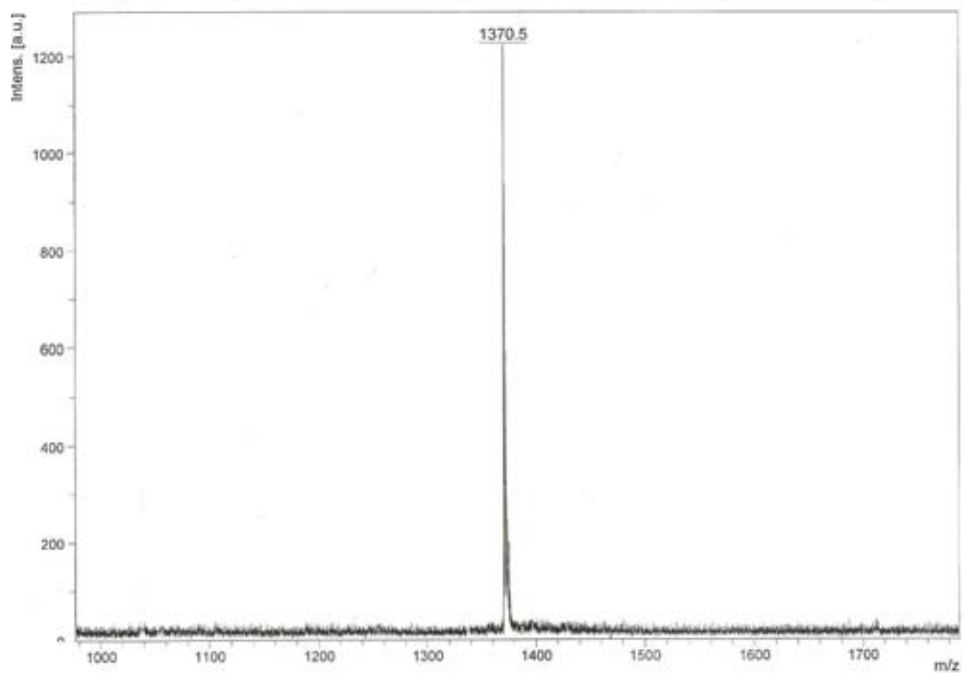
HPLC (analytical)
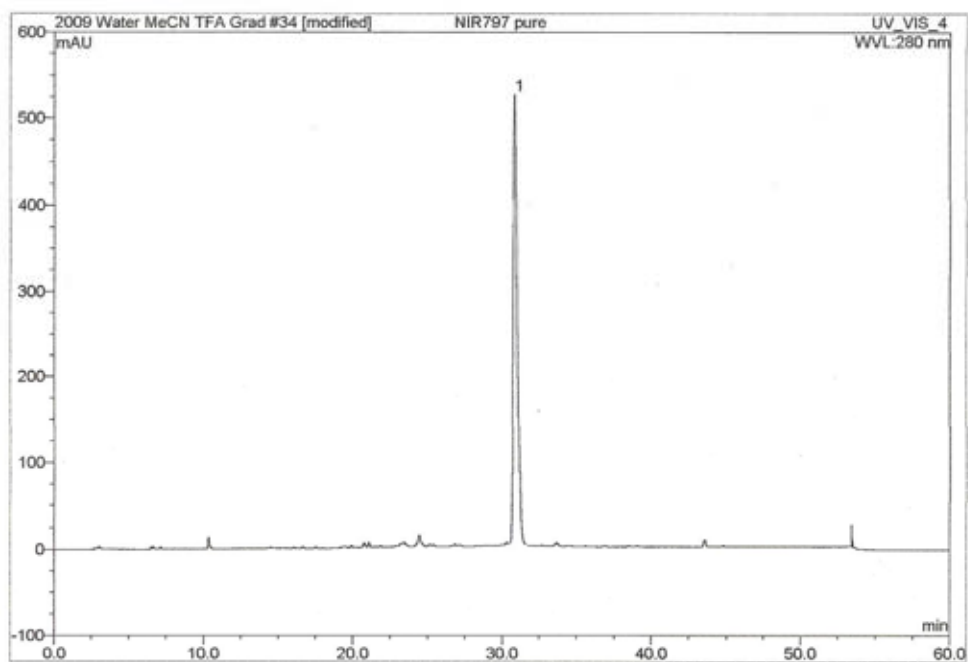
Compound **89**: Disaccharide – Osmium complex
m/z
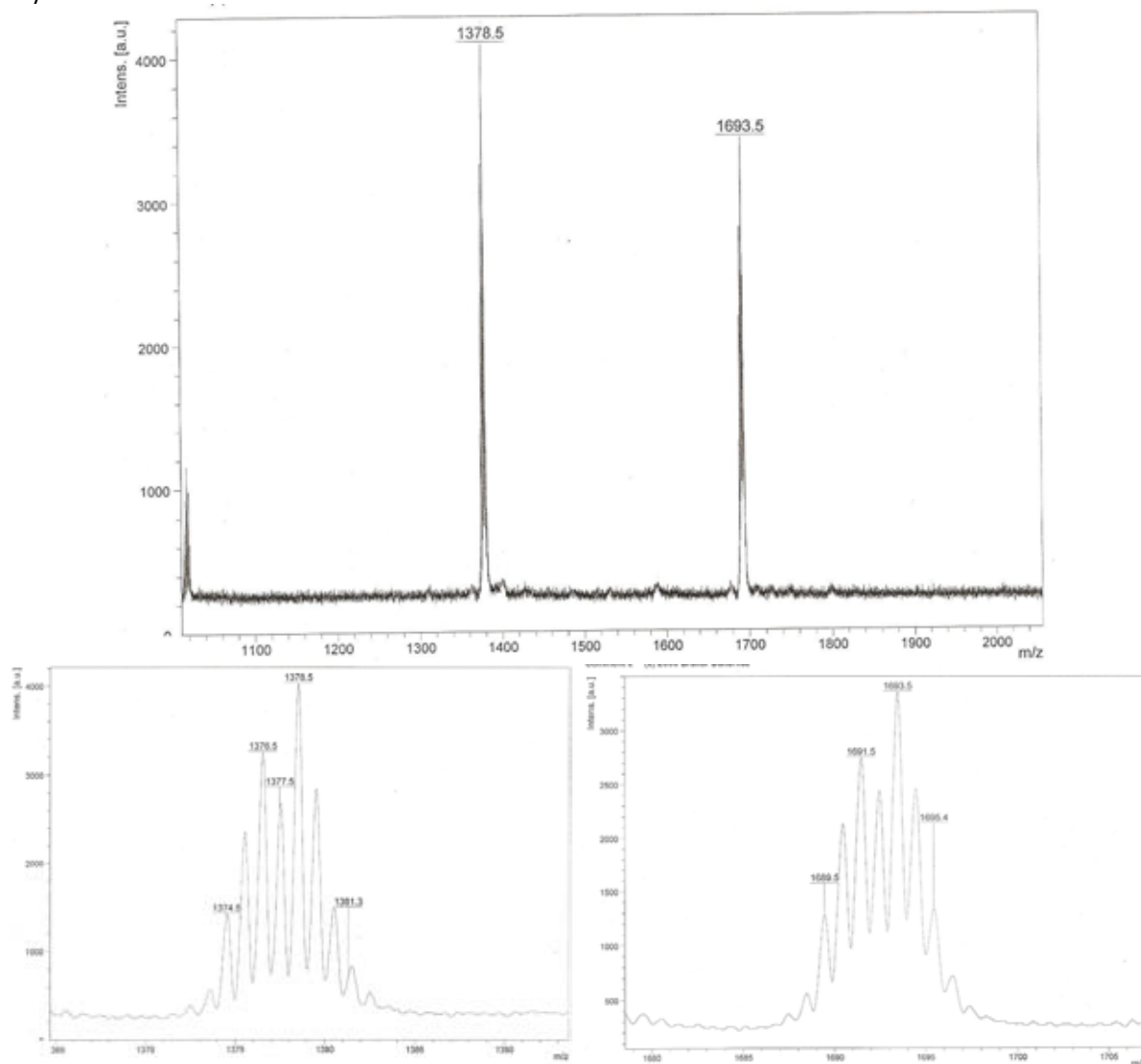
HPLC
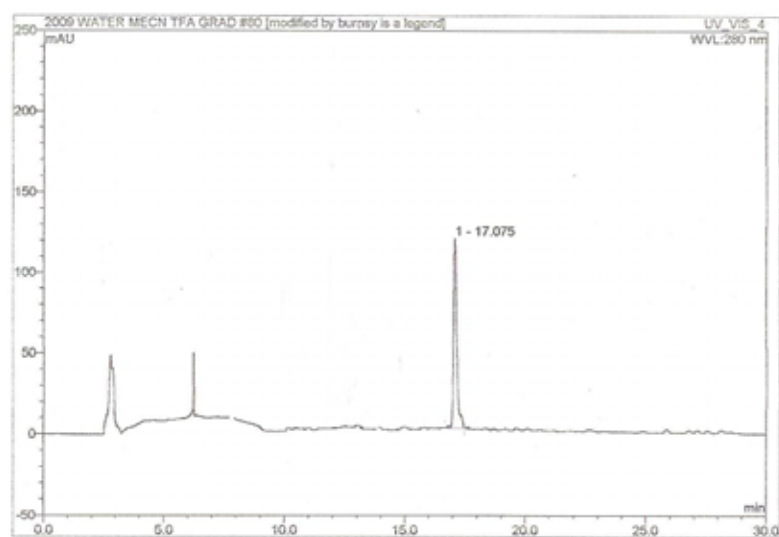
Compound **52**: Osmium complex

HPLC
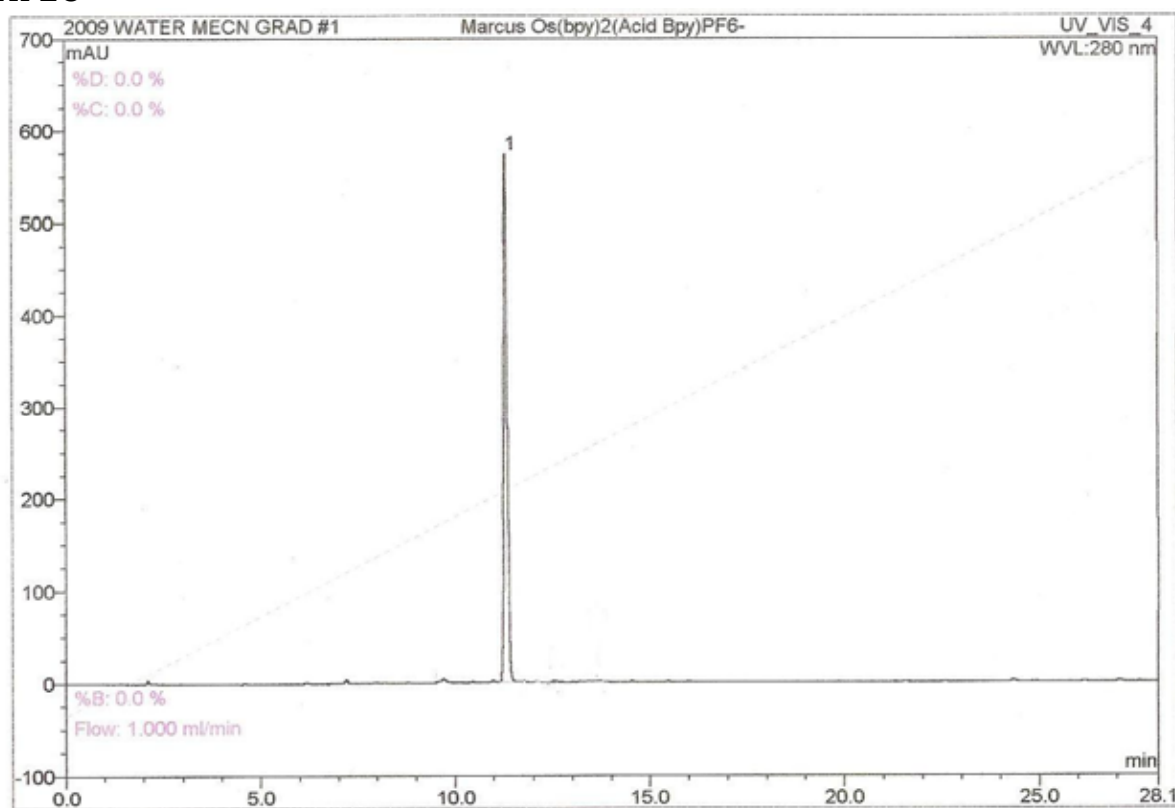
m/z
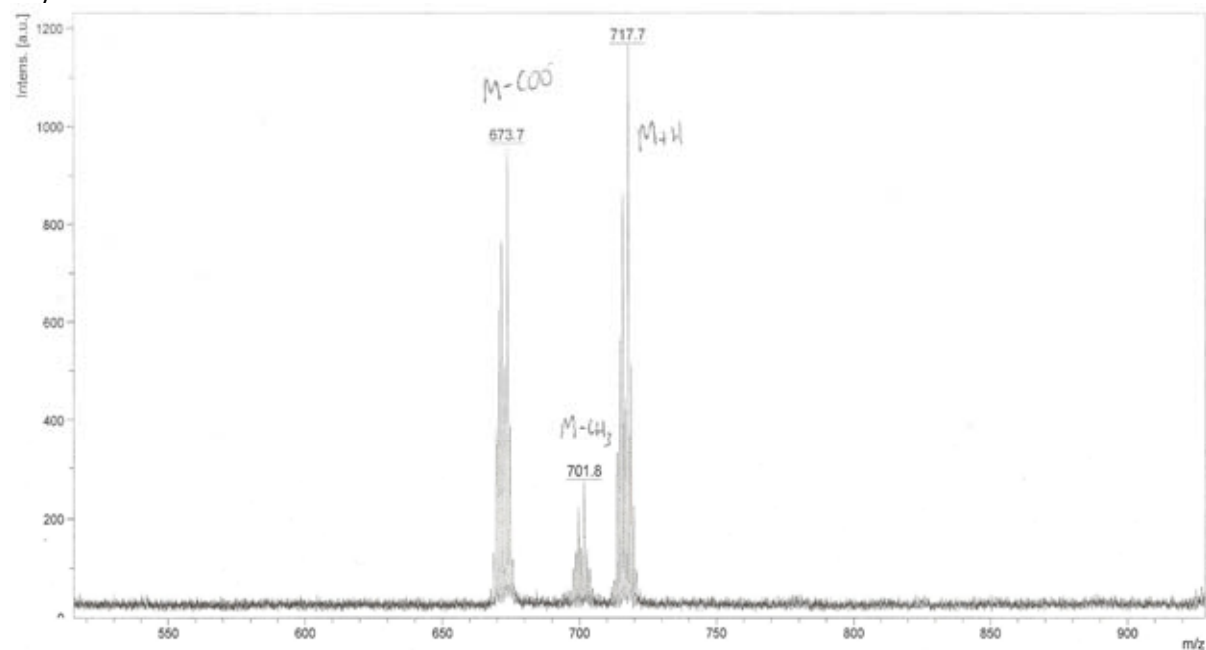
5.6 Appendix 6

2D-nmr data

Compound **71**
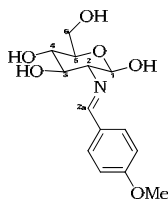
Marcus Main, d6-dmso, +27C, AV400
Barcode label 510, MJ-09

NAME	03-23.31.FAC
EXPNO	10
PROCNO	1
Date_	20090323
Time	14.42
INSTRUM	av400
PROBHD	5 mm BBO BB-1H
PULPROG	zg30
TD	32768
SOLVENT	DMSO
NS	32
DS	2
SWH	8278.146 Hz
FIDRES	0.252629 Hz
AQ	1.9792372 sec
RG	90.5
DW	60.400 usec
DE	6.50 usec
TE	299.8 K
D1	1.00000000 sec
TD0	1

======== CHANNEL f1 ========
NUC1	1H
P1	10.50 usec
PL1	2.00 dB
SFO1	400.0724706 MHz
SI	32768
SF	400.0700000 MHz
WDW	EM
SSB	0
LB	0.30 Hz
GB	0
PC	1.00
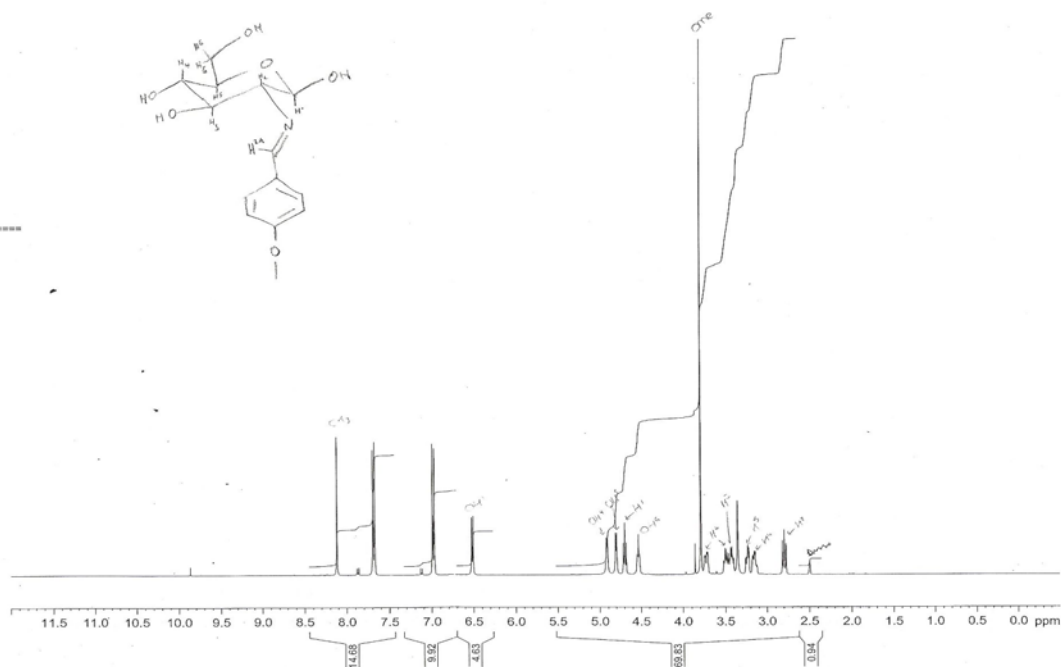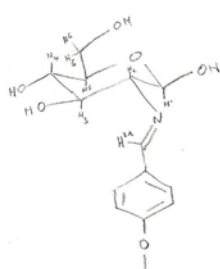
Marcus Main, d6-dmso, +27C, AV300
13C-PENDANT, 1H-decoupled,
Barcode label 514

Marcus Main, d6-dmso, +27C, AV400
Barcode label 510, MJ-09
Gradoent COSY90

Marcus Main, d6-dmso, +27C, AV400
Barcode label 510, MJ-09
Gradoent HSQC

Compound **72**

Marcus Main, CDCl3, +27C, AV400
Barcode label 529

NAME   02-18.31.FAC
EXPNO   10
PROCNO   1
Date_   20090218
Time   15.29
INSTRUM   av400
PROBHD   5 mm BBO BB-1H
PULPROG   zg30
TD   32768
SOLVENT   CDCl3
NS   32
DS   2
SWH   8278.146 Hz
FIDRES   0.252629 Hz
AQ   1.9792372 sec
RG   181
DW   60.400 usec
DE   6.50 usec
TE   300.4 K
D1   1.00000000 sec
TD0   1

======== CHANNEL f1 ========
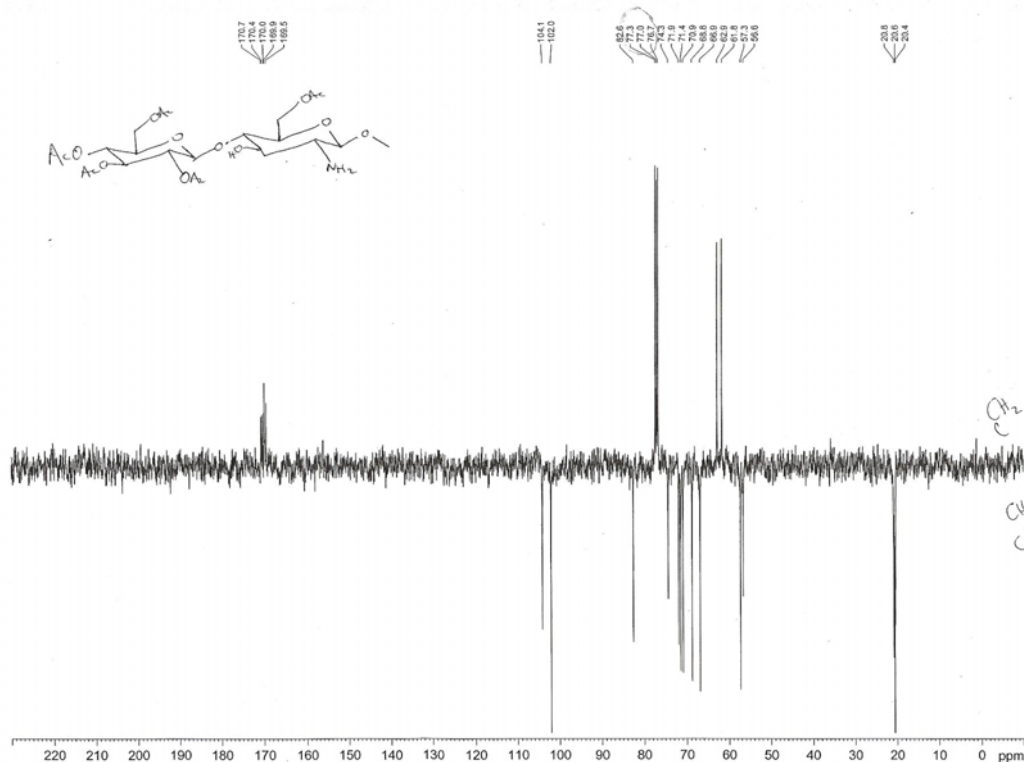NUC1   1H
P1   10.50 usec
PL1   2.00 dB
SFO1   400.0724706 MHz
SI   32768
SF   400.0700000 MHz
WDW   EM
SSB   0
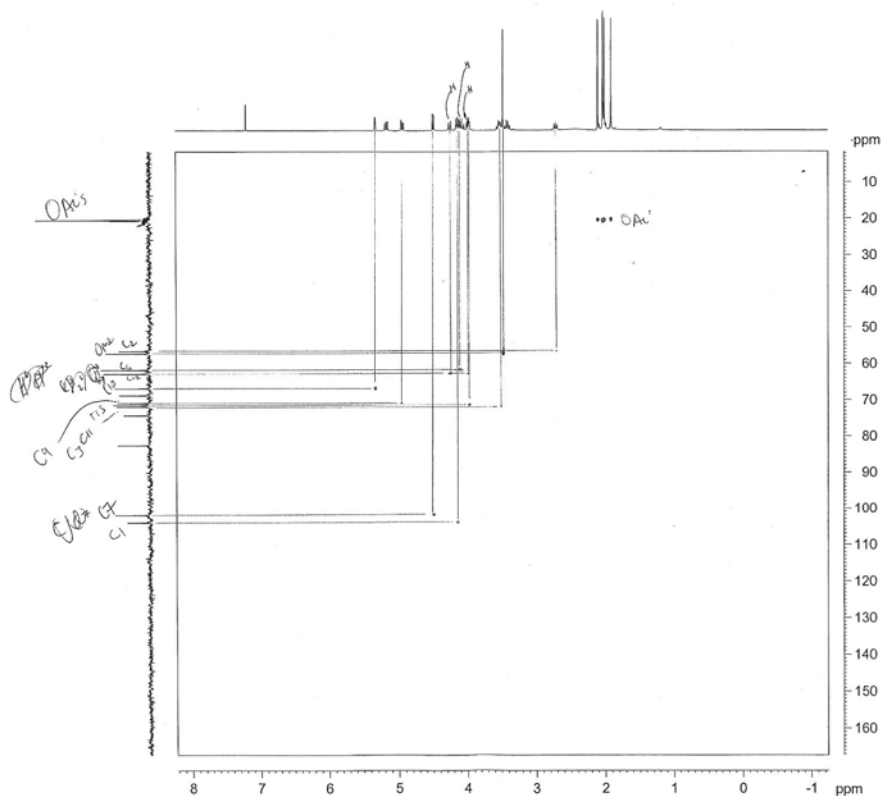LB   0.30 Hz
GB   0
PC   1.00

Marcus Main, CDCl3, +27C, AV300
13C-PENDANT, 1H-decoupled,
Barcode label 505

Current Data Parameters
NAME   12-03.9.FAC
EXPNO   10
PROCNO   1
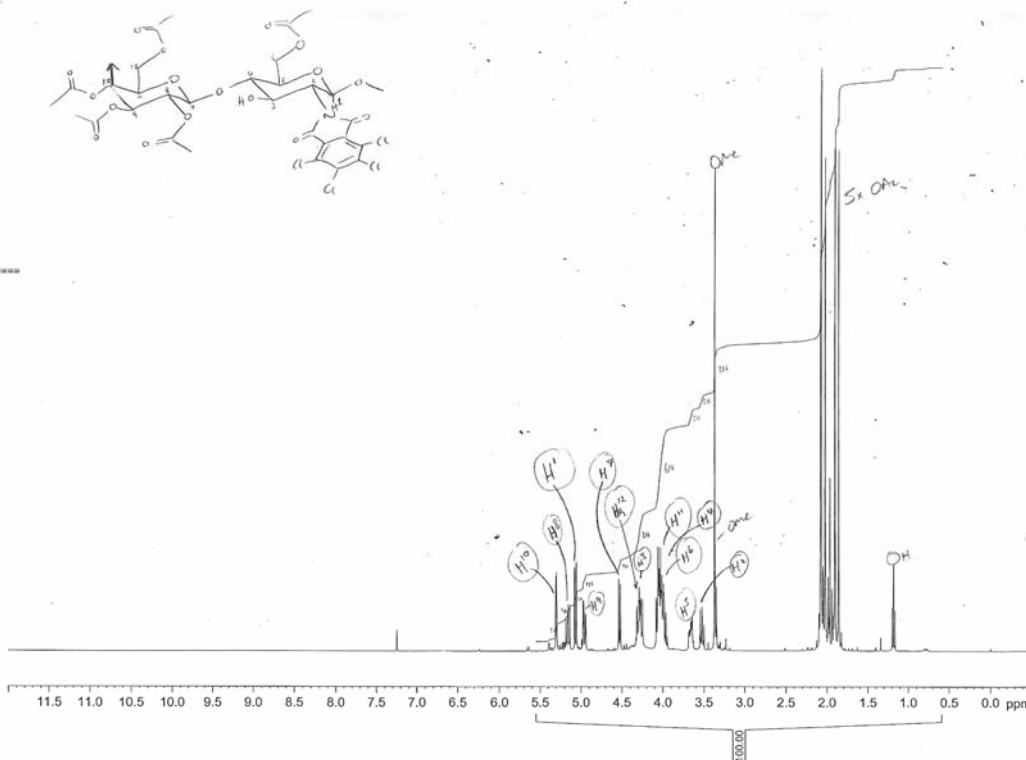
F2 - Acquisition Parameters
Date_   20081203
Time   17.50
INSTRUM   av300
PROBHD   5 mm QNP 1H/1
PULPROG   pendant
TD   65560
SOLVENT   CDCl3
NS   500
DS   4
SWH   18832.393 Hz
FIDRES   0.282939 Hz
AQ   1.7672180 sec
RG   18390.4
DW   26.550 usec
DE   6.00 usec
TE   300.0 K
CNST2   145.0000000
CNST3   1.0000000
CNST4   5.0000000
D1   1.50000000 sec
D2   0.00172414 sec
D3   0.00431034 sec
D12   0.00002000 sec
D13   0.00000400 sec

======== CHANNEL f1 ========
NUC1   13C
P1   10.00 usec
P2   20.00 usec
PL1   7.00 dB
SFO1   75.4618890 MHz

======== CHANNEL f2 ========
CPDPRG2   waltz16
NUC2   1H
P3   10.50 usec
P4   21.20 usec
PCPD2   100.00 usec
PL2   2.00 dB
PL12   22.00 dB
SFO2   300.0765000 MHz

F2 - Processing parameters
SI   65536
SF   75.4637134 MHz
WDW   EM
SSB   0
LB   5.00 Hz
GB   0
PC   0.20

1D NMR plot parameters
CX   40.00 cm
CY   9.00 cm
F1P   230.329 ppm
F1   17379.17 Hz
F2P   -19.213 ppm
F2   -1449.72 Hz
PPMCM   6.23856 ppm/cm
HZCM   470.72217 Hz/cm

Appendix 6: 2D-nmr data

Marcus Main, CDCl3, +27C, AV400
Barcode label 529
Gradient COSY90

Marcus Main, CDCl3, +27C, AV400
Barcode label 529
Gradient HSQC
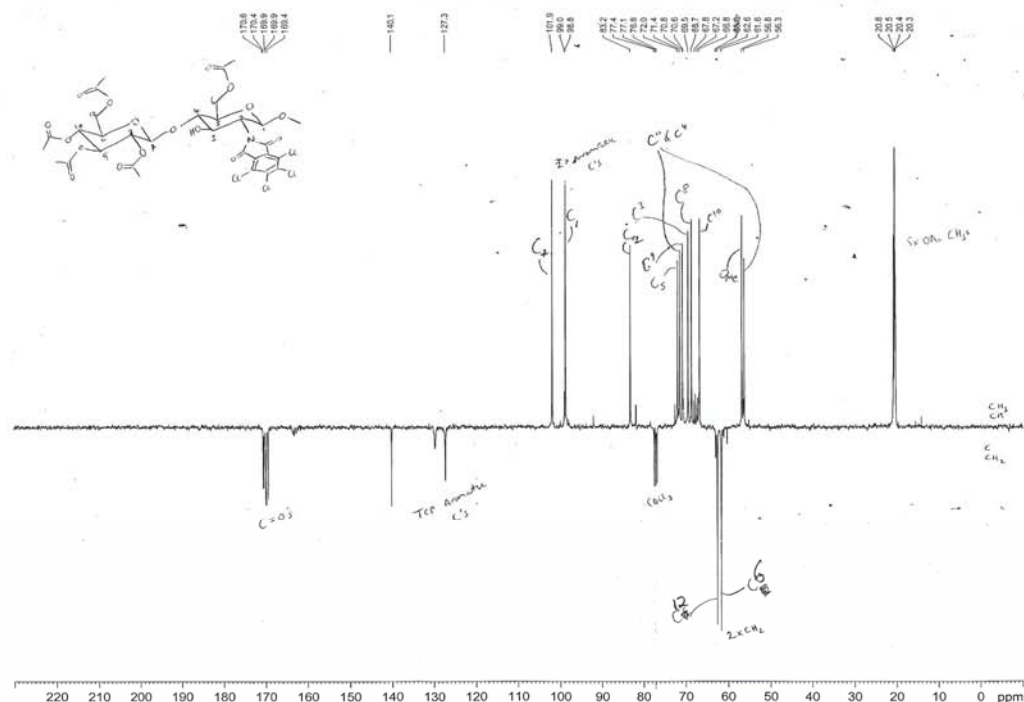
Compound **78**

Appendix 6: 2D-nmr data

John Snaith Project, CDCl3, +27C, AV400
Barcode label 7478
Gradient COSY90
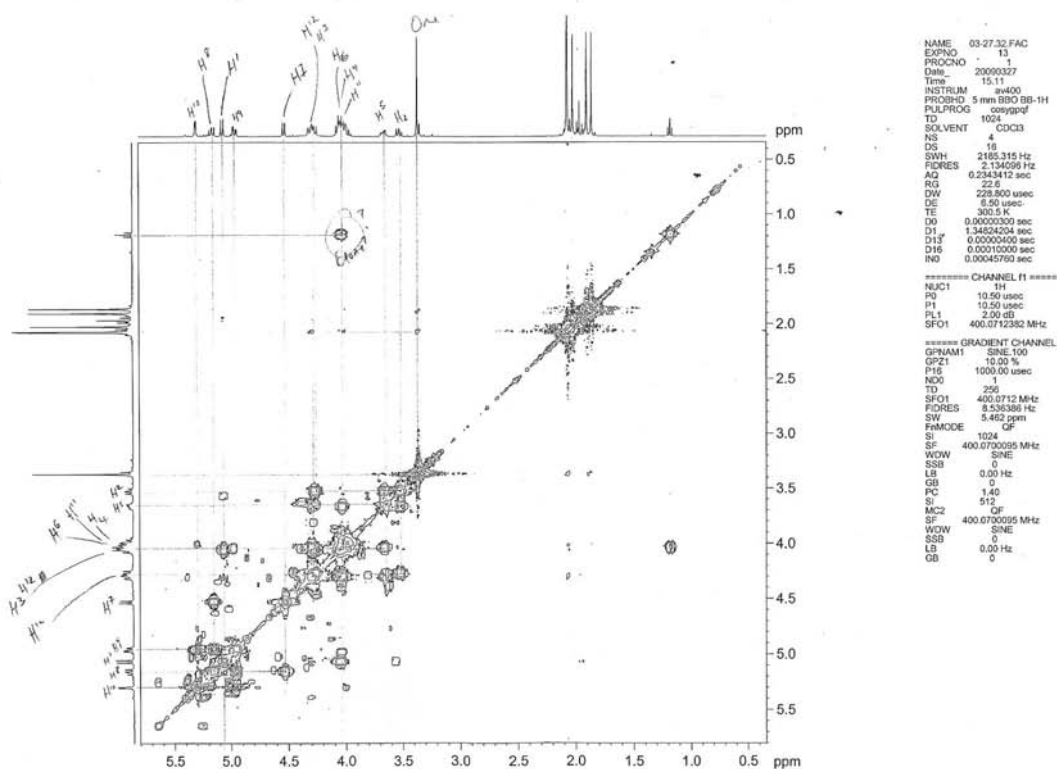
John Snaith Project, CDCl3, +27C, AV400
Barcode label 7478
Gradient HSQC

326

Compound **79**

John Snaith Project, CDCl3, +27C, AV400
Barcode label 7471
Gradient HSQC

John Snaith Project, CDCl3, +27C, AV400
Barcode label 7471
Gradient COSY90

Compound **81**

Marcus Main, CDCl3, AVIII400, NHAcOH

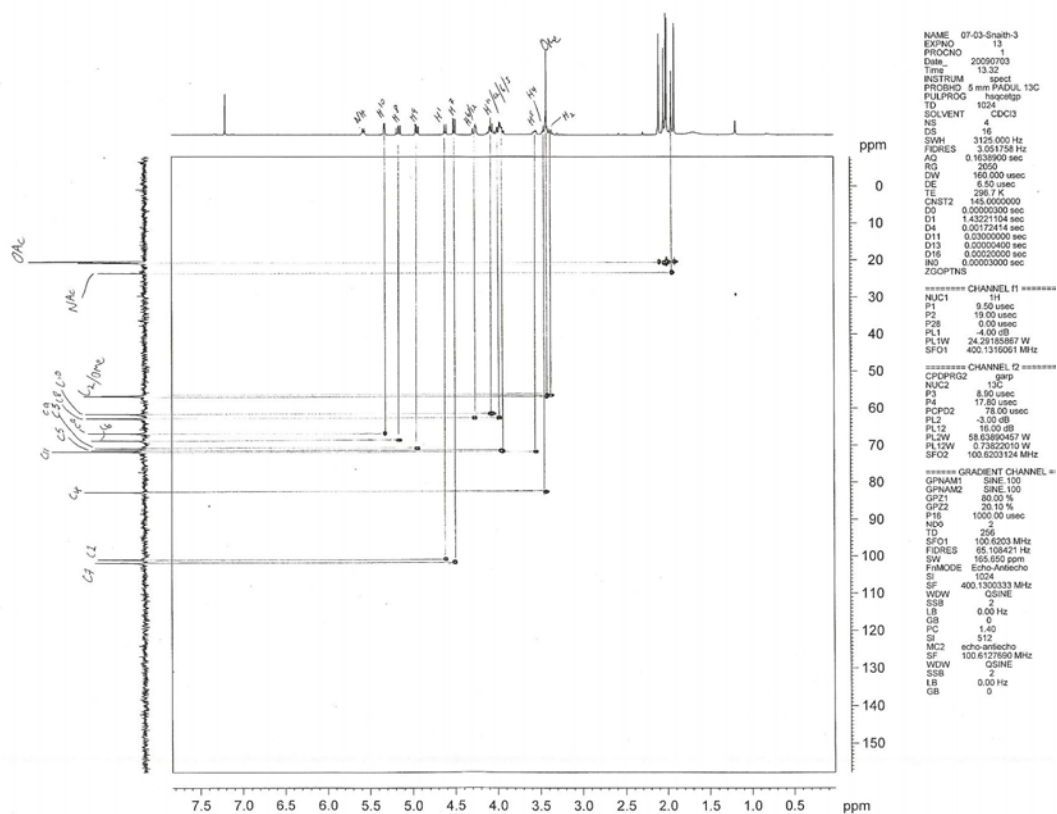Marcus Main, CDCl3, AVIII400, NHAcOH

Appendix 6: 2D-nmr data

Marcus Main, CDCl3, AVIII400, NHAcOH M281M

COSY-90
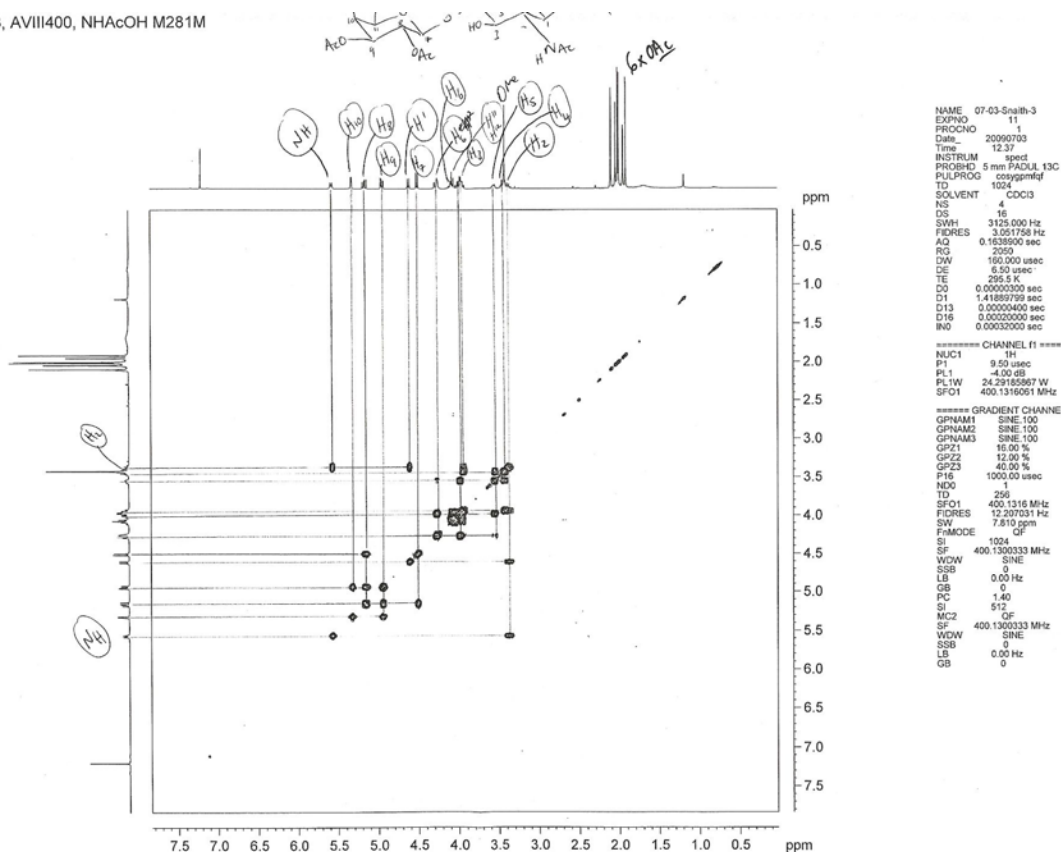
Marcus Main, CDCl3, AVIII400, NHAcOH M281M

HSQC

Compound **80**
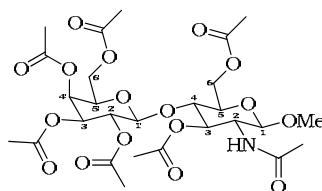
NAME   05-28.31.FAC
EXPNO   10
PROCNO   1
Date_   20090528
Time   14.41
INSTRUM   av400
PROBHD 5 mm BBO BB-1H
PULPROG   zg30
TD   32768
SOLVENT   CDCl3
NS   32
DS   2
SWH   8278.146 Hz
FIDRES   0.252629 Hz
AQ   1.9792372 sec
RG   45.3
DW   60.400 usec
DE   6.50 usec
TE   300.9 K
D1   1.00000000 sec
TD0   1

======== CHANNEL f1 ========
NUC1   1H
P1   10.50 usec
PL1   2.00 dB
SFO1   400.0724706 MHz
SI   32768
SF   400.0700068 MHz
WDW   EM
SSB   0
LB   0.30 Hz
GB   0
PC   1.00

Appendix 6: 2D-nmr data

Marcus Main, CDCl3, +27C, AV400
Barcode label 7451 (JSP label)
Gradient COSY90

Marcus Main, CDCl3, +27C, AV400
Barcode label 7451 (JSP label)
Gradient HSQC

Compound **83**

Appendix 6: 2D-nmr data

Marcus Main, CDCl3, +27C, AV400
Barcode label 508
Gradient COSY90

NAME   05-21.31.FAC
EXPNO   13
PROCNO   1
Date_   20090521
Time   13.19
INSTRUM   av400
PROBHD   5 mm BBO BB-1H
PULPROG   cosygpqf
TD   1024
SOLVENT   CDCl3
NS   4
DS   16
SWH   2913.753 Hz
FIDRES   2.845462 Hz
AQ   0.1757684 sec
RG   64
DW   171.600 usec
DE   6.50 usec
TE   300.0 K
D0   0.00000300 sec
D1   1.40763402 sec
D13   0.00000400 sec
D16   0.00010000 sec
IN0   0.00034320 sec

======== CHANNEL f1 ========
NUC1   1H
P0   10.50 usec
P1   10.50 usec
PL1   2.00 dB
SFO1   400.0717803 MHz

====== GRADIENT CHANNEL =====
GPNAM1   SINE.100
GPZ1   10.00 %
P16   1000.00 usec
ND0   1
TD   256
SFO1   400.0718 MHz
FIDRES   11.381847 Hz
SW   7.283 ppm
FnMODE   QF
SI   1024
SF   400.0700000 MHz
WDW   SINE
SSB   0
LB   0.00 Hz
GB   0
SI   512
PC   1.40
MC2   QF
SF   400.0700000 MHz
WDW   SINE
SSB   0
LB   0.00 Hz
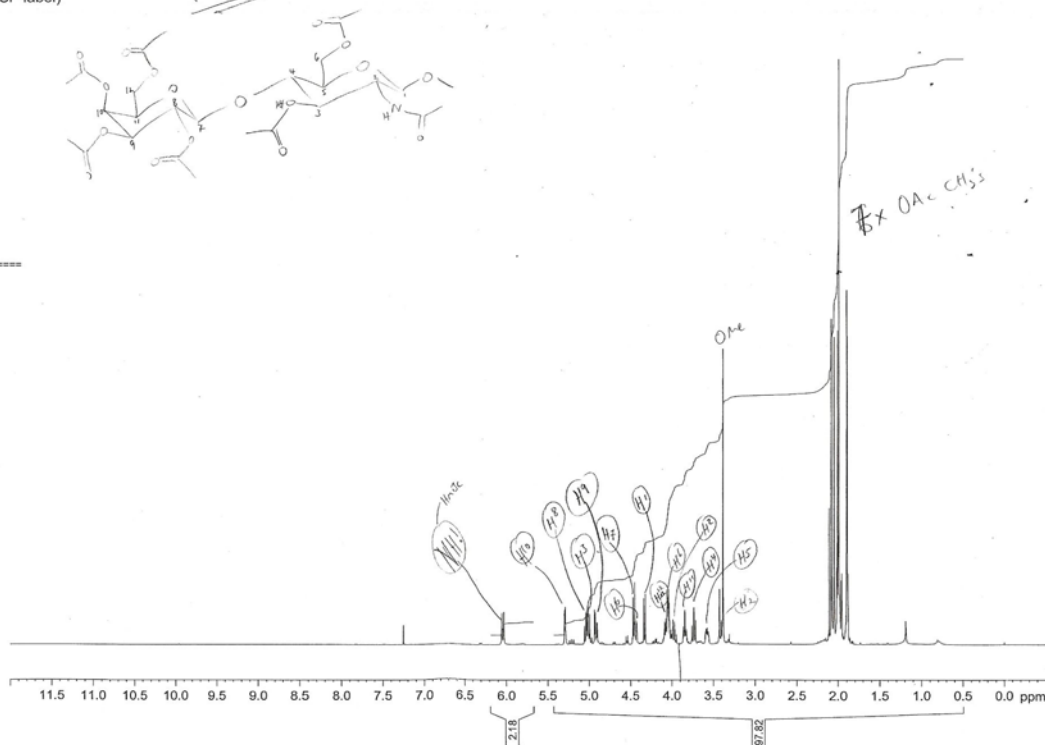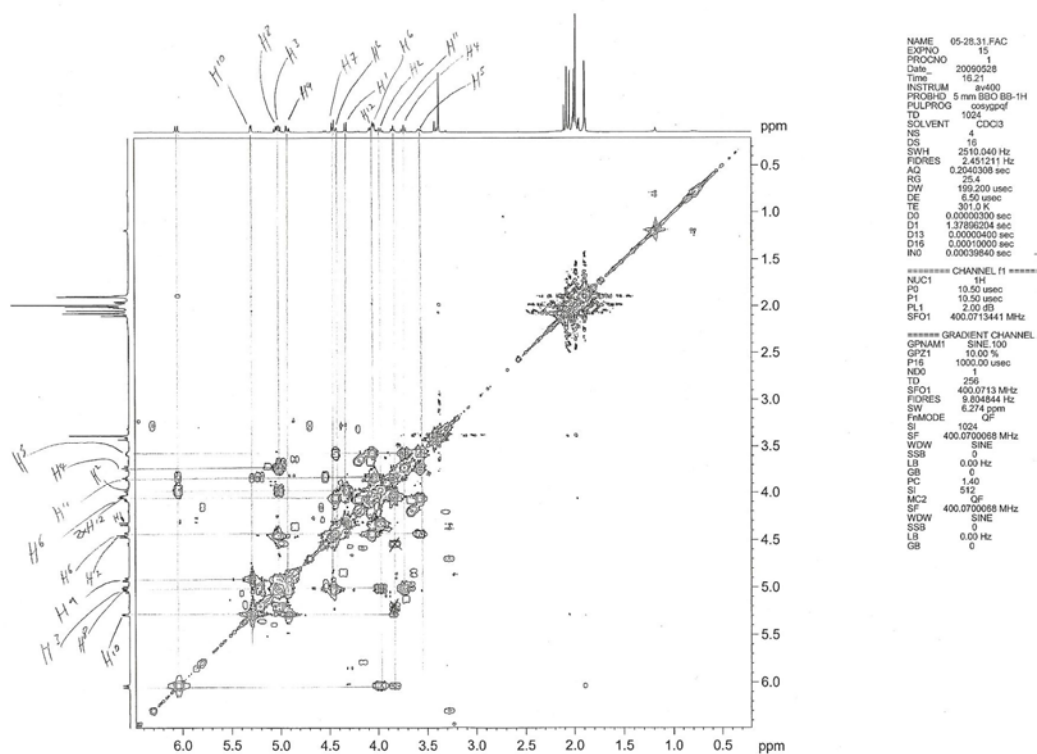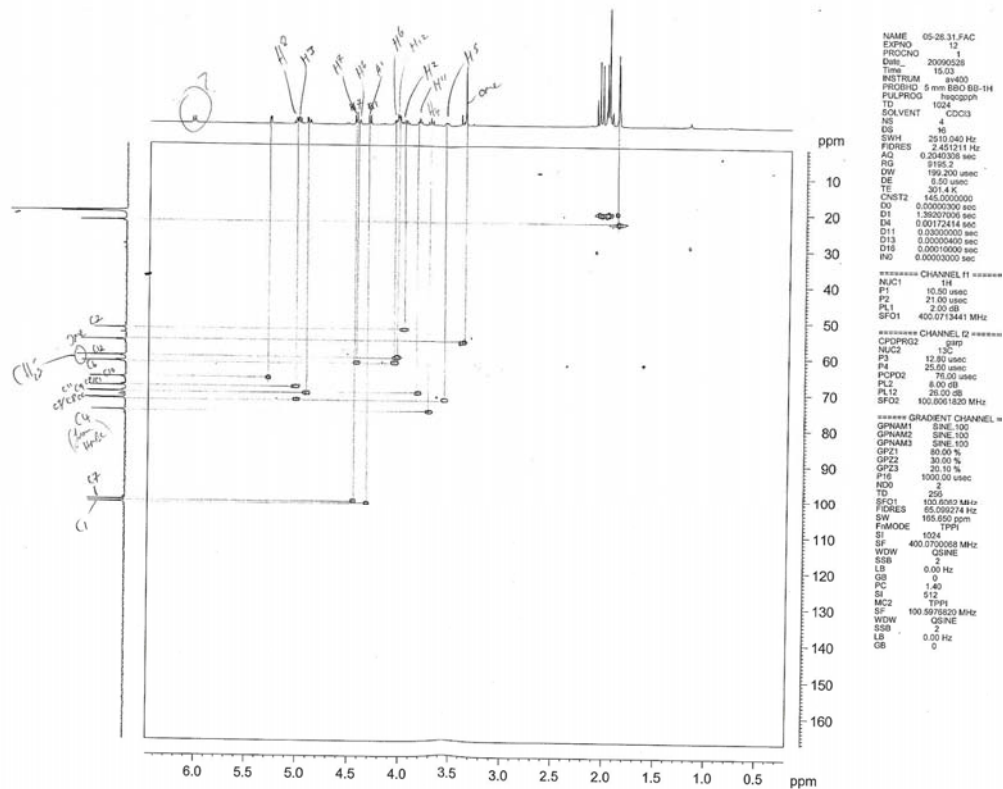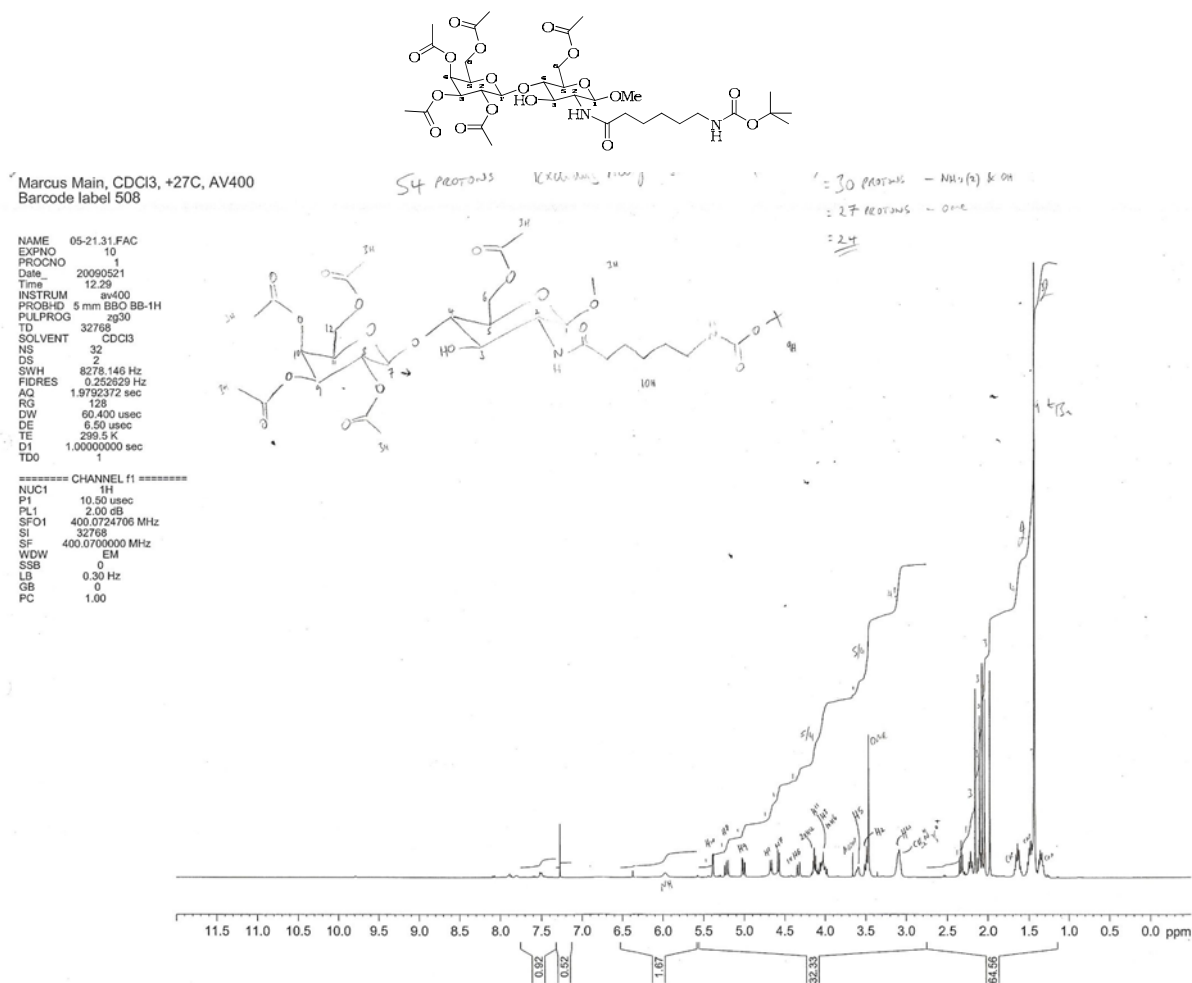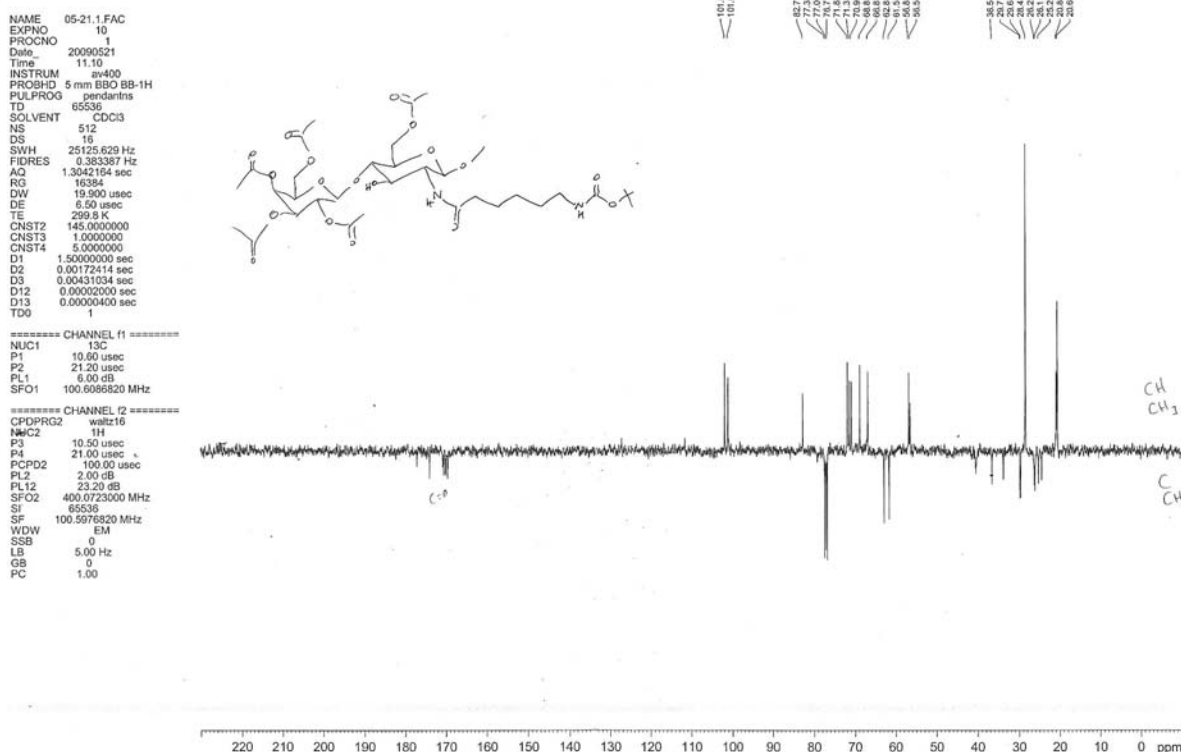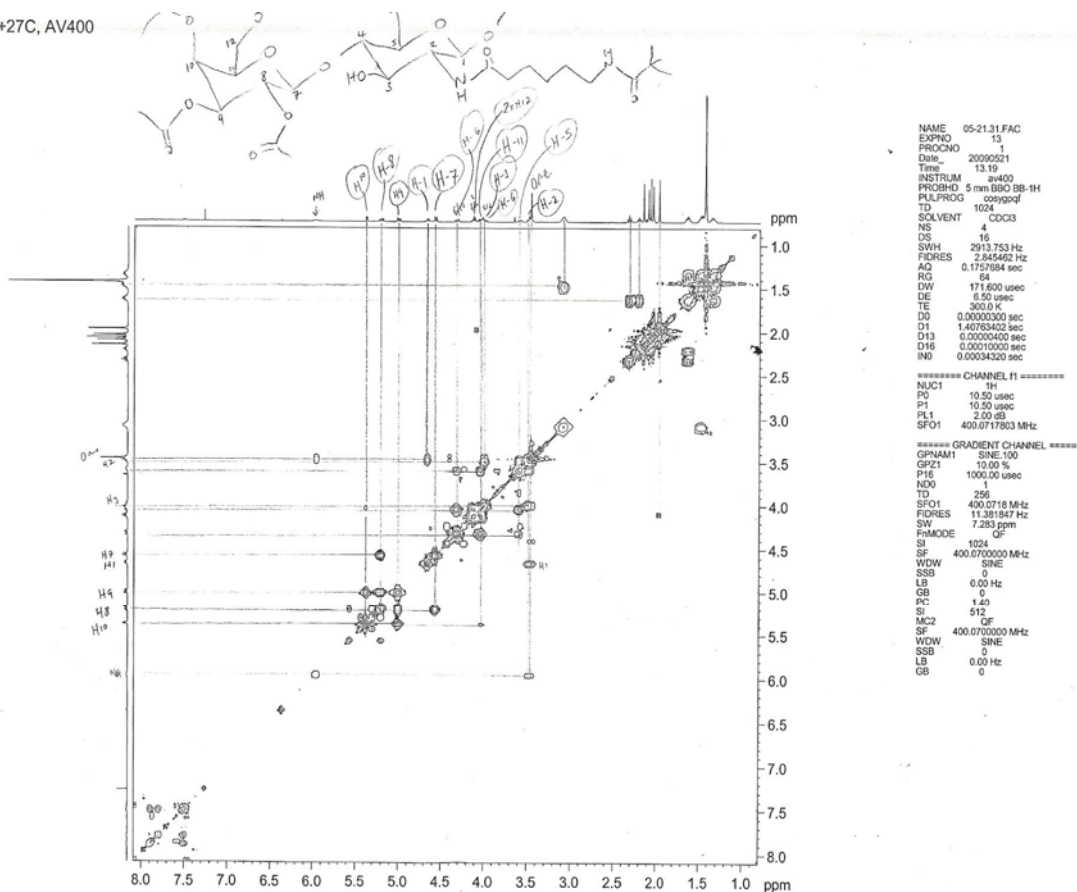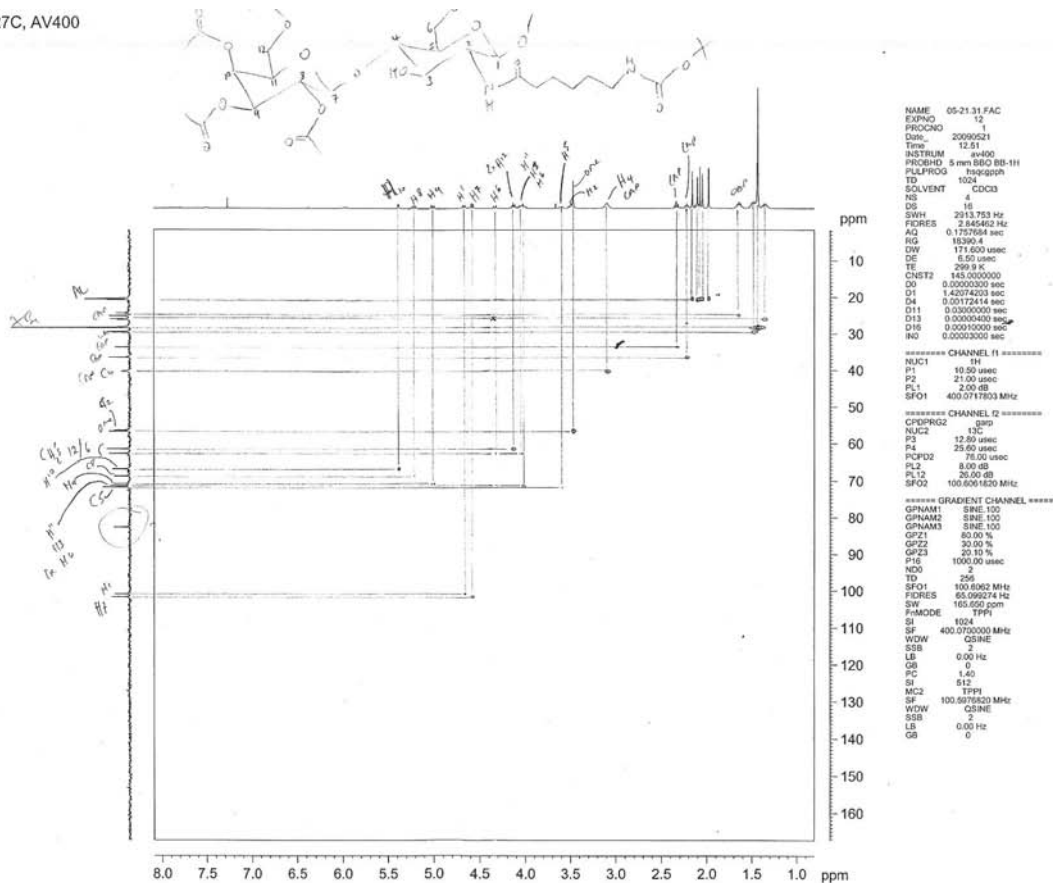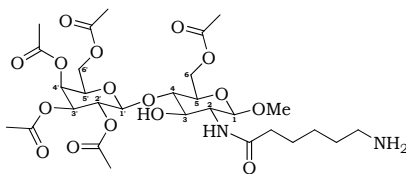GB   0

Marcus Main, CDCl3, +27C, AV400
Barcode label 508
Gradient HSQC

NAME   05-21.31.FAC
EXPNO   12
PROCNO   1
Date_   20090521
Time   12.51
INSTRUM   av400
PROBHD   5 mm BBO BB-1H
PULPROG   hsqcgpph
TD   1024
SOLVENT   CDCl3
NS   4
DS   16
SWH   2913.753 Hz
FIDRES   2.845462 Hz
AQ   0.1757684 sec
RG   16390.4
DW   171.600 usec
DE   6.50 usec
TE   299.9 K
CNST2   145.0000000
D0   0.00000300 sec
D1   1.42074203 sec
D4   0.00172414 sec
D11   0.03000000 sec
D13   0.00000400 sec
D16   0.00000000 sec
IN0   0.00003000 sec

======== CHANNEL f1 ========
NUC1   1H
P1   10.50 usec
P2   21.00 usec
PL1   2.00 dB
SFO1   400.0717803 MHz

======== CHANNEL f2 ========
CPDPRG2   garp
NUC2   13C
P3   12.89 usec
P4   25.60 usec
PCPD2   76.00 usec
PL2   8.00 dB
PL12   26.00 dB
SFO2   100.6061820 MHz

====== GRADIENT CHANNEL =====
GPNAM1   SINE.100
GPNAM2   SINE.100
GPNAM3   SINE.100
GPZ1   80.00 %
GPZ2   30.00 %
GPZ3   20.10 %
P16   1000.00 usec
ND0   2
TD   256
SFO1   100.6062 MHz
FIDRES   65.099274 Hz
SW   165.650 ppm
FnMODE   TPPI
SI   1024
SF   400.0700000 MHz
WDW   QSINE
SSB   2
LB   0.00 Hz
GB   0
PC   1.40
SI   512
MC2   TPPI
SF   100.5976820 MHz
WDW   QSINE
SSB   2
LB   0.00 Hz
GB   0

334

Compound **85**

Biotinylated compound **93**

Marcus Main Biotin derivative, Barcode 505 in CDCl3 at +27C, set temp, drx500

Marcus Main Biotin derivative, Barcode 505 in CDCl3 at +27C, set temp
drx500, Gradient COSY90

Biotinylated compound **94** (tBu cleavage product)

Using the Ugi multicomponent condensation reaction to prepare families of chromophore appended azamacrocycles and their complexes†‡

Marcus Main,[a] John S. Snaith,*[a] Marco M. Meloni,[b] Maite Jauregui,[a] Daniel Sykes,[b] Stephen Faulkner*[b] and Alan M. Kenwright[c]

338