THE UNIVERSITY *of* EDINBURGH

The University of Edinburgh
College of Science and Engineering
School of Chemistry

Cellular analysis and PNA encoded libraries

Nina Svensen

Doctor of Philosophy
2010

Abstract

A peptide nucleic acid (PNA) encoded 1296 member peptide library was synthesised and incubated with a variety of cell types. Library members entering cells were extracted, hybridised onto DNA microarrays and the peptide identity was determined via deconvolution. Global consensus analysis highlighted the tetrapepide, Glu-Llp-Glu-Glu (Llp is 6-hexamine-*N*-aminoacetic acid), a surprise in view of the basic residues typically observed in cell penetrating peptides. When evaluated, Glu-Llp-Glu-Glu revealed cellular uptake comparable to a known basic peptide (tetraLlp). In depth delineation via clustering analysis allowed assessment of differential cellular uptake, with the identified peptides showing clear cellular specificity. This was verified by peptide synthesis and cellular uptake analysis by flow-cytometry, and in all cases an endocytic uptake mechanism was confirmed. This approach establishes a strategy for the identification of short peptides as tools for selective delivery into specific cell types.

The incubation of a 10,000 member PNA-encoded peptide library with D54 and HEK293T transfected with CCR6 cells followed by microarray analysis allowed detailed information on the interaction between peptide-ligands and cell surface receptors to be extracted. This allowed the identification of new ligands for integrins and G-protein coupled receptors and offers a novel approach to ligand discovery allowing the comparative analysis of different cell types for the identification of differences in surface-receptor ligands and/or receptor expression between various cell types. In addition, this work included the development of a novel method for the indirect amplification of a PNA library by amplification of a complementary DNA library hybridised to the PNA.

The generation of 10,000 defined pieces of DNA would have a myriad of applications, not least in the area of defined or directed sequencing and synthetic biology, but also in applications associated with encoding and tagging. By this approach DNA microarrays were used to allow the linear amplification of immobilised DNA sequences on an array followed by PCR amplification. Arrays of increasing sophistication (1; 10; 3875; 10,000 defined oligonucleotides) were used to validate the process, with sequences verified by selective hybridisation to a complementary DNA microarray with DNA sequencing demonstrating error rates of ca $\approx$ 0.2%. This technique offers an economical and efficient way of producing hundreds to thousands of specific DNA primers, while the DNA-arrays can be used as “factories” allowing specific DNA oligonucleotide pools to be generated with or without masking. This study also demonstrated a significant variance observed between the sequence frequencies found via Solexa sequencing compared to microarray analysis.

Declaration

I, Nina Svensen, declare that the thesis entitled “Cellular analysis and PNA encoded libraries” and the work presented herein are my own.

This work was done wholly while in candidature for a research degree at The University of Edinburgh.

None of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution.

Where I have consulted the published work of others it is always clearly attributed.

Where I have reused figures and schemes made by others I have obtained permission from the publishers of the respective work and the source is always given.

Where I have quoted the work of others, the source is always given and with the exception of such quotations, this thesis is entirely my own work.

I have acknowledged all main sources of help.

Where the thesis is based on collaboratory work with others, I have made clear exactly what was performed by others and what I have contributed myself.

Date:_____ Signed:_____

Preface

This report represents, unless otherwise stated, the work carried out between April 2007 and July 2010 at the School of Chemistry, The University of Edinburgh.

Part of the work presented herein has been published as:

- Svensen N. Díaz-Mochón J.J. Bradley M. *Microwave-assisted orthogonal synthesis of PNA-peptide conjugates, Tet. Lett.* **2008**, *49*, 4698.
- Svensen, N; Díaz-Mochón, J.J; Bradley, M. *Microarray generation of 10,000 Oligonucleotides, submitted to Nucl. Acids Res.*
- Svensen, N; Díaz-Mochón, J.J; Dhaliwal, K; Dewar M; Armstrong J. D; Bradley, M. *Targeting specific cellular delivery, submitted to PLoS Biol.*
- Svensen, N; Díaz-Mochón, J.J; Bradley, M. *Screening of an encoded peptide library to discover new cell binding ligands, submitted to J. Am. Chem. Soc.*
- Svensen, N; Díaz-Mochón, J.J; Bradley, M. *Screening of PNA-encoded peptide ligands for cell surface receptors, submitted.*

Acknowledgements

In this context I would like to gratefully thank my supervisor Prof. Mark Bradley for giving me the unique opportunity to work on a fantastic project facing the challenges of both biology and chemistry and for all his endeavour and support. I also deeply thank my supervisor Dr. Rosario M. Sanchez-Martin for encouragement and guidance in biological aspects of this project. A very special thank you to Dr. Juan Jose Diaz-Mochón for his help and guidance in all aspects of the project.

In association with my experimental work I would like to thank the following: A kind thank you to Dr. Kev Dhaliwal for his help with the *in vivo* and primary cell experiments and for the supply of these cells. I also thank Dr. Michael Dewar and Dr. J. Douglas Armstrong for their assistance with the clustering analysis, and Dr. Dominic Campopiano for his guidance in the GPCR world and for the supply of HEK293T-CCR6 cells. Furthermore, I thank all members of the lovely and diverse Bradley Group for creating a great atmosphere in the lab.

On a more personal note to I would like give a very special thank you to Mike for showing me a fantastic Scotland and making my time here so very happy. I am grateful to my sisters and parents for all their support and listening throughout this journey and for their many visits to Scotland.

Contents

Abbreviations

Amino acids

Three-letter code (one-letter code)

Ala (A)	alanine
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
Ile (I)	isoleucine
Leu (L)	leucine
Lys (K)	lysine
Met (M)	methionine
Phe (F)	phenylalanine
Pro (P)	proline
Ser (S)	serine
Tyr (Y)	tyrosine
Val (V)	valine

General

AA	amino acid
Ac	acetyl
A	adenine
Aloc	allyloxycarbonyl
Ar	aryl (NMR)
Azoc	azidomethyloxycarbonyl
Bhoc	benzhydryloxycarbonyl
Boc	*tert*-butoxycarbonyl
br	broad (NMR)
C	cytosine
DAPI	4',6-diamidino-2-phenylindole

CCR6	C-C chemokine receptor type 6
Cbz	carbobenzyloxy
CM	complete medium
CPP	cell penetrating peptide
d	doublet (NMR)
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
DCU	dicyclohexylurea
dd	doublet of doublets (NMR)
Dde	*N*-[1-(4,4-dimethyl-2,6-dioxo*cyclo*hex-1-ylidene)ethyl]
DIC	diisopropylcarbodiimide
DIPEA	*N*,*N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMEM	Dubelco's Modified Eagle Medium
DMF	*N*,*N*-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
ds	double-stranded
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetatic acid
EMEM	Eagle's Minimum Essential Medium
ESAC	encoded self-assembling chemical
FACS	fluorescence activated cell sorting
FAM	5(6)-carboxyfluorescein
FBS	foetal bovine serum
FISH	fluorescence *in situ* hybridisation
FITC	fluorescein isothiocyanate
Fmoc	9-fluorenylmethoxycarbonyl
FRET	fluorescence resonance energy transfer
G	guanine
GDP	guanosine diphosphate
GMEM	Glascow Minimum Essential Medium

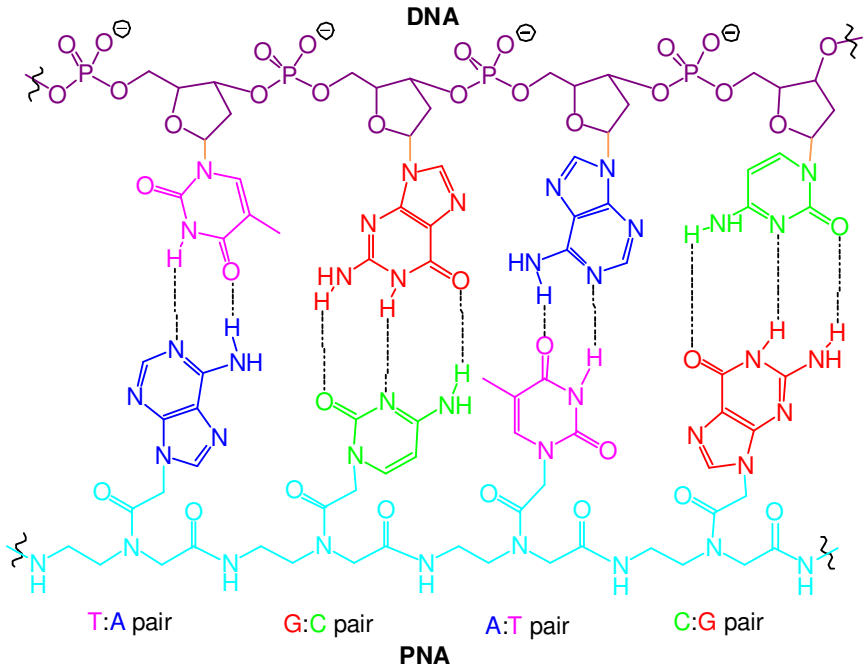
GPCR	G protein coupled receptor
GTP	guanosine triphosphate
HATU	*O*-(7-Azabenzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium hexafluorophosphate
HBTU	*O*-(Benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate
HOBt	1-hydroxy-1*H*-benzotriazole
HPLC	high performance liquid chromatography
IMEM	Improved Modified Eagle Medium
IR	infrared
J	scalar coupling constant (Hz)
Llp	lysine-like peptoid
LRMS	low resolution mass spectroscopy
m	multiplet (NMR) or medium (IR)
MALDI-TOF	matrix-assisted laser desorption/ionisation time-of-flight
Me	methyl
Mmt	monomethoxytrityl
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MTT	(3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NEAA	non-essential amino acids
NEM	*N*-ethylmorpholine
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEGA	polyethylene glycol dimethyl acrylamide
PerCP	peridinin chlorophyll protein complex
PNA	peptide nucleic acid
PS	polystyrene
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
q	quartet (NMR)

quint.	quintuplet (NMR)
Rho	rhodamine
RNA	ribonucleic acid
RP	reverse phase
RPMI	Roswell Park's Memorial Institute
rRNA	ribosomal ribonucleic acid
s	singlet (NMR) or strong (IR)
SDS	sodium dodecyl sulfate
SNP	single-nucleotide polymorphism
SPPS	solid phase peptide synthesis
ss	single-stranded
SSC	saline sodium citrate
T	thymine
t	triplet (NMR)
TAMRA	tetramethyl-6-carboxyrhodamine
TAT	trans-acting activator of transcription
TBE	tris borate ethylenediaminetetraacetatic acid
td	triplet of doublets (NMR)
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIS	triisopropylsilane
TLC	thin layer chromatography
T_m	melting temperature
TRITC	tetramethyl rhodamine isothiocyanate
UV	ultraviolet
w	weak (IR)
δ	chemical shift (NMR)

Chapter 1 Basic concepts

1.1 Peptide nucleic acids

The extensive use of deoxyribonucleic acids (DNAs) or ribonucleic acids (RNAs) in molecular biology has highlighted a number of shortcomings of these natural bases, such as rapid enzymatic degradation and poor mismatch discrimination. This has led to the development of numerous DNA (and RNA) analogues. One very attractive synthetic DNA mimic, the so-called peptide nucleic acids (PNAs), were reported by Nielsen in 1991.[1,2] PNAs have an uncharged, achiral and acyclic polyamide backbone composed of *N*-(2-aminoethyl)glycine units with a side chains comprising one of the four nucleobases, adenine (A), cytosine (C), guanine (G), or thymine (T), linked through a methylene carbonyl spacer (Scheme 1.1). By convention, PNAs are depicted like peptides, with the N-terminus on the left and the C-terminus on the right hand-side.

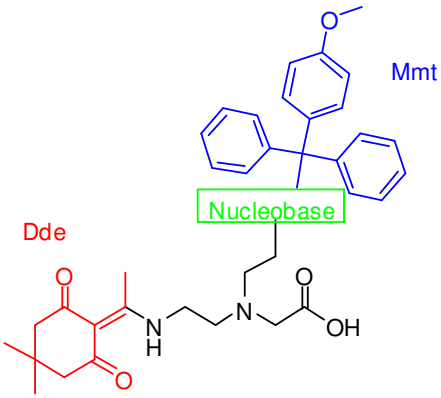


Scheme 1.1: General structure of DNA/PNA duplexes.

1.1.1 PNA synthesis

To prepare full length PNA, PNA monomers (Scheme 1.2) are typically coupled in the same manner as amino acids are in solid-phase peptide synthesis[3,4,5] (SPPS), here the carboxylic acid of the monomer is activated, followed by coupling to the free amino-group of the previous monomer in the growing chain.[6]

PNA is stable to weak bases as well as strong acids and a range of protecting groups has been used for PNA synthesis.[21] In the two most conventional strategies, the backbone-amino group is protected with either an Fmoc[7] or a Boc[8] group, and the nucleobase exocyclic amines are Bhoc- or Cbz-protected respectively. The Fmoc/Bhoc- and Boc/Cbz-protected PNA monomers are commercially available. Various other protecting groups for the backbone amino group have been reported, e.g. Mmt, 4-N_3-Cbz, and Aloc as well as protecting groups for the nucleobase exocyclic amines, e.g. Cbz, Bhoc, bis-Boc, and Mmt.[9,10,11,12] However, during PNA synthesis with the Boc/Cbz-protected PNA monomers repeated trifluoroacetic acid (TFA) mediated Boc-deprotections are required, as well as harsh acidic cleavage conditions (hydrofluoric acid or trifluoromethanesulfonic acid) for Cbz-deprotection. [13] Thus, the mild acid-labile Mmt group is preferred for protection of the nucleobase exocyclic amines.



PNA-monomer

Scheme 1.2: Dde/Mmt-protected PNA monomer.

1.1.2 Orthogonal peptide-PNA conjugate synthesis

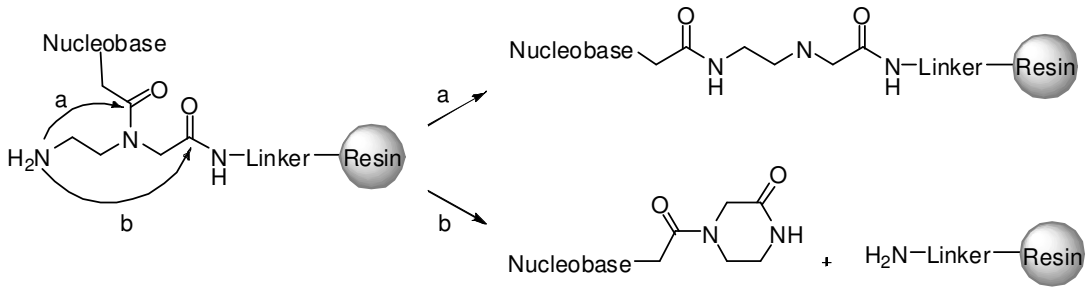
Neither the Fmoc/Mmt, Fmoc/Bhoc, nor Fmoc/bis-Boc-protected PNA monomers would allow orthogonal synthesis of peptide-PNA conjugates with Fmoc/*t*Bu-protected amino acids. This has lead to the development of a number of alternative protection strategies that allow orthogonal synthesis of peptide-PNA conjugates.[9,14] A very attractive alternative emerged with the synthesis of Dde/Mmt-protected PNA monomers as well as alternative Dde-deprotection conditions that are fully

orthogonal to the Fmoc group (Scheme 1.2).[14,15] Traditionally, Dde-deprotection is carried out by transenamination with hydrazine but these conditions also remove the Fmoc-group[16]. However, a mixture of hydroxylamine hydrochloride and imidazole removes the Dde-group but has no reactivity towards the Fmoc-group.[15] Furthermore, the Dde-group is stable to piperidine mediated Fmoc-deprotection as well as TFA. Other Fmoc-orthogonal protecting groups for the backbone-amino group, such as Aloc, Azoc, and 4-N_3Cbz have also been employed.[9] However, Aloc deprotection necessitates the repeated use of a palladium catalysts at neutral conditions, and afforded PNA-peptide conjugates in low yields and purities.[14] The Azoc and the 4-N_3Cbz protecting groups require neutral conditions using phosphines and phosphines followed by TFA respectively.[9] In addition, slightly basic or even neutral deprotection conditions may result in undesirable side reactions (section 1.1.2.1) and result in low yields and purities.[17]

1.1.2.1 Potential side reactions and complications in PNA synthesis

The increased nucleophilicity of the primary amines in PNA (compared to the primary amines in amino acids) can cause attack of the amine on either the nucleobase acetyl moiety[18,19,20] (acyl migration) or on the interunit amide moieties[13,20] (*N*-terminal detachment, Scheme 1.3). Dde protection of the monomers helps limit these potential side reactions, as Dde can be removed at slightly acidic pH, where the primary amine is protonated and non-nucleophillic.

Due to its bulk, the PNA oligomer may aggregate if synthesised on high-loading resins, which would result in incomplete coupling and deprotection.[6] Thus, synthesis of long PNAs (> 5 residues) restrains the choice of resin to those of low loading (0.05-0.1mmol/g).[6]



Scheme 1.3: Potential side reactions in PNA synthesis: acyl migration (a) and *N*-terminal detachment (b).

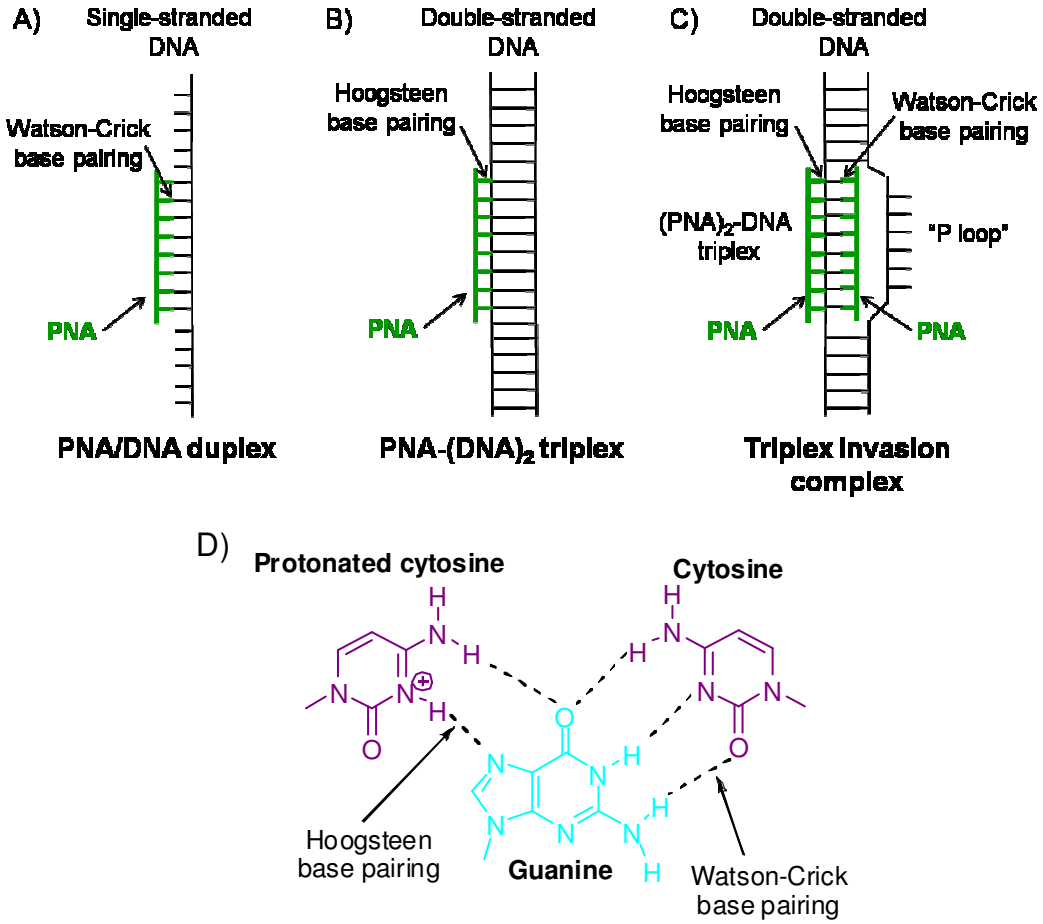
1.1.3 Properties of PNA

PNA hybridises to DNA, PNA, and RNA following standard Watson-Crick base paring[2]. Duplexes can form in both anti-parallel and parallel orientations but the anti-parallel orientation (*N*-terminus of PNA facing the 3'-end of DNA) is strongly preferred[21]. The backbone of PNA does not contain any anionic phosphate groups or other negatively charged groups meaning there is an absence of electrostatic repulsion during and after hybridisation with DNA, which means that the interaction between PNA/DNA is stronger than that between DNA/DNA strands and hybridisation can occur under low salt conditions.[6] The strong PNA/DNA interaction confers higher thermal stability on PNA/DNA duplexes than the corresponding DNA/DNA duplexes leading to significantly higher melting temperatures (T_m) as well as increased mismatch discrimination, with an average decrease of 15 °C in T_m for a single-base mismatch in a 15-mer PNA/DNA duplex, in comparison to a decrease of 11 °C in the T_m of the corresponding DNA/DNA duplex[2].

PNAs can also form PNA/DNA/DNA triplexes as well as other types of structures.[22,23] For example, PNA/DNA/DNA triplexes formed through Hoogsteen base-pairing, are observed between cytosine-rich PNAs and guanine/cytosine-rich DNA duplexes (Scheme 1.4).[24]

Under low salt concentrations double-stranded (ds) PNAs containing a high pyrimidine/purine ratio or homopyrimidine dsPNAs (composed of only thymine and cytosine bases) bind to purine-rich DNA duplexes by strand displacement to form a highly stable PNA/PNA/DNA triplex (Scheme 1.4).[1,25] In this conformation, one

PNA binds to the single DNA strand following Watson-Crick base-pairing rules (by preference in the anti-parallel mode), while the other PNA binds to the PNA/DNA duplex through Hoogsteen base-pairing (preferably in the parallel mode) displacing the other single-stranded (ss) DNA, which forms a P-loop structure (Scheme 1.4).[25] Cytosine mediated triplex invasion is pH-dependent as protonated cytosines are required for Hoogsteen base-pairing (Scheme 1.2d). The ability of PNAs to bind to dsDNA by strand displacement has been successfully used as a powerful tool in molecular biology[26] and functional genomics[27] (section 1.1.4.3 and 1.1.4.4).



Scheme 1.4: Schematic representations of (**A**) PNA/DNA duplexes, (**B**) PNA-$(DNA)_2$ triplexes and (**C**) triplex invasion complexes. (**D**) Hoogsten and Watson-Crick base pairing.

Furthermore, PNA is stable over a wide pH range[6] and is resistant to enzyme degradation because it is poorly recognised by either nucleases or proteases[6]. This resistance to enzymatic degradation extends the possible lifetime of PNAs *in vivo*[28].

However, PNA is not recognised by DNA polymerases and cannot function as a template or primer in the polymerase chain reaction (PCR).[6,21] These characteristics make PNA the molecule of choice for cell assays and DNA microarray screening[29].

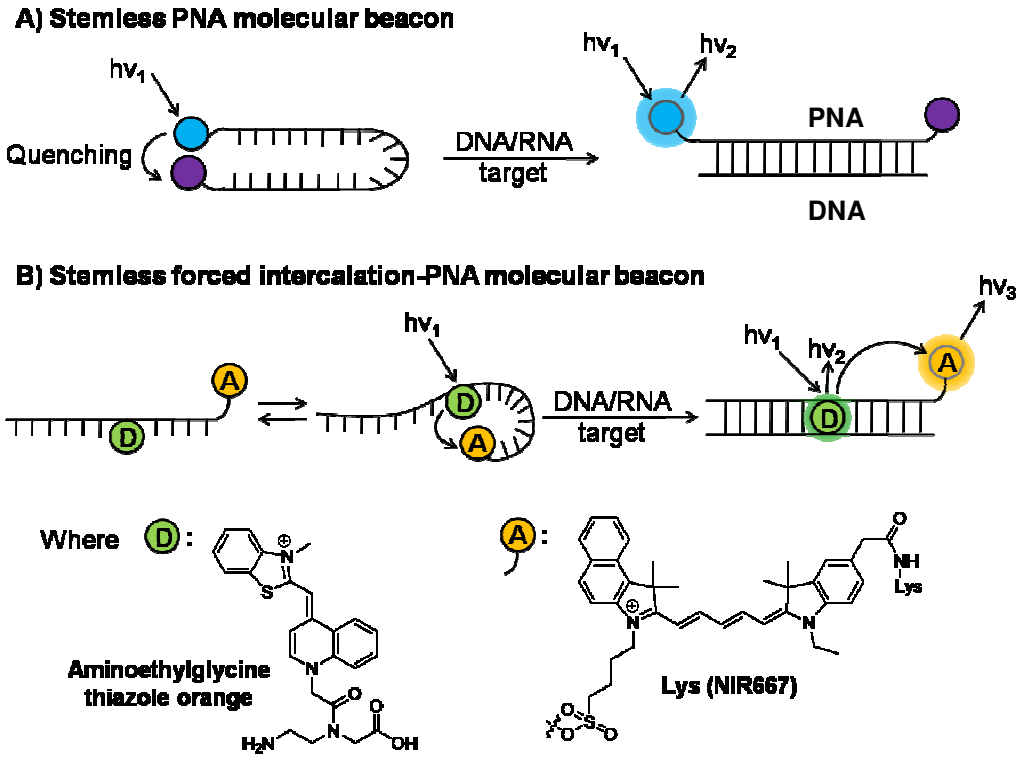
1.1.4 Applications of PNA

The unique characteristics of PNA have given PNA great potential as a tag, in diagnostics, functional genomics, and molecular biology.[21,30]

1.1.4.1 PNAs as a tag

Fluorescence resonance energy transfer[31,32,33] (FRET) in connection with nucleic acid tagging has been improved by the utilisation of PNA. In one example a PNA encoded peptide library is labelled with a FRET-pair and the conjugate library subjected to protease.[33] If the peptide is recognised by the protease it will be degraded, while the PNA remains intact, leading to cleavage of the quencher from the conjugate and fluorescence from the donor. The PNA may then be hybridised to a complementary microarray revealing the sequence of the peptide.

Another PNA-tagging method uses PNA based molecular beacons, which are self-complementary PNAs labelled with both a fluorophore and a quencher.[34] In the absence of a complementary DNA the PNA strand self-hybridises placing the donor and the acceptor in close proximity such that fluorescence is quenched. When hybridised to a complementary DNA or RNA strand, the fluorophore and quencher distance is increased leading to loss of the quench (Scheme 1.5a). This technique has lead to the development of forced intercalation-PNA molecular beacons, which enhance fluorescence rather than quench it while exhibiting increased mismatch discrimination, even at low temperatures. These PNAs have an open conformation and are coupled to an intercalator dye, *e.g.* thiazole orange (donor) and a near infrared dye, *e.g.* NIR667 (acceptor).[35] The PNA is designed such that upon hybridisation with a fully complementary nucleic acid sequence the intercalator dye becomes fluorescent and its fluorescence is subsequently quenched by the near-infrared dye, which emits at a different wavelength (Scheme 1.5b).

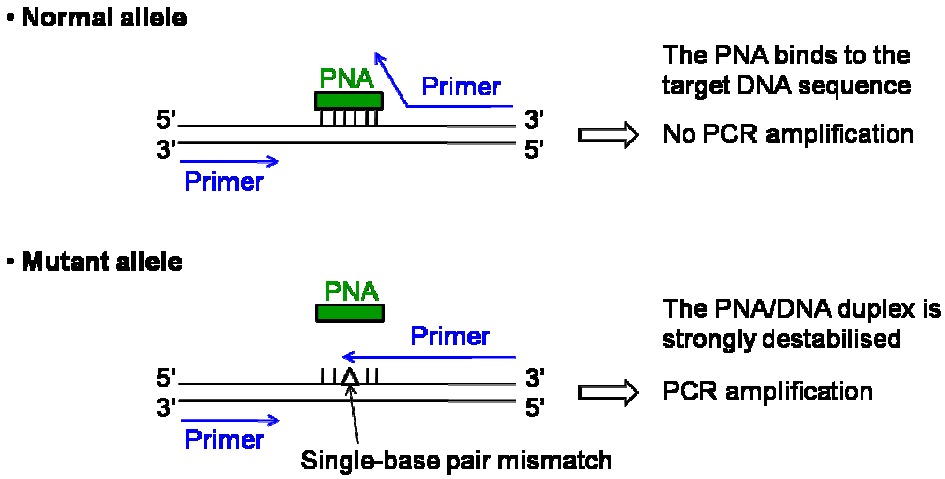


Scheme 1.5: Detection of complementary target nucleic acids using (**A**) PNA molecular beacons and (**B**) forced intercalation-PNA molecular beacons.[35]

1.1.4.2 PNA in diagnostics

PNAs unique hybridisation properties have led to the development of diagnostic assays, such as PNA-fluorescent *in situ* hybridisation (FISH), whereby fluorescently labelled PNA probes are employed to image nucleic acid targets inside cells and tissue.[36] For example, the accurate telomere lengths of chromosomes have been determined using PNA-FISH, and PNA-FISH has applications in cancer and ageing studies.[37]

The detection of single base-pair mutations or single-nucleotide polymorphisms (SNPs) using PNA-directed PCR clamping hold great promise in diagnosis and research of human genetic disease.[38,39] In the PCR clamping technique, one primer site (and hence PCR) is blocked by a fully complementary PNA. Thus a single base-pair mismatch presented by a genetic disease in the target DNA sequence destabilises the PNA/DNA duplex and allows PCR to occur (Scheme 1.6).



Scheme 1.6: The PNA-directed PCR clamping method.[38]

1.1.4.3 Applications of PNA in functional genomics

Southern blotting, which reveals the size and sequence information of DNA, has been improved by the utilisation of PNA.[40] Hereby, biotinylated PNA is hybridised to denatured DNA at low salt concentrations disfavouring formation of DNA/DNA duplexes, followed by size separation of ssDNA by gel electrophoresis. The PNA/DNA duplexes may then be detected on a nylon membrane using chemiluminescence-based methods.

1.1.4.4 PNA in molecular biology

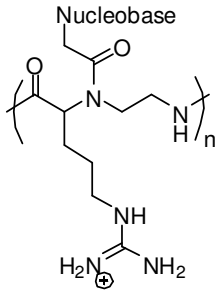
PNA has been shown to have applications in antigene and antisense interference by strand invasion of dsDNA[41], or by binding to messenger RNA[42] (mRNA) respectively. This has lead to advances in therapeutic areas such as cancers[43], and bacterial[44] and viral[45,46] infections. The advantage of employing PNA rather than the natural nucleic acids in antigene and antisense technology is that PNA is not degraded in the cytosol by nucleases in the same way as the natural nucleic acids[47].

As an antigene, PNA can inhibit transcription by strand-invasion of chromosomal-DNA forming triple-helixes, which sterically hinders and blocks transcription.[48] For example, PNA targeting the transcription-initiation sites for the human progesterone receptors A and B, have been reported to inhibit formation of the transcription

complex.[49] Chromosomal-DNA replication has also been inhibited by PNAs by formation of triple helixes[50], and plasmid-DNA replication can be inhibited in the same manner *in vivo*.[51]

PNAs have also been reported in antisense interference by hybridising to complementary mRNA to form triple helixes, which sterically hinder and block translation of the target genes.[42] In addition, antisense PNAs have been shown to inhibit microRNAs.[52]

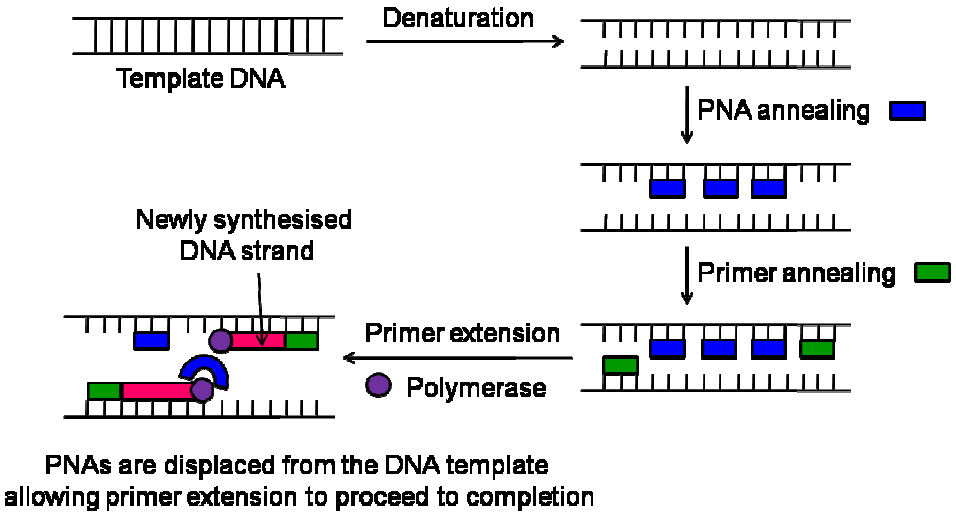
The main complication of the application of PNA in antigene and antisense research is the poor cellular penetration of PNA due to its high molecular weight and its lack of ability to function as a carrier system or interact with cell-surface receptors.[53] Moreover, traditional transfection agents developed for ionic nucleic acids (*e.g.* polymers and cationic lipids) are often incapable of delivering the uncharged PNA. Cellular delivery of PNA by other methods, such as conjugation to steroids[54] or cell-penetrating peptides[55,56] (*e.g.* arginine-rich peptides, Penetratin), alteration of the structure of PNA to increase its cell-permeability[57,58], co-transfection with DNA[59], and electroporation[60] has been reported. For example, the Tat_{48-57} decapeptide or a hepta-arginine peptide delivered conjugated PNA with a fluorescein-label into cells with high efficiency.[55] Similarly, incorporation of D-arginine residues into the PNA backbone instead of glycine side chains produced a cell-permeable PNA analogue, guanidine-PNA (GPNA), which inhibited protein translation (Scheme 1.7).[57] PNA has also been conjugated to a triphenylphosphonium carbamate, which functions as a prodrug such that it improves the cell permeability of the construct but is degraded in the cytosol by glutathione reductase leading to the release of the unmodified PNA contruct.[58]



Scheme 1.7: Structure of a GPNA-monomer.[56]

PNA has also been employed in advances in nucleic acid purification, whereby duplex formation between hexahistidine-PNA conjugate and complementary DNA in combination with nickel affinity chromatography has been reported as a nucleic acid purification strategy.[61]

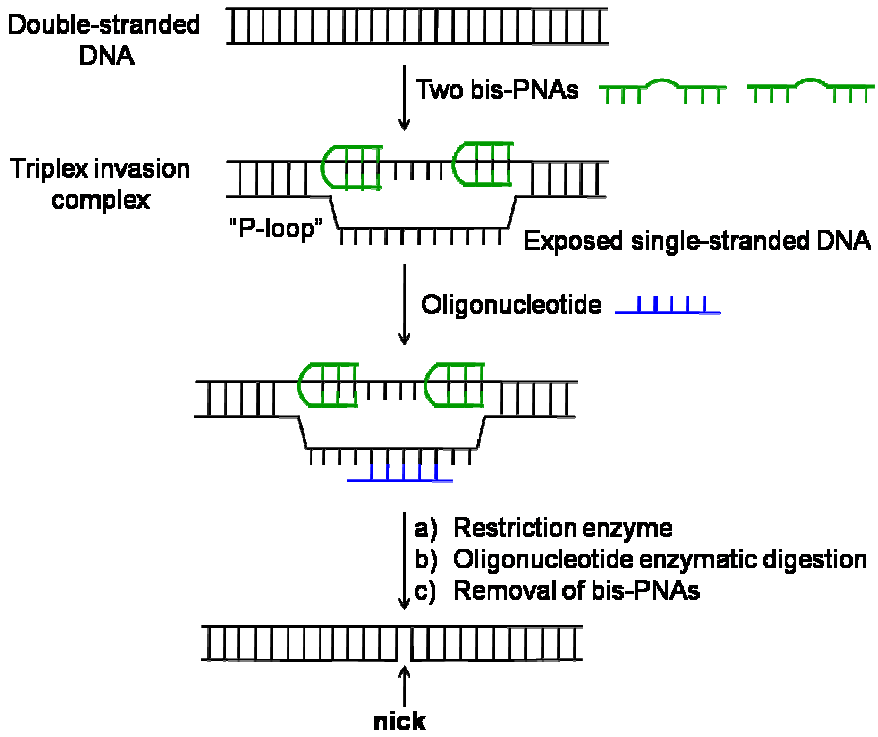
Genetic typing often relies on PCR amplification of variable numbers of tandem repeats. However, in this technique, small allelic products of heterozygous individuals are preferentially amplified compared to large alleles, which can lead to erroneous genotyping patterns.[62] PNA hybridised to the DNA template during PCR has been demonstrated to render the template unavailable for inter-strand and intra-strand interactions but not inhibit the primer extension as the polymerase displaces the PNA leading to reliable PCR amplification of different size alleles (Scheme 1.8).[62]



Scheme 1.8: Genotyping PCR using PNA for reliable amplification of different size alleles.[62]

PNA has also been applied to enhance the specificity of restriction enzyme digestions of genomic DNA.[26] A short bis-PNA (Scheme 1.9) is hybridised to a specific site of complementary DNA, where after DNA not protected by PNA is methylated. Following PNA dehybridisation, the PNA complementary DNA is digested by a methylase-sensitive restriction enzyme, allowing the PNA to function as a sequence-directing genome cutter. Bis-PNA mediated dsDNA strand invasion

has been exploited in development of a method for constructing DNA nickases.[27] Hereby, bis-PNA is hybridised to complementary DNA by strand-invasion exposing a ssDNA loop, which is hybridised with a complementary oligonucleotide. The resulting dsDNA is digested by a restriction enzyme, the dsDNA is degraded, and the bis-PNA is removed allowing the DNA duplex to rehybridise (albeit with a nick, Scheme 1.9).



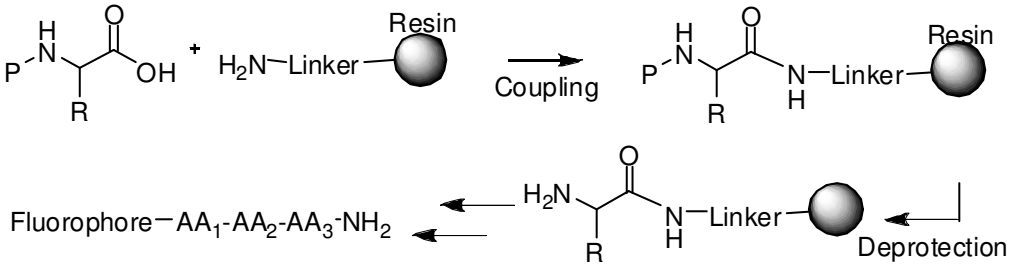
Scheme 1.9: Construction of site-directed DNA nickases using PNA mediated strand invasion.

PNA-FISH is also employed in commercial diagnosis systems for bacterial infections (AdvanDx, Inc.Woburn, MA, USA),[63] where a fluorescently labelled complementary PNA probe is in-situ hybridised to species-specific ribosomal RNA (rRNA) of the target bacteria. This allows direct identification of harmful bacteria in whole blood and other body secretes using fluorescent microscopy.

1.2 Solid-phase peptide synthesis

Numerous advances have been made in the formation of amide bonds; and one of the most prominent was the development of SPPS by Merrifield 1963 (Scheme 1.10).[5] In

addition to the synthesis of common peptides, SPPS allows for synthesis of peptide derivatives including the coupling of unnatural amino acids and peptides with backbone modifications. Here solid-supported synthesis will be discussed with a focus on peptides, PNA, and poly-*N*-alkylglycines.



Scheme 1.10: SPPS of peptides. R is a side chain of any amino acid. P is a protecting group. AA is any amino acid.

In SPPS an insoluble gel based resin is coupled to a linker, on which the peptide chain is elongated by repeated cycles of coupling and deprotection and the peptide is cleaved off the solid support at the end of the synthesis. The strength of the SPPS strategy is the ease of purification (simply washing the supported peptide) and an opportunity for achieving high yields over numerous steps. However, homogenous products can only be achieved with high conversion in each step (>99%). A potential downside to SPPS is the requirement to use high concentrations and excess reagents (often between 2-10 equiv.) to force reactions to completion.

1.2.1.1 Parallel SPPS methods

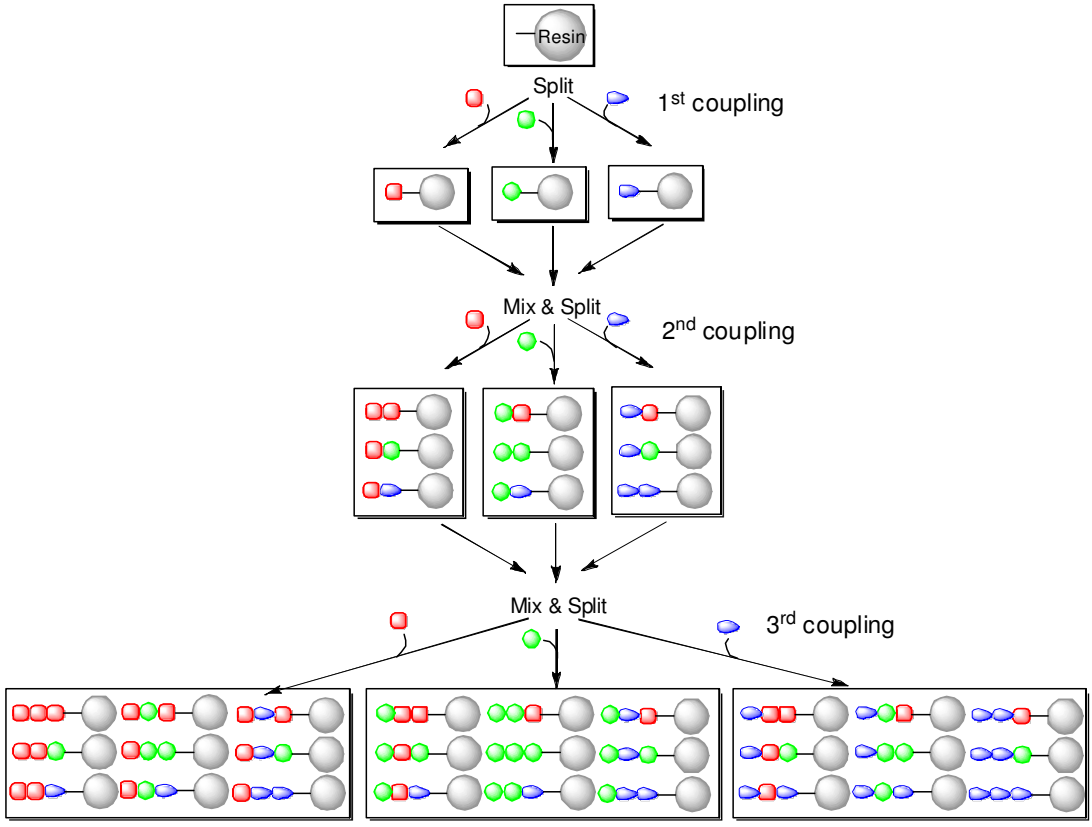
Geysen *et al.* (1984) developed SPPS into a parallel synthesis method using the so-called multipin method for antibody epitope mapping.[64] Synthesis of hundreds of individual peptides supported on polyacrylic acid-grafted polyethylene pins, which were arrayed in a 96 well-plate, was achieved. Coupling, washing, and neutralisation steps were carried out by immersing the pins into specific solutions. The use of a range of grafts (such as polyacrylamide and polystyrene) and linkers were reported. The main downside to this effective technique was the small amounts of material obtained compared to standard solid supports.

The so-called "tea-bag" method, developed by Houghten (1985), is another parallel approach to SPPS, where functionalised resin beads are sealed into small polypropylene mesh packets (like tea-bags).[65] Individual peptide synthesis is then carried out by immersing the "tea-bags" into separate specific amino acid coupling solutions producing yields and purities comparable to traditional SPPS techniques. Based on the "tea-bag" approach other molecular resin-based systems have been designed. For example, using Irori resin carriers compound libraries were synthesised in polypropylene "mesh" microreactors using optical[66] or radiofrequency[67] encoding.

Parallel synthesis of peptides on cellulose membranes, termed SPOT synthesis, was reported by Frank (1992).[68] Using this technique activated Fmoc-protected amino acids were coupled at specific locations on amino functionalised membranes by manual or automatic dispension. Thereafter, unreacted amino groups could be capped by acetylation and the peptides released from the support by employing a cleavable linker on the derivatised membrane. This technique is cost effective and uses only small amounts of reagents, which gives the method many applications in high-throughput screening and in molecular biology, such as antibody or enzyme substrate screening.

A great expansion of parallel SPPS was achieved with the emergence of the split and mix method first reported by Furka *et al.* (1988) with the synthesis of equimolar peptide library mixtures.[69,70] By this method, the solid-support is divided in separate fractions and one or more reactions are carried out on each separate fraction. This is followed by mixing of all the reaction products in one pool, then division and another set of separate reactions are carried out and the process is repeated as many times as desired. Different building blocks can be employed in each step, leading to a great variety and number of products (Scheme 1.11). The maximum number of products that can be achieved by split and mix synthesis is: the number of building blocks in each step raised to the power of the number of steps (reaction cycles). Thus, a pentapeptide library with the 20 natural amino acids as building blocks would generate 3,200,000 (20^5) different compounds with just 100 coupling reactions. If

there is no competition between the introduced reactants, each library member will be synthesised in approximately equimolar quantities in the final mixture. However, split and mix reactions often produce product mixtures, which necessitates deconvolution or encoding strategies to identify individual products. The technology has been used to produce so-called "one-bead-one-compound" libraries.[71]



Scheme 1.11: Schematic representation of split and mix synthesis.

1.2.2 Encoding strategies

Numerous chemical and non-chemical encoding strategies for combinatorial libraries have been developed and the description of all of these lies beyond the scope of this thesis. As such the discussion of encoding strategies will be limited to DNA- and PNA-encoding.

1.2.2.1 DNA-encoding

There are many examples of the application of DNA libraries being used as an encoding tag for peptides or small molecules[72-80], enabling the high-throughput screening of peptide interactions with biological targets or screening of drug candidates. The concept of DNA-encoding was first demonstrated by Brenner (1992), who reported split-and-mix synthesis of a combinatorial DNA-encoded peptide library by alternating amino acid and nucleic acid couplings[72]. This library was employed in ligand/protein affinity studies and the DNA-encoding allowed PCR-amplification followed by DNA-sequencing in order to identify the active peptides.

Affymax and Nielsen simultaneously reported the bead supported synthesis of DNA-encoded libraries composed of up to 823,543 heptapeptides[73,74]. This synthesis approach has been employed in other libraries[76].

DNA encoded self-assembling chemical (ESAC) libraries have been reported as an efficient tool to identify molecules that bind target proteins.[77] This technique allows hybridisation of specific small molecule-linked DNA oligonucleotides from different sub-libraries to generate DNA-duplexes synergistically encoding two compounds leading to large libraries with diverse chemical groups that can be screened against biological targets.

Other examples of DNA-encoded libraries involving enzymatic DNA ligation have been patented by Nuevolution[78] and Praecis[79]. Recently, synthesis of million to billion member small molecule libraries encoded by double stranded DNA were reported by combination of enzymatic and chemical synthesis in a split-and-pool format, allowing screening against two kinase targets.[80]

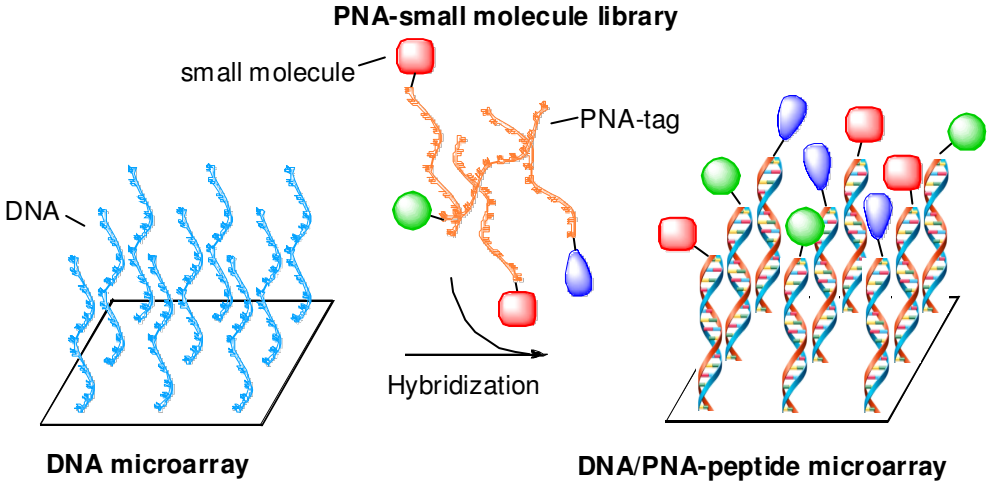
DNA barcoding is a taxonomic method for rapid species identification based on DNA sequences[81,82,83], whereby a short genetic marker (a barcode, 400-800 bp) in an organism's DNA is used to identify the respective organism as belonging to a

particular species, such that the information of one or a few gene regions can be used to identify all species of life.

1.2.2.2 PNA-encoding

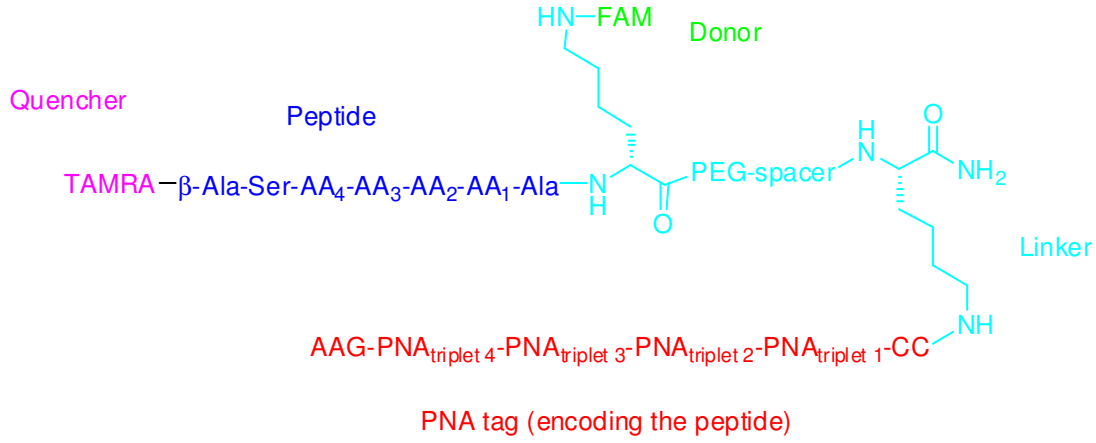
The greater chemical stability of PNA, its resistance to biological degradation and strong binding affinity for DNA (section 1.1.3) renders PNA a superior encoding alternative to DNA.[33,84-88] PNA-encoding has been employed in small molecule libraries, where each molecule is linked to a unique PNA tag, which allows identification of the small molecules by hybridisation of the PNA-tag on a predefined microarray location.[85] Furthermore, it has been shown that PNA does not alter or otherwise interfere with the interaction between the small molecules and a protease.

PNA-encoded peptide-libraries have been employed to identify optimal substrates and inhibitors for proteases. In one example, following incubation with cathepsin the enzyme bound PNAs were purified by size exclusion followed by DNA microarray hybridisation to decode the small molecule structures.[86,88] The approach was further developed to study cysteine protease activity on complex mixtures of proteins from crude cell lysates and led to the identification of human caspase inhibitors.[86]



Scheme 1.12: Schematic representation of DNA microarray hybridisation of a PNA-encoded small molecule library.

PNA-encoded split-and-mix peptide libraries have been employed to study the substrate specificity of proteases as well as tyrosine kinases.[33,87] This approach included the synthesis of a FRET-based 10,000-member PNA-encoded peptide library labelled with 5(6)-carboxyfluorescein (FAM) and tetramethyl-6-carboxyrhodamine (TAMRA, Scheme 1.13).[33] Upon enzymatic cleavage of the peptide (B-Ala-Ser-AA_4-AA_3-AA_2-AA_1-Ala, Scheme 1.13), the quencher (TAMRA) was removed from its close proximity to the donor (FAM, Scheme 1.13) allowing FAM fluorescence to be visualised. After enzymatic incubation the PNA tags were hybridised onto a complementary DNA microarray, which allowed not only the identification of peptides recognised by the protease but also quantification of the relative degree of cleavage by calculation of the FAM//TAMRA intensity ratio at each spot.



Scheme 1.13: Structure of the FRET-labelled PNA-encoded peptide library.

1.2.3 The combinatorial chemistry approach adopted in this project

This project presents: 1) Screening of a combinatorial library of potential delivery agents and extraction of those with enhanced cellular uptake. 2) Screening of a combinatorial library of potential peptide ligands for cell surface receptors controlling cell signalling pathways and thereby moderating cell behaviour from the outside of the cell.

The libraries herein were synthesised by split-and-mix combinatorial chemistry with PNA tags along with each amino acid or peptoid monomer coupling. Thus each peptide was covalently linked to a unique PNA tag much like a barcode, from which the peptide identity could be derived by hybridising the PNA-tags to a complementary DNA microarray (section 1.3). This one-spot-one-compound approach circumvents laborious structure determination after library screening by allowing the PNA tag to be read like a barcode. In addition, this approach allows the target biological assays to be carried out before or after hybridisation of the library to the microarray.

1.3 DNA microarrays

DNA microarrays are comprised of DNA oligonucleotides chemically attached to a solid surface, usually a glass slide, in a predefined location. DNA microarrays were first developed to assay global gene expression.[89-92] Microarray technology established a method for the simultaneous study of the relative expression of thousands of genes in a single experiment.

The advantage of microarrays allows opportunities for the high throughput screening of tens of thousand of library members at once. This permits easy identification of "hits" as well as multiplexed analysis of positive and negative responses. Importantly, the multiple assays that one microarray experiment comprises are carried out under the same reaction conditions, often with internal controls. This allows the discovery of unique patterns in addition to mere identification of positive responds. However, this multidimensional analysis produces a vast amount of data, which poses a number of statistical problems, including evaluation of the standard deviations, as well as the normalisation and standardisation of the data.

Other advantages of microarray technology are the use of microliter reaction volumes and nano- and picomolescale amount of samples required. Microarray technology has been applied in the high-throughput screening of a variety of biomolecules, such as DNA[90], peptides[93], proteins[94], live cells[95], tissue[96],

carbohydrates[97], polymers[98], and small molecules[99]. Microarrays have been established as a vital tool in many research areas including proteomics, genetic screening, diagnostics, and drug discovery. [100]

DNA microarrays are classically applied in gene expression profiling, where the presence and relative amount of a gene in a control and a sample of interest is identified.[101] This method allows rapid identification of specific genes expressed at high levels in a disease state.

DNA microarrays are also employed in genome-wide genotyping as well as detection of mutations or SNPs, identifying susceptibility to development of a specific disease.[102]

Microarray technology has led to the development of the so-called "comparative genomic hybridisation" technique, where two sets of genomes are compared in a single experiment to determine if expression of specific genes has been up or down regulated establishing a powerful diagnostic tool.[103]

1.3.1 DNA microarray fabrication

DNA microarrays can be custom synthesised to contain high numbers (up to 176,000; Agilent) of relatively long (up to 200 bp) DNA oligonucleotides.[92] DNA microarrays are typically prepared by *in-situ* DNA synthesis either by: photolithography, where photolithography-masks (real or virtual) are applied to direct oligonucleotide synthesis[104-107]; by inkjet printing mediated synthesis[108,109]; by semiconductor directed synthesis, where an array of individually controlled microelectrodes embedded in a fluidic chamber selectively generate chemical reagents by means of an electrochemical reaction[110]; or by laser facilitated synthesis using oligonucleotides with photolabile protecting groups[111].

The spotting of pre-synthesised DNA onto a surface, such as a bead (e.g. Illumina) or a glass surface is more expensive and laborious than *in-situ* DNA array synthesis.

However, this approach does allow DNA sequences to be purified and characterised prior to printing, thereby eliminating base deleted oligonulceotides which arise during *in situ* synthesis.

1.3.2 Surfaces and immobilisation

The surface material and chemistry should be optimised for immobilisation of DNA probes, but resist unspecific association of target molecules. Moreover, the surface may also influence the density of probe molecules, the reaction efficiency of biological assays and the data quality, as the background must be compatible with the detection method. Most microarrays are comprised of modified glass slides as these are flat, planar, resistant to high temperatures, easy to handle and are transparent affording low light absorption.[112] Classical, commercially available surface modifications are amine, aldehyde, epoxide, and activated ester. These modifications can be achieved by reacting a glass surface with the desired functional group displayed on the end of propyltrimethoxysilane.[100]

Biomolecules may be immobilised on microarrays *via* covalent bonds or *via* adsorption, which are non-covalent interactions between the biomolecules and the slide surface. For example, the attachment of amino-modified DNA to a succinimidyl ester surface forms a covalent amide bond. An example of absorption is DNA immobilisation on an amine surface through electrostatic interactions between the amine groups and the phosphate groups of DNA. However, this technique results in a probe orientation parallel to the surface rather than an upright orientation, which affords higher hybridisation efficiency.[112]

1.4 Cell surface receptors

Cell surface receptors are glycoproteins embedded in the cell membrane of all mammalian cells and they have specific ligand-binding moieties exposed to the extracellular matrix (ECM)[113]. Following binding of a ligand to a cell surface receptor, a biological signal is propagated across the membrane to the intracellular matrix mediating cellular responses such as differentiation, proliferation, and

migration etc[113]. There are hundreds of known cell surface receptors and discussion of all of these lies beyond the scope of this thesis. Focus will be on two of the largest and most diverse families of cell surface receptor: integrins and G-protein coupled receptors (GPCRs).[114,115,116]

1.4.1 Integrins

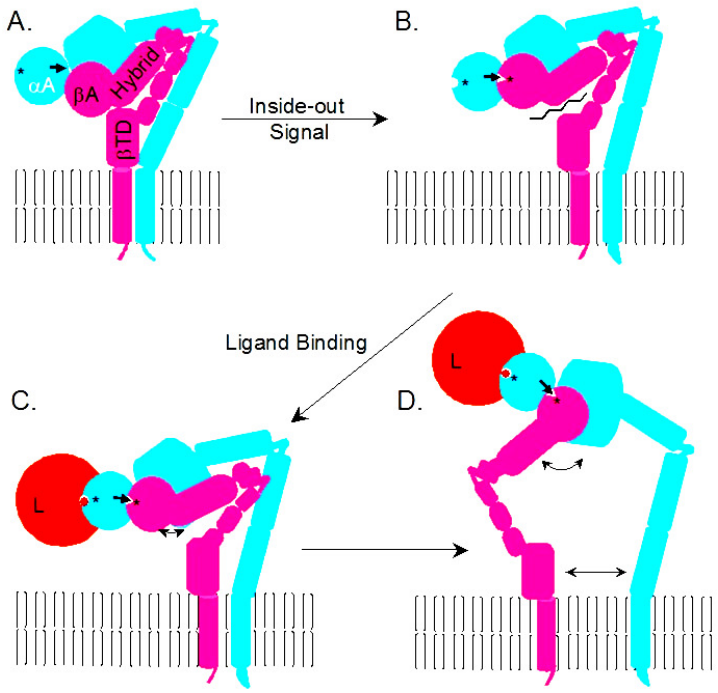
Integrins are a large family of heterodimeric transmembrane receptors that govern ECM and cell-cell adhesion.[114] In mammals 18 α and 8 β-integrin genes encode proteins that combine and form 24 αβ receptors[117]. The extracellular domains of these receptors interact with a range of ECM proteins and cell surface proteins (e.g. fibronectin, collagen, laminin, and vitronectin; Table 1.1) while the cytoplasmic domain is linked to the actin-cytoskeleton.[118,119]

Table 1.1: Integrins found in mammalian cells.[120]

Name	Synonyms	Distribution	Ligands
$\alpha_1\beta_1$		many	collagens, laminins
$\alpha_2\beta_1$		many	collagens, laminins
$\alpha_4\beta_1$	VLA-4	hematopoietic cells	fibronectin, VCAM-1
$\alpha_5\beta_1$	fibronectin receptor	widespread	fibronectin and proteinases
$\alpha_6\beta_1$	laminin receptor	widespread	matrix macromolecules laminins
$\alpha_L\beta_2$	LFA-1	T-lymphocytes	ICAM-1, ICAM-2
$\alpha_M\beta_2$	Mac-1, CR3	neutrophils and monocytes	serum proteins, ICAM-1
$\alpha_{IIb}\beta_3$		platelets	fibrinigen, fibronectin
$\alpha_V\beta_3$	vitronectin receptor	activated endothelial cells, melanoma, glioblastoma	vitronectin,[7] fibronectin, fibrinogen, osteopontin, Cyr61
$\alpha_V\beta_5$		widespread, esp. fibroblasts, epithelial cells	vitronectin and adenovirus
$\alpha_V\beta_6$		proliferating epithelia, esp. lung and liver	fibronectin; TGFβ1+3
$\alpha_6\beta_4$		epithelial cells	laminin

Most integrins are Ca^{2+} or Mg^{2+} dependent, i.e. only when an extracellular Ca^{2+} or Mg^{2+} ion is bound to the receptor, will it adopt the correct conformation for ligand binding. Common peptidic features found in many ligands including vitronectin, laminin, and fibronection for a range of integrins are Arg-Gly-Asp (RGD) and to a lesser extent Leu-Asp-Val (LDV), which have long been exploited in designing probes and therapeutics for modulating specific cellular function.[117]

Integrins are activated by conformational changes induced either by ligands binding to their extracellular moiety or by conformational changes of their cytoplasmic domain[121]. This latter form of signalling, termed "inside-out-signalling", ensures rapid and flexible responses as it enables the cell to change the receptor conformation from within the cell thereby allowing binding or release of ligands (Scheme 1.14).



Scheme 1.14: Schematic representation of the conformational changes during integrin activation and signalling. (**A**) The low affinity leg-bent structure. (**B**) Inside-out activation disrupts the βTD/βA complex, allowing stable αA/βA complex formation and thereby stabilising αA in the open, high affinity state. (**C**, **D**) Binding of ligand (L, red) to αA stabilises the occupancy of βA, which opens up the βA/hybrid hinge and forces genuextension* and leg separation resulting in outside-in signalling.[122]

Through the cell surface moiety, integrins attach the cell to the ECM through multiple receptors with low binding affinity: a type of binding, which ensures that the cell does not become irreversibly attached to the ECM and retains highly controlled and fast responses as well as the ability to migrate. Thus, integrins comprise a main mode of responce to the ECM and play an important role in many

* Comes from the Integrin's ability to bend and stretch like a knee (Genu).

cellular processes including adhesion, definition of cellular shape, cell mobility, endocytosis, and signal transduction both from the ECM to the cell and from the cytoplasm through the receptor to the ECM.[114-117]

The cell signalling pathways that integrins govern, amongst others, control cell growth, cell division, apoptosis, proliferation, migration, differentiation as well as viral binding and entry into the cell.[117,123] These processes, if not tightly regulated, can lead to down-regulation of vital genes, which is the cause of a variety of diseases including cancer and thrombosis.

1.4.2 Syndecans

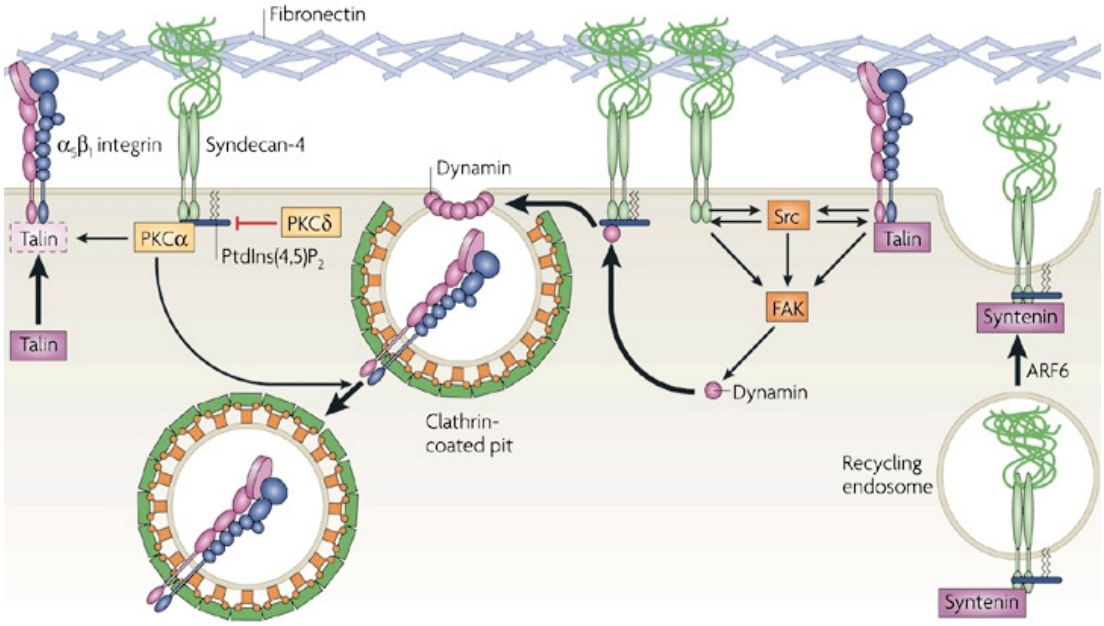
Syndecans act as co-receptors especially for integrins and GPCRs, whereby they modulate the ligand-dependent activation of primary cell surface signalling receptors.[124] Syndecans are thus involved in the same cellular functions as primary receptors they cooperate with, such as in signal transduction, cell proliferation, migration, and ECM and cell-cell adhesion. They consist of a single hydrophobic transmembrane domain, a short C-terminal cytoplasmic domain that interacts with the actin-cytoskeleton, and an extracellular domain comprising a protein core covalently attached to chodroitin sulphate chains and 3-5 heparin sulphate chains[125,126]. These chains interact with various ligands such as fibronectin, fibroblast growth factors, antithrombin-1, transforming growth factor-beta, and vascular endothelial growth factor. In addition, the core proteins themselves can act as integrin ligands.[126] Mammals express four syndecans, three of which (syndecan 1-3) have restricted tissue distribution and syndecan-4 is expressed ubiquitously in all cell types.[124]

1.4.2.1 Integrin and syndecan synergy

The syndecan co-receptor function is maintained by core proteins, which target the large heparin sulphate chains to the appropriate plasma membrane compartment, which makes them ideal receptors for dilute ligands distant from the membrane.[127] However, much of the mechanism of the synergistic action of syndecans and

integrins remains to be elucidated.[127] Links between integrin and syndecan-4 indicate a pathway via active PKCα, which activates integrin promotion of migration of cells on ECM substrates (e.g. fibronectin, collagen and laminin)[128,129].

There is also evidence that this is a highly specific system whereby the integrin and syndecan synergy vary depending on the integrin that is involved; e.g. $\alpha_5\beta_1$ integrin-mediated spreading of myoblasts on fibronectin is PKC-dependent, whereas $\alpha_7\beta_1$ integrin-mediated spreading on laminin is PKC-independent[127].



Scheme 1.15: Integrin and syndecan-4 synergy involving protein kinase signalling and regulates endocytosis. From left: Syndecan-4 binds protein kinase Cα (PKCα) and phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P2), which is prevented by phosphorylation of syndecan-4 by PKCδ[130]. PKCα activity and binding of talin is necessary for integrin activation[129]. From right: Endocytosis and recycling of syndecan-4 is dependent on ARF6 regulated association of the cytoplasmic domain of syndecan and one of the PDZ domains of syntenin (scaffolding protein)[131]. $\alpha_5\beta_1$-integrin and syndecan-4 binding of fibronectin activates the Tyr kinase, focal adhesion kinase (FAK) and the Src kinases, which also contribute to FAK activation. By enabling dynamin binding to the cytoplasmic domain of syndecan-4, FAK regulates focal-adhesion disassembly[132], which activates integrin endocytosis. Thin arrows = indirect signalling links. Thick arrows = translocation. Fibronectin represents the ECM.[127]

1.4.3 G-protein coupled receptors

GPCRs are transmembrane receptors associated with a G-protein and consist of

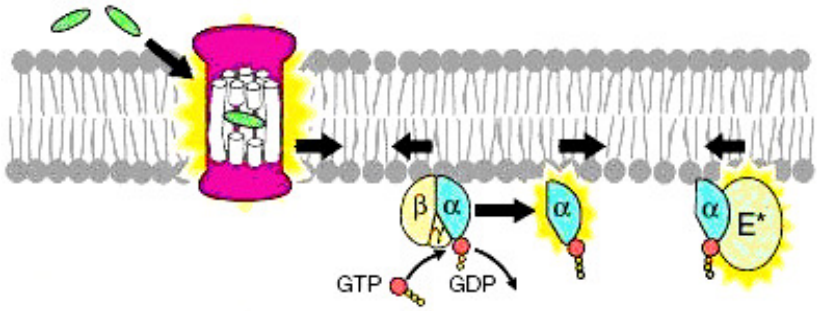
different subunits dependent of specific protein function[133]. GPCRs have an extracellular N-terminus and an intracellular C-terminus combinded by 7 transmembrane domains (helices) that form a barrel cavity, which constitute the binding site for smaller ligands. GPCRs also have 3 extracellular loops, where larger ligands bind as well as 3 intracellular loops that are associated with a G-protein. Ligands interaction with protease-activated GPCRs differ from other GPCRs in that these are activated by cleavage of part of the extracellular domain

Ligand interaction with the GPCRs causes conformational change in the receptor, causing the receptor to act as a guanine nucleotide exchange factor. This allows the GPCR to activate an associated G-protein by catalysing exchange of its bound guanosine diphosphate (GDP) for guanosine triphosphate (GTP). Hereby, the G-protein's α subunit dissociates along with the GTP, which will further affect intracellular signaling proteins or target functional proteins directly depending on the α subunit type (Scheme 1.16).[133] A single GPCR can activate many molecules of G-protein because it is released unaltered after the interaction.

GPCRs bind an array of ligands from small peptides to large proteins reflecting their wide range of functions. In fact, there are approximately 750 GPCR genes in the human genome comprising 3% of the entire genome. Although, many GPCRs are orphan[†] the GPCR used in this study, the C-C chemokine receptor type 6 (CCR6)[134], has a well-known natural ligand (CCL20 also know as MIP-3α).[135,136] GPCRs are also in synergy with the co-receptor, the syndecans.[124,125]

Through signalling pathways GPCRs govern some of the most vital cellular processes including: binding of neurotransmitters regulating the nervous system; controlling blood pressure; heart rate; digestive processes; binding of chemokines regulating inflammatory responses; binding of opsins (visual sense), odorants and pheromones (sense of smell); as well as regulating cell density sensing.[115] These essential biological functions make GPCRs amongst the most heavily investigated drug targets in the pharmaceutical industry[115,137].

[†] Orphan receptors are receptors for which no ligand has been identified.



Scheme 1.16: Schematic representation of the GPCR signal transduction cascade. Upon ligand binding the receptor (pink) is activated, which causes activation a coupled G-protein by catalysing the exchange of GDP for GTP. Hereby, Gα-GTP is released and can bind to an effector protein E, causing it to switch states to an activated form, E^*.[138]

Thus, research into selective peptide-ligands for GPCRs, as well as syndecans and integrins, could lead to: advances in understanding of cell surface receptors and provide a means for directing their actions; the identification of new ligands; and possibly improve the design of drug targets for these receptors.

1.4.3.1 Chemokine receptor CCR6

There are 19 chemokine receptors that bind chemokines, which are small secreted proteins that regulate leukocyte trafficking, a process that is essential for lymphoid organogenesis and inflammatory responses, and host defence. Chemokine receptors contain two cysteines in the N-terminal domain, which do not form disulfide bridges (except CCR6 which only contain one cysteine).[136] All chemokine receptors also contain cysteines in all three extracellular loops, which form disulfide bridges and are essential for receptor function.[136]

The natural CCR6 ligand, CCL20 is expressed in lymphoid, liver, and lung tissue as well as on activated monocytes.[134,135] The CCR6 receptor is expressed on lung memory T and B cells and immature dendritic cells in the lymphoid tissue and CCR6 plays a role in inflammatory responses in the skin and gut.[134,135] Besides, CCL20 the CCR6 receptor also binds defensins, which are small cysteine rich cationic peptides

expressed in immune cells, where they assist in killing phogocytosed bacterias by binding to the bacteria membrane leading to pore formation.[136]

1.5 Endocytosis in mammalian cells

Endocytosis is the process by which cells take up material from the ECM by engulfing it with their membrane.[139,140,141] This technique is used by almost all eukaryotic cells to transport large polar molecules, which cannot pass the hydrophobic plasma membrane. Endocytosis happens by one of three processes: phagocytosis, pinocytosis and receptor mediated endocytosis (Scheme 1.17).

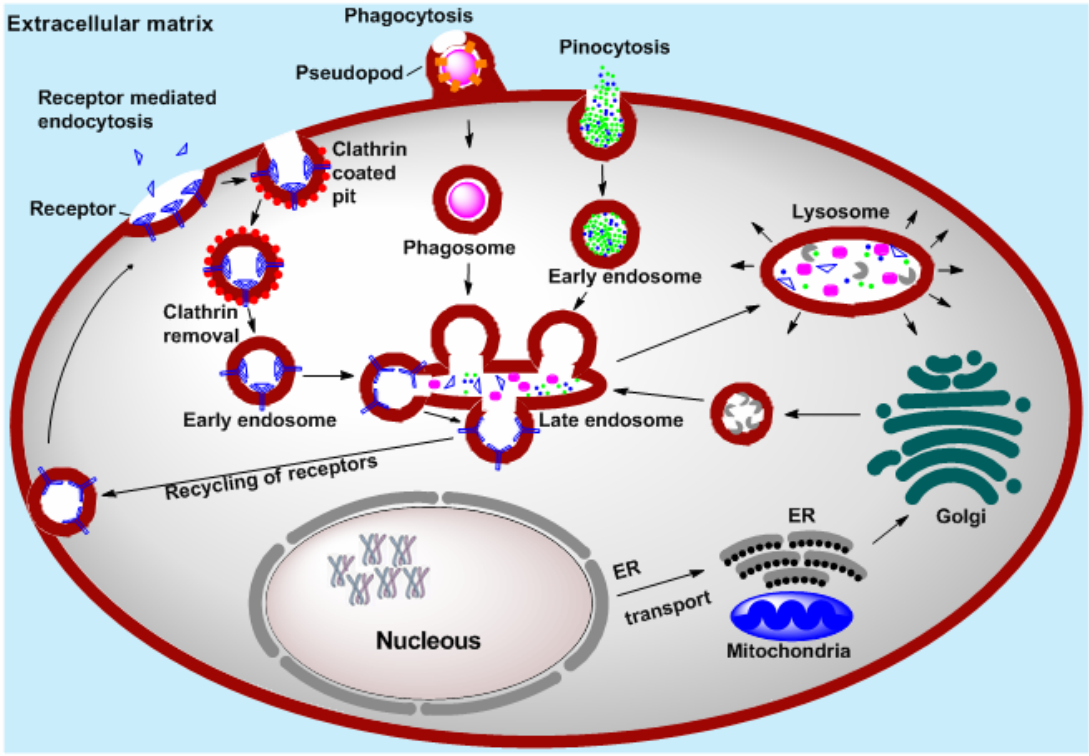
By phagocytosis the cell membrane folds around an object, such as microorganisms, cell debris and apoptotic cells and ingests this in a phagosome (i.e. a large vacuole made up by the plasma membrane; Scheme 1.17).[142]

During pinocytosis part of the cell membrane invaginates in a non-specific manner and forms a pocket, which buds off into the cell forming a vesicle filled with solutes and other small molecules dissolved in the ECM.[143]

During receptor-mediated endocytosis ligands bind to transmembrane receptors, which accumulate locally and this part of the cytoplasmic membrane is then associated with either caveolin or clathrin. Caveolin or clathrin-coating mediates formation of a small pit, which encloses the ligand-receptor complexes and the surrounding ECM followed by budding off the pit into the cytoplasm (Scheme 1.17).[140,144] Hereafter, clathrin is removed forming the early endosome. This form of transport provides a specific and concentrating mechanism for uptake of ligands without taking in correspondingly large amounts of the extra cellular fluid. Thus, the process of receptor mediated endocytosis both up and down regulates transmembrane signal transduction.

Common for all types of endocytosis is that the early endosomal vesicles or phagosomes fuse and release their cargo into the late endosome. In the case of

receptor-mediated endocytosis many ligands dissociate from their receptor here and the receptor is recycled to the cell surface (Scheme 1.17).[139] Enzymes from the trans-Golgi network are deposited into the late endosome followed by delivery of the late endosomal material into the lysosome, where the material is hydrolysed and released into the cytosol.[145]



Scheme 1.17: Schematic representation of different forms of endocytosis.

1.6 Cell penetrating peptides

Cell penetrating peptides (CPPs) are typically highly cationic peptides[146], with a high abundance of amino acids such as Arginine or Lysine, or alternating charged/hydrophobic amino acids residues.[146] Perhaps the most highlighted CPP is the so-called TAT peptide, named after the trans-acting activator of transcription (TAT) which demonstrates a remarkable ability to penetrate cells via binding to the TAT-receptor during human immunodeficiency virus infection.[147,148] The so-called TAT peptide (GRKKKRQRRR) has subsequently been shown to promote cellular delivery or uptake of conjugated proteins, phage, liposomes, small molecules and

nanoparticles,[146,149-155] with the peptide able to traverse almost all tissues via receptor mediated-endocytosis, including the brain.

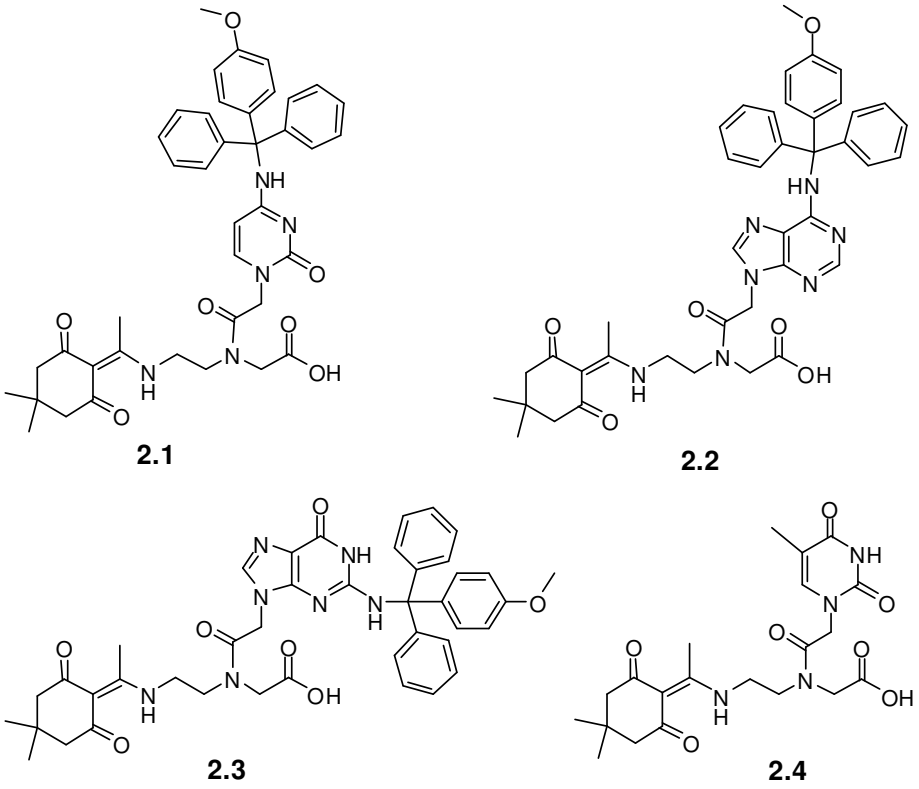
The uptake mechanism of CPPs (and any conjugated cargo) into mammalian cells has been the subject of some controversy and both energy-dependent and independent pathways have been reported.[150,155-164] Through the energy-dependent pathway CPPs traverse the cell membrane by endocytosis,[163,164] although, the exact mechanism of endosomal uptake is poorly defined with no single uptake pathway identified, indeed several endocytosis pathway may be operating depending on the class of structure and cell type.[163,164] It has also been suggested that CPPs enter cells by interaction with cell surface polysaccharides such as heparin sulphate, however, this too is the subject of some debate.[165,166]

CPPs are readily synthesised and typically have low immunogenicity[156,167] and offer numerous opportunities for the delivery of a variety of cargos.[155,156] Furthermore, CPPs have been shown to selectively deliver therapeutic cargos by exploiting peptide affinity for tissue-specific markers, which comprise the so-called vascular ‘‘zip code’’ system. [168,169] Thus a chimeric peptide comprising of a tumor homing peptide combined with a cell-penetrating peptide has been reported to yield a peptide with tumor homing specificity that could efficiently deliver cargo molecules inside cells.[170] Other CPPs such as RGD (Arg-Gly-Asp) and NGR (Asn-Gly-Arg) have also entered clinical trials and represent the first generation tumor homing peptides that do not home to the corresponding normal organ but seem to recognize molecules that are upregulated during tumor progression.[171,172,173] This peptide-based targeted delivery increases the efficacy and decreases the side effects caused by the overall toxicity of the corresponding therapeutic compound.[174]

Chapter 2 Microwave-assisted orthogonal synthesis of PNA-peptide conjugates

2.1 Introduction to PNA-peptide conjugate synthesis

The orthogonal, solid phase synthesis of peptide-PNA conjugates using a Dde/Fmoc strategy with Dde/Mmt protected PNA monomers **2.1–2.4** (Scheme 2.1) and Fmoc/*t*Bu protected amino acids has been reported previously[6,14,15, 29]. However, this approach is time consuming and non-automated, limiting wider applications of PNA-peptide conjugates in chemistry and biology. As described in section 1.1, PNA has unique characteristics making it a very interesting and *in vivo* applicable DNA mimic. Thus, improved and faster synthesis of PNA could lead to more extensive applications of PNA in microarray technology and gene therapy.



Scheme 2.1: The four Dde/Mmt-protected PNA building monomers employed in this project: (**2.1**) cytosine, (**2.2**) adenine, (**2.3**) guanine, and (**2.4**) thymine.

To speed-up and automise PNA synthesis, microwave heat acceleration of both PNA monomer coupling and Dde deprotection was developed. To ensure that this method was applicable to orthogonal PNA-peptide conjugate synthesis, the previously reported orthogonality between Dde- and Fmoc-protection groups needed to be kept.[6,14,15] This strategy was developed with the synthesis of two specific receptor-

based, cellular binding, PNA–encoded peptide ligands, Arg-Gly-Asp-12-mer-PNA and Arg-Gly-Glu-12-mer-PNA.

2.2 Choice of resins

In previously reported procedures, polyethylene glycol dimethyl acrylamide (PEGA) resin was the resin of choice because of its high swelling-properties and low loading, which enables complete coupling of the bulky PNA chain. However, its large swelling ratio increases the amount of monomer required in each reaction as a larger reaction volume is required to suspend the resin. Although, microwave-aided synthesis is less laborious and time consuming than general reaction conditions, the microwave heating should be used with care with PEGA resin, as this resin is fragile and excessive stirring fragments the resin. During synthesis of PNA sequences (up to 20 PNA/amino acids), this problem can be circumvented by reducing the stir rate to a maximum 300 rpm.

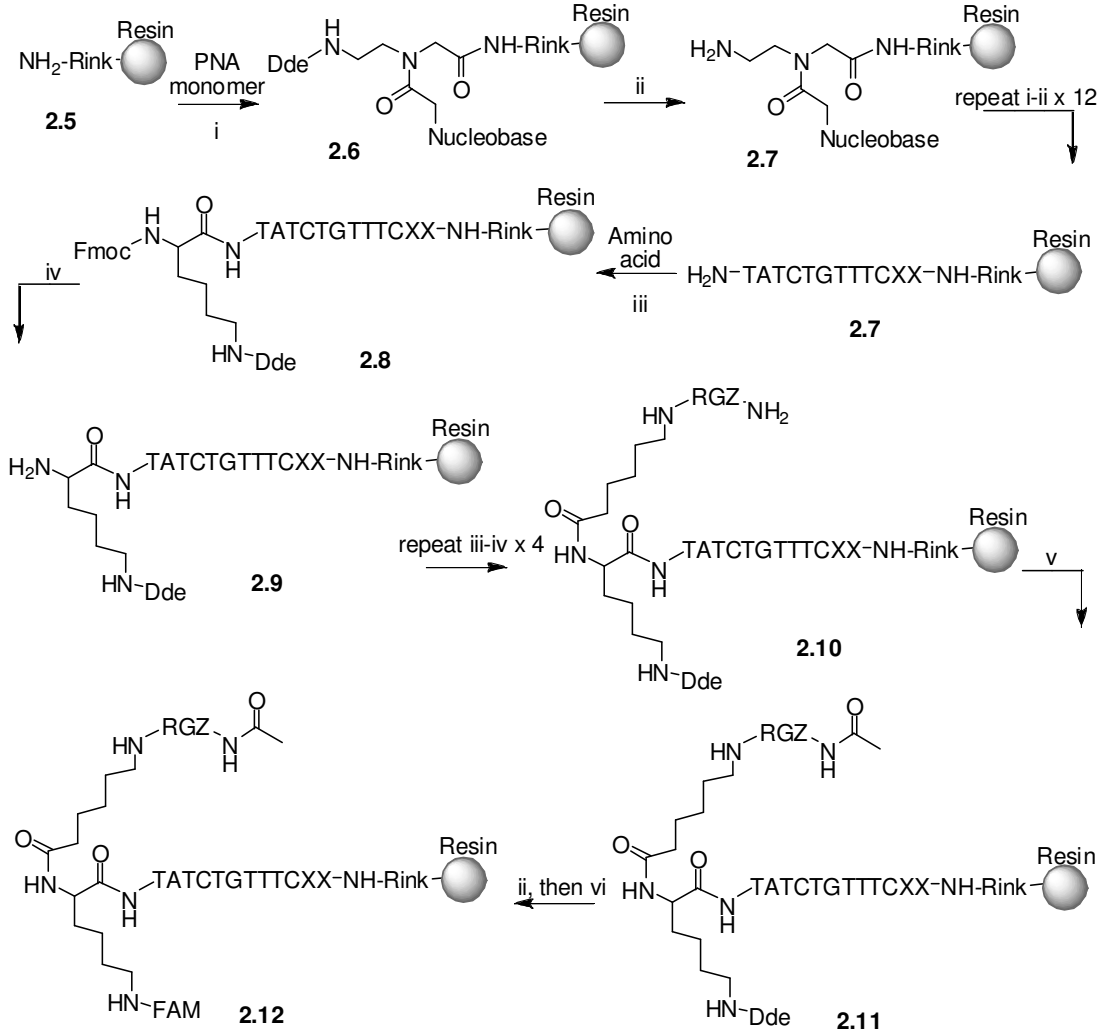
2.3 Synthesis of PNA-peptide conjugates

Resin **2.5** (Scheme 2.2) and the PNA monomers **2.1–2.4** were synthesised according to previously reported procedures,[14] with the sole exception that the coupling of the functionalised nucleobases to the backbone was carried out with solid-supported dicyclohexylcarbodiimide (PS-DCC) rather than conventional DCC.[175,176]

PNA-peptide synthesis was carried out in an automated CEM peptide synthesiser. PNA monomer coupling (Scheme 2.2) was carried out using pre-activated monomers (0.22 M Dde-monomer, 5.5 equiv.) in DMF, mixed with 0.20 M benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP, 5 equiv.) also in DMF and with *N*-ethylmorpholine (NEM; 11 equiv.) for 1 min followed by microwave irradiation for 20 min at 60 ^{o}C. After PNA monomer coupling the filtrate was collected and unreacted monomers were recovered. Dde-deprotection was carried out under acidic Dde-deprotection conditions (20% $NH_2OH{\cdot}HCl$/imidazole 1:0.75 in NMP/DMF 5:1). Here the key advantage of the $NH_2OH{\cdot}HCl$ deprotection conditions

is highlighted, namely limiting the *N*-terminal detachment side reaction, which makes PNA synthesis problematic.

Fmoc-deprotection and amino acid couplings were carried out according to standard literature procedures[3,14] (Scheme 2.2). Following amino acid addition, the peptide was capped with acetic anhydride/pyridine (1:1) for 20 min, before the PNA was capped with 5(6)-carboxyfluorescein as described above for PNA monomer couplings.[14] The resin was treated with piperidine to cleave any fluorescein ester dimers.[14]



Scheme 2.2: Orthogonal coupling of Dde/Mmt protected PNA monomers and Fmoc/*t*Bu-protected amino acids. (i) 5.5 equiv. Dde-monomer, 5 equiv. PyBOP, and 10 equiv. NEM in DMF, microwave irradiation, 20 min, 60 °C; (ii) 20% $NH_2OH{\cdot}HCl$/imidazole 1:0.75 in NMP/DMF 5:1, 1 h, room temperature or

microwave irradiation, 10 min, 60 °C; (iii) 5 equiv. amino acid, 5 equiv. HTBU/HOBt, 10 equiv. DIPEA in DMF, microwave irradiation, 20 min, 60 °C; (iv) 20% piperidine in DMF, 2 x 10 min, room temperature; (v) acetic anhydride/pyridine (1:1), 20 min, (vi) 5.5 equiv. 5(6)-carboxyfluorescein (FAM), 5 equiv. PyBOP, and 10 equiv. NEM in DMF, 20 min, 60 °C. Resin = PEGA (0.04 mmol/g). Rink = Rink amide linker. X = A or T. Z = Asp or Glu.

The peptide–PNA conjugates were cleaved from the solid support with 5% triisopropylsilane (TIS) in TFA and purified by reverse phase high performance liquid chromatography (RP-HPLC) and analysed by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF): Arg-Gly-Glu-Ahx-Lys(FAM)-TATC-TGTT-TCTA-NH_2: 12.42 min (95.2% purity), MALDI-TOF m/z (%) 4206.33 $[M+H]^+$ (100). Arg-Gly-Asp-Ahx-Lys(FAM)-TATC-TGTT-TCAT-NH_2: 12.37 min (95.2% purity), MALDI-TOF: m/z (%) 4193.04 $[M+H]^+$ (100).

Using these acidic Dde-deprotection conditions, the peptide-PNA constructs Arg-Gly-Asp-Ahx-Lys(FAM)-TATC-TGTT-TCAT-NH_2 and Arg-Gly-Glu-Ahx-Lys(FAM)-TATC-TGTT-TCTA-NH_2 (Scheme 2.2, Figure 2.1) were synthesised with all PNA and amino acid coupling, and deprotection cycles carried out in an automated microwave peptide synthesiser (CEM). These conjugates would be synthetically difficult to obtain via previous methods, as their sequences contain multiple poly-thymine residues[6]. However, this improved microwave-assisted synthesis yielded the conjugates Arg-Gly-Asp-Ahx-Lys(FAM)-TATC-TGTT-TCAT-NH_2 and Arg-Gly-Glu-Ahx-Lys(FAM)-TATC-TGTT-TCTA-NH_2 with respective overall isolated yields of 14% and 16% (over 36 steps) and purities of 95.2% and 93.7% (RP-HPLC, Figure 2.1).

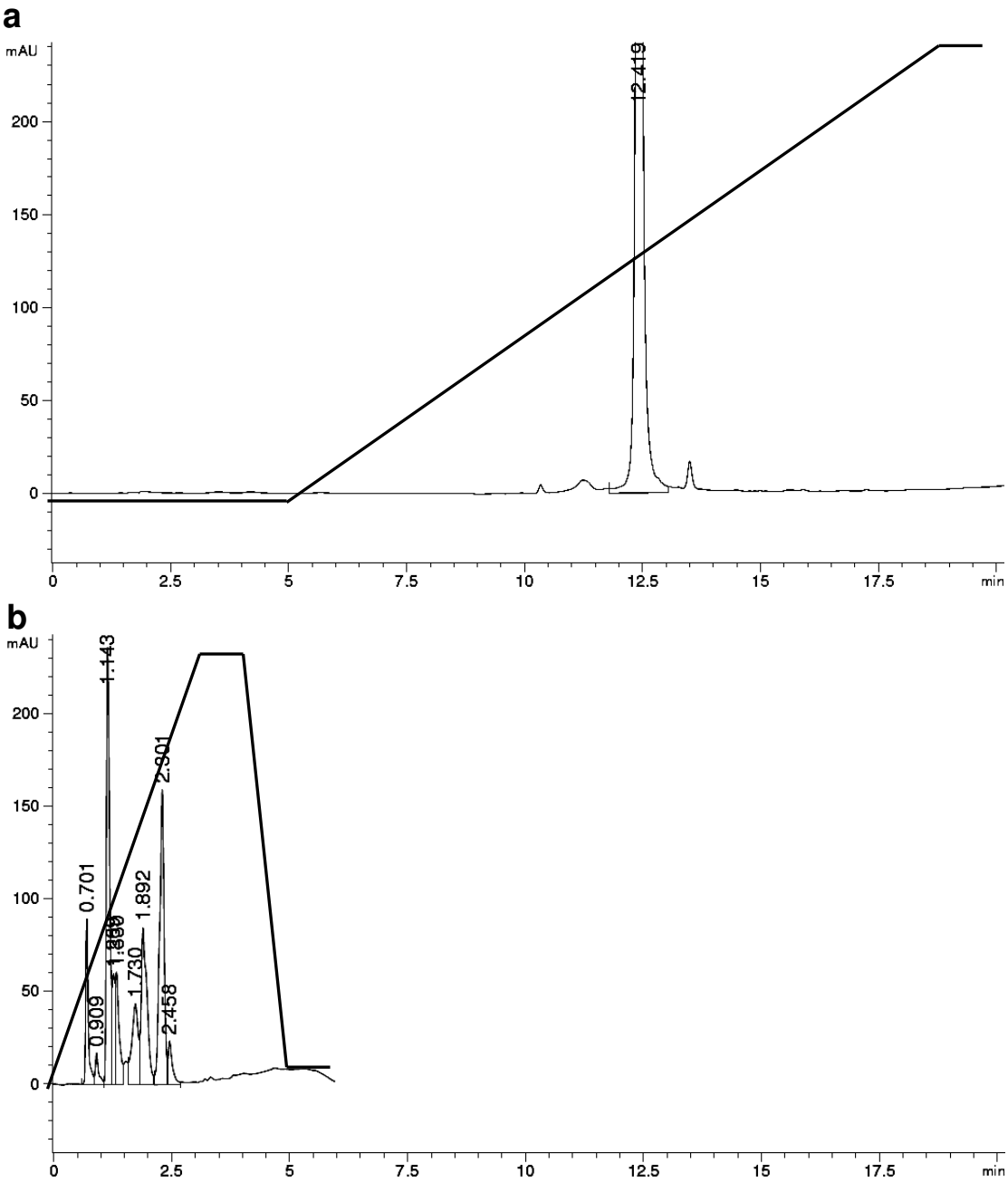


Figure 2.1: (**a**) RP-HPLC of Arg-Gly-Glu-12-mer-PNA synthesised under acidic Dde-deprotection conditions (20% $NH_2OH{\cdot}HCl$/imidazole 1:0.75 in NMP/DMF 5:1). Eluting with A for 5 min; 100% A to 100% B over 20 min; 100% B for 5 min. (**b**) RP-HPLC of PNA-TGTT TCAT synthesised under basic Dde- deprotection conditions (20% $NH_2OH{\cdot}H_2O$/imidazole 1:0.75 in NMP/DMF 5:1). Column: Phenomenex Luna, C18, 15 cm x 1.00 cm, 5 cm.; λ = 220 nm; solvent A: H_2O with 0.1% TFA; solvent B: CH_3CN with 0.1% TFA, eluting with 95% A to 95% B over 3 in; 100% B for 1 min.

2.4 Conclusion

Mild thermal effects achieved via microwave heating accelerated PNA synthesis allowed the automated synthesis of peptide-PNA conjugates while improving synthetic efficiency. This approach achieved PNA synthesis in a manner orthogonal to Fmoc chemistry, and in a way that prevented unwanted PNA side reactions, which shows Dde-chemistry is the method of choice for PNA-peptide conjucate synthesis. Hereby two PNA-encoded peptides, Arg-Gly-Asp-Ahx-Lys(FAM)-TATC-TGTT-TCAT-NH_2 and Arg-Gly-Glu-Ahx-Lys(FAM)-TATC-TGTT-TCTA-NH_2 were synthesised by this method with overall isolated yields of 14% and 16% respectively (40 steps). Moreover, this strategy allowed synthesis of traditional synthetically difficult PNA-sequences containing multiple poly-Thymine residues.

Chapter 3 Targeting specific cellular delivery

3.1 Introduction to cell penetrating peptides

Selective delivery of cargos into specific cell types has the capacity to ensure optimal distribution of therapeutic entities into diseased cells while limiting possible adverse, off-target effects.[174] One possible route whereby this might be achieved is via receptor mediated endocytosis, which provides opportunities for targeted delivery where specific receptors are (over) expressed with so-called cell penetrating peptides (CPPs) that have been shown to facilitate cell penetration and delivery of a variety of covalently linked cargos.[149-153, 174]

As described in section 1.6, CPPs are readily synthesised and typically have low immunogenicity[156] and offer numerous opportunities for the delivery of a variety of cargos.[155,156] Hereby, this peptide-based targeted delivery increases the efficacy and decreases the side effects caused by the overall toxicity of the corresponding therapeutic compound.[174] The aim of this part of the project was to develop a strategy for the high-throughput screening of a PNA-encoded peptide/peptoid hybrid library to allow the identification of versatile, and human cell selective cell penetrating peptides (Figure 3.1).

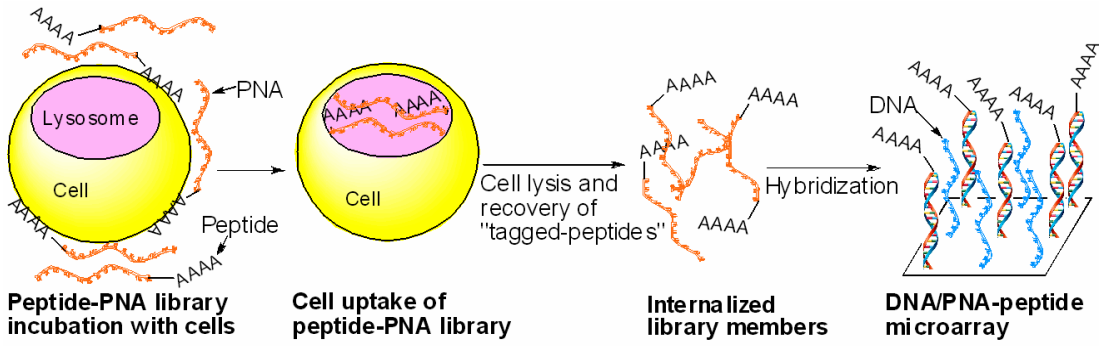
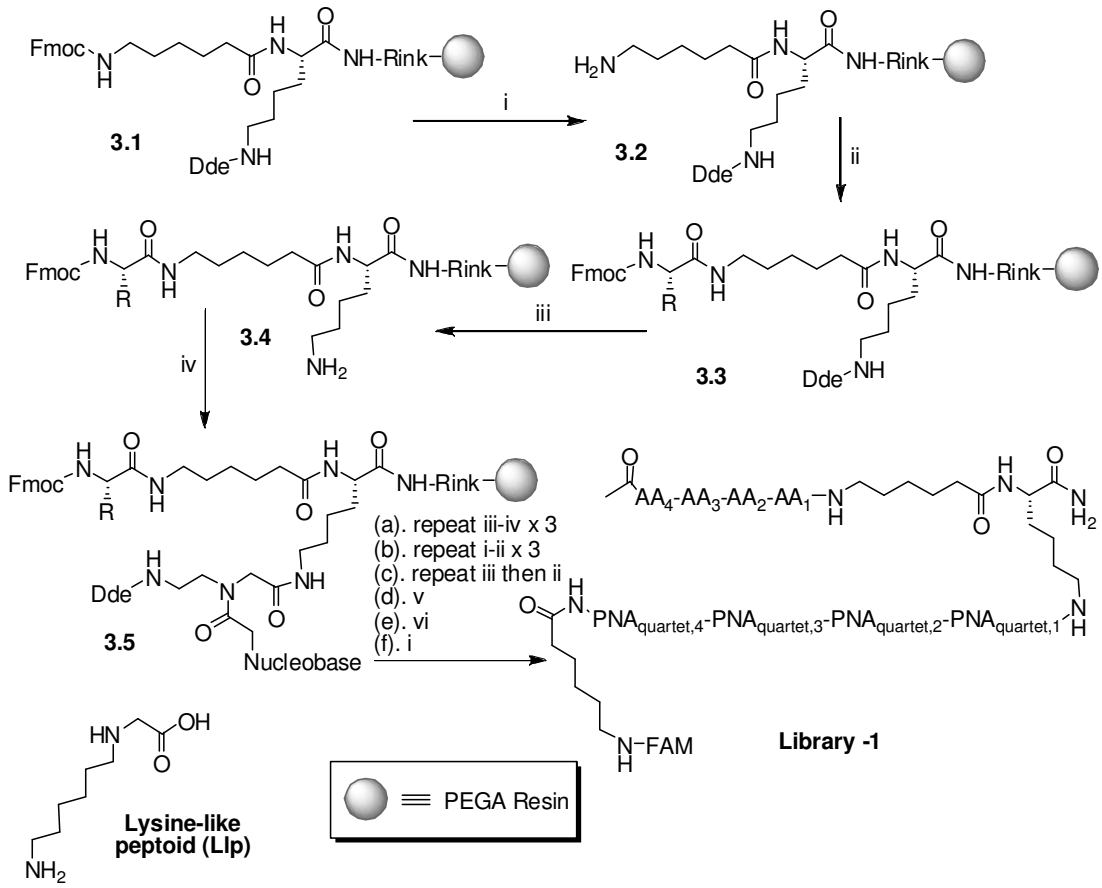


Figure 3.1: The strategy to allow the identification of cell selective penetrating peptides. An encoded 1296 member peptide library was incubated with cells. Any extra-cellular peptides or peptides binding to the surface of the cells were released with trypsin and washed away. Cells were lysed, intracellular PNA extracted and hybridised onto a complementary 44,000 DNA microarray.

3.2 Encoded Library Synthesis

A 1296 member fluorescein-labelled PNA encoded[17,72] tetrapeptide-library (Library1) was designed and synthesised using split and mix methods[69,70] (Scheme 3.1). Six amino acids were used: Pro (turn), Glu (acidic), Leu (hydrophobic), Lys (basic), Tyr (aromatic), and a Lysine-like peptoid monomer (Llp, basic; Scheme 3.1) to represent the overall classes of the natural amino acids as well as a moiety, which is known to enhance cellular delivery and resist enzymatic degradation.[177] Each amino acid was encoded by a PNA-quartet (Table 3.1) with each quartet designed to have maximum 50% similarity with other tags and maximum 50% purine content with the PNA tags designed to have uniform T_m and not include palindrome sequences or poly-thymine moieties (maximum 3 consecutive thymines; see table 1 and appendix I).



Scheme 3.1: Synthesis of a PNA encoded 1296 member peptide library: (i) 20% piperidine in DMF, 2 x 10 min; (ii) amino acid [Fmoc-Pro-OH, Fmoc-Leu-OH, Fmoc-Glu(O*t*Bu)-OH, Fmoc-Tyr(*t*Bu)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Llp(Boc)-OH, 5 equiv]. HTBU/HOBt (5 equiv.). DIPEA (10 equiv.) in DMF, 20 min, 60 ^{o}C;

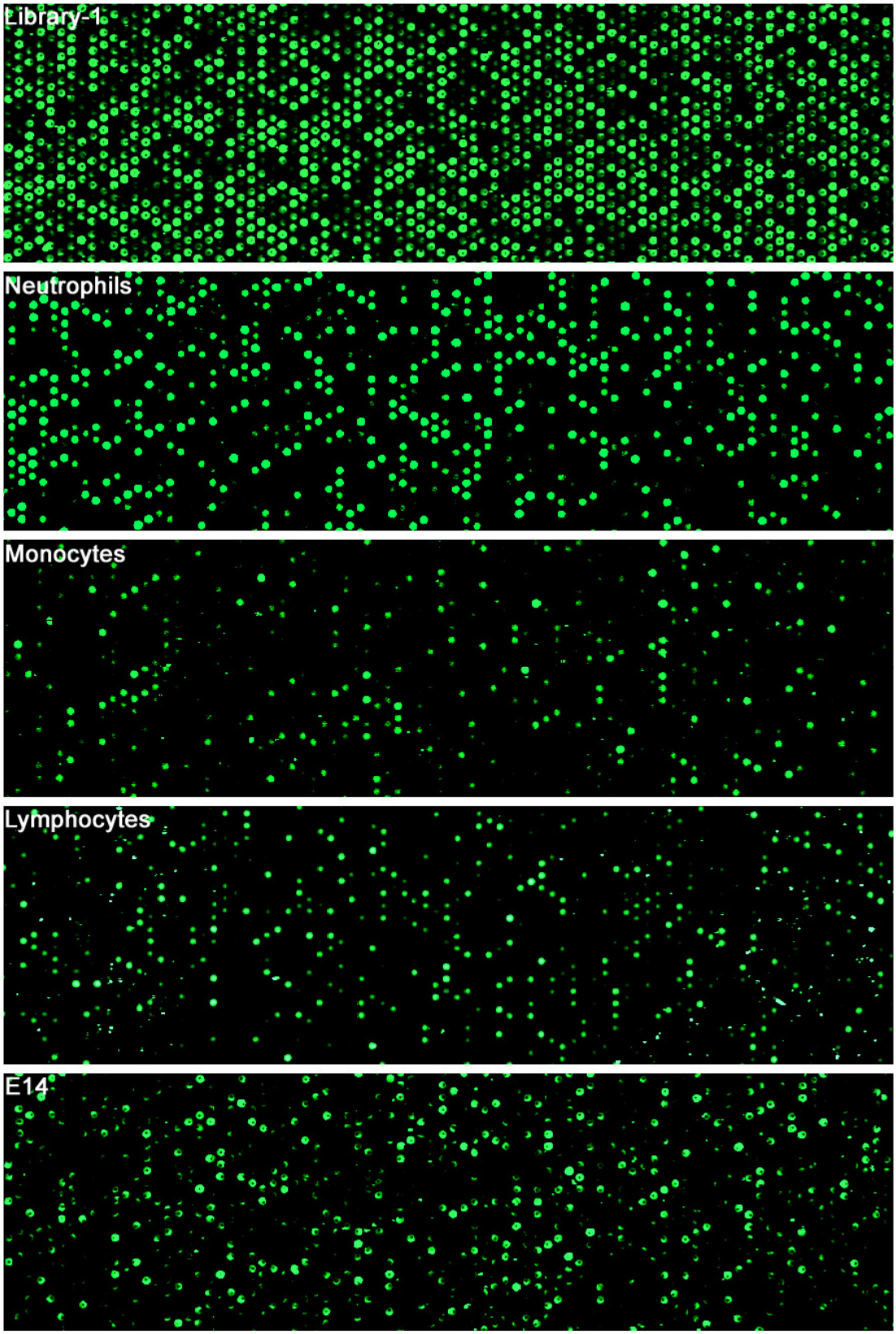
(iii) 20% NH2OH·HCl/imidazole 1:0.75 in NMP/DMF 5:1, 1 h or 10 min, 60 °C; (iv) Dde/Mmt-protected PNA monomer (5.5 equiv.), PyBOP (5 equiv.), and N-ethylmaleimide (NEM, 10 equiv.) in DMF, 20 min, 60 °C; (v) acetic anhydride/pyridine (1:1), 20 min, (vi) 5(6)-carboxyfluorescein (FAM, 5.5 equiv.), PyBOP (5 equiv.), and NEM (10 equiv.) in DMF, 20 min, 60 °C. Resin = PEGA (0.04 mmol/ml). Rink = Rink amide linker.[178]

Table 3.1: PNA-encoding of the 6 amino acids.

Amino acid	PNA Tag (N-C)
Llp	TTAC
Pro	AAAC
Leu	TACA
Glu	ATCT
Tyr	ACAA
Lys	TCAT

3.3 Library screening and hybridisation

HeLa (human cervix epitheloid carcinoma), HEK293T (transformed human embryo kidney cells), SH-SY5Y (human neuroblastoma), K562 (human erythroleukemic cells), and ARN8 (human amelanotic melanoma cells) as well as mouse embryonic stem cells, E14 and human primary lymphocytes, human primary monocytes, and human primary neutrophils were incubated with the Library (100 μM) to ensure efficient exposure of each library member to the cells (this corresponds to 77 nM of each member) as well as yielding sufficient material following cell lysis for microarray analysis. Following incubation, washing and treatment with trypsin, the cells were lysed, the library members were extracted and purified by filter centrifugation (between 3,000 and 10,000 Da) and hybridised onto custom designed microarrays with 4 subarrays of 44,000 features each with 33 replicates of each oligonucleotide complementary to each member of the library as well as 1232 non-coding negative controls (Agilent). Microarray scanning with data analysis (BlueFuse, BlueGenome) was used to extract the intensity of the FAM label, thereby determining the relative amount of PNA hybridised to each spot and the identify the peptide (Figure 3.2 and appendix II).


Library-1
Neutrophils
Monocytes
Lymphocytes
E14

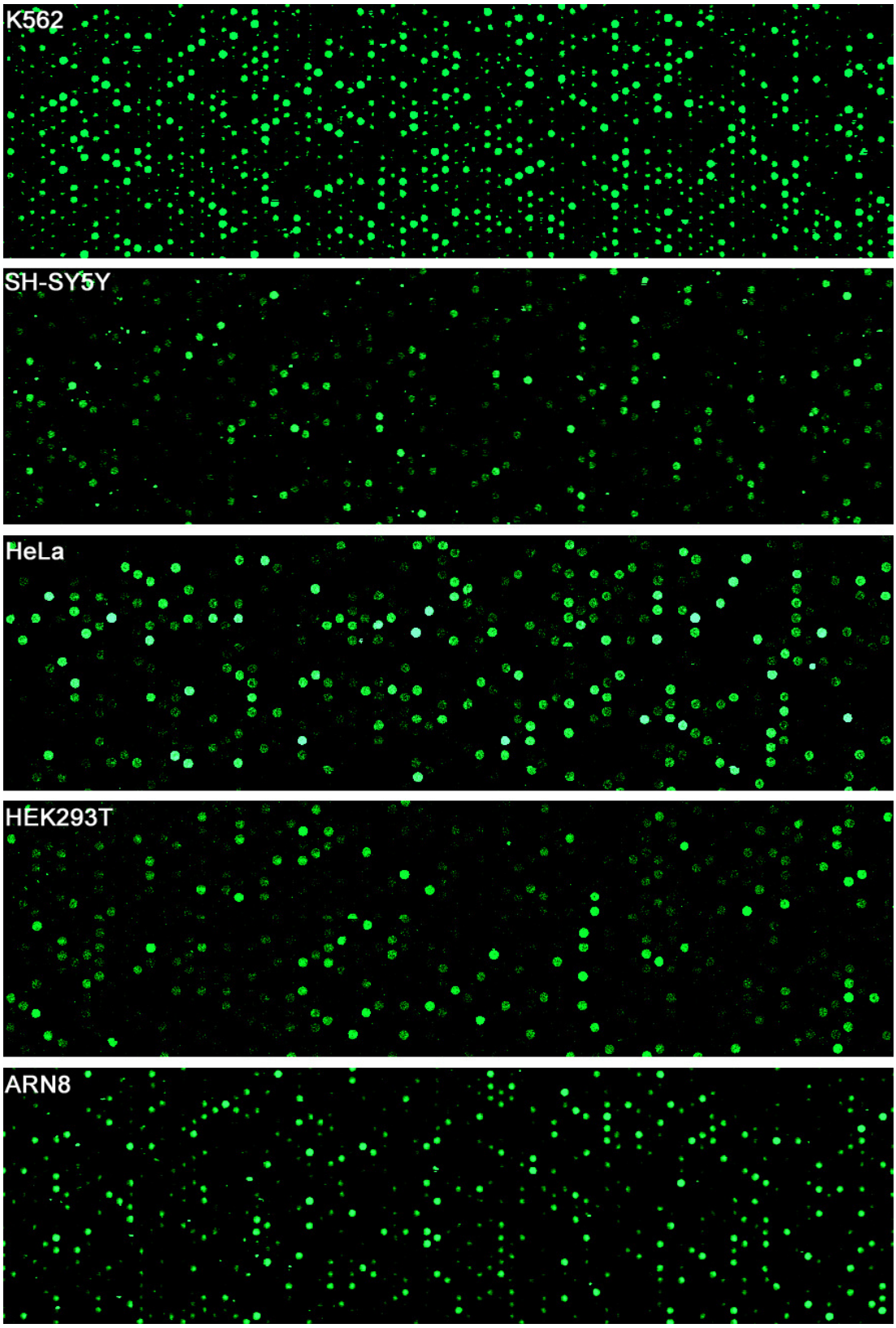


Figure 3.2: Sections of FITC-filtered fluorescence image of 44,000 feature DNA microarrays with 33 replicates of each oligonucleotide complementary to each member of the Library-1 library as well as 1232 non-coding negative controls

hybridised with the Library-1 library or PNA extracted from cells in 110 µL 1 x GenHyb (Genetix) buffer overnight from 50 °C to 32 °C.

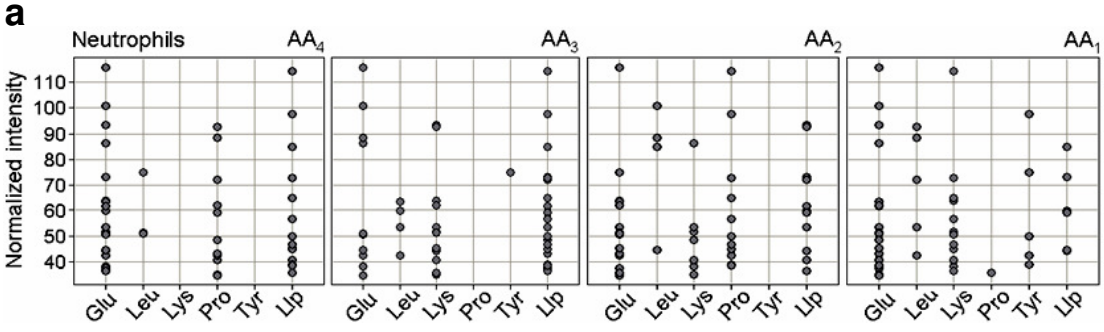
3.3.1 Microarray data and normalisation

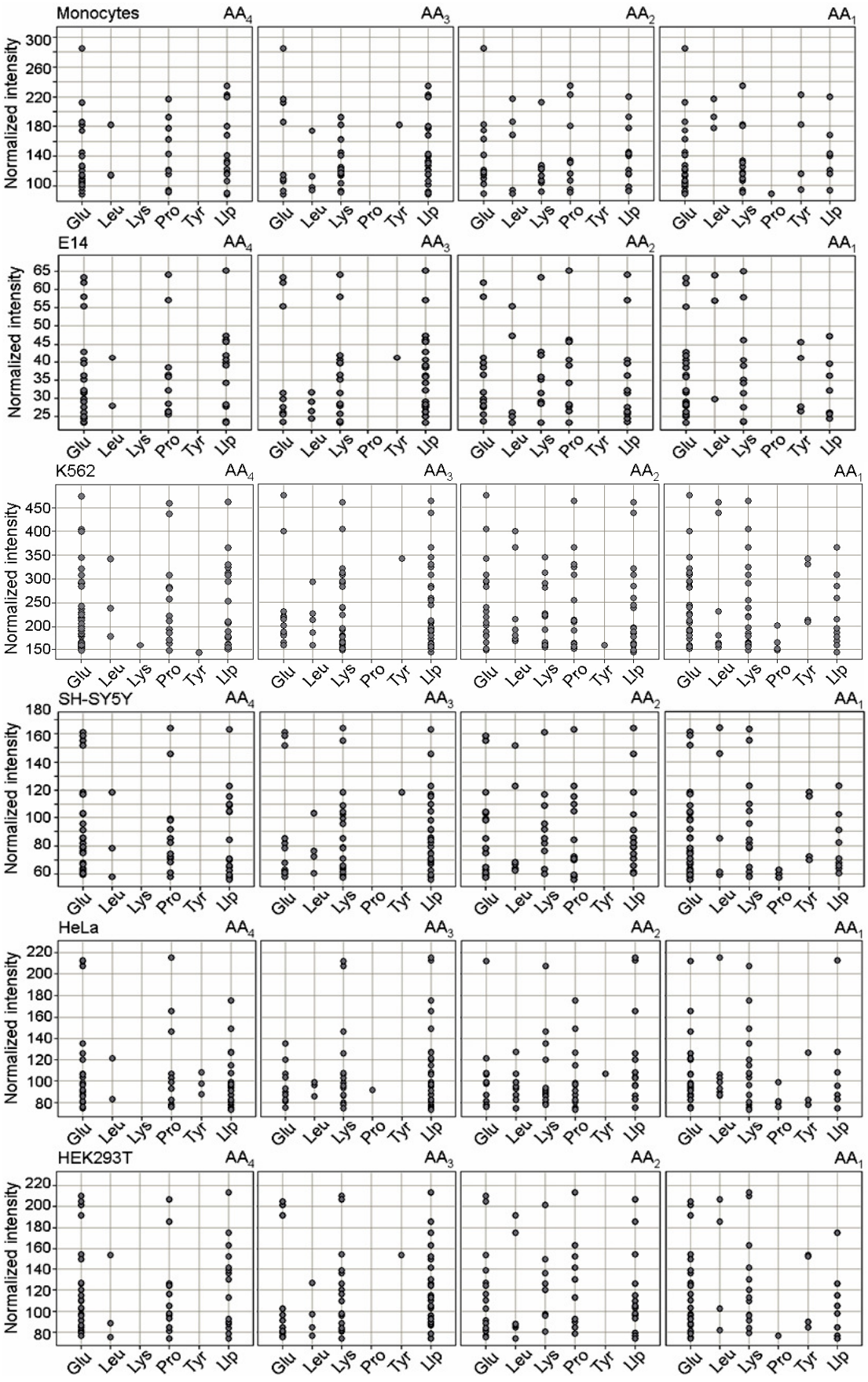
Raw microarray data was obtained from Bluefuse, which allows grid alignment and signal estimation, in a Microsoft Excel format. In Microsoft Excel the top ~5% and the bottom ~5% of each of the 33 replicate-sets were removed as outliers (erroneous values caused by dust, scrapes etc[179,180]). The mean fluorescence intensity was calculated over each of the remaining 29 replicates and over the non-complementary DNA features (negative controls).

The 1296 mean intensities were corrected for the microarray background by subtracting the mean intensity of the non-complementary negative control features. Hereafter, the data was normalised for the control hybridisation by calculation the ratio between the background corrected mean intensities of the data sets obtained from hybridisation of PNA extracted from cells and the data set obtained from the hybridisation of the untreated 1296-peptide/PNA library (appendix II).

Initially a generic consensus uptake sequence for all cell types was sought by analysis of scatter plots of the normalized mean intensities versus the variable monomers for positions AA_1 to AA_4 (Figure 3.3a). This revealed a generic global consensus sequence for all the cell types of Glu-Llp-Glu-Glu; a surprise in view of the typically basic residues traditionally observed (Figure 3.3b).

a




Monocytes
E14
K562
SH-SY5Y
HeLa
HEK293T
AA4
AA3
AA2
AA1
Normalized intensity
Glu
Leu
Lys
Pro
Tyr
Llp

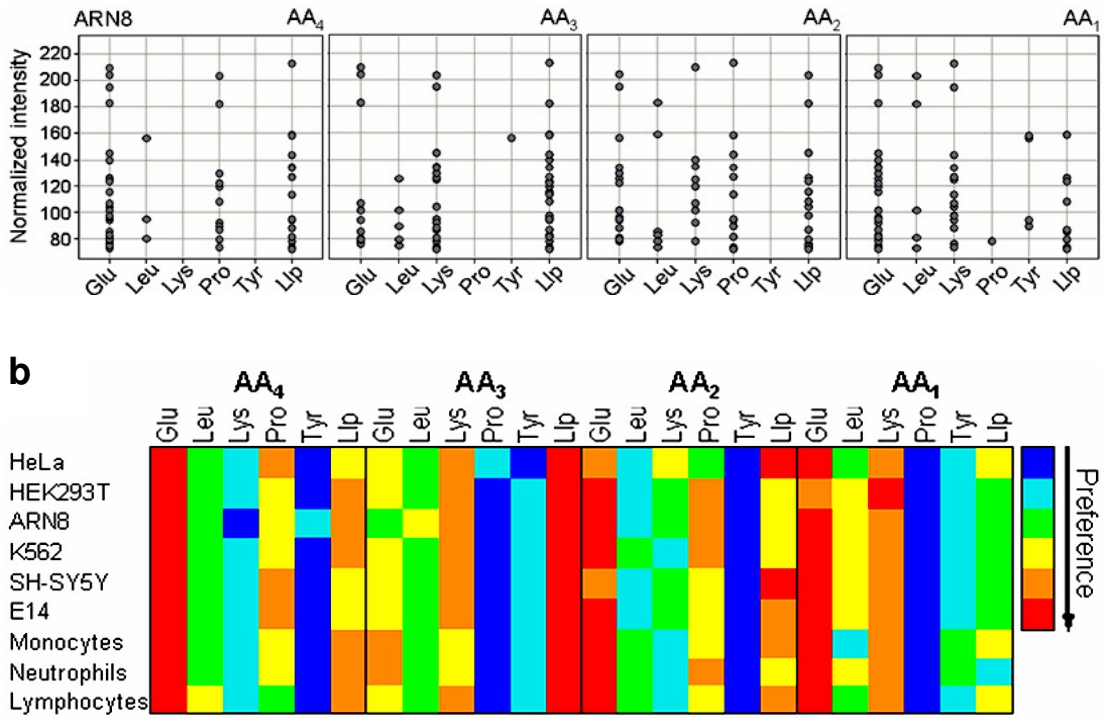


Figure 3.3: DNA microarray analysis of a peptide-PNA library identifies a generic consensus sequence of highly cell-penetrating peptide. (**a**) Scatter plots of the background-corrected and normalised mean intensities of the FAM-labelled PNA derived from microarray hybridisation of PNA extracted from cells onto a complementary DNA microarray, plotted versus the monomers for positions AA_1 to AA_4. (**b**) Heat-map showing consensus sequences of the best cell-penetrating peptides extracted from the scatter plots.

3.3.2 Quantitative delivery of the consensus sequence

To verify and quantitatively compare the cellular delivery abilities of the consensus sequence, Glu-Llp-Glu-Glu, this peptide and the positive control, tetra-Llp were synthesised with a fluorescent tag at the *N*-terminus: FAM-Ahx-AA_4-AA_3-AA_2-AA_1-NH_2 (X-AA_4-AA_3-AA_2-AA_1-NH_2).[15,17] and incubated with the cells and analysed by flow cytometry.

The identified consensus sequence, X-Glu-Llp-Glu-Glu-NH_2 had uptake levels similar to or greater than X-Llp-Llp-Llp-Llp-NH_2 in K562, SH-SY5Y, HEK293T, HeLa, and ARN8 cells indicating that this novel highly anionic delivery agent could in some cases result in better delivery than the known, highly basic, peptide-based delivery agent (Figure 3.4).

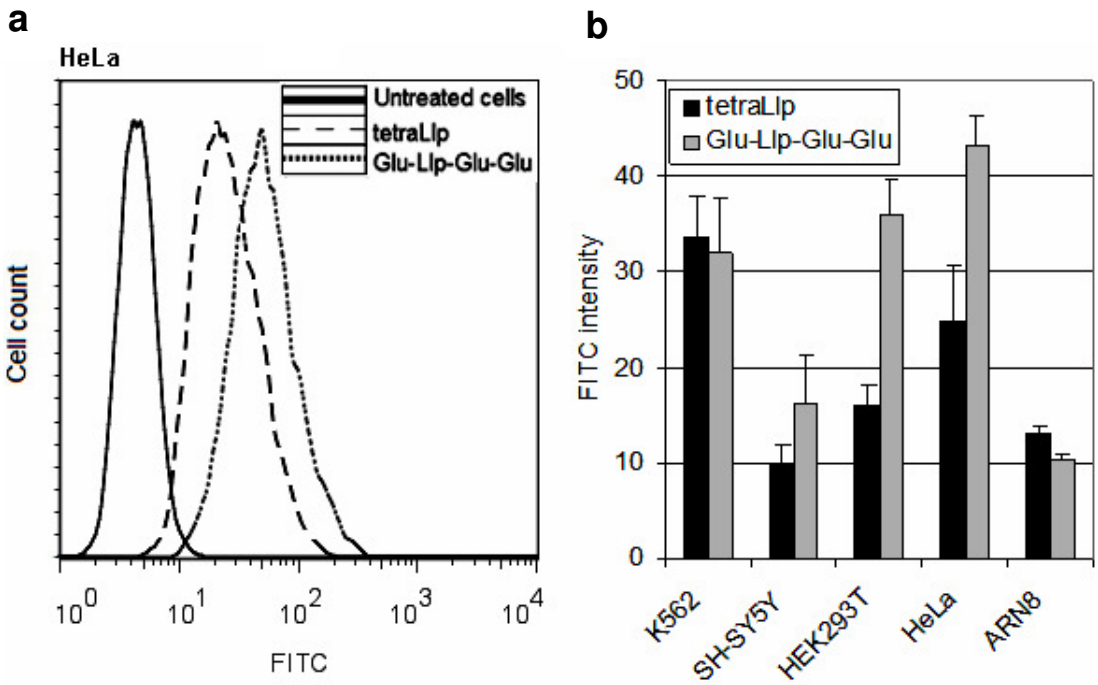


Figure 3.4: Flow cytometry analysis of the fluorescein-labelled X-Glu-Llp-Glu-Glu-NH_2 (written as Glu-Llp-Glu-Glu) and X-Llp-Llp-Llp-Llp-NH_2 (written as tetraLlp) using a FACS Calibur cytometer with FITC and PE/PI filters (10,000 populations, n=3). Any extracellular peptide was released by treatment with trypsin (which also served to detached the cells) while trypan blue was used to quench any extracellular fluorescence[181], ensuring that any observed increase in fluorescence was from just the intracellular peptide. (**a**) FITC filtered flow cytometry histograms gated for live, single cells following incubation with "hit"-tetramer as well as propidium iodide (selectively stains the nuclei of necrotic and apoptotic cells)[182] to exclude dead cells (data not shown). (**b**) Mean FITC filtered fluorescence of histograms versus the "hit"-tetramers. Error bars indicate ± s.d.

3.3.3 Clustering analysis‡

To allow comparison of the data sets from different arrays, the 9 data sets were normalised so that the data sets had the same mean intensity arbitrarily set to 100 and the 1296 mean intensities in each data set were multiplied with respective normalisation factors (appendix II).

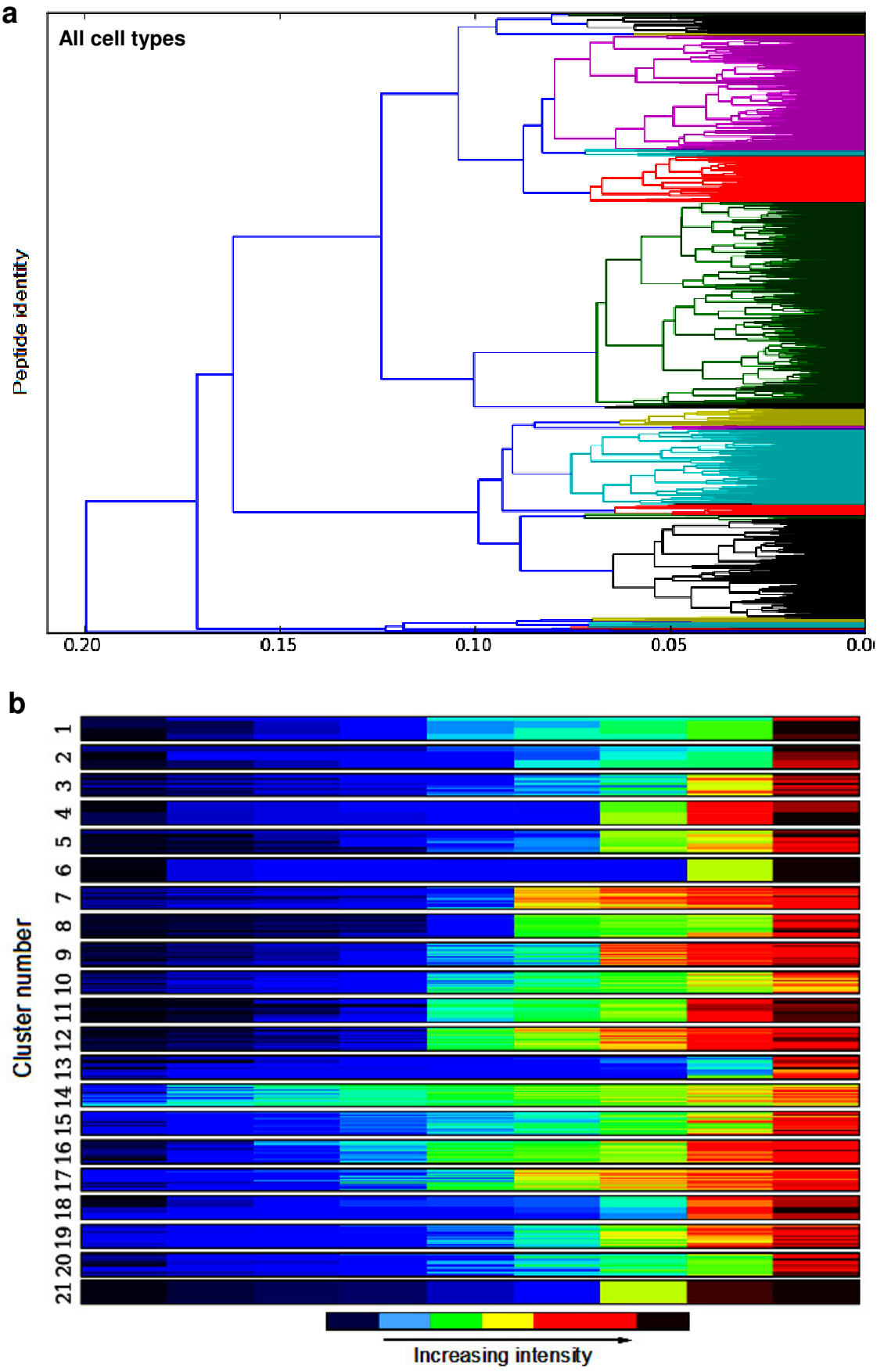
The data set for each cell type were normalised allowing comparison of the data sets derived from different arrays (appendix II) and the data for the 9 cell types were clustered using Enthought Python and a Euclidean distance method (Figure 3.5a-b).

‡ The clustering analysis was performed in collaboration with Dr. Michael Dewar and Dr. J. Douglas Armstrong. Dr. Michael Dewar wrote the code using the Enthought Python software package and a Euclidian distance method.

This resulted in a dendrogram and a heat map for each cluster, where the highest ratio between the maximum intensity and the remaining intensities represent the most selective peptide for the respective cell type (Figure 3.5a-b). The most selective clusters were selected and column diagrams of each of the peptides versus the cell types were generated. From this initial clustering it was evident that most of the highly selective peptide had maximum uptake in primary neutrophils. This was expected as this cell type is a 'professional' phagocyte[183], which would lead to high uptake.

In order to find the most selective peptide for each cell type, domination of the data set by highly endocytotic cell types must be avoided. This was achieved by dividing the 1296 normalised intensities into 9 groups whereby peptides with maximum uptake in one cell type were separately compiled (appendix II). This allowed clustering analysis to be performed on the basis of each individual cell type leading to a dendrogram and heat diagram for each cell type (Figure 3.5c-d; example of dendrograms and heat-maps shown for lymphocytes). The most selective tetramer for each cell type was identified from the heat maps by selecting the cluster with most significant "hot/cold" interface (Figure 3.5c, highlighted oval). In clusters with more than one peptide the peptide leading to the highest uptake level in the intended cell type was selected leading to a selective "hit"-peptide for each cell type (Figure 3.6 and Table 3.2).

Noticebly, the "hit"-tetramers identified for the neutrophils and the lymphocytes showed excellent selectivity, and the "hit"-tetramers identified for the monocytes, the SH-SY5Y, and the HEK293T cells showed moderate selectivity (Figure 3.6). The "hit"-tetramers identified for the E14, the K562, the HeLa and the ARN8 cells showed little selectivity (Figure 3.6).


a
All cell types
Peptide identity
0.20
0.15
0.10
0.05
b
Cluster number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Increasing intensity

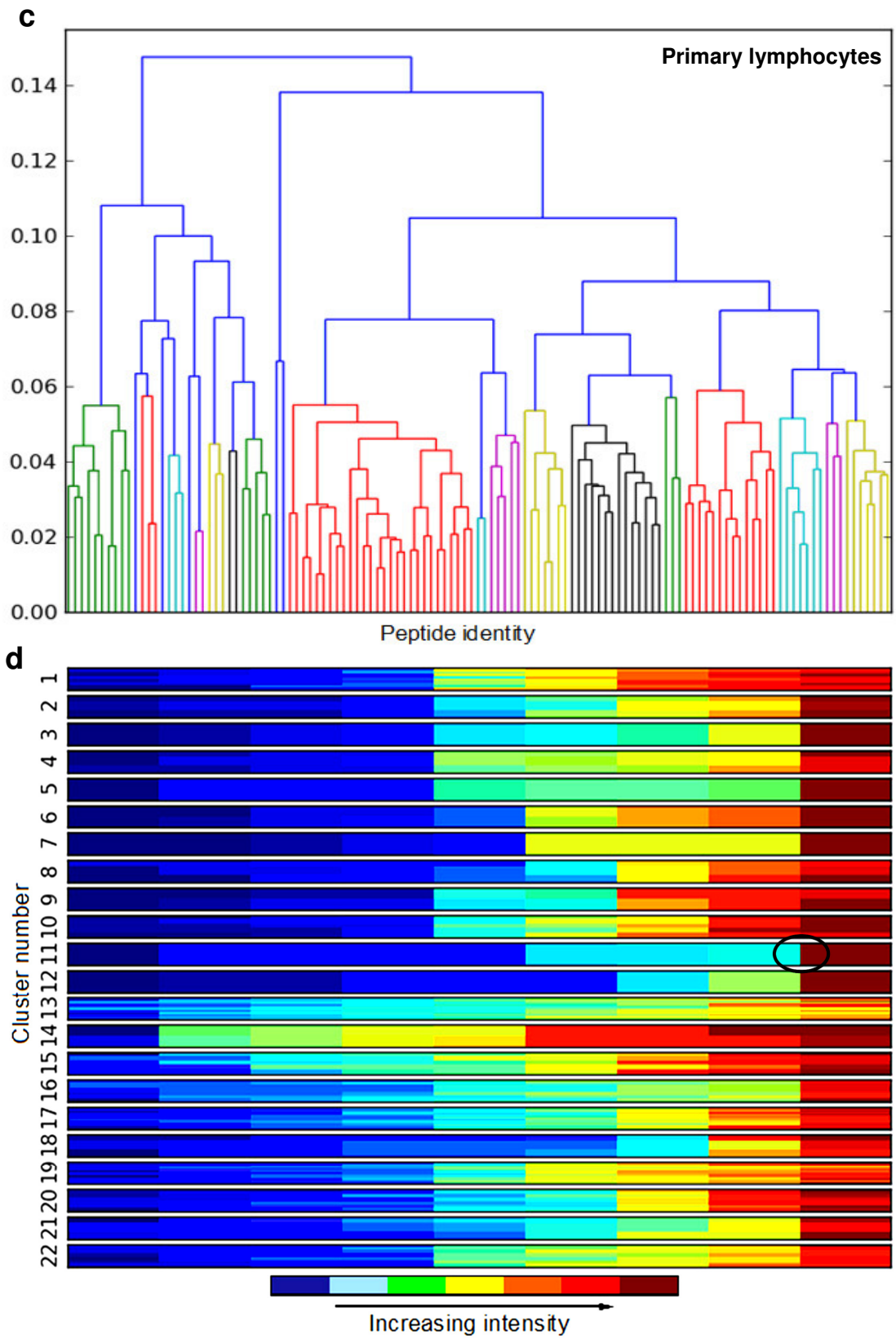


Figure 3.5: Clustering analysis of peptide-PNA/DNA microarray data identifies cell selective delivery agents. Dendrograns were derived from clustering analysis of the 1296 internally normalised intensities of FAM-labelled PNA extracted from the 9

cell types using Enthought Python and a Euclidean distance method. The dendrograms were cut such that ~20 clusters were produced. (**a**) Dendrogram obtained for combined data for all cell types. (**b**) Heat maps of clusters derived from the dendrogram. (**c**) Example of the dendrogram obtained for maximal uptake in primary lymphocytes. (**d**) Example of heat maps of clusters derived from the lymphocyte dendrogram. Red represents high relative intensity and blue represents low relative intensity. The number of horizontal bars across each cluster (different shade of coloured lines) depicts the number of peptides the respective cluster. In this heat map, cluster 11 (lymphocytes), which appears to have only one peptide, displays excellent selectivity with a distinct hot/cold interface (highlighted oval).

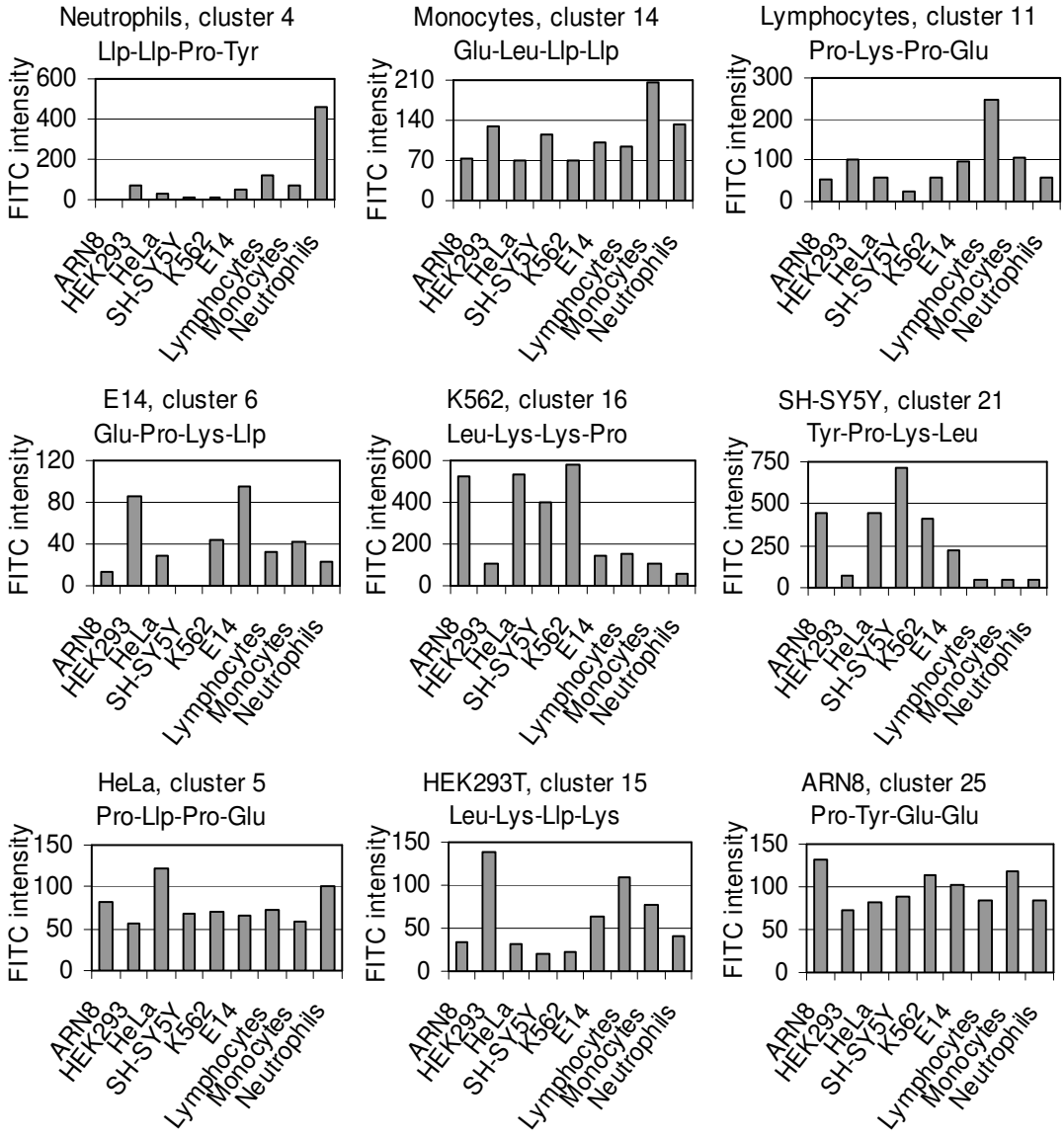


Figure 3.6: Column diagram of each of the peptide versus the cell types from the clusters exhibiting the greatest selectivity for each cell type using the original microarray data (before the internal normalisation).

Table 3.2: Cell selective consensus sequences derived from clustering analysis.

	AA4	AA3	AA2	AA1
Neutrophils	Llp	Llp	Pro	Tyr
Monocytes	Glu	Leu	Llp	Llp
Lymphocytes	Pro	Lys	Pro	Glu
E14	Glu	Pro	Lys	Llp
K562	Leu	Lys	Lys	Pro
SH-SY5Y	Tyr	Pro	Lys	Leu
HeLa	Pro	Llp	Pro	Glu
HEK293	Leu	Lys	Llp	Lys
ARN8	Pro	Tyr	Glu	Glu

3.3.4 Quantitative delivery of "hit" peptides

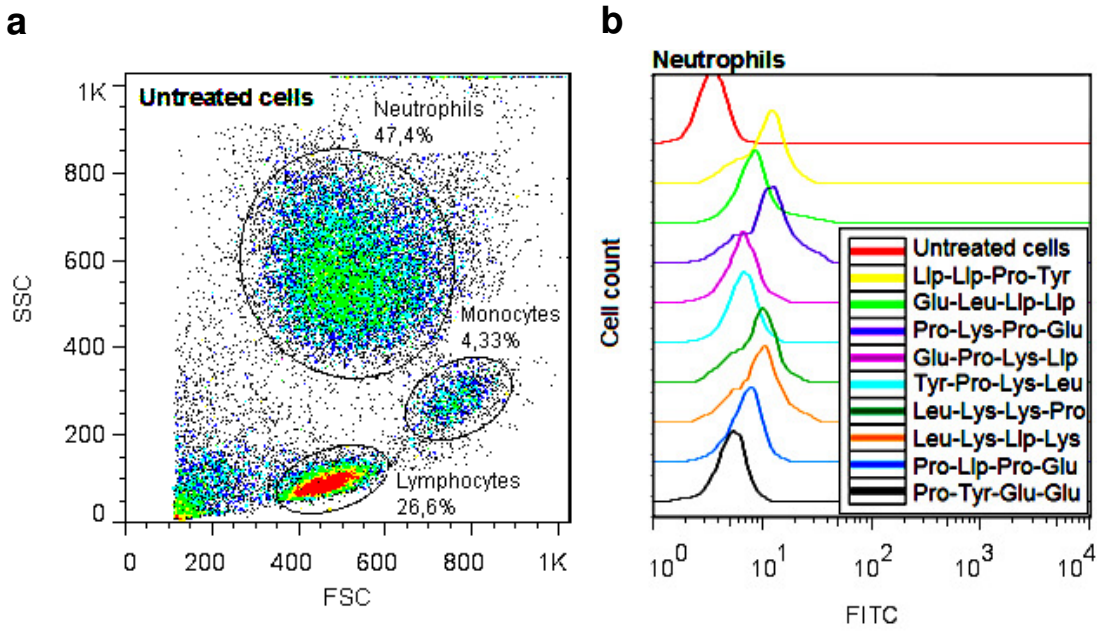
To verify and quantitatively compare the cellular delivery abilities of the selected peptides, the sequences identified based on the microarray data were synthesised with a fluorescent tag: FAM-Ahx-AA_4-AA_3-AA_2-AA_1-NH_2 (X-AA_4-AA_3-AA_2-AA_1-NH_2).[15,17] All the cell lines and the stem cells were incubated with the potential selective "hit"-tetramers and analysed by flow cytometry. To ensure that the experimental conditions are as close to *in vivo* conditions as possible the potential selective delivery agents were incubated with anticuagulated whole blood, followed by collection of the leukocytes and the neutrophils, monocytes and lymphocytes were separated by flow cytometry (Figure 3.7a).

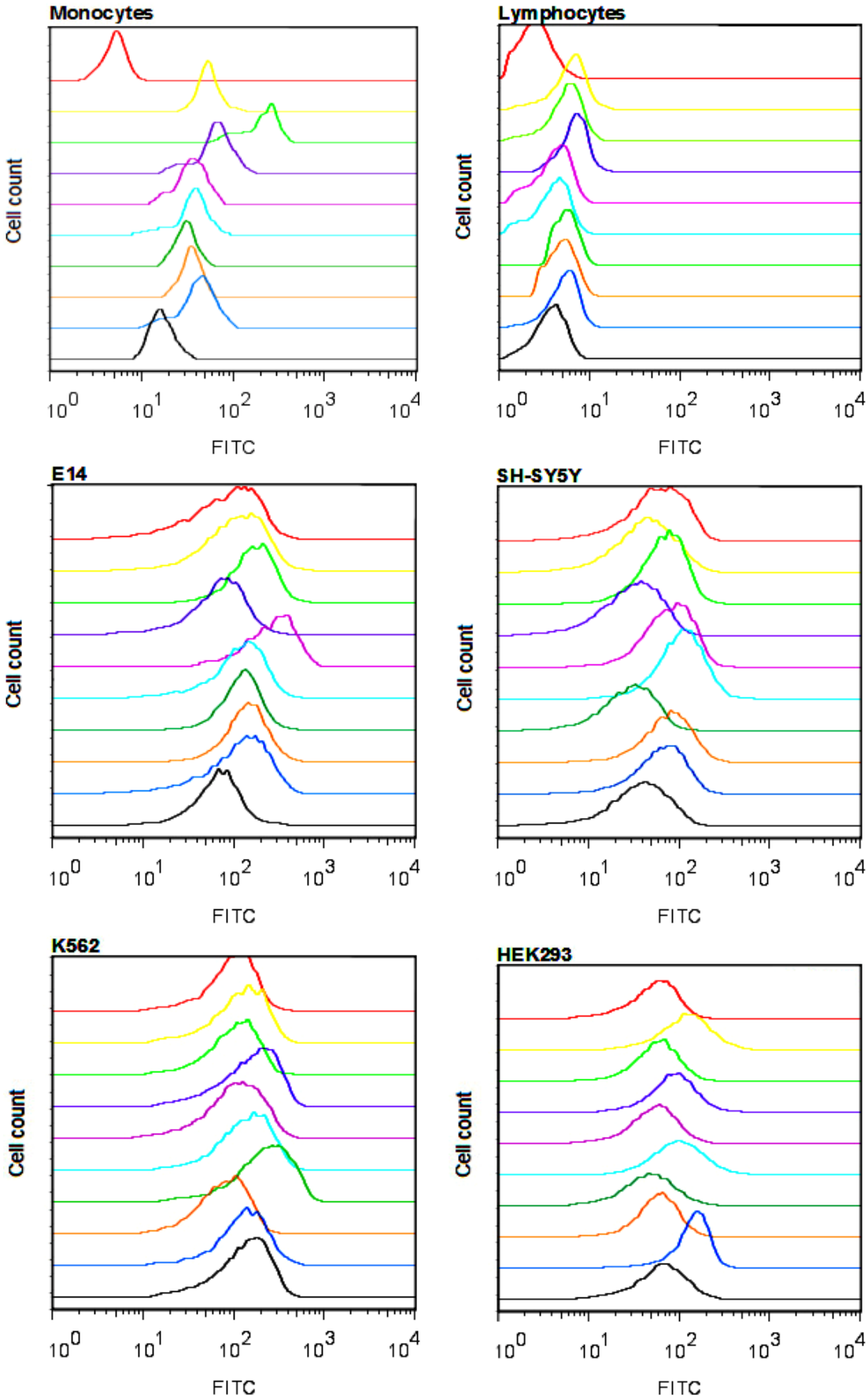
As expected, each of the "hit"-tetramers resulted in 100% uptake, i.e. the entire population is shifted in the histogram illustrating that all cells have internalised the respective peptide. In addition, a higher shift in fluorescence in the targeted cell type compared to the remaining 8 tetramers (Figure 3.7b).

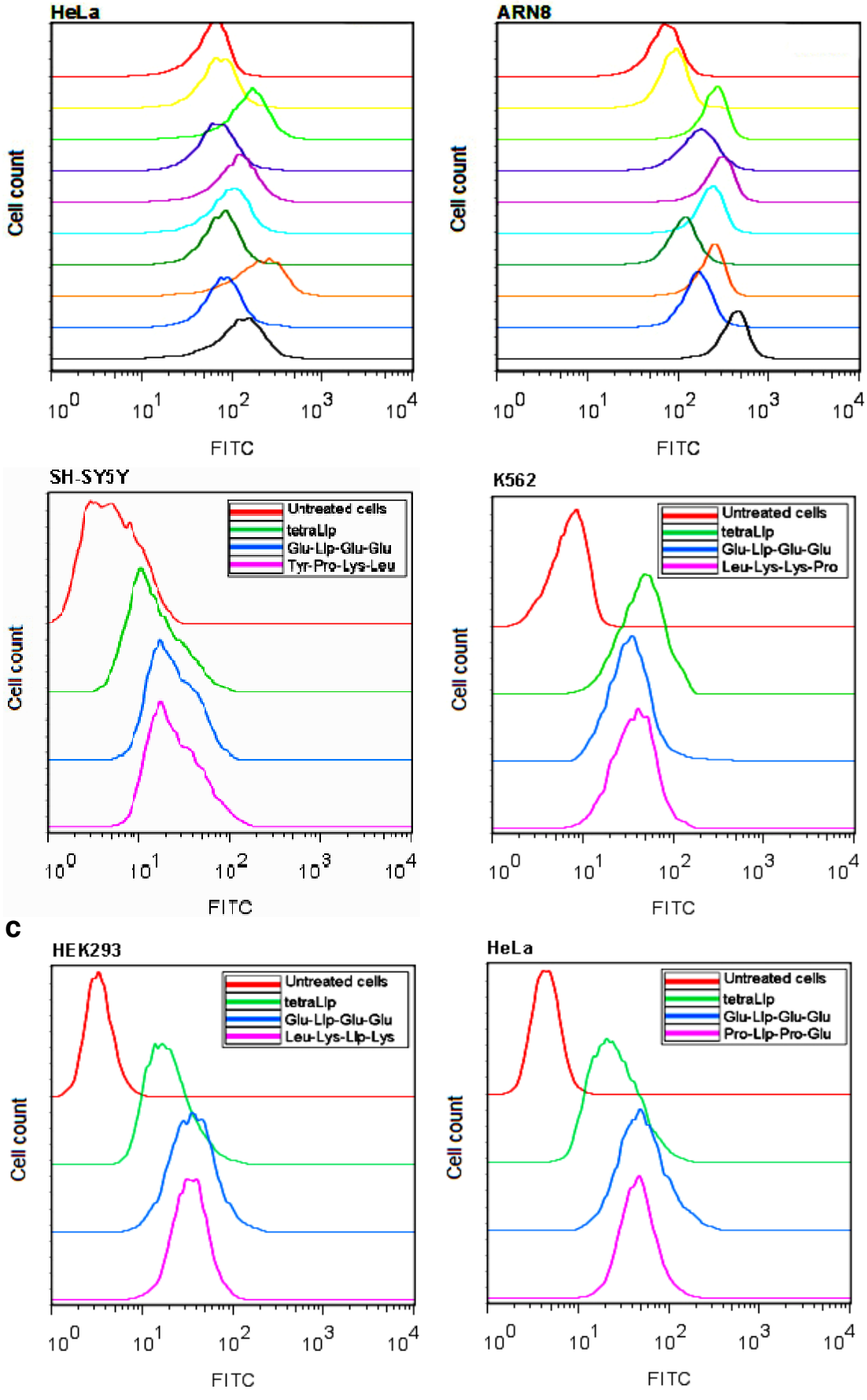
Comparison of each of the "hit"-tetramers across the cell types clearly demonstrated selectivity for the targeted cell type (Figure 3.8a). High-lighting this, X-Leu-Lys-Lys-Pro-NH_2 induced a 5-fold higher uptake in K562 cells compared to its uptake in all other cell types; X-Pro-Tyr-Glu-Glu-NH_2 showed a 4-fold higher uptake in ARN8 than in all other cell types; while X-Glu-Pro-Lys-Llp-NH_2 under went a 3-fold higher uptake in E14 than in all other cell types (Figure 3.8a); wherease the remaining "hit"-tetramers showed less selectivity for the targeted cell type (Figure 3.8a).

The myeloid leukaemia cell line, K562 has been reported to share some similar phenotypic characteristics to primary monocytes[184,185]. Selective uptake between circulating malignant and primary immune cells would be of great interest in targeting haematological cancers. In this regard, X-Leu-Lys-Lys-Pro-NH_2 uptake in K562 cells was 6-fold higher than uptake in primary monocytes (Figure 3.8a).

The uptake of some of the more unexpected "hit"-tetramers (X-Tyr-Pro-Lys-Leu, X-Pro-Llp-Pro-Glu-NH_2, and X-Pro-Tyr-Glu-Glu-NH_2) were directly compared to the consensus sequence, X-Glu-Llp-Glu-Glu-NH_2 and the positive control, X-Llp-Llp-Llp-Llp-NH_2 (Figure 3.7c and Figure 3.8b). The "hit"-tetramers had uptake levels equal to or higher than the positive control and also the consensus sequence in SH-SY5Y and HeLa cells (Figure 3.7c and Figure 3.8b). X-Pro-Tyr-Glu-Glu-NH_2 had an uptake lever equal to or slightly lower than the positive control and the consensus sequence in ARN8 cells (Figure 3.7c and Figure 3.8b). This comparison illustrates that these "hit"-tetramers are not only selective but also very efficient cell delivery agents.




Monocytes
Lymphocytes
E14
SH-SY5Y
K562
HEK293
Cell count
FITC
10^0
10^1
10^2
10^3
10^4


HeLa
ARN8
SH-SY5Y
K562
HEK293
HeLa
Cell count
FITC
Untreated cells
tetraLip
Glu-Lip-Glu-Glu
Tyr-Pro-Lys-Leu
Leu-Lys-Lys-Pro
Leu-Lys-Lip-Lys
Pro-Lip-Pro-Glu
C

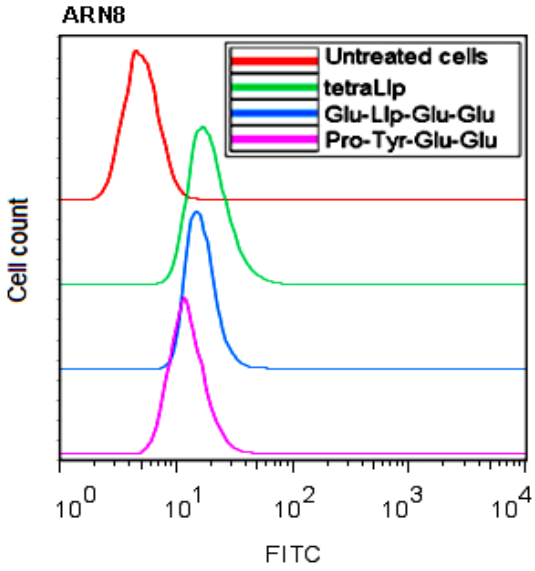


Figure 3.7: Flow cytometry analysis of the fluorescein-labelled "hit"-tetramers (X-AA_4-AA_3-AA_2-AA_1-NH_2 written as AA_4-AA_3-AA_2-AA_1) using a FACS Aria cytometer with FITC and PE-Texas-Red filters (**a**-**c**, 10,000 populations, n=3,) or a FACS Calibur cytometer with FITC and PE/PI filters (**d**, 10,000 populations, n=3). Any extracellular peptide was released by treatment with trypsin (which also served to detached the cells) while trypan blue was used to quench any extracellular fluorescence[181], ensuring that any observed increase in fluorescence was from just the intracellular peptide. (**a**) Gating strategy for whole blood samples selecting populations of live single lymphocytes, monocytes and neutrophils. FITC filtered flow cytometry histograms gated for live, single cells following incubation with propidium iodide (selectively stains the nuclei of necrotic and apoptotic cells)[182] to exclude dead cells as well as (**b**) cell selective "hit"-tetramers and (**c**) "hit"-tetramers and the positive control: X-Llp-Llp-Llp-Llp-NH_2 (written as tetraLlp).

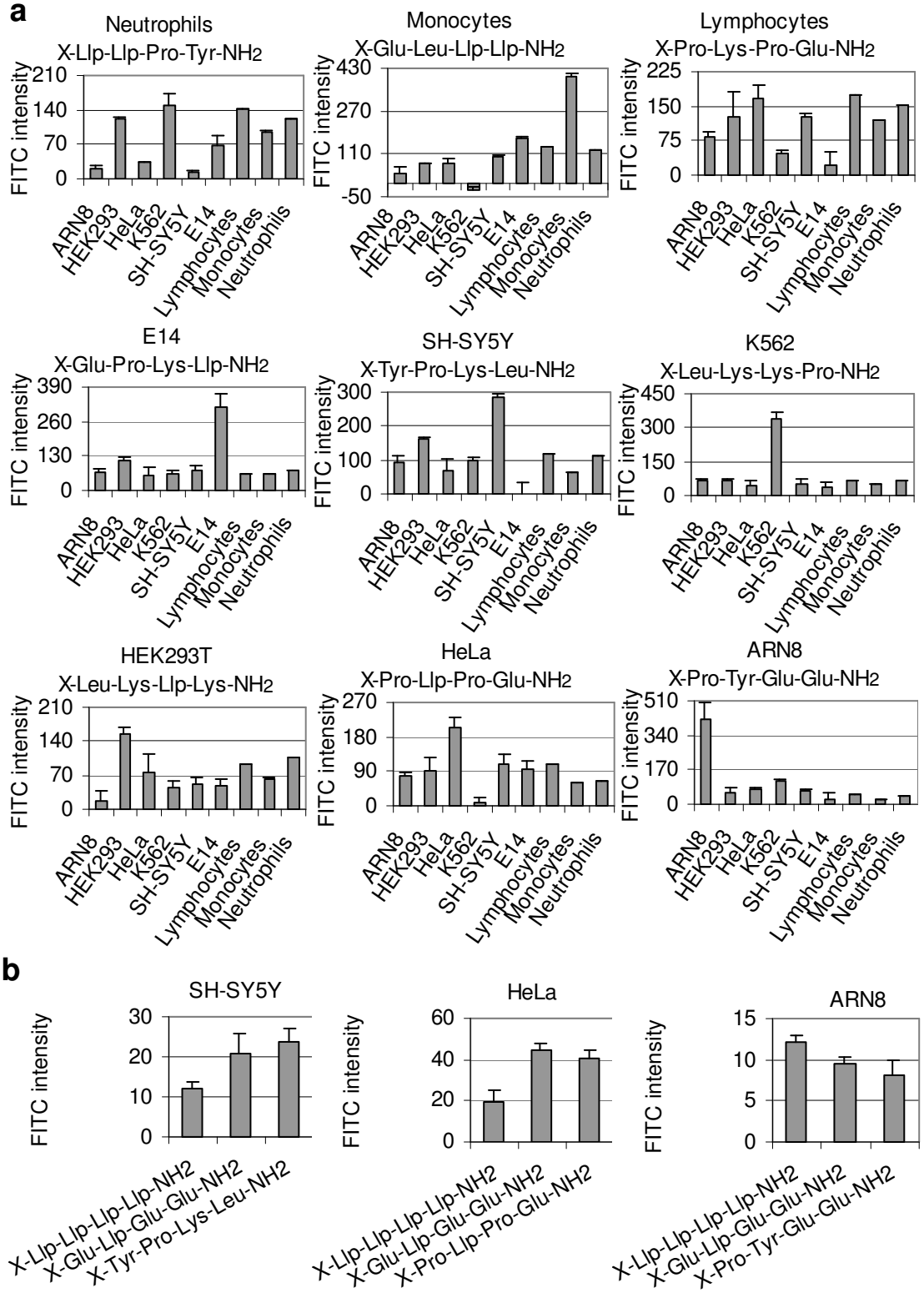
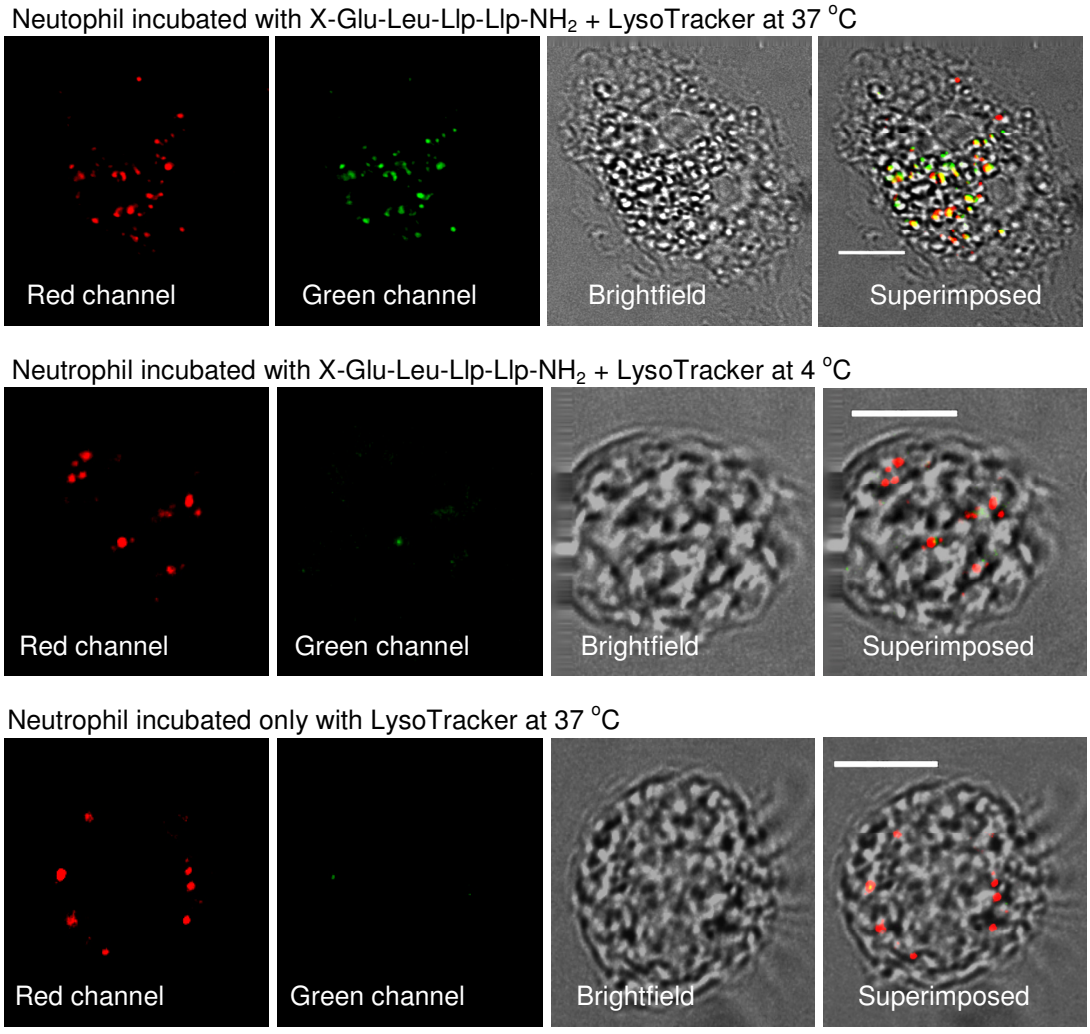


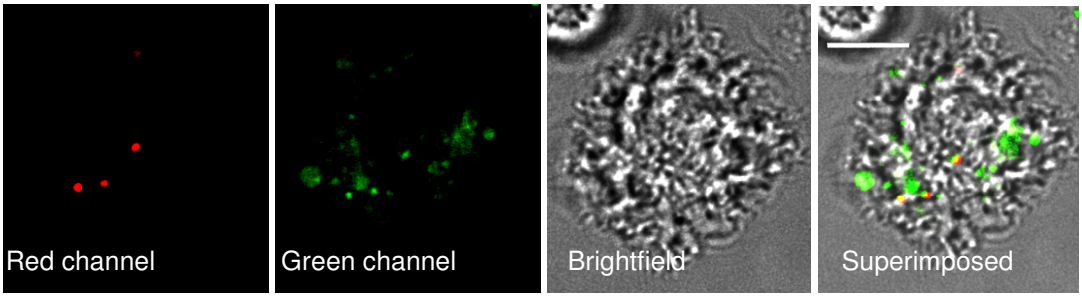
Figure 3.8: (**a**) Mean FITC filtered fluorescence of histograms versus the cell selective "hit"-tetramers (X-AA_4-AA_3-AA_2-AA_1-NH_2). (**b**) Mean FITC filtered fluorescence of histograms versus the generic "hit"-tetramer compared to the cell selective "hit"-tetramers and positive control, X-Llp-Llp-Llp-Llp-NH_2. Error bars indicate ± s.d.

3.4 Tetramer delivery is mediated by endocytosis

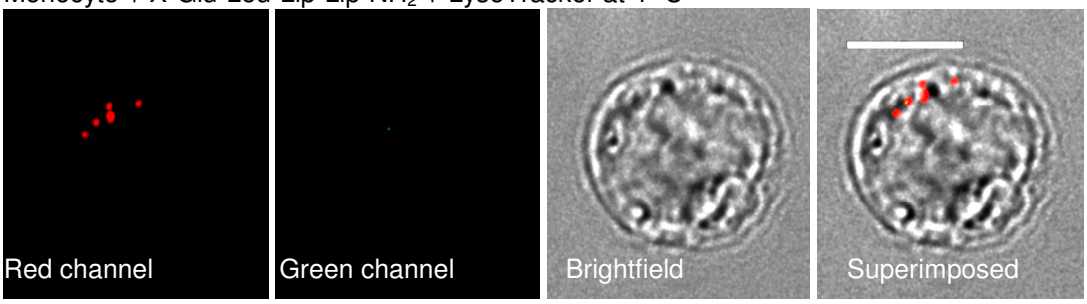
To study compartmentalisation, the cells were incubated with their respective 'hit' tetramer and/or stained with LysoTracker red, and analysed by confocal microscopy. As a negative control for endocytosis, cells were also incubated with their respective 'hit' tetramer and LysoTracker at 4 °C. Superimposed images of the red and green channels revealed co-localisation of the fluorescein label and LysoTracker red (yellow) illustrating that all the "hit"-tetramers incubated at 37 °C were compartmentalised into the lysosome. In contrast, only the LysoTracker fluorescence in the red channel and no fluorescence in the green channel were observed at 4 °C illustrating that internalisation of "hit"-tetramers was diminished by inhibition of active cellular uptake processes (Figure 3.9). Lysosome localisation and active-transport dependence verifies that the all the "hit"-tetramers were taken up by endocytosis (Figure 3.9).



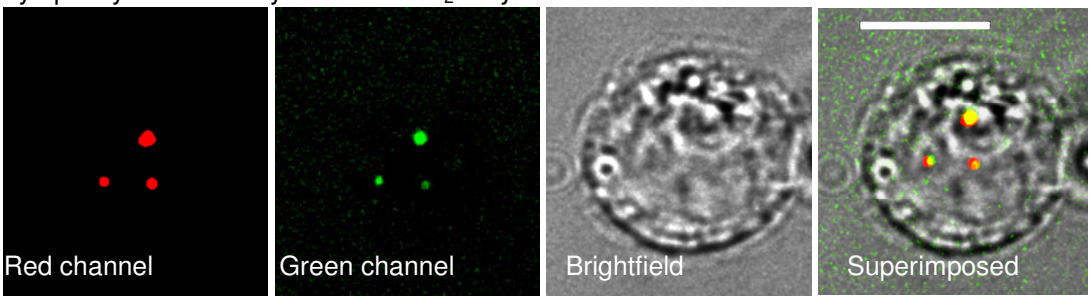
Monocyte + X-Glu-Leu-Llp-Llp-NH_2 + LysoTracker at 37 °C



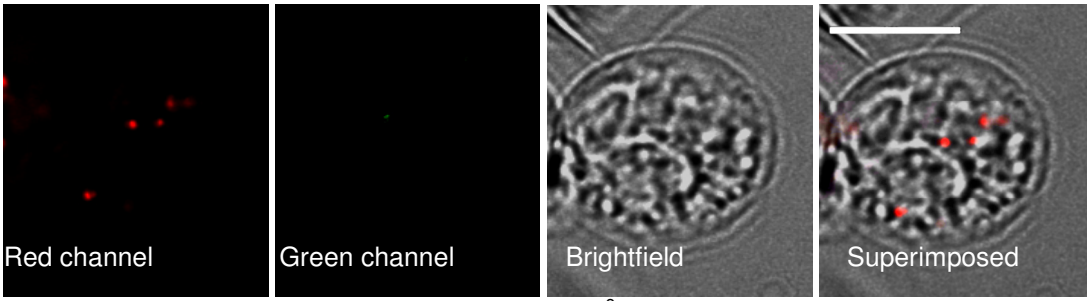
Monocyte + X-Glu-Leu-Llp-Llp-NH_2 + LysoTracker at 4 °C



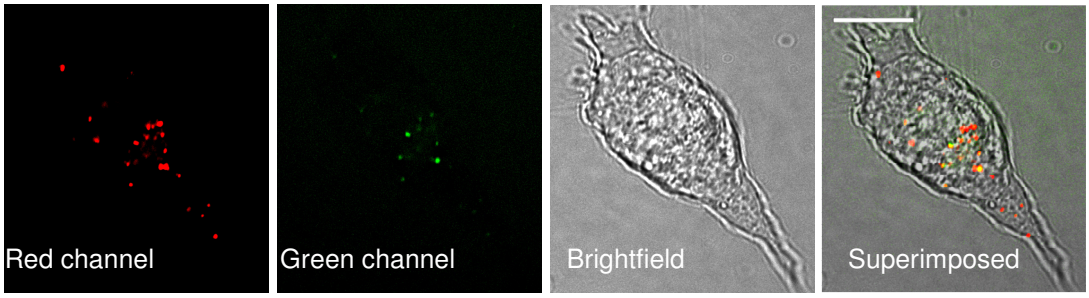
Lymphocyte + X-Pro-Lys-Pro-Glu-NH_2 + LysoTracker at 37 °C



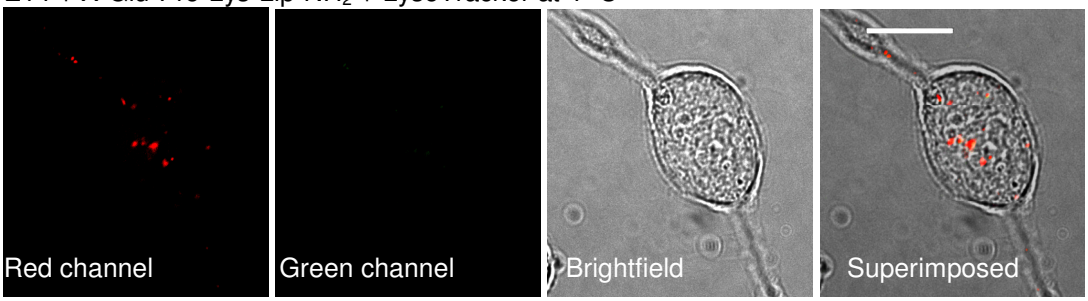
Lymphocyte + X-Pro-Lys-Pro-Glu-NH_2 + LysoTracker at 4 °C



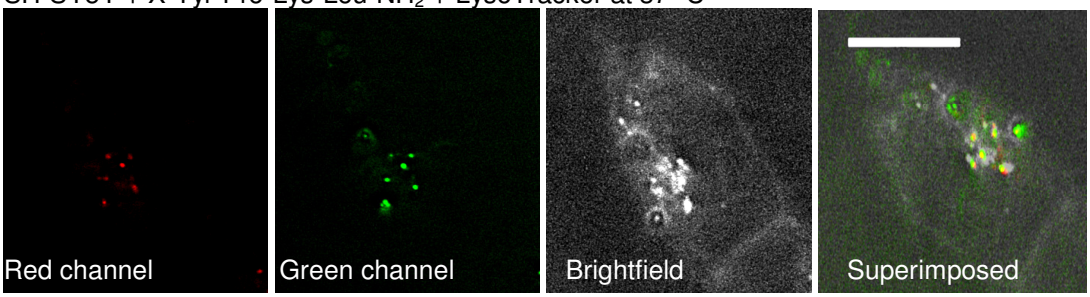
E14 + X-Glu-Pro-Lys-Llp-NH_2 + LysoTracker at 37 °C



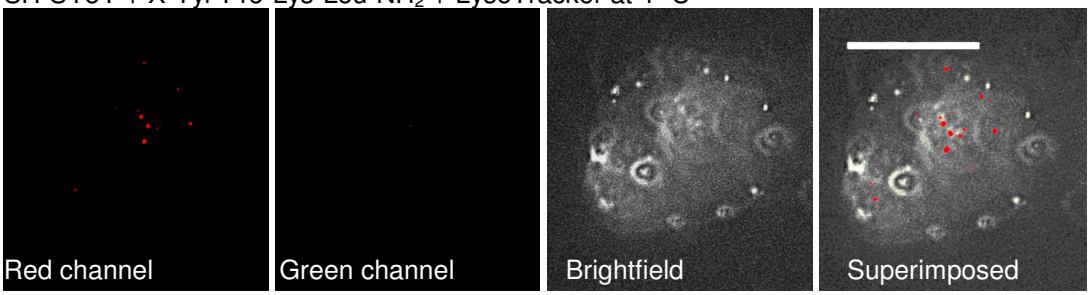
E14 + X-Glu-Pro-Lys-Llp-NH_2 + LysoTracker at 4 °C



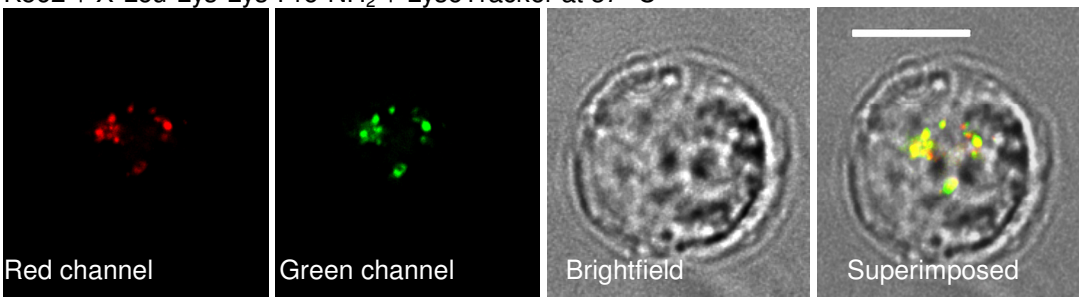
SH-SY5Y + X-Tyr-Pro-Lys-Leu-NH_2 + LysoTracker at 37 °C



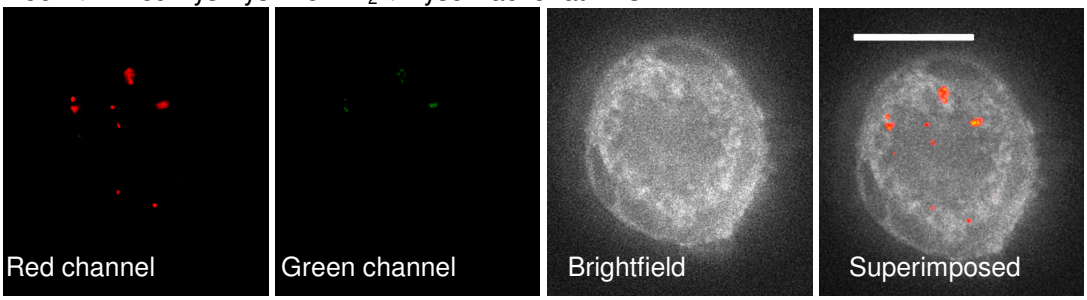
SH-SY5Y + X-Tyr-Pro-Lys-Leu-NH_2 + LysoTracker at 4 °C



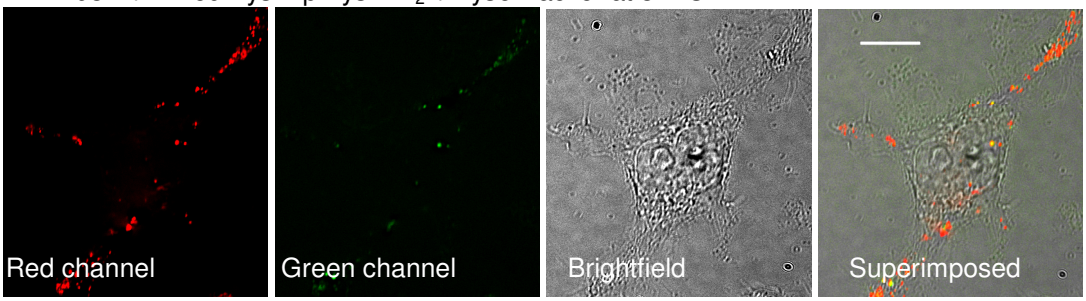
K562 + X-Leu-Lys-Lys-Pro-NH_2 + LysoTracker at 37 °C



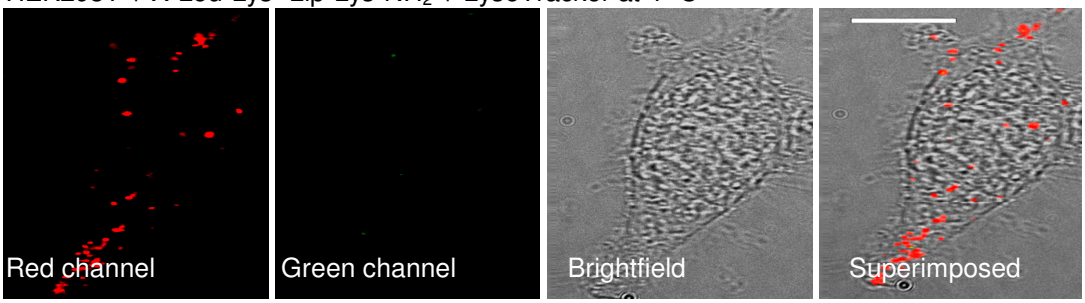
K562 + X-Leu-Lys-Lys-Pro-NH_2 + LysoTracker at 4 °C



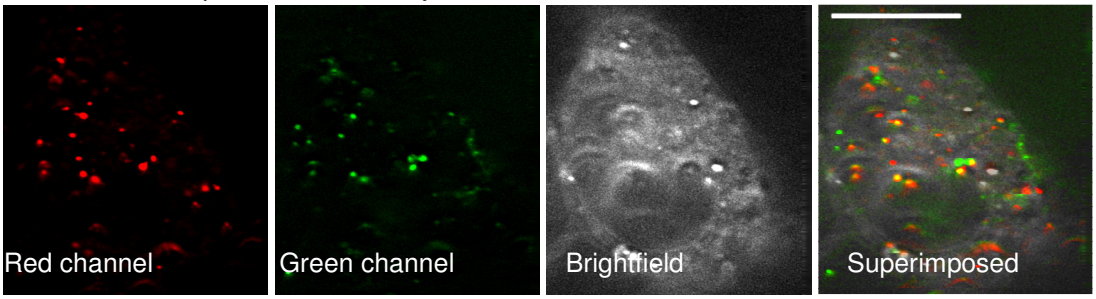
HEK293T + X-Leu-Lys-Llp-Lys-NH_2 + LysoTracker at 37 °C



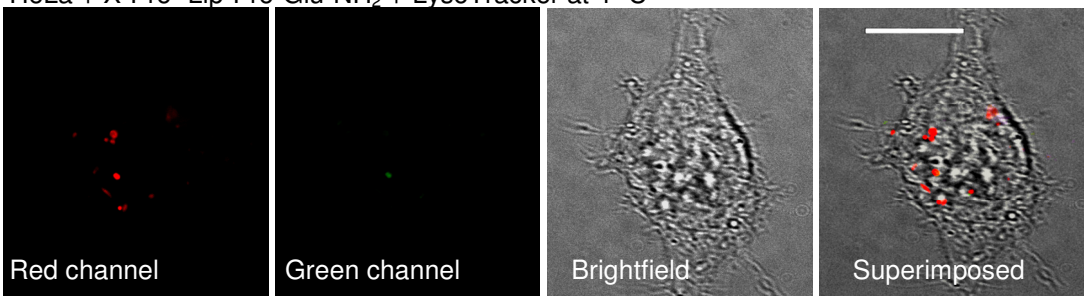
HEK293T + X-Leu-Lys- Llp-Lys-NH_2 + LysoTracker at 4 °C



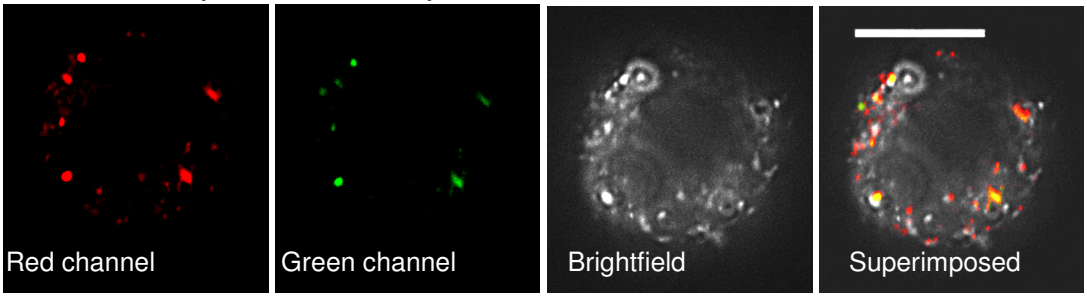
HeLa + X-Pro- Llp-Pro-Glu-NH_2 + LysoTracker at 37 °C



HeLa + X-Pro- Llp-Pro-Glu-NH_2 + LysoTracker at 4 °C



ARN8 + X-Pro-Tyr-Glu-Glu-NH_2 + LysoTracker at 37 °C



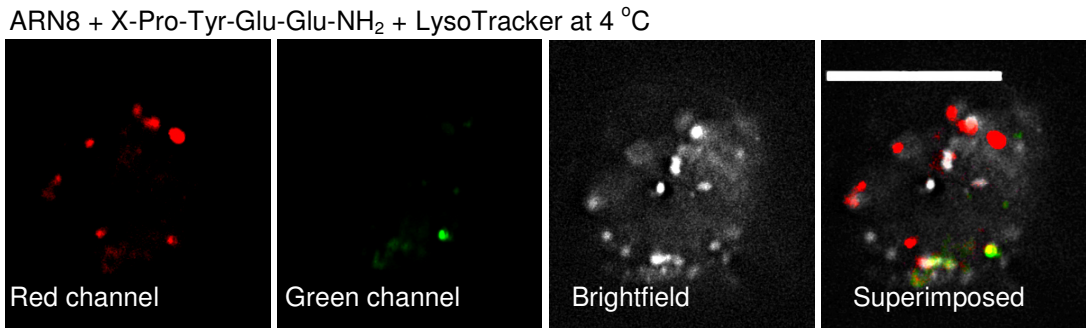


Figure 3.9: Selected Z-stacks of cells incubated with FAM-labelled "hit"-tetramer ($X\text{-}AA_4\text{-}AA_3\text{-}AA_2\text{-}AA_1\text{-}NH_2$), followed by staining with LysoTracker Red, which specifically stains the lysosome. Confocal microscopy was performed using a Deltavision Tollervey microscope with FITC and TRITC filters. Red = LysoTracker red fluorescence, green is the fluorescein-labelled peptide. Yellow/orange colour is co-localisation of red and green fluorescence. Scale bars = 5 µm.

3.5 Cytotoxicity

In order for the peptides to have biological applications it is vital that they do not exhibit cellular toxicity. Cell viability upon treatment with the "hit"-tetramers ($X\text{-}AA_4\text{-}AA_3\text{-}AA_2\text{-}AA_1\text{-}NH_2$) was assessed with (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT)[186] assays. In living cells MTT is reduced to the purple formazan by mitochondrial reductase. As this enzyme is degraded in dead cells the colour intensisty is an expression of cell viability.

Untreated cells were assumed to be 100% viable and the viability of peptide-treated cells was expressed as percentage live cells compared to untreated cells. All the "hit"-tetramers exhibited no cytotoxicity in all the tested cell types at 10 times the concentration of agent used for cell delivery (100 µM; Figure 3.10a).

Additionally membrane toxicity was assessed using erythrocyte haemolysis assays. These demonstrated no evidence of haemolysis at 10 times the concentration of agent used for cell delivery (100 µM; Figure 3.10b).

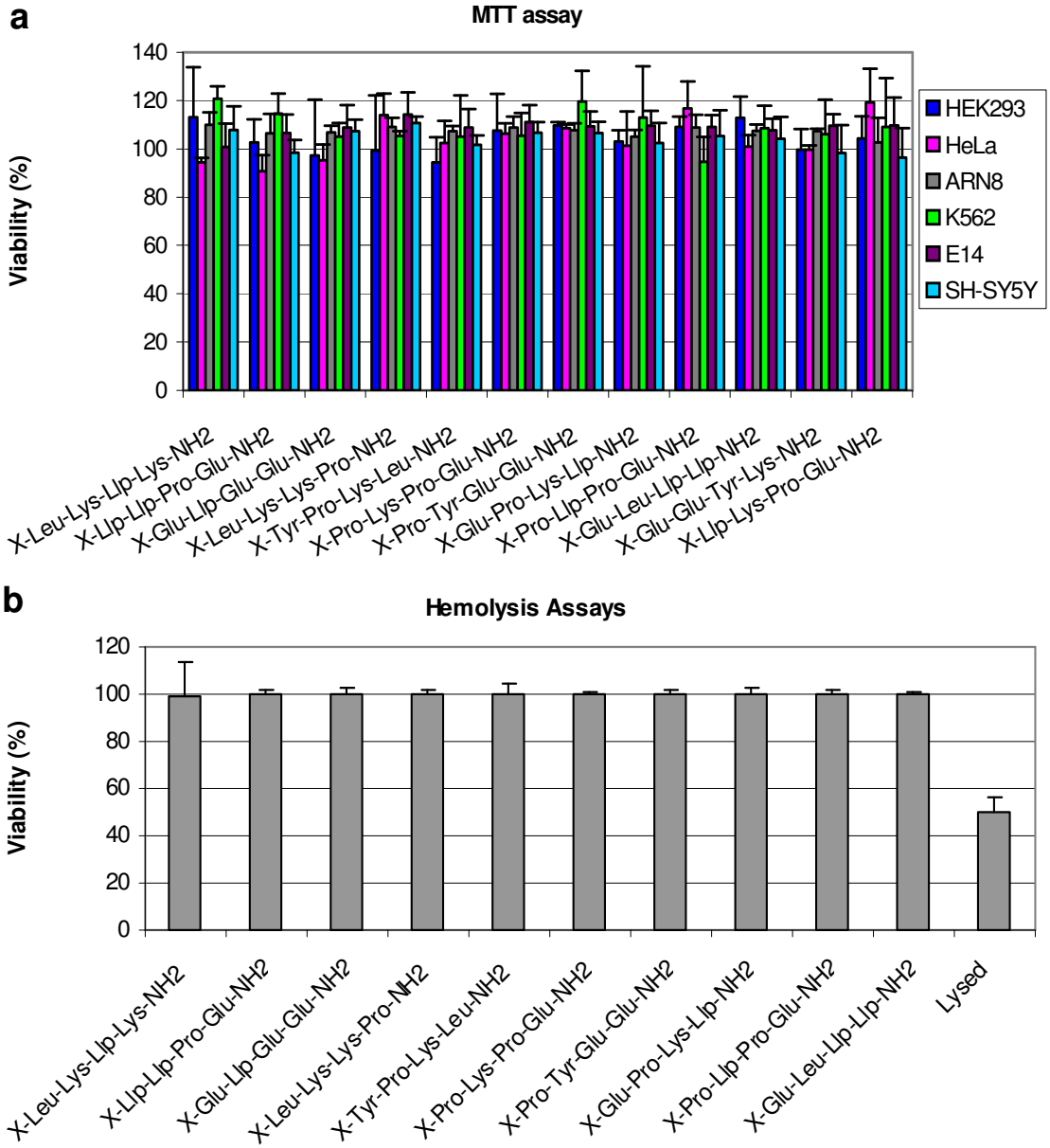


Figure 3.10: (**a**) Cytotoxicity of hit-peptides assessed as cell viability by MTT assays (cell lines and stem cells). Cells were incubated with 100 µM "hit"-tetramer (X-AA_4-AA_3-AA_2-AA_1-NH_2; 24 h) the media was changed and the incubation continued for 24 h and absorbance was measured at 580 nm. Untreated cells are assumed to be 100% variable. (**b**) Haemolysis assays showing no evidence of haemolysis at 100 µM of 'hit' tetramers on human erythrocytes. Untreated cells are assumed to be 100% variable. Error bars indicate ± s.d. n = 3.

3.6 Conclusions

Microarray analysis of a 1296 member PNA encoded peptide library identified an efficient generic delivery agent (Glu-Llp-Glu-Glu) for cervix epitheloid carcinoma (HeLa), erythroleukemic (K562), embryonic kidney cancer (HEK293T),

neuroblastoma (SH-SY5Y), amelanotic melanoma (ARN8) cells, stem cells (E14), and human primary lymphocytes, monocytes, and neutrophils.

Clustering analysis of the microarray data identified selective delivery agents for all of the above cell types, excellent selectivity predicted for neutrophils and lymphocytes, moderate selectivity predicted for monocytes, SH-SY5Y, and HEK293T cells, and little selectivity predicted for E14, K562, HeLa and the ARN8 cells.

Flow cytometry analysis of the identified "hit"-peptides showed that X-Leu-Lys-Lys-Pro-NH_2 induced a 5-fold higher uptake in K562 cells compared to its uptake in all other cell types; X-Pro-Tyr-Glu-Glu-NH_2 showed a 4-fold higher uptake in ARN8 than in all other cell types; while X-Glu-Pro-Lys-Llp-NH_2 under went a 3-fold higher uptake in E14 than in all other cell types; whereas the remaining "hit"-tetramers showed less selectivity for the targeted cell type.

Comparison of the uptake of some of the more unexpected "hit"-peptides (X-Tyr-Pro-Lys-Leu, X-Pro-Llp-Pro-Glu-NH_2, and X-Pro-Tyr-Glu-Glu-NH_2) with the consensus sequence, X-Glu-Llp-Glu-Glu-NH_2 and the positive control, X-Llp-Llp-Llp-Llp-NH_2 revealed that the "hit"-peptides had uptake levels close to or higher than the positive control and the consensus sequence in SH-SY5Y, HeLa, and ARN8 cells. This illustrates that these "hit"-peptides are not only selective but also very efficient cell delivery agents.

Confocal microscopy revealed that the "hit"-peptides were taken up by a lysosome pathway, which verifies an endosomal mechanism for transfection of PNA by small peptide. The observed selectivity indicates that receptor mediated endocytosis is the mechanism for uptake of these peptides. Furthermore, the identified peptides showed no toxicity in the tested cell types.

Noticeably, selective uptake between circulating malignant and primary immune cells, which would be of great interest in targeting haematological cancers, was

observed with 6-fold higher uptake of Leu-Lys-Lys-Pro in K562 cells compared to uptake in primary monocytes.

This approach establishes a general strategy for the identification of small peptides as tools/reagents that allow selective delivery into any desired cell type. In addition, this technology presents and efficient and targeted delivery of PNA, this can enhance PNA's antigene and antisense applications. This technology is not only limited to PNA delivery and will find applications in targeted delivery of potential drugs that do not have properties previously expected of drugs such as small size and cell permeability.

Chapter 4 Screening of an encoded peptide library to discover new cell binding ligands

4.1 Introduction to receptor mediated cellular attachment

Through cell signalling pathways the integrins and GPCRs govern many vital cellular processes (section 1.4); thus integrins are involved in cell growth, division, apoptosis, survival, differentiation as well as viral binding and entry into cells.[114,123] GPCRs are involved in a wide range of functions[115] such as binding of neurotransmitters regulating the nervous system controlling blood pressure, heart rate, and digestive processes, binding of chemokines regulating inflammatory responses, binding of opsins (visual sense) and odorants and pheromones (sense of smell), as well as sensing and regulating cell density (section 1.4). These essential biological functions make Integrins and GPCRs amongst the most heavily investigated drug targets in the pharmaceutical industry[115,187,188]. Thus, identification of new ligands could lead to the possibility of improved therapeutic targeting against these receptors.

4.1.1 Previously reported integrin binding assays

Several binding assays for integrins have been described to develop ligand arrays for integrin receptors and include the work by C. M. Niemeyer[189], who recently developed a platform for DNA-directed immobilization of peptides, which allowed integrin mediated adhesion of fibroblast cells on RGD-features. Moreover, this method enables implementation of DNA microarrays as decoding tools in combinatorial synthesis and screening of cell surface receptor ligands.[189] Simultaneously in another example of integrin binding assays, L.L. Kiessling *et al.* developed peptide surfaces that support human embryonic stem cell growth and self-renewal; although, no receptors responsible for the cell adhesion could be identified.[190] However, Niemeyer and Kiessling demonstrated that peptide arrays supporting cell growth can used to identify ligands inducing cell migration, proliferation, and differentiation.

C. Lethias *et al.* mapped cell adhesion sites and identified integrin receptors by cell binding to tenascin-X[191]. This demonstrated induction of cell attachment via binding of RGD (in tenascin-X) to integrin receptors.

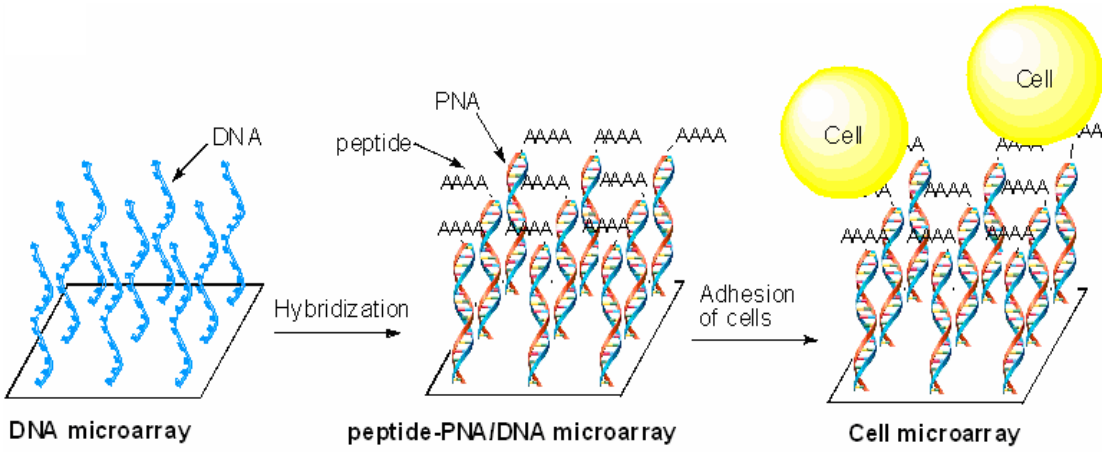
K. S. Lam. *et al*. identified peptide ligands for cell surface receptors by the adhesion of human lung cancer cells onto a bead-based combinatorial library that allowed identification of a consensus sequence (cNGRGEQc) for the $\alpha_3\beta_1$ integrin.[192] This work established the strategy of screening libraries for cell surface receptor binding; however, it lacks the advantages of high-throughput screening of an array. Y. Yamada and M. Nomizu *et al*. identified several metastasis promoting sequences by screening potential integrin peptide ligands.[193,194]

L. G. Griffith *et al*. computed the effect of clusters of integrin ligands on integrin signaling and linkage between integrins and intracellular structures.[195] This demonstrated that higher ligand densities increase adhesion strength by locally concentrating receptor in the membrane.

4.2 Strategy of cell membrane-receptor screening

Expanding on the previously reported integrin binding assays the aim of this project was to develop a microarray based high-throughput screening method for the identification of peptide ligands for cell surface receptors. This would achieve screening of human cells in a selective "cell-growth-on-microarray" design[189] (Scheme 4.1), allowing information on the interaction between the peptide-ligands and the cell surface receptors to be extracted on a single spot-by-spot (compound-by-compound) basis. This technique would facilitate screening with cells with variations in surface-receptor ligand densities and/or receptor expression.

Initially the PNA-encoded peptides Arg-Gly-Glu (RGD) and Leu-Asp-Val (LDV), which are perhaps the most studied integrin ligands[118], well as Gly-Gly-Gly (GGG) were synthesised by microwave aided PNA and peptide couplings (Chapter 2). The PNA-tags were hybridised to complementary DNA microarrays specially designed with a “cell-repellent” surface and screened against HeLa cells (human epithelial cervix cancer). In addition, ligand binding was verified by flow cytometry. The overall strategy is shown in Scheme 4.1.



Scheme 4.1: Strategy of cell based microarray screening of a library of PNA-encoded peptide ligands for binding to cell surface receptors.

4.3 Cell microarray nomenclature

The following nomenclature is assigned to cell binding on microarrays: specific binding refers to the attachment of cells only in printed areas, i.e. no unspecific binding to the microarray surface; selective binding refers to the attachment of cells only to receptor ligands, i.e. no unselective binding to negative controls.

4.4 Production of cell specific microarray surface

peptide-PNA/DNA microarrays were manufactured by contact printing amino modified DNA complementary the RGD-, LDV- and GGG-encoding PNA (Chapter 2), TCTTA–PNA (negative control), and non-coding fluorescent DNA (positive control for printing and negative control for cell binding) on Codelink microarrays. A range of reaction conditions and compounds were tested to generate a novel cell repellent microarray surface (data not shown) and this was achieved by fabrication of fluorinated glass surfaces by capping unreacted active esters on the Codelink surface with perfluorooctyl propylamine.

The PNA-encoded peptides were hybridised to their complementary partners on the array prior to screening with cells followed by high content microscopy and fluorescent scanning. This revealed a highly selective surface with attachment on RGD and LDV features with no cell attachment observed outside the printed regions

(Figure 4.1). Following adhesion, cells were observed to spread over the entire printed area of 80% of all RGD features, 70% of all LDV features, and 55% all GGG features (72 h, Figure 4.1). Slight cell attachment was also observed 10% of all features of non-coding DNA and PNA (72h, Figure 4.1). Following analysis the cells could be removed from the array by sonication in 10 mM Tris buffer (pH 8) for 10 min, a gentle cell stripping procedure that removes all attached cells without removing the PNA or damaging the DNA microarray, allowing the cell binding microarrays to be re-used.

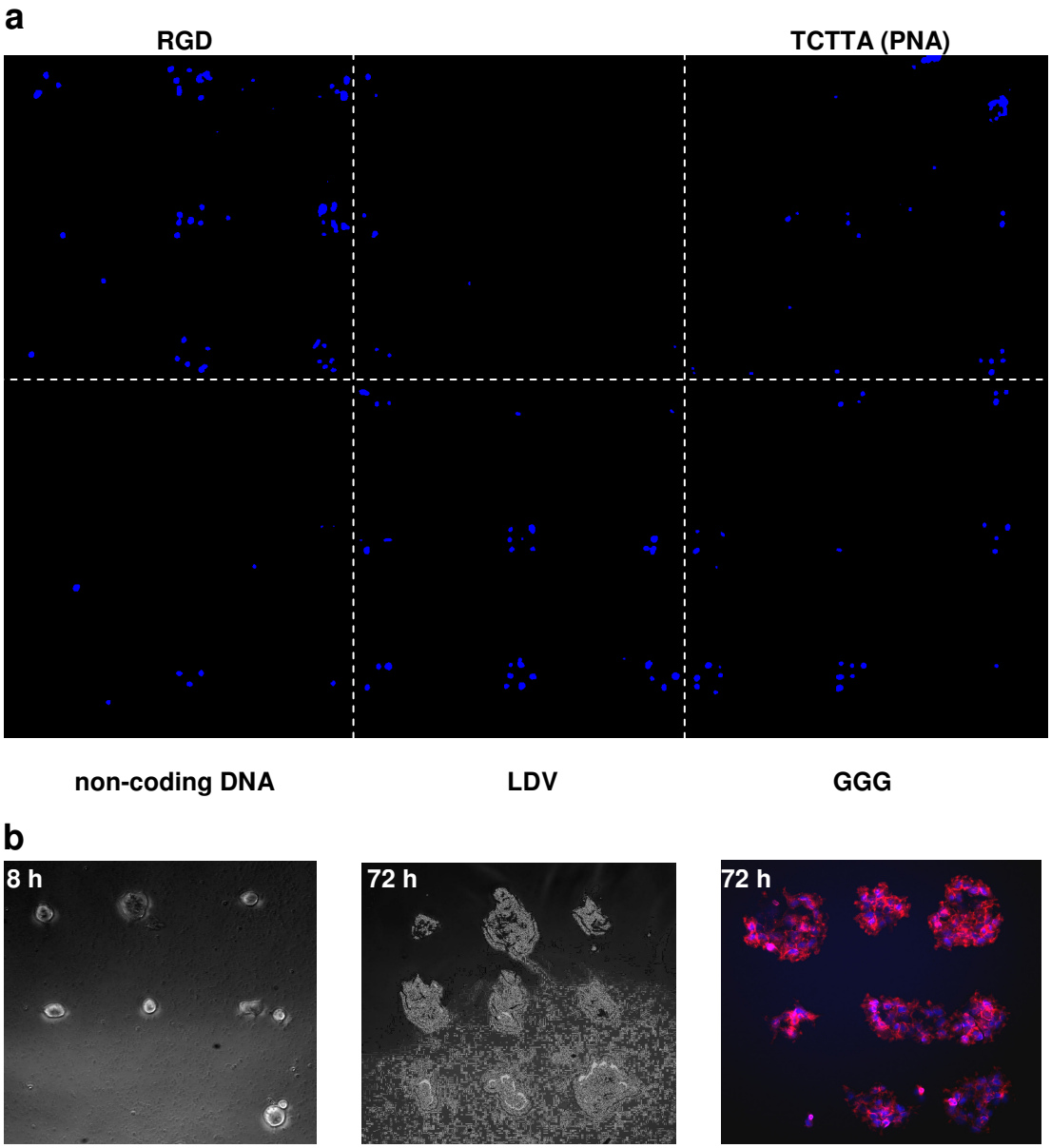


Figure 4.1: Microarray analysis of cell binding onto PNA-encoded peptides. DNA microarrays designed with 3 x 3 x 3 x 3 matrices of DNA oligonucleotides

complementary to and hybridised with RGD-PNA, LDV-PNA, GGG-PNA, TCTTA (PNA, negative control), non-coding DNA (negative control) and fluorinated surface modification. (**a**) HeLa cells were incubated 72 h and the cells fixed (4% paraformaldehyde) and stained with DAPI (blue) and phalloidin (red). (**b**) Brightfield and DAPI and phalloidin images of RGD features at 8 and 72 h.

4.4.1 Verification of integrin-ligand binding by flow cytometry

HeLa cells were analysed by flow cytometry to measure the relative binding affinity and to verify surface binding of the fluorescent peptide-PNA conjugates. Histograms of cell count vs. fluorescein isothiocyanate (FITC) fluorescence were created based on scatter plots gated for single, FITC-only positive, live cells (Figure 4.2). A similar increase in fluorescence was observed with the RGD-PNA and GGG-PNA treated cells compared to the untreated cells, whereas, LDV-PNA caused a slightly smaller increase in fluorescence. The relatively high level of fluorescence observed with GGG-PNA may be due to cellular uptake and/or unspecific cell surface interactions rather than integrin dependent binding. This would also explain the discrepancy between flow cytometry and microarray screening, where GGG features showed less cellular binding than RGD and LDV. Trypsin treatment after RGD-PNA incubation reduced the fluorescence signal to that of the untreated cells (Figure 4.2). Trypsin degrades surface receptor moieties and any bound peptide, which shows that the RGD-PNA is surface bound and not taken up by the cells.

The cytotoxicity was evaluated by comparing the percentage of dead cells during flow cytometry in untreated cells and peptide-PNA conjugate treated cells. The relatively large number of dead cells (12%) in the untreated sample is expected due to the harsh detachment by scraping. RGD-PNA and LDV-PNA induced limited cytotoxicity (22% dead cells), whereas GGG-PNA had a greater cytotoxicity (26% dead cells).

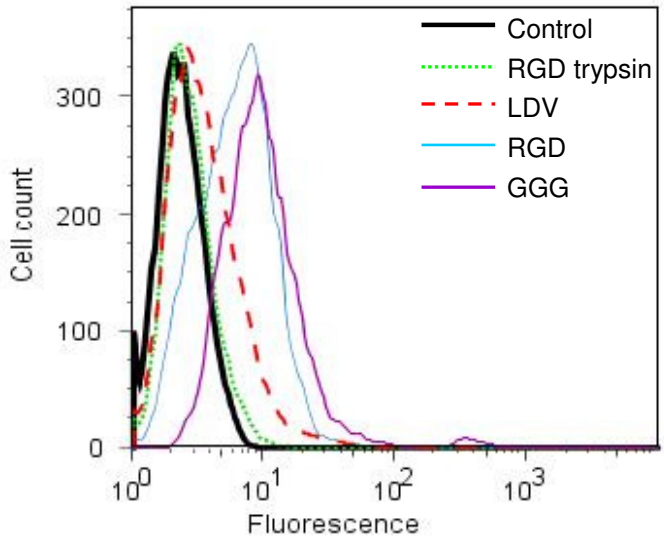


Figure 4.2: Histograms of number of fluorescent cells versus fluorescence intensity of HeLa cells after incubation with 10 µM, RGD-, LDV-, GGG-PNA conjugates in DMEM CM + $CaCl_2$ (50 µM) and $MgCl_2$ (50 µM) at 37 ^oC for 30 min.

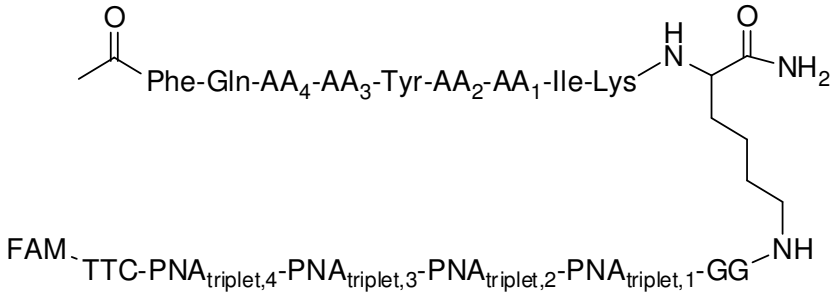
4.4.2 Conclusion

The fluorinated surface provides a specific microarray platform for receptor mediated cell binding. This method of sorting of a library by the specific DNA and PNA-amino acid hybridisation enables screening of thousand member libraries with internal controls. In addition, this approach facilitates the study different cell types simultaneously and thereby screening for differences in surface-receptor ligands and/or receptor expression between diseased and normal cells or between various cell types.

4.5 High throughput screening of receptor mediated cellular binding on microarray

The aim of this project was screening of an immobilised 10,000 member PNA-encoded peptide library (Library-2[29], Scheme 4.2) with human cells in a selective "cell-growth-on-microarray" design[189]. The PNA encoded peptide library was based on sequences reported by Y. Yamada *et al.*[193] and designed to target integrins as well as kinases. Y. Yamada screened 208 overlapping peptides covering the short and

long arms of mouse laminin α_1 and found Phe-Gln-Val-Ala-Tyr-Ile-Ile-Ile-Lys (FQVAYIIIK) demonstrating strong B16-F10 cell adhesion.[193] Consequently, Library-2 consisted of 10,000 nonapeptide-PNA conjugates with four variable positions each containing 10 different amino acids in the common sequence: Ac-Phe-Gln-AA_4-AA_3-Tyr-AA_2-AA_1-Ile-Lys-PNA_{18}-Fluorescein with each variable amino acid encoded by a PNA-triplet.[29] The arrays used each contained 4 x 44K features, allowing 16 copies of each library member to be analysed per array. Cell lines used were D54 ($\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrin overexpressing)[196,197] and HEK293T-CCR6 (CCR6 GPCR overexpressing)[134] allowing screening with cells with variations in surface-receptor ligand densities and/or receptor expression.



Scheme 4.2: Generic structure of Library-2. AA_1, AA_2, and AA_4 were Ile, Val, Phe, Pro, Arg, Glu, Lys, d-Pro, Ser or *d*-Val and AA_3 Ile, Val, Phe, Ala, Pro, Arg, Glu, Lys, *d*-Ala, Ser or Pro. FAM is 5(6)-carboxyfluorescein amide.

4.5.1 High throughput screening of cell/peptide-microarrays

Library-2 was hybridised onto custom DNA microarrays (Agilent, 44k x 4, Figure 4.3); which thereby renders the library of peptides available for receptor binding. Fluorescent microarray imaging and data analysis (BlueFuse; BlueGenome) confirmed the efficient hybridisation of the library onto the array (Figure 4.3). D64 and HEK293T-CCR6 cells were incubated with the PNA hybridised microarrays in serum-free media to reduce non-specific adsorption of cell adhesive proteins. To minimize interference from cell secreted proteins incubation times were kept short (2 h), while the surface of non-printed areas of the microarray was very hydrophobic surface and thus prevents cellular attachment and growth. The cells were fixed and nuclei stained with DAPI and fluorescent imaging was used to extract the intensity of the DAPI label (appendix III). This choice of fluorescent labels allows relative

quantification of the amount of peptide/PNA and cells (DAPI) on each spot, which enables accurate comparison of the receptor affinity for the peptides (Appendix III). Microarray imaging showed that attached D54 and HEK293T-CCR6 cells appeared to have a normal and healthy morphology (elongated and spread out shape, Figure 4.3).

4.5.2 "Hit" sequences allowing selective cell attachment

Raw microarray data was obtained from Bluefuse, which allows grid alignment and signal estimation, in a Microsoft Excel format. In Microsoft Excel the top ~5% and the bottom ~5% of each of the replicate-sets were removed as outliers (erroneous values caused by dust, scrapes etc[179,180]). The average fluorescence intensities were calculated over each of the remaining 14 replicates and over the non-complementary negative control features. The average intensities were corrected for the background by subtracting the average intensity of the non-complementary negative control features. Thereafter, the data was normalised for the PNA hybridisation by calculating the ratio between the background corrected average intensities of the DAPI-label and the FAM-label for each of the background-corrected averages (appendix III).

Scatter plots of the ratios of the DAPI and FAM background corrected average microarray intensities (cell binding) versus the variable amino acids for positions AA_4 to AA_1 were constructed (Figure 4.4) and showed that cells showed preference for some peptides over others and allowed consensus sequences for each cell line to be derived by combining the most preferred amino acids for each of the variable positions (Table 4.1). In the same manner, consensus sequences of non-binding peptides (negative controls) for each cell line were obtained.

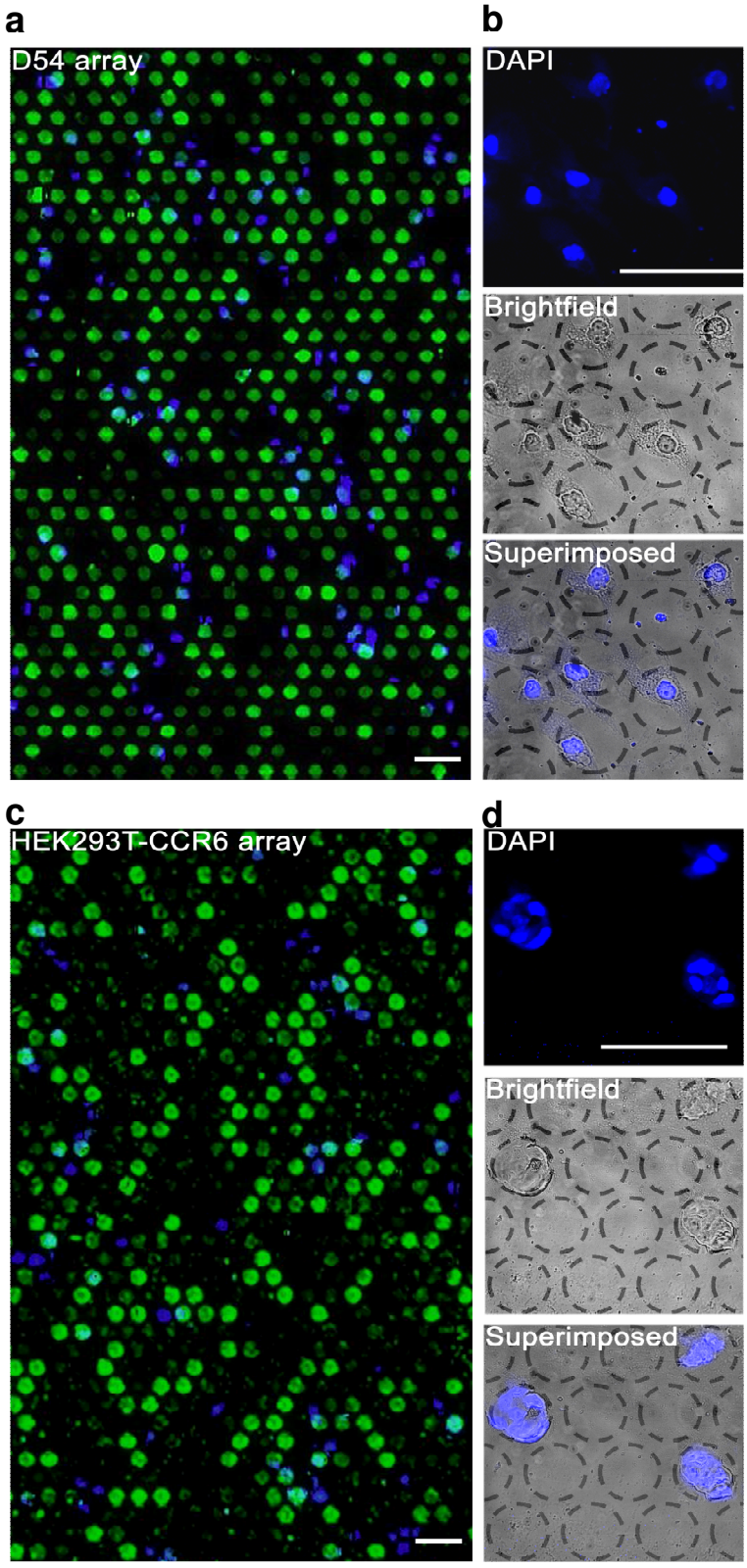


Figure 4.3: Microarray analysis of cell binding onto PNA-encoded peptides. DNA microarrays designed with 4 sub-arrays of 44,000 features each with 4 replicates of each oligonucleotide complementary to each member of the 10,000 member library as well as 4,000 non-coding negative controls. D54 and HEK293T-CCR6 cells incubated on Library-2 hybridised microarrays in serum free media (2 h) followed by

incubation in complete media (16 h) and fixation (4% paraformaldehyde) and nuclei staining with DAPI (blue). (**a**, **c**) FAM and DAPI fluorescent imaging and (**b**, **d**) Brightfield and DAPI fluorescent imaging of D54 and HEK293T-CCR6 cell arrays. Scale bars = 20 µm.

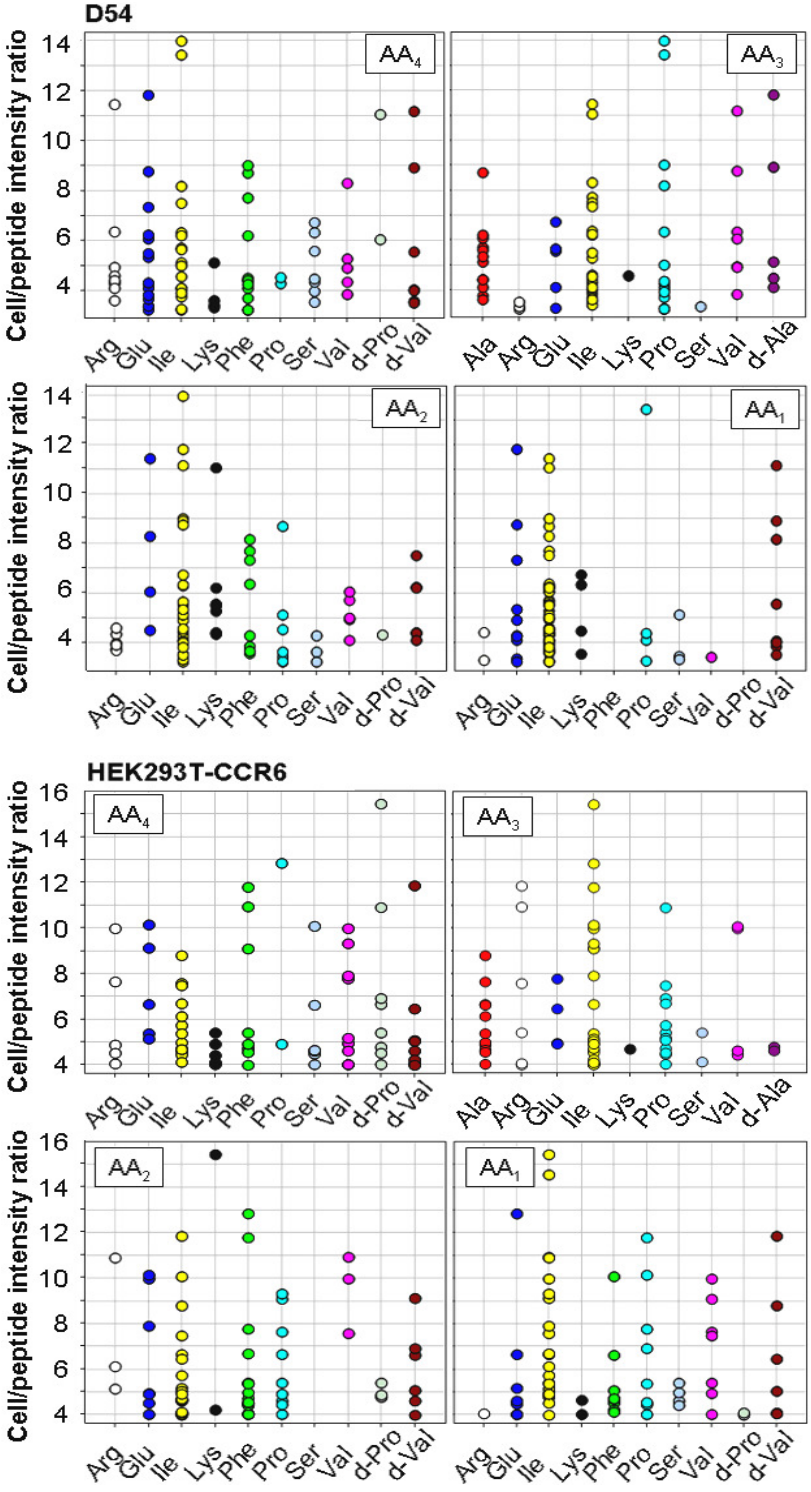


Figure 4.4: Scatter plots of relative microarray intensities derived from the binding of D54 and HEK293T-CCR6 on Library-1 microarrays versus the variable amino acids for positions AA_{1-4}.

Table 4.1: Consensus sequences derived from scatter plots based on microarray analysis by combining the most or least preferred amino acids for each position (AA_{1-4}).

Binding	AA_4	AA_3	AA_2	AA_1
HEK293TCCR6	Ile	Pro	Ile	Ile
D54	pro	Ile	Ile	Ile

Non-binding	AA_4	AA_3	AA_2	AA_1
HEK293TCCR6	Pro	Lys	Ser	Phe
D54	Lys	Lys	Ser	Arg

4.5.3 Synthesis and microarray printing of hit peptides

To verify and quantitatively compare cellular affinities of the binding and non-binding sequences the fluorescein (FAM) labelled peptides, FAM-Ahx-Phe-Gln-*d*Pro-Ile-Tyr-Ile-Ile-Ile-Lys-NH_2 (X-FQ*p*IYIIIK-NH_2), FAM-Ahx-Phe-Gln-Ile-Pro-Tyr-Ile-Ile-Lys-NH_2 (X-FQIPYIIIK-NH_2), FAM-Ahx-Phe-Gln-Lys-Lys-Tyr-Ser-Arg-Ile-Lys-NH_2 (X-FQKKYSRIK-NH_2), and FAM-Ahx-Phe-Gln-Pro-Lys-Tyr-Ser-Phe-Ile-Lys-NH_2 (X-FQPKYSFIK-NH_2) were synthesised.[15,17] As a positive control for integrin binding FAM-Ahx-Arg-Gly-Asp-NH_2 (X-RGD-NH_2) was also synthesised.

Peptide arrays were generated by contact printing the FAM-labelled peptides onto succinimidyl-ester activated slides (750-900 μm spots) with a highly cell-repellent surface produced elsewhere on the slide by blocking unreacted regions with perfluorooctyl propylamine as described before ensuring that any cellular attachment on the arrays was peptide-mediated.

4.5.4 Cell adhesion through specific peptide interactions

D54 cells were incubated on the X-FQ*p*IYIIIK-NH_2 and X-FQKKYSRIK-NH_2 displaying array, whereas HEK293T-CCR6 cells were incubated on a X-FQIPYIIIK-NH_2 and X-FQPKYSFIK-NH_2 displaying array as described above. As expected cells attached only on the printed regions of array and appeared to have a normal and healthy morphology (elongated and spread out shape, Figure 4.5). Substantially

greater cellular adhesion was observed on the "hit" binding peptides than on the non-binding peptide controls (92% of all adhered D54 cells were on X-FQ*p*IYIIIK-NH_2 compared to only 8% of all adhered cells on X-FQKKYSRIK-NH_2, Figure 4.5).

Likewise, more than 99% of all adhered HEK293T-CCR6 cells were on X-FQIPYIIIK-NH_2 compared to less than 1% of all adhered cells on X-FQPKYSFIK-NH_2 (Figure 4.5). This verifies that novel cell binding surface substrates can be identified by high-throughput screening of microarray supported PNA-encoded peptides and that cell adhesion is indeed peptide specific.

4.5.5 Validation of receptor mediated cell adhesion

To verify that cell adhesion was receptor mediated, a soluble modulator for the respective receptor was added to the media during microarray incubation. The D54 cells over-express $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrins[196,197,198], which can be blocked by the natural ligand, RGD; whereas, the CCR6 receptor can be blocked by the agonist, anti human CCR6. Thus, D54 cells were incubated in the presence of X-RGD-NH_2 (100 μM) on the X-FQ*p*IYIIIK-NH_2-surface, which blocked cell adhesion by 98% in the number of cells bound compared to incubation without X-RGD-NH_2, demonstrating that D54 binding to X-FQ*p*IYIIIK-NH_2 is integrin mediated (Figure 4.5).

Similarly, HEK293T-CCR6 cells were incubated in the presence of peridinin chlorophyll protein complex (PerCP) anti human CCR6 on the X-FQIPYIIIK-NH_2 surface, which blocked cell adhesion by 98% demonstrating that HEK293T-CCR6 binding to X-FQIPYIIIK-NH_2 is GPCR mediated (Figure 4.5).

As a negative control for CCR6 binding HEK293T cells were incubated on a X-FQIPYIIIK-NH_2 and X-FQPKYSFIK-NH_2 surfaces, in the same manner. As expected, HEK293T adhesion was greatly reduced (~90%) compared to HEK293T-CCR6 adhesion on the X-FQIPYIIIK-NH_2 features, further illustrating that X-FQIPYIIIK-NH_2 binding is CCR6 mediated (Figure 4.5).

This study illustrates that peptide-ligands for cell surface receptors can be identified by high-throughput screening of microarray supported PNA-encoded peptides.

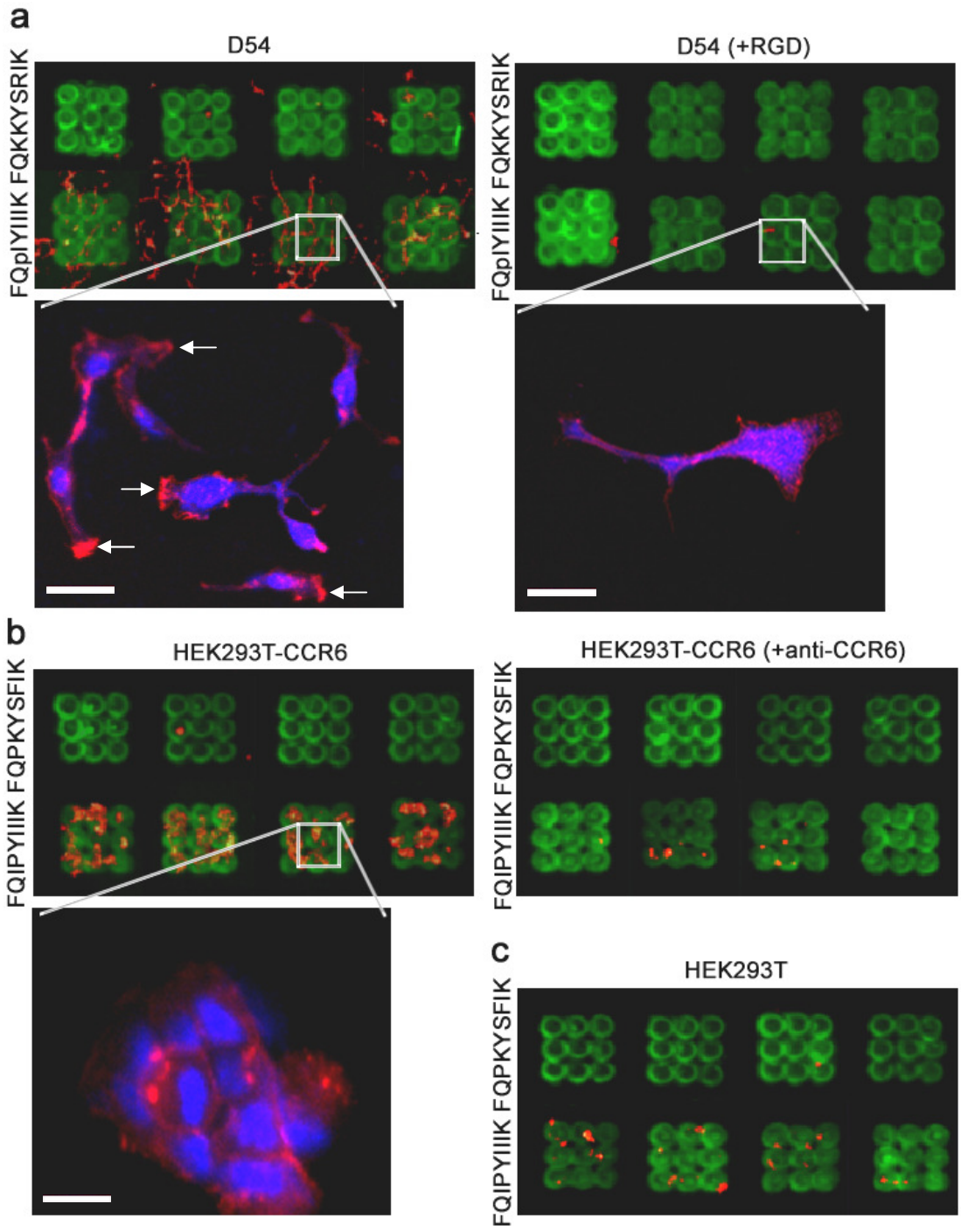


Figure 4.5: Verification of ligand binding by microarray analysis. Green (FITC) = peptide ligand, blue (DAPI) = nuclei staining of attached cells, and red (phalloidin) = actin-filament staining of attached cells visualization of cells incubated in serum free media (2 h) followed by incubation in complete media (16 h): (**a**) D54 cells incubated on X-FQ*p*IYIIIK-NH_2 (written as FQ*p*IYIIIK) and X-FQKKYSRIK-NH_2

(written as FQKKYSRIK; left) and in the presence of soluble X-RGD-NH_2 (right). (**b**) HEK293T-CCR6 cells incubated on X-FQIPYIIIK-NH_2 (written as FQIPYIIIK) and X-FQPKYSFIK-NH_2 (written as FQPKYSFIK, left) and in the presence of soluble PerCP anti-human CCR6 (right). (**c**) HEK293T cells incubated on FQIPYIIIK and FQPKYSFIK. The cells were fixed (4% paraformaldehyde), the actin-filaments stained with phalloidin (red) and the nuclei stained with DAPI (blue). Scale bars = 10 μm.

4.5.5.1 The agonist/antagonist binding-mode of the identified integrin ligand

Agonist binding to integrins causes a conformational change in the receptor allowing phosphorylation and activation of focal adhesion kinase (FAK), which leads to formation of the adhesion complex attached to the actin-cytoskeleton[117]. Hereby, the actin filamentous system can be organized to induce cell migration by continuously binding new ligands at the front of the cell and releasing those at the rear[199].

Staining of the F-actin filaments in D54 revealed that cells attached to X-FQ*p*IYIIIK-NH_2 had a high concentration of actin in one end of the axon and less in the remainder of the cell (Figure 4.5, white arrows). Contrarily, D54 cells adhered to X-FQKKYSRIK-NH_2 did not show the same localization of actin filaments towards either end of the cell. Noticeably, cells attached to X-FQ*p*IYIIIK-NH_2 in competition with soluble X-RGD-NH_2 had a round morphology with no axon formation and no actin localization illustrating that the X-FQ*p*IYIIIK-NH_2 failed to induce focal adhesion formation when inhibited by X-RGD-NH_2. These differences in cell morphology indicate that X-FQ*p*IYIIIK-NH_2 is binding as an agonist inducing conformational change in the receptor and ultimately linking the actin filaments.

4.5.6 Cytotoxicity

In order for the peptides to have biological applications it is vital that they do not exhibit cellular toxicity. Cell viability upon treatment with the "hit"-peptides was assessed with MTT[186] assays as described before (section 3.5). Untreated cells were assumed to be 100% viable and the viability of peptide-treated cells was expressed as

percentage live cells compared to untreated cells. None of the "hit"-peptides showed cytotoxicity in any of the tested cell lines at 100 μM (Figure 4.6).

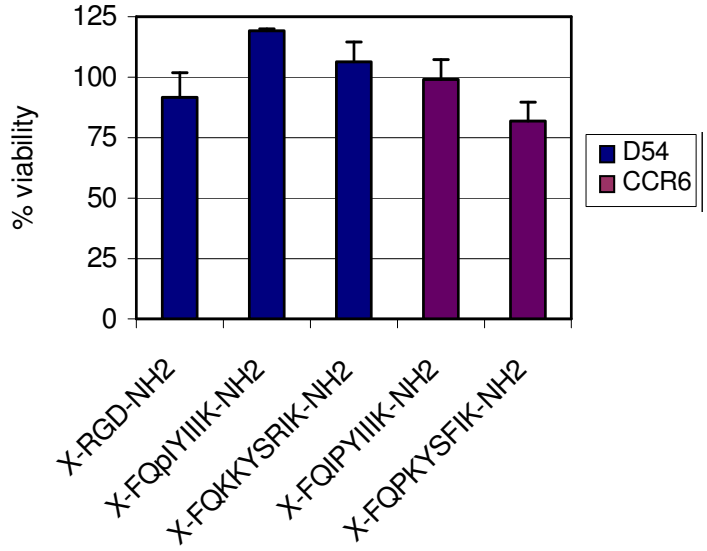


Figure 4.6: Cytotoxicity of hit-peptides assessed as cell viability by MTT assays. Cells were incubated with X-FQAA$_4$AA$_3$YAA$_2$AA$_1$IK-NH$_2$ (100 μM) or X-RGD-NH$_2$ (100 μM; 24 h), the media was changed and the incubation continued for 24 h and absorbance was measured at 580 nm. Untreated cells are assumed to be 100% variable. Error bars indicate ± s.d. (n = 3).

4.5.7 Conclusion

Microarray analysis of a 10,000 member PNA encoded peptide library identified new peptide ligands for $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrin and the CCR6 GPCR, namely FQ*p*IYIIIK and FQIPYIIIK that demonstrated low cytotoxocity. Scatter plots of the microarray intensity versus the varying amino acid position in Library-2 showed preference for some amino acids over others, allowing consensus sequences for both receptor binding (FQ*p*IYIIIK and FQIPYIIIK) and non-binding peptides (FQKKYSRIK and FQPKYSFIK) to be extracted for $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrins and CCR6 respectively.

Synthesis followed by surface immobilisation of these "hit"-peptides with a fluorescent label allowed verification of receptor binding. Competition studies with the natural ligands for receptors demonstrated that binding of X-FQ*p*IYIIIK-NH$_2$ and

X-FQIPYIIIK-NH_2 is integrin and CCR6-mediated respectively, whereas X-FQKKYSRIK-NH_2 and X-FQPKYSFIK-NH_2 do not bind these receptors.

This approach establishes a strategy for high-throughput screening for the identification of ligands for integrins and GPCRs offering and efficient approach to de-orphanise the many GPCRs, for which there are no known ligands. As integrins and GPCRs amongst the most heavily investigated drug targets in the pharmaceutical industry today[115,137] identification of new agonist or antagonists of these receptors by this method can help identify new drug-targets as well as leads.

The technology would also allow screening of a wide variety of cell surface receptors and offers a tool for investigating differences in membrane receptor expression between different cell types including diseased cells.

Furthermore, this approach facilitates simultaneous screening of different cell types in a competition study simply by staining the different cells with different cell tracker dyes prior to screening.

Chapter 5 Microarray generation of 10,000 oligonucleotides

5.1 Generation of DNA libraries

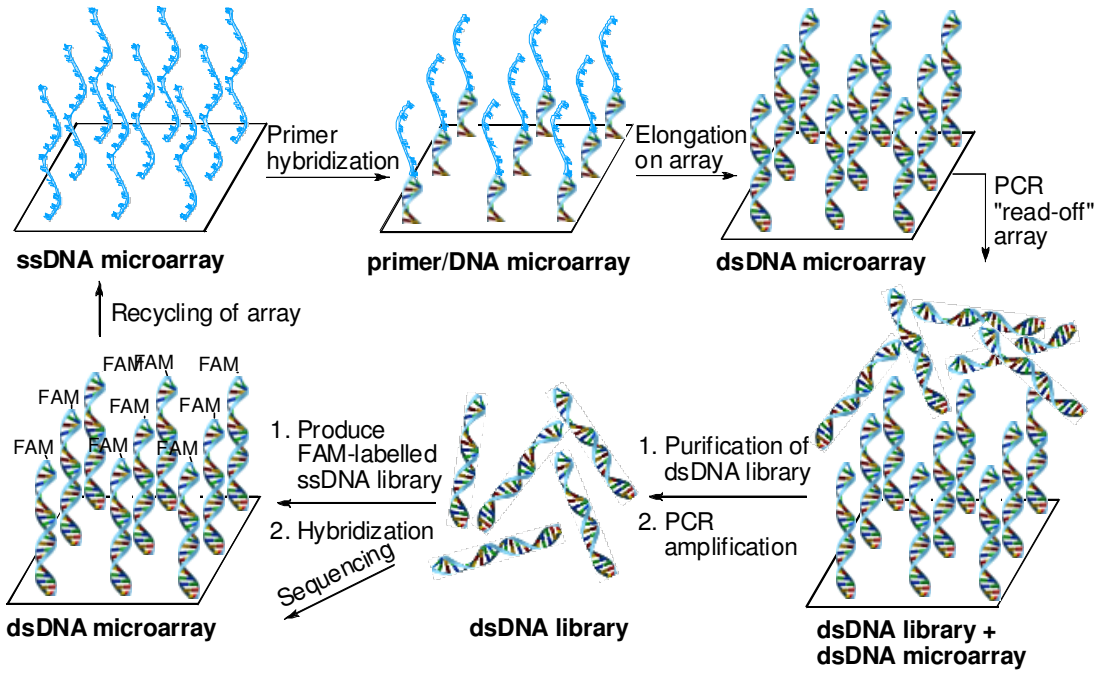
As discussed in section 1.2.2.1 there are many examples of DNA being used as an encoding device for peptides or small molecules, enabling the high-throughput screening of peptide interactions with biological targets or screening of drug candidates. These DNA-encoding methods (section 1.2.2.1) rely on synthesis of DNA oligonucleotide libraries and development of methods for effective and economical generation of DNA libraries would be of great interest to this field.

Microarray based DNA libraries have been employed in gene expression profiling studies[89-201], whereby cellular mRNA is extracted and the corresponding cDNA is hybridised onto a complementary DNA microarray. Hereby, the expression of thousands of genes is measured at once, to create a global picture of cellular function including identification of genes whose expression is changed in response to pathogens or other organisms[200].

Examples of the fabrication of DNA libraries include "PCR" on solid supported primers to enable high-throughput production of oligonucleotides[202] Hereby, primers are covalently attached to microarrays allowing hybridisation of specific DNA targets and elongation of the primers generating microarray supported DNA libraries with high density of oligonucleotides of any length.[202,203] This technique has been shown to reduce non-selective DNA oligonucleotide amplification and thereby enhancing identification of diagnostic targets[204] and improving SNP detection[205].
Efforts have been made to synthesise oligonucleotide libraries from microarrays simply by cleaving the oligonucleotides off the array followed by PCR amplification of the cleaved oligonucleotides in solution. This generated multiplex DNA libraries for parallel genomic assays[206], however, the technique is sacrificial, offering no means of reuse of the DNA array and it allows amplification of oligonucleotide libraries as mixtures only.

However, the use of oligonucleotide libraries in chemical, biological and genetic research, including gene expression profiling, has its limitations in the expense of

vast numbers of DNA oligonucleotides. Thus, the aim of this project was to develop an approach to the generation of DNA libraries from DNA microarrays, while keeping the array intact and useable for subsequent applications, allowing the efficient and inexpensive production of custom made thousand-member DNA libraries. This was achieved using 10,000 oligonucleotides, which were "read-off" the array by a polymerase and subsequently amplified by PCR (Scheme 5.1). The generated DNA libraries were evaluated by Solexa sequencing as well as conventional microarray analysis.



Scheme 5.1: Concept of generation of DNA oligonucleotides from microarrays and parallel analysis. FAM = 5(6)-carboxyfluorescein.

5.2 Microarray design

In order to explore the fidelity of the approach microarrays were designed to contain an increasing number of different DNA oligonucleotides (1, 10, 3875, or 10,000) and were based on the 17 base pair (bp) sequences complementary to a previously reported 10,000 member PNA-encoded peptide library[29].

The 1-member oligonucleotide array was designed to contain just one sequence (Table 5.1a), which included two domains complementary to primer-1 and primer-2

(Table 5.1a, red domains), in a 10 x 10 pattern. The 10-member oligonucleotide array was designed with the variable domain (17 bp; Table 5.1a, black) flanked by domains complementary to primer-1 and primer-2 (Table 5.1a, red). The oligonucleotides were randomly organised with 4,000 replicates in 4 x 44,000 subarrays. In addition, each subarray included 4,000 non-complementary DNA oligonucleotides as negative controls.

The 3,875 and 10,000-member oligonucleotide arrays were designed with the variable domain (17 bp; Table 5.1a, black) flanked by domains complementary to Solexa-primer-1 and 2 (Table 5.1a, blue/green) to allow subsequent DNA sequencing. In order to quantitatively assess the amplification of each oligonucleotide on the array the 3,875 oligonucleotide array was designed with scaling of the content of the oligonucleotides with either 1, 2, 4, 8, or 16 replicates of each oligonucleotide in each of the 4 x 44,000 subarrays (Table 5.1b). In addition, each subarray included 1,375 non-complementary DNA oligonucleotides as negative controls. The 10,000 oligonucleotides were organised randomly with 4 replicates of each in 4 x 44,000 subarrays and each subarray included 4,000 non-complementary DNA oligonucleotides as negative controls.

Table 5.1: (**a**) General microarray supported oligonucleotide sequences and primer sequences. X is A, C, or T. (**b**) Number of replicates of oligonucleotides on the scaled content 3,875-member oligonucleotide microarray.

a

1-member oligonucleotide microarray
5'-TCCCAGGGAAAGCATGGAAGAAGGAGAACCTTCTCTCTCTCTCTCTCTCTCT-3'
10-member oligonucleotide microarray
5'-TCCCAGGGAAAGCATGGXXXXXXXXXXXXCTTCTCTCTCTCTCTCTCTCT-3'
3,875 and 10,000-member oligonucleotide microarrays
5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGG
XXXXXXXXXXXXCTTAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG 3'
Primer-1
5'-TCCCAGGGAAAGCATGG-3'
Primer-2
5'-AGAGAGAGAGAGAGAGAGAAG-3'
Primer-2-FAM
5'-FAM-AGAGAGAGAGAGAGAGAGAAG-3'
Solexa-primer-1
5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3'
Solexa-primer-2
5' CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT-3'
Primer-3
5' CTACACGACGCTCTTCCGATCTGG 3'
Primer-4-FAM
FAM- 5' GCATACGAGCTCTTCCGATCTAAG 3'

b

Number of oligonucleotides x number of replicates	
2000 oligonucleotides x 16:	32,000
1000 oligonucleotides x 8:	8,000
500 oligonucleotides x 4:	2,000
250 oligonucleotides x 2:	500
125 oligonucleotides x 1:	125
3,875 oligonucleotides in total:	42,625

5.3 PCR "read-off" microarrays

The first steps in the process involved primer hybridisation and elongation on the solid support and required extended hybridisation and elongation times for efficient production of a double stranded DNA microarray, with one DNA strand covalently attached to the surface. The newly synthesized DNA strand could then function as a template for normal solution phase PCR carried out over the microarray (Scheme 5.1).

PCR "read-off" the 1-member oligonucleotide array gave a 50 bp band by DNA gel electrophoresis (Figure 5.1a). Conventional Sanger sequencing of the PCR amplified product showed the expected oligonucleotide sequence (Table 5.1a). PCR "read-off" the 10-member oligonucleotide microarray (Table 5.1a) also gave the expected 50 bp band by gel electrophoresis (DNA-10), while the larger 3,875 and 10,000-member oligonucleotide microarrays gave the expected 107 bp band (DNA-3,875 and DNA-10,000 respectively; Figure 5.1a). Furthermore, enzymatic digestion with EcoICRI (recognition sequence: 5'-GAG▼CTC-3') of DNA-3,875 and DNA-10,000 resulted in the two expected fragments (85bp and 22bp; Figure 5.1c).

PCR on the 10,000-member oligonucleotide microarray was repeated 5 times after the initial round of primer hybridisation, extension, and washing but without stripping the newly synthesized DNA. This resulted in similar isolated yields (39-40%, equation 7.1) illustrating that PCR can be performed multiple times without damaging the array (Figure 5.1b). No DNA was observed when the PCR "read-off" the 10,000 oligonucleotide array was carried out without primers (blank PCR; Figure 5.1b).

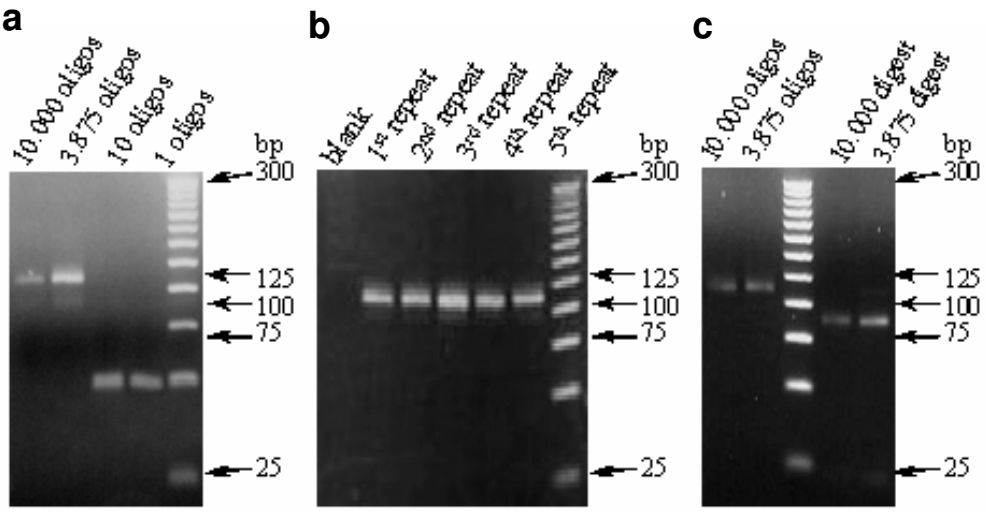


Figure 5.1: DNA gel electrophoresis of: (**a**) PCR products from 1, 10, 3,875, 10,000-member oligonucleotide arrays; (**b**) 5 repeats of PCR products from the 10,000-member oligonucleotide array; (**c**) DNA-10,000 and DNA-3875 and EcoICRI digestion.

5.4 Microarray hybridisation of PCR products

To allow microarray quantification of the DNA microarray "read-off" libraries these were further amplified by PCR with a FAM labelled primer and an unlabelled primer (primer-1 and primer-2-FAM for DNA-10; primer-3 and primer-4-FAM for DNA-3,875 and DNA-10,000) producing FAM-labelled dsDNA libraries (DNA-10-FAM, DNA-3,875-FAM, DNA-10,000-FAM; Table 5.1 and Figure 5.2).

The doublestranded DNA-3,875 and DNA-10,000 libraries were sought to be hybridised to DNA microarrays that encode only the variable domain of the DNA-10,000 library (17 bp; Table 5.1a, black). However, hybridisation of the 17 bp microarray supported oligonucleotides with the 107 bp dsDNA libraries is very challenged due to the competition between the non-microarray complementary 107 bp DNA strands and the 17 bp strands. Thus, to improve the efficiency of microarray hybridisation single stranded DNA libraries were synthesised by PCR with a single primer (primer-4-FAM) producing microarray complementary ssDNA libraries (ssDNA-3,875-FAM and ssDNA-10,000-FAM; Scheme 5.1 and Figure 5.2).

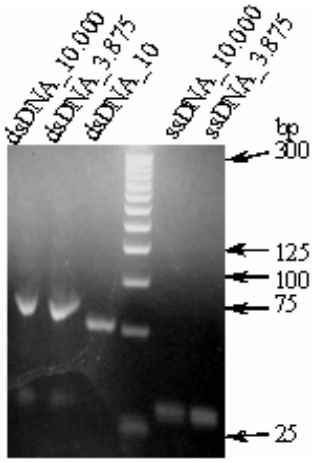


Figure 5.2: DNA gel electrophoresis of products from PCR with two primers producing dsDNA-10,000-FAM and dsDNA-3,875-FAM (primer-3 and primer-4-FAM), and dsDNA-10-FAM (primer-1 and primer-2-FAM) and products from PCR with a single primer producing ssDNA-10,000-FAM and ssDNA-3,875-FAM (primer-4-FAM).

dsDNA-10-FAM was then hybridised onto a complementary DNA microarray identical to the "read-off" DNA microarray (above); whereas the ssDNA-3,875-FAM

and ssDNA-10,000-FAM libraries were hybridised onto microarrays that were complementary the 17 bp variable domain (Table 5.1a, black). Fluorescent microarray imaging in combination with BlueFuse technology (BlueGenome) was used to quantify the intensity of the FAM-label and thereby determine the amount of DNA hybridised to each spot (Figure 5.3 and appendix IV). Note that the observed differences in spot intensities on the 10 and 10,000-member oligonucleotide microarrays are due the differences in hybridisation efficiencies of the oligonucleotides caused by different T_m and secondary structures (Figure 5.3).

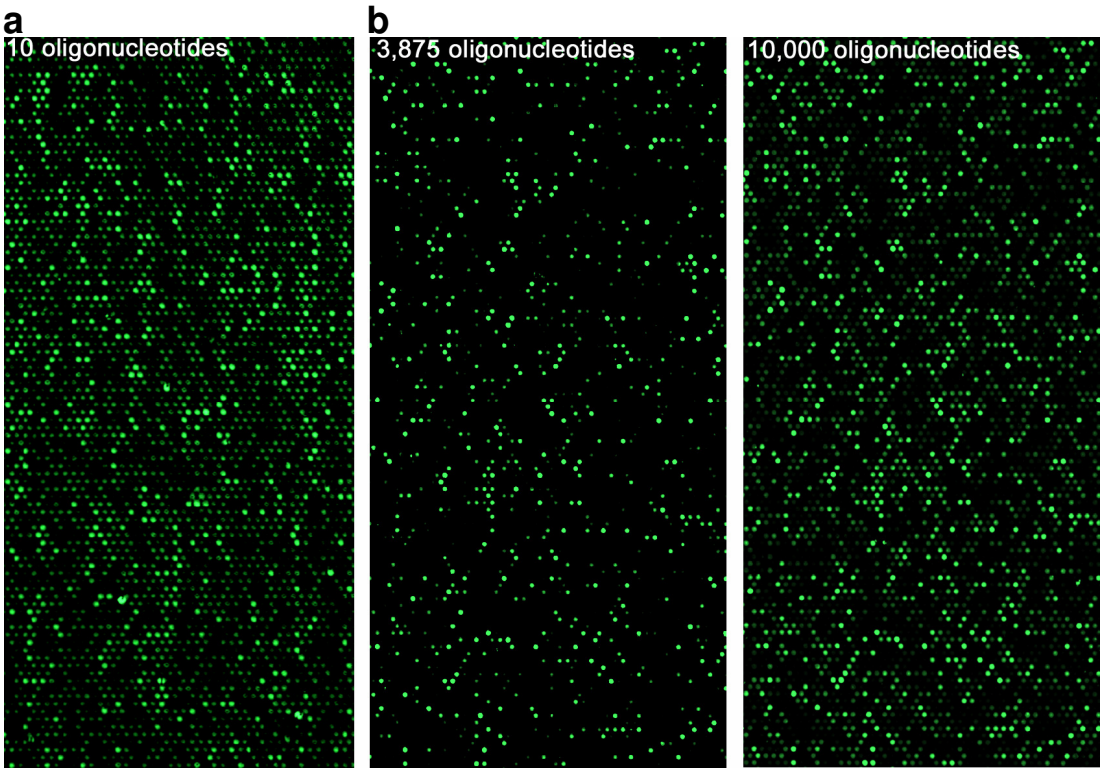


Figure 5.3: Microarray analysis of FAM-labelled oligonucleotide libraries generated by "read-off" DNA microarrays. FITC-filtered fluorescence of 44,000 feature DNA microarrays with variable number of replicates of each oligonucleotide as well as non-coding negative controls. (**a**) 10 oligonucleotide array hybridised with dsDNA-10-FAM. (**b**) Array complementary to the varying 17 bp region of the DNA-10,000 library hybridised with ssDNA-3,875-FAM. (**c**) Array complementary to the varying 17 bp region of the DNA-10,000 library hybridised with the dsDNA-10,000-FAM. Hybridisations were in 0.1% SDS in 4 x SSPE buffer (110 μL) from 65 ^{o}C to 27 ^{o}C.

5.4.1 Quantification of microarray hybridisations

Raw microarray data was obtained from Bluefuse, which allows grid alignment and signal estimation, in a Microsoft Excel format. In Microsoft Excel the top ~5% and the bottom ~5% of each of the replicate-sets were removed as outliers (erroneous values caused by dust, scrapes etc[179,180]). The average intensity was calculated for all of the replicate sets and of the intensity of the non-coding negative control features on each microarray. The average intensities were corrected for the background by subtracting the average intensity of the non-coding negative control features (appendix IV). In order to assess the efficiency of the microarray "read-off" and subsequent PCR amplification the background-corrected average microarray intensities were plotted versus the number of replicates (Table 5.1 and Figure 5.4).

The slight differences in background-corrected average intensities in the 10 oligonucleotide graph (Figure 5.4a) arise from the differences in secondary structures and T_m of the oligonucleotides as these properties greatly affect the hybridisation efficiency. Thus, a curved normal distribution of microarray intensity versus the oligonucleotide sequences is expected.[207] The relatively similar average intensities and curved distribution of the 10 oligonucleotide graph illustrate that the amounts of DNA targets in the hybridisation mixture were close to equivmolar and that the microarray "read-off" had occurred uniformly over the whole array.

The graph for the 3,875 oligonucleotides shows a linear relationship between the microarray intensities versus the number of replicates illustrating that the 3,875 DNA templates had been "read-off" and amplified relative to the number of replicates of probes on the microarray (Figure 5.4b). Each data point in the 3,875 oligonucleotides graph represents the average of many different oligonucleotides (Table 5.1b) with different T_m and secondary structures. Consequently, the effects of T_m and secondary structures on the hybridisation efficiency are evened-out resulting in a smoother graph of the 3,875 oligonucleotides compared to that of the 10 oligonucleotides.

The average intensity versus the number of replicates graph for the 10,000 oligonucleotides shows a curved normal distribution similar to that of the 10

oligonucleotides illustrating that also the microarray "read-off" occurs uniformly over high-content arrays with few replicates of each probe (Figure 5.4c).

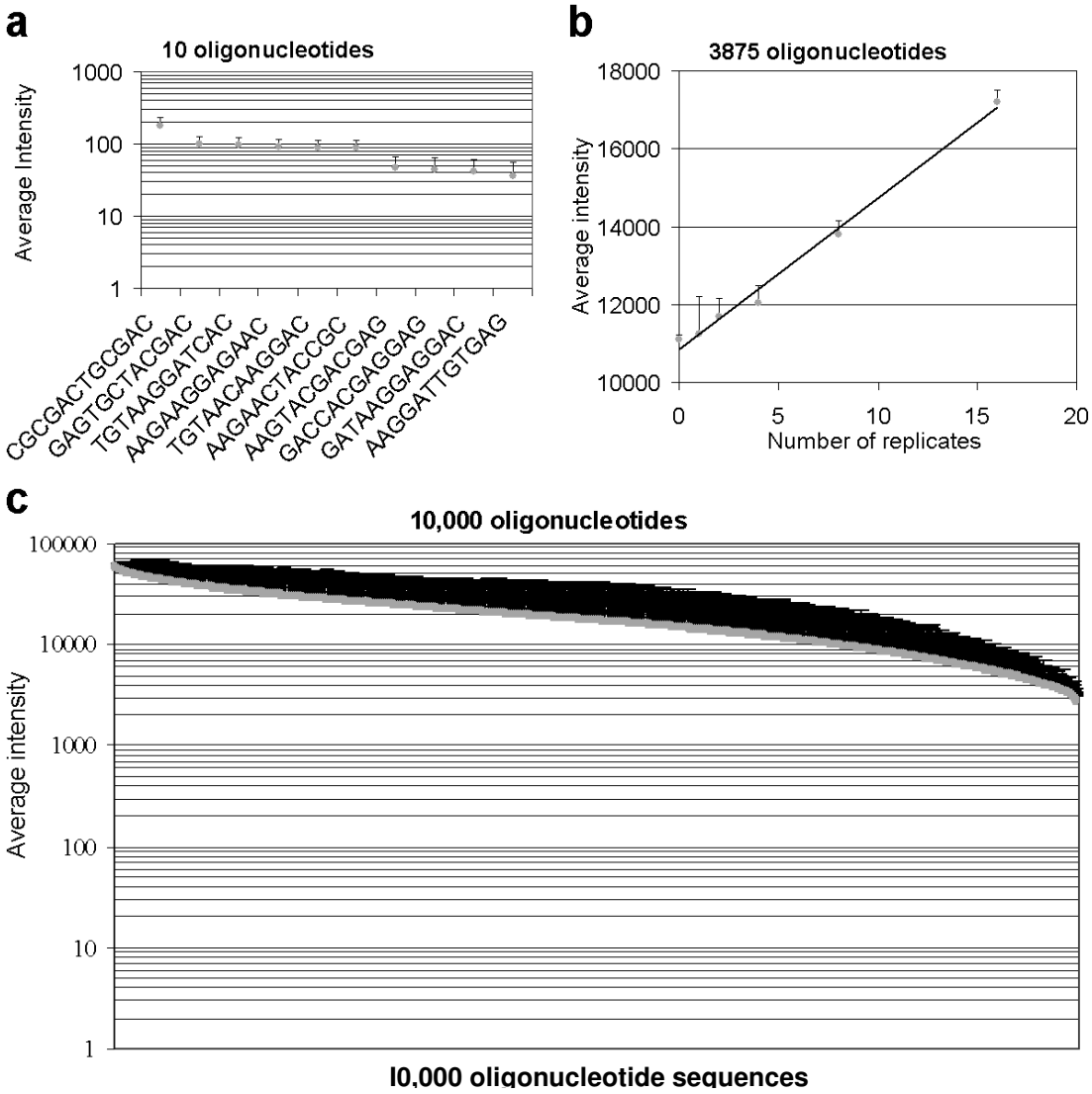


Figure 5.4: The background corrected average intensities were plotted versus the number of replicates for: (**a**) the dsDNA-10-FAM library, (**b**) the ssDNA-3,875-FAM library, and (**c**) the ssDNA-10,000-FAM library. Error bars indicate ± s.d.

5.4.2 Solexa sequencing

Solexa sequencing of the DNA-10,000 oligo-pool identified 9976 sequences from the possible 10,000 DNA oligonucleotides synthesised on the DNA microarray giving an error rate of 0.2% (24 oligonucleotides out of 10,000 not seen; Table 5.2). Noticeably, the oligonucleotides not seen via sequencing all had one of the following

consensus sequences (X = any base): CGC-XXX-XXX-CGC, CGC-XXX-CGC-XXX, CGC-CGC-XXX-XXX, CAC-GAX-XAG-TGC (Table 5.2).

The 9976 sequences were each seen between 1 to 4837 times (Figure 5.5). This significant difference in the number of reads of each oligonucleotide was initially thought to correspond to an unexpected large difference in the actual amount of the respective oligonucleotide in the library. Closer examination of the sequences revealed that the oligonucleotides that had poor frequencies of observation had the same consensus sequences as the non-identified oligonucleotides (Table 5.2). All of the oligonucleotides not seen by sequencing were observed by microarray hybridisation in substantial amounts, i.e. microarray intensities in the range of 6,000-38,000 compared to the full intensity range of the microarray of 2,700-55,000 (Table 5.2 and appendix V).

Table 5.2: Oligonucleotide sequences not seen by Solexa sequencing and their background-corrected average microarray intensities (10,000-member oligonucleotide microarray).

Sequence	Microarray intensity
CACGACGAGTGC	15289
CACGAGAAGTGC	26436
CACGATAAGTGC	11159
CACGATGAGTGC	6376
CGCCACAAGCGC	15261
CGCCACGAGCGC	23314
CGCCGCCGCCGC	38132
CGCCGCGAGCGC	36391
CGCGACGAGCGC	16935
CGCGAGCGCCAC	20588
CGCGAGGAGCGC	17958
CGCGAGGATCGC	25375
CGCGATAAGCGC	35700
CGCGATGAGCGC	23669
CGCTACAAGCGC	34653
CGCTACGAGCGC	23022
CGCTGCAAGCGC	32531
CGCTGCGAGCGC	32798
CGCTGTAAGCGC	23710
CGCTGTCGCCGC	20946
GAGCGCAAGCGC	37537
GAGCGCCGCCGC	35227
GAGCGCCGCGAC	21341
AAGCGCAAGCGC	14723

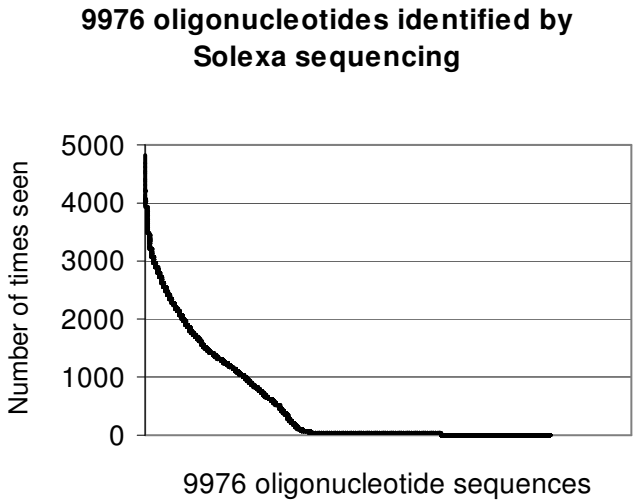


Figure 5.5: Graph of the number of times each of the 9976 sequences was seen by Solexa sequencing versus the oligonucleotide sequences. 36-bp reads of the Solexa primer of the dsDNA-10,000 oligo-pool generated by "read-off" the 10,000 oligonucleotide microarray.

Based on the data from sequencing and the microarray screening it can be assumed that the relative amounts observed by sequencing are an effect of the *actual* amounts of the oligonucleotides in the sample, but that this is probably secondary to the efficiency of the base calling of the respective sequence.

5.5 Conclusion

Four microarrays with 1, 10, 3875 or 10,000 different oligonucleotide sequences were utilised to determine whether they could be used as platforms for large scale DNA synthesis. Hereby a novel microarray "read-off" technology was established that allows high-throughput amplification of microarray supported DNA probes producing tens of thousands member DNA libraries.

Microarray hybridisation of 1, 10, 3875, and 10,000 DNA oligonucleotide "read-off" libraries illustrated that microarray "read-off" had occurred uniformly over the whole of high-content DNA microarrays with only few replicates of each probe, and that the amount of oligonucleotide in the library mixture was relative to the content of each oligonucleotide on the "read-off" array.

Solexa sequencing identified 9976 sequences (error rate of 0.2%). However, there were significant differences in the number of reads of each oligonucleotide. Examination of the oligonucleotides with poor frequencies of observation revealed a set of consensus sequences, indicating that the relative amounts observed by sequencing are an effect of the *actual* amounts of the oligonucleotides in the sample, but that this is probably secondary to the efficiency of the base calling of the respective sequence. This also explains the discrepancy observed between DNA sequencing and microarray data.

This technique establishes an economical and efficient way of producing thousand member DNA oligonucleotide libraries. In addition, this technique would also allow production of DNA libraries as separate oligonucleotides rather than mixtures simply by employing an appropriate microarray design. For example, a microarray with 100 subarrays, each with repeats of a single oligonucleotide, would enable synthesis of 100 separate oligonucleotides using matching coverslide with 100 “wells”.

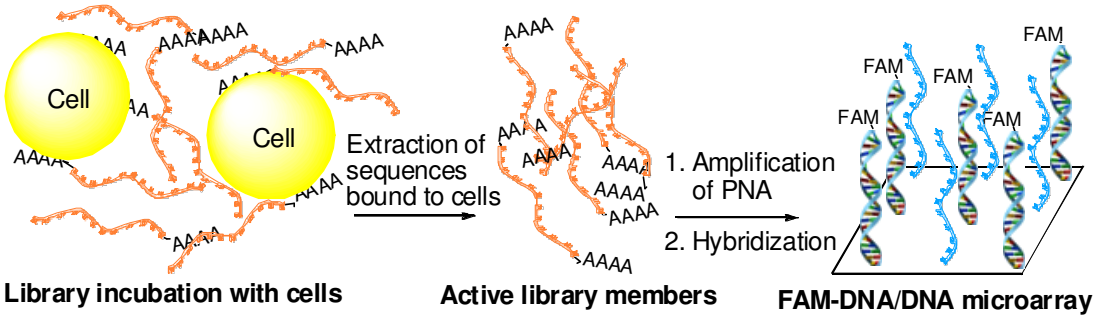
This process offers an efficient technique for the generation of thousands of oligonucleotides allowing the rapid synthesis of specific primers for use in genome sequencing and genotyping assays.[208] Another application of the technique is the synthesis of siRNA by employing an RNA polymerase[209] rather than DNA polymerase, which would allows pools of siRNA to be synthesised from DNA microarrays.[210-214]

Chapter 6 Screening of an encoded peptide library to discover new ligands for cell surface receptors

6.1 Introduction to microarray screening *post* cellular selection

In the approach described in Chapter 4, "screening of an encoded peptide library to discover new cell binding ligands" and several other integrin binding assays[189-195] two potential complications were identified: 1) unspecific binding of the cells to the surface mediated by cell secreted adhesive proteins and 2) degradation of the microarray due to exposure to cell adhesion and proteases. The aim of this project was to avoid false positives from unspecific cell-surface adhesion and to prevent microarray degradation by development of an alternative approach to the high throughput screening of PNA-encoded peptide libraries against cell surface receptors.

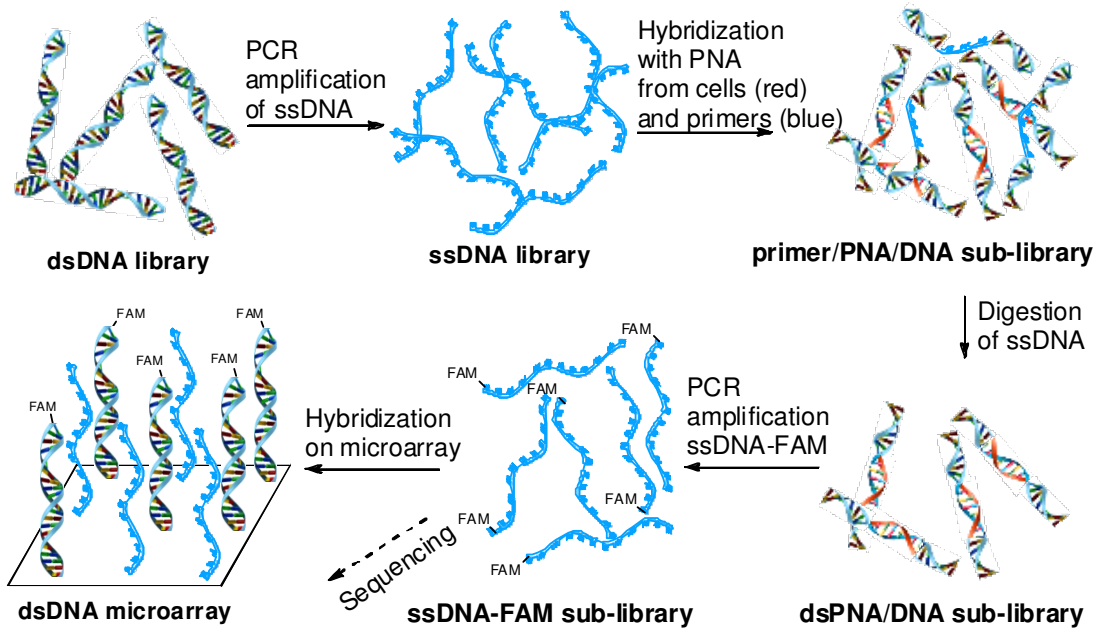
The 10,000 member PNA-encoded peptide library described in section 4.2 (Library-2 [29]) was screened against human cells followed by extraction of the surface bound library members and microarray hybridisation of the PNA tags (Scheme 6.1). Hereby, allowing information on the interaction between the peptide-ligands and the cell surface receptors to be extracted on a single spot-by-spot (compound-by-compound) basis. The arrays used each contained 4 x 44K features, allowing 16 copies of each library member to be analysed per array. Cell lines used were D54 ($\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrin overexpressing)[196,197] and HEK293T-CCR6 (CCR6 GPCR overexpressing)[134] as well as unmodified HEK293T allowing screening with cells with variations in surface-receptor ligand densities and/or receptor expression.



Scheme 6.1: Concept of identification of ligands for cell surface receptors by microarray analysis of a library of PNA-encoded peptide ligands.

6.2 Amplification of the PNA tags

Following cellular incubation with Library-2 the cell surface bound peptide-PNAs were extracted by pronase degradation of the extracellular receptor moieties as well as the bound peptide ligands releasing the PNA tags (Scheme 6.1). The extracted PNA was purified by membrane filtration but produced only small amounts of PNA (pmol scale) due to the limited number of cells and receptors expressed in the cell surface. Theoretically, the maximum amount of PNA-peptide that could be obtained from the 10^6 cells (assuming that there are ~3 x 10^6 target receptors per cell and 100% yield) is ~5 pmol. This comprises just 1% of the amount of Library-2 required for efficient hybridisation (~500 pmol; section 0), thus, necessitating the development of a PNA amplification methodology. PNA is not recognized by DNA polymerase or other enzymes,[6,28] which complicates its amplification by traditional DNA amplification methods, and no amplification method of PNA has been described in the literature. The approach developed in this context was indirect amplification of PNA by amplification of a complementary DNA strand (Scheme 6.2). In this approach the double stranded (ds) DNA library with one strand complementary to the desired PNA library was obtained as discussed in Chapter 5 (dsDNA-Library-2).



Scheme 6.2: Concept of indirect amplification of a PNA library by amplification of the complementary DNA library.

Hybridisation of the 17 bp PNA-Library-2[29] with the 64 bp dsDNA-Library-2 (Table 6.1) is very challenged due to the competition between the non-PNA complementary 64 bp ssDNA and the 17 bp PNA. Thus to improve the efficiency of PNA/DNA hybridisation the single stranded (ss) DNA library complementary to the PNA-Library-2 was synthesised by PCR with a single primer (ssDNA-Library-2; Scheme 6.2).

To investigate the efficiency of the PNA amplification strategy a single DNA oligonucleotide (oligonucleotide-1; Table 6.1) was hybridised with a previously reported single strand of PNA[17] (TGTTTGTTTGTT; Chapter 2; Table 6.1 black) as well as oligonucleotides-2 and 3 (Table 6.1, red; Scheme 6.2). Hereby, a fully double stranded DNA/PNA duplex was formed. The unhybridised ssDNA could then be degraded by a single strand specific DNase, eliminating PCR amplification of any ssDNA not associated with PNA. Oligonucleotide-2 and 3 protected the primer binding domains of oligonucleotide-1 (Table 6.1, red) from nuclease degradation, rendering these domains intact for PCR amplification. DNA gel electrophoresis identified oligonucleotide-2 and 3 (below 25 bp band, Figure 6.1), oligonuleotide-1 (~30 bp band, Figure 6.1), the ssDNA/dsDNA hybrid complex (~45 bp band, Figure 6.1; Scheme 6.2), and the dsDNA/PNA complex (~45 bp band, Figure 6.1). After nuclease degradation the ssDNA/dsDNA hybrid complex band had disappeared leaving the only the dsDNA/PNA complex. This demonstrates that DNA hybridised to PNA can be separated from unhybridised DNA.

Table 6.1: General oligonucleotide sequences and primer sequences.

dsDNA-Library-2
5' CTACACGACGCTCTTCCGATCTGGXXXXXXXXXXXXCTTAGATCGGAAGAGCTCGTATGC 3'
Oligomucleotide-1
5' CTACACGACGCTCTTCCGATCTGGAACAAACAAACAAACACTTAGATCGGAAGAGCTCGTATGC 3'
Oligonucleotide-2
5' CCAGATCGGAAGAGCGTCGTGTAG 3'
Oligonucleotide-3
5' GCATACGAGCTCTTCCGATCTAAG 3'
Primer-3
5' CTACACGACGCTCTTCCGATCTGG 3'
Primer-4-FAM
FAM- 5' GCATACGAGCTCTTCCGATCTAAG 3'

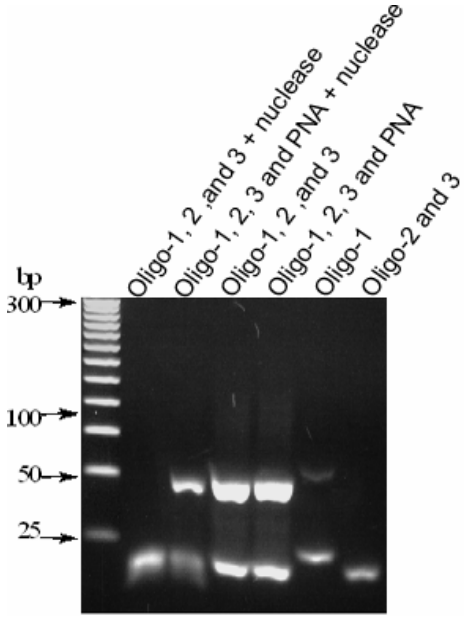


Figure 6.1: DNA gel electrophoresis of hybridisation of oligonucleotide-1 with oligonucleotide-2 and 3 with or without peptide-TGTTTGTTTGTT. Half of each hybridisation solution was treated with single strand specific nuclease.

Similarly, PNA purified from cells and the PNA-Library-2 was hybridised to the ssDNA-Library-2 as well as oligonucleotide-2 and 3, followed by nuclease degradation, and purification of the dsDNA/PNA complexes by gel column filtration. The purified dsDNA/PNA complexes were amplified by PCR with a FAM-labelled primer (Table 6.1) as previously described (section 5.4) producing FAM-labelled ssDNA libraries complementary to custom made DNA microarrays (ssDNA-Library-2-FAM, ssDNA-D54-FAM, and ssDNA-HEK293T-CCR6-FAM; Figure 6.2).

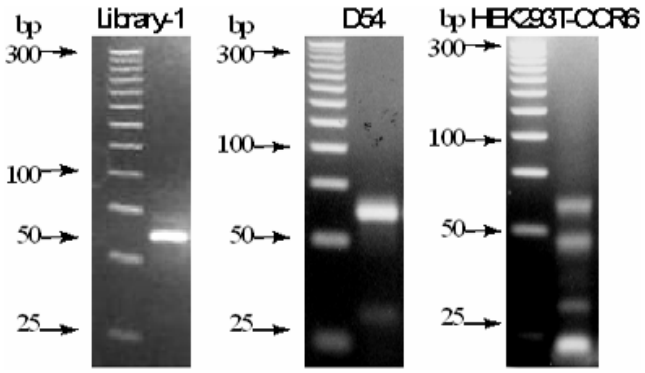


Figure 6.2: DNA gel electrophoresis of PCR amplification of dsDNA-Library-2-FAM, dsDNA-D54-FAM, and dsDNA-HEK293T-CCR6-FAM with a single primer producing ssDNA-Library-2-FAM, ssDNA-D54-FAM, and ssDNA-HEK293T-CCR6-FAM respectively.

6.3 Microarray hybridisation of amplified ssDNA

The ssDNA-FAM libraries were hybridised to a DNA-microarray complementary to the 17 bp PNA encoding domains. Fluorescent microarray imaging in combination with BlueFuse (BlueGenome) technology was used to extract the intensity of the FAM-label and thereby determine the amount of FAM labelled ssDNA hybridised to each spot (Figure 6.3 and appendix VI). The FITC-filtered images of the Library-2 microarray illustrates that the ssDNA-Library-2-FAM hybridisation had been efficient (i.e. most spots are very intense, Figure 6.3) demonstrating efficient indirect amplification of the PNA-Library-2 through amplification of the complementary ssDNA-Library-2. The FITC-filtered images of the D54 and HEK293T-CCR6 microarrays illustrate that only some spots had been hybridised, indicating that the cells had selected certain peptide-PNA conjugates and that the indirect amplification of these had been efficient (Figure 6.3).

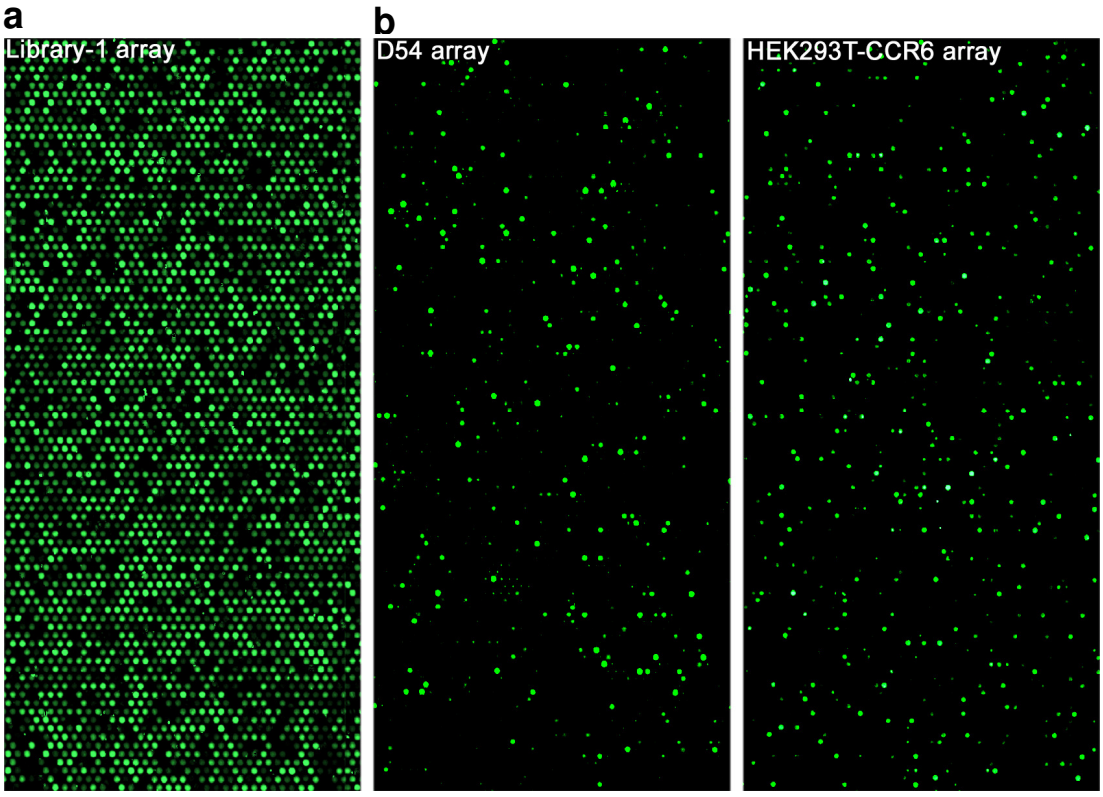


Figure 6.3: Microarray analysis of receptor ligand PNA-encoded peptides. DNA microarrays were designed with 4 replicates of each oligonucleotide complementary to each member of the 10,000 member library as well as 4,000 non-coding negative controls resulting in a 44,000 feature microarray. Microarrays hybridised with (**a**) ssDNA-Libaray-2 and (**b**) ssDNA-FAM from amplification of PNA signal from D54

and HEK293T-CCR6 cells in 110 µL 0.1% SDS, 4 x SSPE buffer from 65 ^{o}C to 27 ^{o}C overnight.

6.3.1 Quantification of microarray hybridisations

Raw microarray data was obtained from Bluefuse, which allows grid alignment and signal estimation, in a Microsoft Excel format. The average fluorescence intensity was calculated over each of the 4 replicates and over the non-complementary negative control features. The average intensities were corrected for the background by subtracting the average intensity of the non-complementary negative control features. Thereafter, the D54 microarray data was normalised for the Library-2 hybridisation by calculating the ratio between the background corrected average intensities of the ssDNA-D54-FAM hybridisation and the background corrected average intensities of the ssDNA-library-2-FAM hybridisation (appendix VI). The HEK293T-CCR6 microarray data was normalised in the same manner.

Scatter plots of the normalised microarray intensities versus the variable amino acids for positions AA_4 to AA_1 were constructed (Figure 6.4) allowing cellular preference for specific peptides to be elucidated, or indeed consensus sequences of non-binding peptides (negative controls) for each cell line to be obtained (Table 6.2).

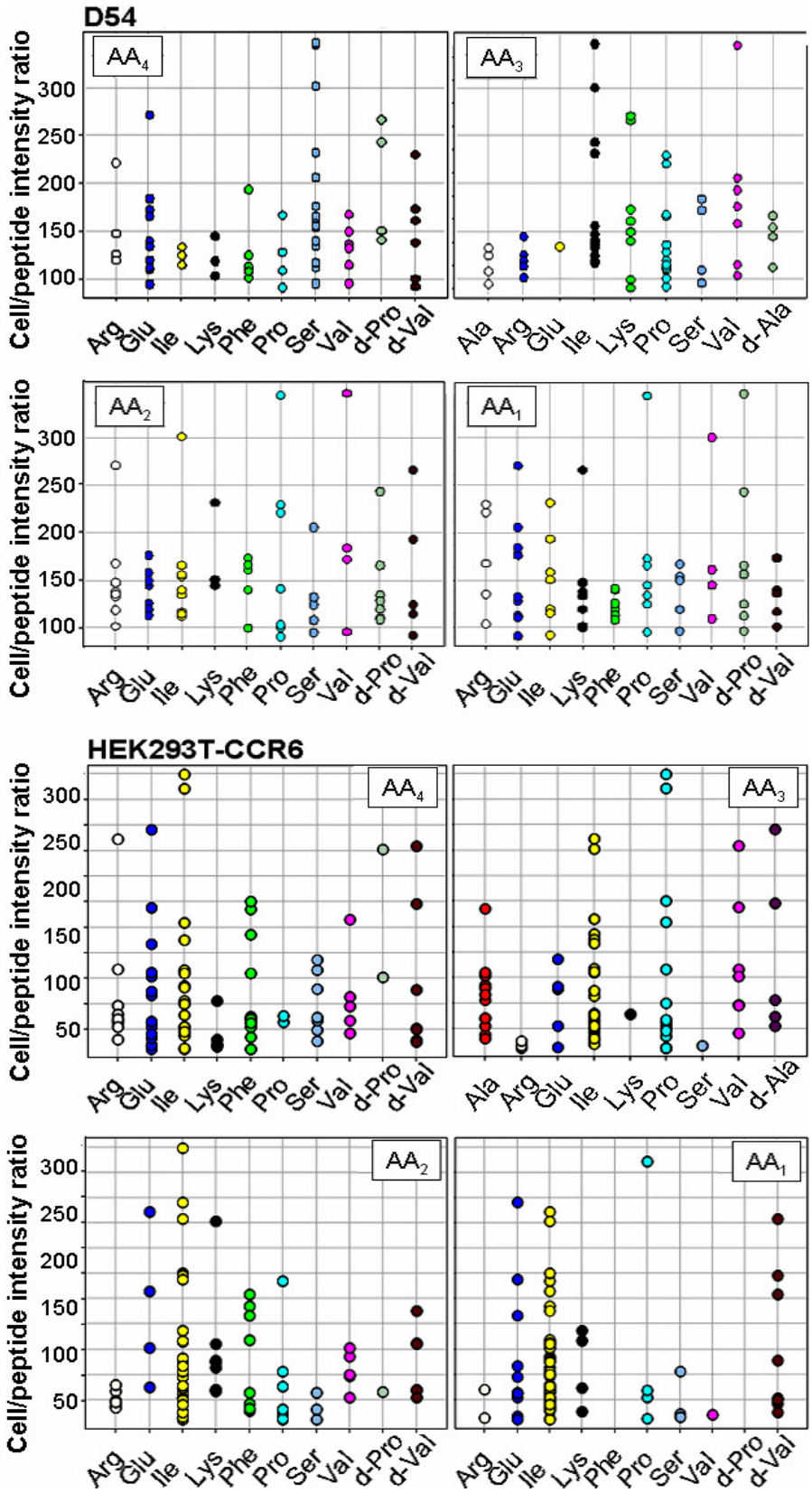


Figure 6.4: Scatter plots of the background-corrected mean FITC microarray intensities versus the variable amino acids for position $AA_{1\text{-}4}$ of PNA derived from D54 and HEK293T-CCR6 cells hybridised on complementary DNA microarrays.

Table 6.2: Consensus sequences derived from scatter plots based on microarray analysis by combining either the most or the least preferred amino acid in each position (AA_{1-4}).

Binding	AA_4	AA_3	AA_2	AA_1
D54	Ser	Ile	Pro	*d*Pro
HEK293T-CCR6	Ile	Pro	Ile	Ile

Non-binding	AA_4	AA_3	AA_2	AA_1
D54	Lys	Glu	Lys	Phe
HEK293T-CCR6	Pro	*d*Ala	Pro	Phe

6.4 Synthesis of "hit" peptides

To verify and quantitatively compare cellular affinities of the identified binding and non-binding sequences the fluorescein labelled peptides, FAM-Ahx-Phe-Gln-Ser-Ile-Tyr-Pro-*d*Pro-Ile-Lys-NH_2 (X-FQSIYP*p*IK-NH_2), FAM-Ahx-Phe-Gln-Lys-Glu-Tyr-Lys-Phe-Ile-Lys-NH_2 (X-FQKEYKFIK-NH_2), and FAM-Ahx-Phe-Gln-Pro-*d*Ala-Tyr-Pro-Phe-Ile-Lys-NH_2 (X-FQP*a*YPFIK-NH_2) were synthesised.[17,87] As a positive control for integrin binding Rhodamine-Ahx-Arg-Gly-Asp-NH_2 (Z-RGD-NH_2) was also synthesised.

6.5 Evaluation of selective cell binding of "hit" peptides

D54, HEK293T, and HEK293T-CCR6 cells were analysed by flow cytometry following incubation with the fluorescently labelled peptides or a CCR6 antibody, anti-human CCR6.

6.5.1 Peptide-cell interactions are peptide specific

In D54 cells the identified peptide integrin ligand, X-FQSIYP*p*IK-NH_2 induced a significantly higher shift in the level of fluorescence compared to that of the untreated cells similar to the shift induced by the positive control, X-RGD-NH_2 (section 4.5.3; Figure 6.5 and appendix VII). The identified non-binding peptide, X-FQKEYKFIK-NH_2 caused a smaller increase in fluorescence compared to X-FQSIYP*p*IK-NH_2. This illustrates that binding of X-FQSIYP*p*IK-NH_2 is peptide

specific and that this peptide has a similar affinity for the binding receptors as X-RGD-NH_2. Treatment with pronase/EDTA greatly reduced the fluorescence signal of all peptides illustrating that the peptides are associated with the cell surface and not taken up by the cells (Figure 6.5 and appendix VII).

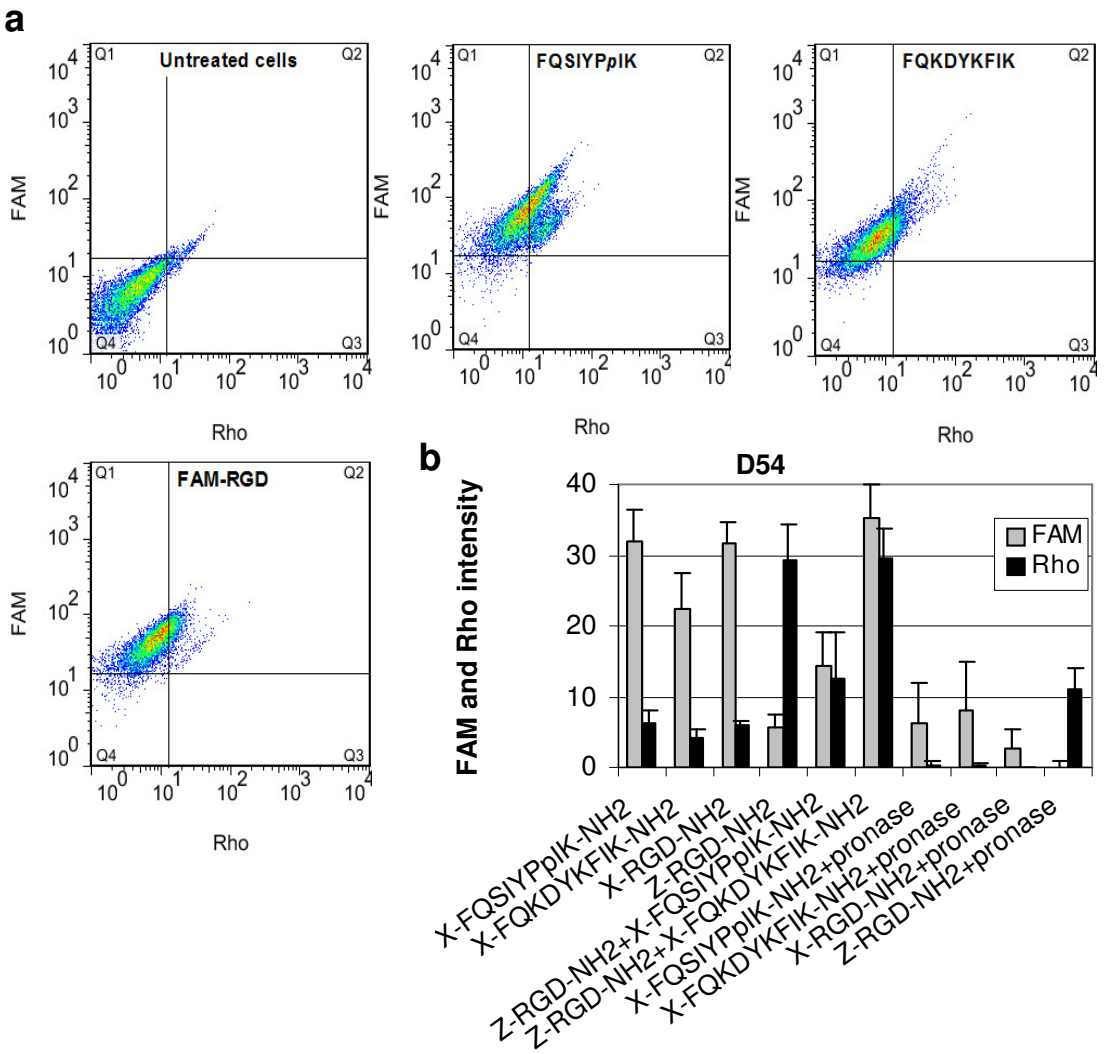


Figure 6.5: Flow cytometry analysis of peptide ligands. (**a**) FAM versus Rho-filtered scatter plots gaited for live, single cells following incubation with X-$FQAA_4AA_3YAA_2AA_1IK$-NH_2, X-RGD-NH_2, and Z-RGD-NH_2 using a FACS Aria cytometer with FITC and PE-Texas-Red filters (10,000 populations, n=3). D54 cells were detached by scraping for minimal destruction of the extracellular receptor moieties or by trypsin/EDTA as a control for extracellular binding of peptides. Peptides are written as simplified sequences. (**b**) Mean FAM-filtered fluorescence of scatter plots versus the labelled peptides and antibodies. Compensation was applied between the FAM/Rho fluorophore pair. Error bars indicate ± s.d. (n=3).

In HEK293T-CCR6 cells the identified peptide CCR6 ligand, X-FQIPYIIIK-NH_2 (section 4.5.3) induced a significantly higher shift in the level of fluorescence compared to that of the untreated cells but slightly smaller than the shift induced by the positive control, FAM-anti-human CCR6 (Figure 6.6 and appendix VII). The identified non-binding peptide, X-FQP*a*YPFIK-NH_2 caused a smaller increase in fluorescence compared to X-FQIPYIIIK-NH_2 (Figure 6.6 and appendix VII). This illustrates that binding of X-FQIPYIIIK-NH_2 is peptide specific and that this peptide has a similar affinity for the binding receptor as FAM-anti-human CCR6. Treatment with pronase/EDTA greatly reduced the fluorescence signal of all peptides illustrating that the peptides are associated with the cell surface and not taken up by the cells (Figure 6.6 and appendix VII).

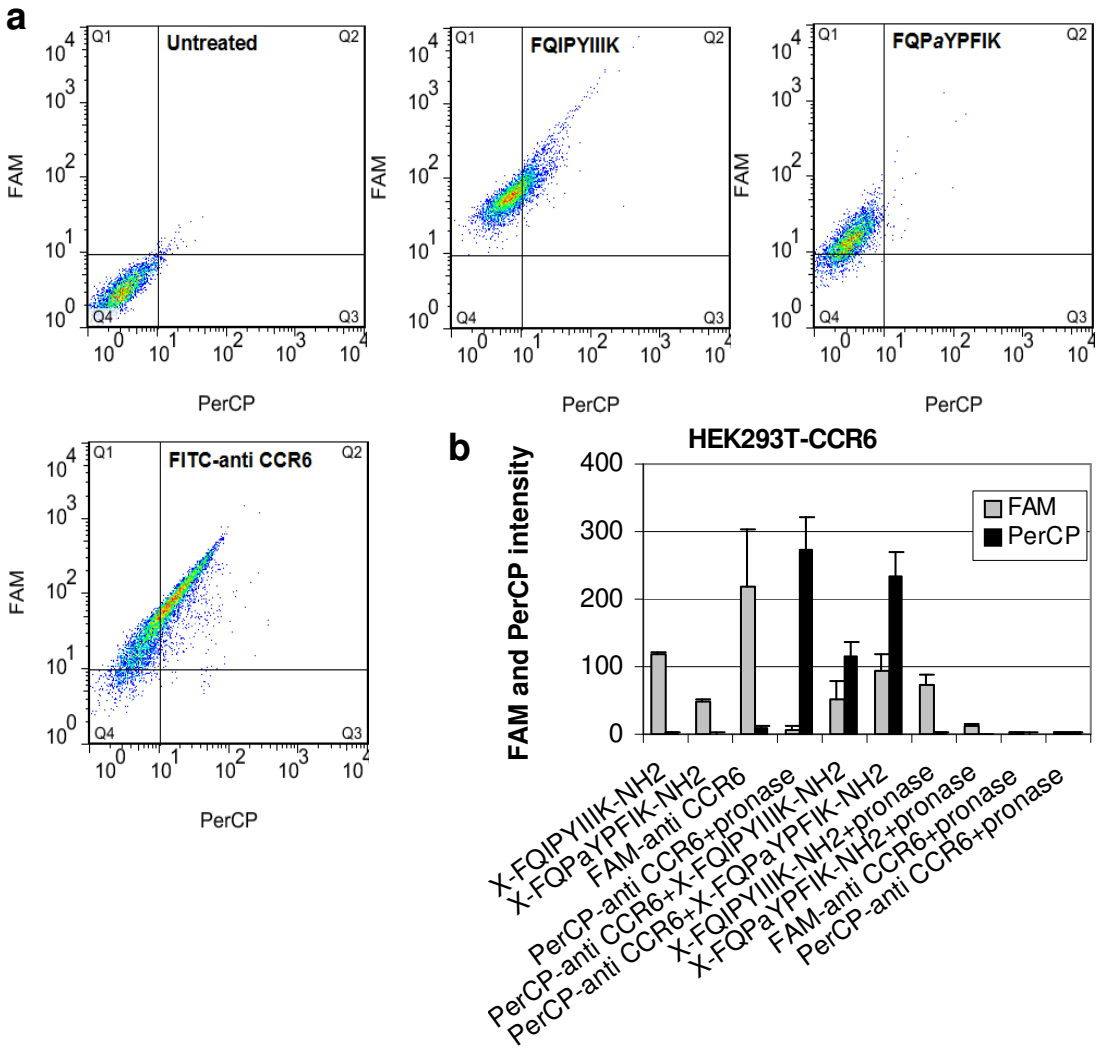


Figure 6.6: Flow cytometry analysis of peptide ligands. (**a**) FAM versus PerCP-filtered scatter plots gaited for live, single cells following incubation with X-

$FQAA_4AA_3YAA_2AA_1IK\text{-}NH_2$ or anti-human CCR6 using a FACS Aria cytometer with FITC and PE-Texas-Red filters (10,000 populations, n=3). HEK293T-CCR6 cells were detached by scraping for minimal destruction of the extracellular receptor moieties or by trypsin/EDTA as a control for extracellular binding of peptides. Peptides are written as simplified sequences. (**b**) Mean FAM and PerCP-filtered fluorescence of scatter plots versus the labelled peptides and antibodies. Compensation was applied between FAM/PerCP fluorophore pair. Error bars indicate ± s.d. (n=3).

6.5.2 Peptide-cell binding is receptor mediated

To verify that binding of the identified ligands was mediated by the receptor in question, a soluble inhibitor for the respective receptor labelled with a different fluorophore than the peptides were co-incubated with the peptides. The D54 cells overexpress $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrins[196,197,198], which can be blocked by the natural ligand, RGD; whereas, the CCR6 receptor can be blocked by the agonist, anti-human CCR6.

Thus, D54 cells were incubated with equimolar Z-RGD-NH_2 and X-FQSIYP*p*IK-NH_2, which blocked X-FQSIYP*p*IK-NH_2 binding by 55% and Z-RGD-NH_2 binding by 57% (Figure 6.5 and appendix VII). Z-RGD-NH_2 binding could almost be completely blocked (80%) by addition of 10 equivalents of X-FQSIYP*p*IK-NH_2 (appendix VII). Whereas, incubation with X-FQKEYKFIK-NH_2 and Z-RGD-NH_2 had no effect on binding of either peptide (Figure 6.5 and appendix VII). This illustrates that X-FQSIYP*p*IK-NH_2 binding to D54 is integrin mediated, whereas X-FQKEYKFIK-NH_2 binding is not.

Similarly, HEK293T-CCR6 cells were incubated with equimolar PerCP-anti human CCR6 and X-FQIPYIIIK-NH_2, which inhibited X-FQIPYIIIK-NH_2 binding by 56% and PerCP-anti human CCR6 binding by 58%, and 10 equivalents of X-FQIPYIIIK-NH_2 completely blocked anti human CCR6 binding (Figure 6.6 and appendix VII). Whereas, incubation with X-FQP*a*YPFIK-NH_2 and PerCP-anti human CCR6 had no effect on binding of either peptide (Figure 6.6 and appendix VII). This demonstrates that X-FQIPYIIIK-NH_2 binding to HEK293T-CCR6 is CCR6 mediated, whereas X-FQP*a*YPFIK-NH_2 binding is not.

As a negative control for CCR6 binding unmodified HEK293T cells were also incubated with X-FQIPYIIIK-NH_2, X-FQP*a*YPFIK-NH_2, and PerCP-anti human CCR6 in the same manner as the HEK293T-CCR6 cells. As expected, much smaller shifts in the levels of fluorescence was induced in the HEK293T cells compared to the HEK293T-CCR6 cells and no competition between the peptides and the antibody was observed (Figure 6.7 and appendix VII). In addition, the fluorescence signals could not be significantly reduced by pronase/EDTA treatment (Figure 6.7 and appendix VII). These differences in binding of X-FQIPYIIIK-NH_2 in HEK293T-CCR6 and HEK293T cells further illustrates that X-FQIPYIIIK-NH_2 binding is CCR6 mediated.

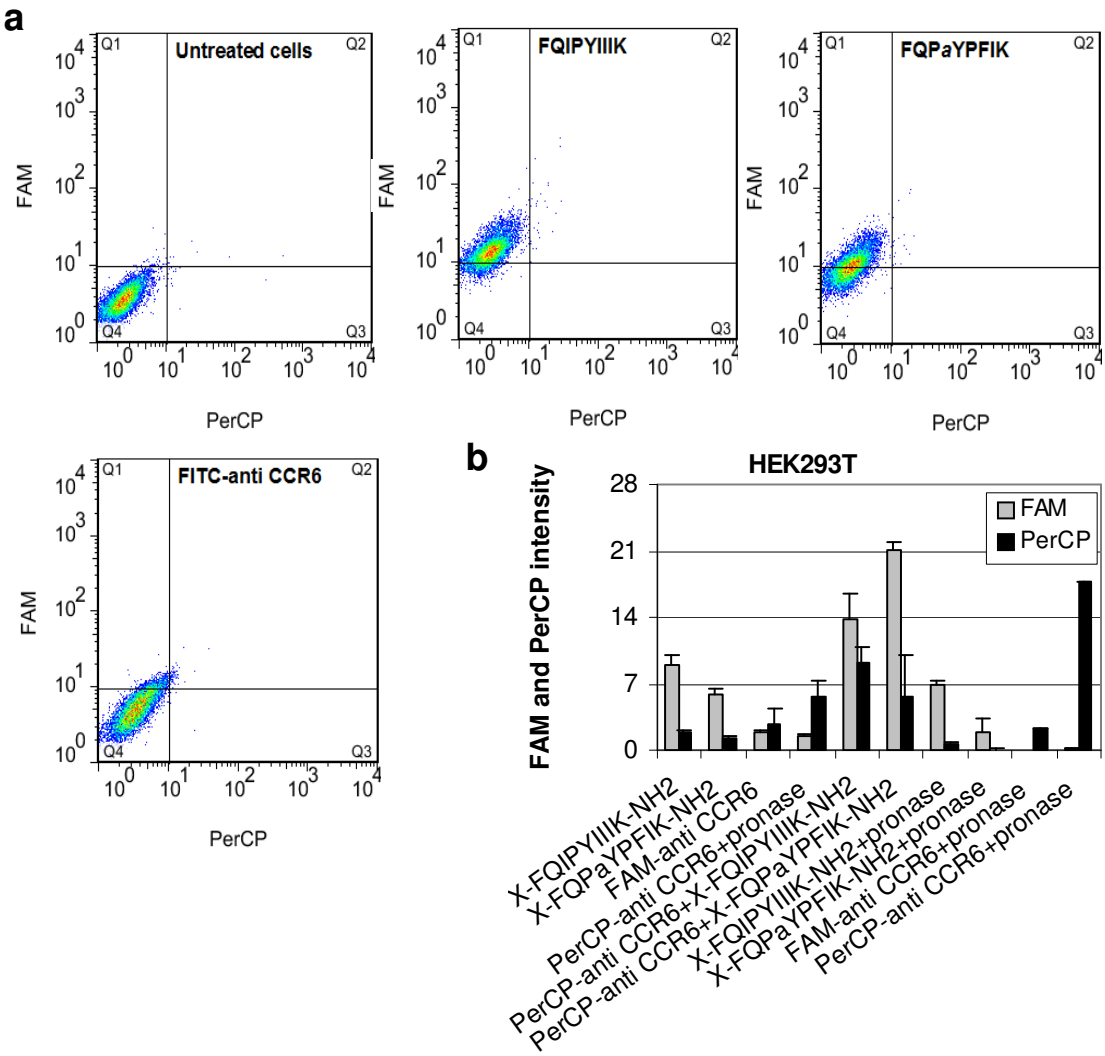


Figure 6.7: Flow cytometry analysis of peptide ligands. (**a**) FAM versus PerCP-filtered scatter plots gaited for live, single cells following incubation with X-FQAA_4AA_3YAA_2AA_1IK-NH_2 or anti-human CCR6 using a FACS Aria cytometer

with FITC and PE-Texas-Red filters (10,000 populations, n=3). HEK293T cells were detached by scraping for minimal destruction of the extracellular receptor moieties or by trypsin/EDTA as a control for extracellular binding of peptides. Peptides are written as simplified sequences. (**b**) Mean FAM and PerCP-filtered fluorescence of scatter plots versus the labelled peptides and antibodies. Compensation was applied between FAM/PerCP fluorophore pair. Error bars indicate ± s.d. (n=3).

6.6 Cytotoxicity

In order for the peptides to have biological applications it is vital that they do not exhibit cellular toxicity. Cell viability upon treatment with the "hit"-peptides was assessed with standard MTT[186] assays demonstrating no cytotoxicity was exhibited by any of "hit"-peptides in the tested cell lines at 100 μM (Table 6.3).

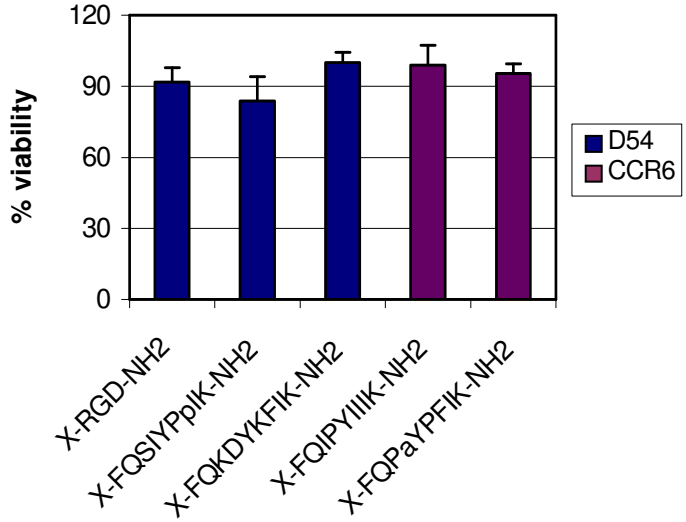


Table 6.3: Cytotoxicity of "hit"-peptides assessed as cell viability by MTT assays. Cells were incubated with X-FQAA_4AA_3YAA_2AA_1IK-NH_2 (100 μM) or X-RGD-NH_2 (100 μM; 24 h) the media was changed and the incubation continued for 24 h and absorbance was measured at 580 nm. Untreated cells are assumed to be 100% variable. Error bars indicate ± s.d. (n = 3).

6.7 Conclusion

Microarray analysis of a 10,000 member PNA encoded peptide library followed by scatter plot analysis allowed consensus sequences for both receptor binding (FQSIYP*p*IK and FQIPYIIIK) and non-binding peptides (FQKEYKFIK and

FQP*a*YPFIK) to be extracted for $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrins and CCR6 respectively that demonstrated low cytotoxicity.

Synthesis followed by flow cytometry analysis of these "hit"-peptides allowed verification of receptor binding. Competition studies with the natural ligands for integrins and CCR6 demonstrated that binding of X-FQ*p*IYIIIK-NH_2 and X-FQIPYIIIK-NH_2 is $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrin and CCR6-mediated respectively, whereas X-FQKKYSRIK-NH_2 and X-FQPKYSFIK-NH_2 do not bind to these receptors.

This "microarray-screening-*post*-cellular-selection" approach to screening of encoded peptide ligands for cell surface receptors has a number of advantages: 1) exclusion of false positives from to unspecific cellular attachment to solid supported ligands; 2) the method allows ligand binding to both adherent and suspension cells. However, this technique is based of screening of ligands in suspension, and hence, does not allow screening for ligands promoting cell adhesion and cell migration.

In contrast, the previously reported "microarray-screening-of-cell-binding-ligands" approach to screening of encoded peptide ligands for cell surface receptors (Chapter 4) has the advantages of allowing: 1) screening of ligands that promote cell adhesion; 2) screening of ligands that induce cell migration (by measuring the exact microarray location of each adhered cell at different time points); 3) screening of both adherent and suspension cells. Though, this method is based on screening of microarray supported ligands, which may lead to false positives due to unspecific cellular adhesion caused by cell secreted proteins.

In conjunction, the two approaches can complement each other as illustrated with the identification of two *different* ligands for $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrins, X-FQ*p*IYIIIK-NH_2 and X-FQ*p*IYIIIK-NH_2 by "microarray-screening-*post*-cellular-selection" and "microarray-screening-of-cell-binding-ligands" respectively.

Moreover, the two methods independently identified X-FQIPYIIIK-NH_2 as the best CCR6 ligand, which further establishes both of these screening strategies as reliable

methods for high-throughput screening for the identification of ligands for cell surface receptors and offers a tool for investigating differences in membrane receptor expression between different cell types including diseased cells.

This approach establishes a strategy for high-throughput screening for the identification of ligands for integrins and GPCRs offering and efficient approach to de-orphanise the many GPCRs, for which there are no known ligands. As integrins and GPCRs amongst the most heavily investigated drug targets in the pharmaceutical industry today[115,137] identification of new agonist or antagonists of these receptors by this method can help identify new drug-targets as well as leads.

The technology would also allow screening of a wide variety of cell surface receptors and offers a tool for investigating differences in membrane receptor expression between different cell types including diseased cells.

Chapter 7 Experimental

7.1 General reaction conditions

All reactions were carried out at room temperature under N_2 unless otherwise stated. All solvents including those for flash chromatography and crystallisation were HPLC grade. Commercially available reagents were used as delivered unless otherwise described.

^{1}H nuclear magnetic resonance (NMR) and ^{13}C NMR spectra were obtained for all isolated organic compounds using a Bruker DP360 spectrometer operating at 360 MHz for ^{1}H NMR and 90 MHz for ^{13}C NMR, chemical shifts, δ are in ppm relative to internal standard $CHCl_3$ (residual, 7.27 ppm for ^{1}H NMR) and $CDCl_3$ (77.0 ppm for ^{13}C NMR) in the solvents indicated at 298 K. Coupling constants (*J*) are reported in Hz.

Infrared (IR) spectroscopy was recorded on a Fourier transform Bruker Tensor 27 fitted with a Specac single reflection diamond attenuated total reflection Golden Gate. All samples were run neat, frequencies are reported in cm^{-1} and only frequencies corresponding to significant functional groups are reported.

Thin layer chromatography (TLC) analysis was performed with aluminium backed silica plates (Merk: Silica gel 60 F_{254}). The TLC plates were visualised by ultraviolet (UV) detection at 254 nm or by dipping in 5 % phosphormolybdic acid in ethanol (PMA-DIP), or 0.6 (w/w)% ninhydrin in ethanol (ninhydrin-DIP), or 1 % $Ce(SO_4)_2$, 2.5% $(NH_4)Mo_6O_{24}$, and 10% H_2SO_4 (CEMOL-DIP), or 1% $KMnO_4$, 6.7% K_2CO_3, and 0.08% NaOH in water ($KMnO_4$-DIP) and developed with a heat-gun.

Flash chromatography was carried out as described by Still *et al.*[215] by use of silica gel 60 Å (Amicon 85040, 35-70 μ). Solvents were evaporated by concentration *in vacuo* at ~ 2-800 mmHg and 40 ^{o}C. Single peptides and peptide-PNA conjugates were characterised by RP-HPLC performed on an Agilent Technologies 1100 modular HPLC system with detection by UV-absorbance according to the following methods:

HPLC method 1: Supelco Discovery C18 Column, 5 cm x 4.6 nm x 5 μm; solvent A: H_2O with 0.1% TFA; solvent B: CH_3CN with 0.1% TFA, eluting with 10% B to 90% B over 3 min; 90% B for 1 min;
HPLC method 2: Phenomenex Luna Column, C18, 15 cm x 1.00 cm, 5 μm; λ = 220 nm; solvent A: H_2O with 0.1% TFA; solvent B: CH_3CN with 0.1% TFA, eluting with A for 5 min; 100% A to 100% B over 20 min; 100% B for 5 min;
HPLC method 3: Phenomenex Luna Column, C18, 15 cm x 1.00 cm, 5 μm.; λ = 254 nm; solvent A: H_2O with 0.1% formic acid; solvent B: MeCN with 0.1% formic acid, eluting with 95% A to 95% B over 3 min; 100% B for 1 min.

Electronspray mass spectrometry (MS) performed on an Agilent Technologies LC/MSD Series 1100 quadrupole mass spectrometer, equipped with an electronspray ion source. MALDI-TOF spectrometry performed on an Applied Biosystems Voyager DESTR MALDI-TOF mass spectrometer (matrix: sinapinic acid in CH_3CN/H_2O 1:1).

UV-visible spectroscopy was performed on an 8453 Agilent UV-Visible spectrophotometer. Fluorescence measurements were performed on a Jobin Yvon Spex Fluoromax spectrofluorometer or a FS900 spectrofluorometer.

Microwave irradiated reactions were carried out in a Biotage Initiator Sixty microwave chamber. Peptides, and in some cases PNA, were synthesised in a CEM Liberty Microwave peptide synthesiser.

PCR was performed in a Techne TC-312 PCR cycler unless otherwise stated. DNA gel electrophoresis was carried out on a Bio-Rad PowerPac 300 at 90 V. PCR on microarrays was performed using a GeneMachines Hyb4 automated station (Genomic Solutions, UK). DNA Illumina Solexa Sequencing was performed by the GenePool facility at The University of Edinburgh.

7.1.1 Qualitative Ninhydrin test

5-20 resin beads were washed with MeOH (x 3) and 6 drops of reagent A [50 mg/mL of ninhydrin (1,2,3-indanetrione monohydrate, Sigma-Aldrich) in EtOH] was added followed by 3 drops of reagent B [2 g/mL phenol (Sigma-Aldrich) and 10 μM KCN (Sigma-Aldrich) in EtOH/pyridine 1:1] and the sample was heated to 100 ^{o}C for 5 min. A blue colour solution was indicative of primary amines (positive) and a yellow colour indicative of no primary amines present (negative).

7.1.2 Preparation of samples for MALDI-TOF

Samples in H_2O (1 μL of ~1 μg/mL) was mixed with matrix [1 μL, 15% sinapic acid (Sigma-Aldrich) in H_2O/AcN 1:1] and the resulting suspension was loaded onto a stainless steel MALDI-TOF plate and allowed to dry for at least 1 h before running MALDI-TOF.

7.1.3 DNA gel electrophoresis and purification of DNA

Samples (20-30 μL) were prepared with 6 x Blue/Orange Loading Dye (5 μL, Promega) and DNA grade H_2O were run on a 5 (wt)% agarose gel (Promega Preparative grade for small fragments) with either 600 ng/mL ethidium bromide (EtBr, for dsDNA) or 1 x SYBR Green II RNA Gel Stain (for ssDNA) in 1 x Tris Borate Ethylenediaminetetraacetatic acid (pH 8.3, TBE, Fisher Scientific) buffer for approximately 1 h. The gel was analysed under UV light and the appropriate bands were excised with a scalpel.

DNA was purified using a QIAEX II Agarose Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. 600 μl of buffer QX1 per each 100 mg of agarose gel was added to the gel followed by addition of 30 μl QIAEX II per 10 μg DNA and the suspension was incubated at 50°C for 10 min and vortexed every 2 min for 15 s to solubilise the agarose and bind the DNA. The samples were centrifuged at maximum speed for 30 s and the supernatant discarded. The pellet was washed with 500 μl of buffer QX1 and 500 μl of buffer PE x 2, during which the beads were pelleted by centrifugation at maximum speed and the pellet was resuspended by vortexing for 15

s. The pellet was air dried for 30-60 min and the DNA was eluted with 2 x 20-40 mL Tris·Cl (10 mM), pH 8.5 following 5 min incubation and recovery of the supernatant after centrifugation at maximum speed. Samples were stored at -20 ^{o}C.

7.2 General microarray procedures

All buffers and washing solutions were made up with cell tissue culture H_2O and all DNA work was carried out with DNA grade H_2O (Promega). Printing was carried out on a Genetix QArray Mini with Genetix Qsoft Microarraying software in 70% humidity. Hybridisations were carried out in 1 x GenHyb buffer (Genetix) in an Agilent Hybridisation Chamber in the oven *or* in an automated GeneMachines Hyb4 automated hybridiser with Hyb4 Editor (Genomic Solutions) software. PCR on microarrays were also carried out in the GeneMachines Hyb4 automated hybridiser. Customised microarrays were purchased from OGT. All arrays were sterilised under UV light for 5 min before cellular incubation. The microarrays were imaged using a Nikon Eclipse 50i microscope with FITC, DAPI, and tetramethyl rhodamine isothiocyanate (TRITC) filters and Pathfinder software (Imstar) *or* a Tecan LS Reloaded microarray scanner with ArrayPro 32 Analyser (Media Cybernetics Inc) software, *or* a CCD based microarray scanner (BioTEc LaVision BioAnalyser) using FITC, DAPI and Cy_3 filters. Image analysis was performed using Bluefuse software (BlueGnome, Cambridge) and data analysis was performed using Microsoft Excel and clustering analysis was performed using Enthought Python software.

7.2.1 Stringency washes after DNA microarray hybridisation

OGT arrays were washed according to standard protocols: 1) Washed with OGT recommended buffer [20 mL, 100 mM NaCl (Fisher Scientific), citric acid (40 mM, Acros), 0.7% (w/v) *N*-lauroylsarcosine sodium salt (Aldrich), EGTA (0.1 mM, Fisher Scientific), pH 7.5] at the hybridisation temperature for 10 min, rinsed with H_2O (20 mL x 3), rinsed with Tris buffer at pH 8.0 (20 mL, 50 mM, Fisher), and dried under a N_2 flow; *or* 2) washed with Gene Expression Wash buffer I (20 mL, Agilent) at room temperature for 2 min, washed with Gene Expression Wash buffer II (20 mL,

Agilent) at 37 ^{o}C for 2 min, rinsed with H_2O (20 mL x 3), rinsed with Tris buffer at pH 8.0 (20 mL, 50 mM, Fisher), and dried under a N_2 flow; *or* 3) washed with 0.2% Sodium Dodecyl Sulphate (SDS, Promega) in 2 x Saline-Sodium Citrate (SSC, 20 mL, Promega) for 5 min, 0.2 x SSC (20 mL) for 5 min, 0.1 x SSC (20 mL) for 5 min, and briefly rinsed in DNA grade H_2O (20 mL) and Tris buffer at pH 8.0 (20 mL, 10 mM) and dried under a N_2 flow. Codelink (Amersham Biosciences) arrays were washed acetonitrile (20 mL x 2) briefly, H_2O (20 mL x 2) briefly, 0.1% SDS in 4 x (20 mL) at 50 ^{o}C for 30 min, H_2O briefly (20 mL x 2), 10 mM Tris pH 8 (20 mL) briefly, and dried under a N_2 flow.

7.3 General biological work

Cells were cultured in a Heraeus HERAcell 150 incubator and cell experiments were performed in a Heraeus HERAsafe KS 18 class II negative flow cabinet. Flow cytometry was performed on a BD Biosciences FACSAria system with the BD FACSDiva software *or* a BD Biosciences Calibur system with BD CellQuest software and FlowJo software was used for analysis. Confocal microscopy was performed on a Deltavision Tollervey microscope with a Photometrics Coolsnap HQ camera, SoftWoRx (Applied Precision) software using FITC (Ex = 490/20 nm, Em = 528/38 nm) and TRITC (Ex = 545/30 nm, Em = 610/75 nm) filters (100 x magnification). Cell viability was assessed recording absorbance at 570 nm using a Bio-Rad 1.15 microplate reader.

7.3.1 Cell culture

HeLa and K562 (courtesy of Prof. Tim Elliot, Cancer Research UK, University of Southampton) cells were maintained in Roswell Park's Memorial Institute (RPMI) 1640 (Sigma-Aldrich) complete media [CM: 10% foetal bovine serum (FBS, Biosera), 2 mM L-glutamine, 100 mg/mL streptomycin, and 100 U/mL penicillin (Gibco)]. HEK293T (ECACC), HEK293TCCR6 (courtesy of Dr. Dominic Compopiano, University of Edinburgh, UK), D54 (courtesy of Dr. Darell D. Bigner, Duke University Medical Centre, California), and ARN8 (courtesy of Prof. Ted Hupp, The Edinburgh Cancer Research Centre, The University of Edinburgh) were

maintained in Dubelco's Modified Eagle Medium CM (DMEM CM, Sigma-Aldrich). SH-SY5Y (European Collection of Cell Cultures) were maintained in Ham's F12: Eagle's Minimum Essential Medium (EMEM, Sigma-Aldrich) (1:1) supplemented with 15% FBS (Biosera); 2 mM L-glutamine, 100 mg/mL streptomycin, 100 U/mL penicillin (Gibco); and 1% Non Essential Amino Acids (NEAA, Gibco). E14 cells (mouse embryonic stem cells, courtesy of Prof. Joshua Brickman, The Institute for Stem Cell Research, The University of Edinburgh, UK) were grown on gelantine-coated polystyrene (polystyrene was treated with 3% gelatine in phosphate buffered saline (PBS) for 5 min prior to seeding the cells) in Glasgow Minimum Essential Medium (GMEM, Sigma-Aldrich) CM supplemented with 1 mM sodium pyruvate (Gibco), 1% NEAA (Gibco), 0.1 mM β-mercaptoethanol, and 100 U/mL leukaemia inhibitory factor (Gibco). All cells were grown in T-75 or T-25 flasks (Nunc) in 5% CO_2 at 37 ^oC.

Adherent cell lines were passaged by removing the old media from the flask, washing the surface with PBS (5 mL, Oxoid), and detached by standard trypsination (incubating the cells with a minimal amount of 0.05% trypsin and 0.53 mM ethylenediaminetetraacetatic acid (EDTA) in PBS, Fisher Scientific), or by scraping. Suspension cell lines were passaged by collecting the cells by centrifugation (1100 rpm for 5 min) and the old media was removed. Hereafter, the cells were suspended in appropriate media and transferred to a sterile tube, a fraction was left in the flask and fresh media was added. The cells were used within an hour after they were transferred to the tube. Before microarray incubation, the cells were counted following 4 x dilution in trypan blue (0.4 mM, Sigma-Aldrich).

Primary cells were isolated and purified as previously reported[216] and subsequently maintained in Improved Modified Eagle Medium (IMEM, Invitrogen). Briefly, with local ethical approval, freshly obtained anticoagulated blood was centrifuged, plasma removed and the white cell fraction was separated using Percoll density gradient centrifugation. Neutrophils were obtained from the granulocyte layer and washed and confirmed to be 95% pure by flow cytometry. The mononuclear cells were enriched for monocytes by adherence for one hour on treated sterile culture plates.

Lymphocytes were obtained from the non-adherent fraction. Flow cytometry confirmed that the monocytes were >80 % pure and the lymphocytes >80% pure.[§]

7.3.2 Fixing and microscopy analysis of cells

Cells were fixed with 4% *para*-formaldehyde (Sigma-Aldrich) in PBS for 15-30 min and washed with PBS x 3. The attached cells were stained with DAPI (20 μL/mL, Sigma-Aldrich) in PBS for 15 min and washed with PBS and/or stained with AlexaFluor 568 phalloidin (1 μL/mL, Invitrogen) in PBS for 15 min and washed with PBS x 3.

7.3.3 MTT Cytotoxicity assays

Cells were grown to ~50% confluence (to avoid cells reaching complete confluence during the 48 h experiment) in 96-well plates and incubated with FAM-labelled peptide (100 μM) in appropriate growth media for 24 h at 37 ^{o}C and the media was removed and fresh media was added and the incubation continued for 24 h with the last row was used as a blank (no cells seeded). Cells were washed with PBS and 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, 50 μL, 1mg/mL, Sigma-Aldrich) in Iscove's Modified Dulbecco's Media (IMDM) was added and incubated for 4 h at 37 ^{o}C. Hereafter, MTT solubilising solution [60 μL, 10 % Triton-X 100 (Sigma-Aldrich), 0.1 N HCl (Fisher Scientific) in isopropanol] was added and the well-plate was shaken until complete dissolution of the formazan crystals. Absorbance was measured at 570 nm in a Bio-Rad Benchmark microplate reader.

Untreated cells were considered to be 100% viable and cell viability was calculated from equation 7.1:

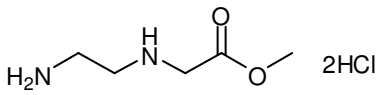
$$\%\ \text{Viable Cells} = \frac{\text{Abs}}{\text{Abs}_{\text{untreated\_cells}}} \cdot 100\% \qquad \text{(Equation 7.1)}$$

[§] Performed by Dr. Kevin Dhaliwal

7.4 Experimental for Chapter 2

All counpounds described in the following sections (7.4.1-7.4.14) were synthesised according to previously reported procedures[14] and unless otherwise described the obtained TLC, NMR, HPLC, and LCMS data of the products is consistent with the published data.

7.4.1 Synthesis of methyl *N*-aminoethylglycinate·2HCl salt[14]



7.1

Chloroacetic acid (105.82 mmol, 10.00 g, Sigma-Aldrich) was added portion-wise to ethylenediamine (60 mL, Sigma-Aldrich) at 4 ^{o}C and the reaction was allowed to warm to room temperature while stirring for 16 h. The solvent was removed by concentration *in vacuo* (<50 ^{o}C). This gave a crude yellow oil, which was further purified by crystallisation from dimethylsulfoxide (DMSO). The resulting white crystals were collected by filtration and washed with DMSO (3 x 50 mL), THF (3 x 50 mL), and diethyl ether (3 x 50 mL) and dried *in vacuo* at 50 ^{o}C overnight to give a white crystalline compound 10.56 g (84%), which was carried directly into the next step without further purification.

The product (84.65 mol, 10.00 g) was added to MeOH (300 mL) at 0 ^{o}C followed by drop wise addition of $SOCl_2$ (846.5 mol, 61.8 mL, Sigma-Aldrich) and the reaction was warmed to reflux temperature and stirred for 16 h. The product was filtered and washed with MeOH and dried in a vacuum oven overnight to yield **7.1** as a white solid 10.84 g (62%).

*R*f (EtOAc) = 0.31.

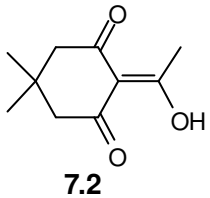
^{1}H NMR (250 MHz, DMSO): δ 9.82 (s, 1H, NH), 8.50 (s, 2H, NH_2), 4.06 (s, 2H, CH_2CO), 3.75 (s, 3H, CH_3), 3.41-3.11 (m, 4H, 2 CH_2).

^{13}C NMR (62.5 MHz, DMSO): δ ppm 167.28 (C=O), 52.97 (CH_2C=O), 46.93 (C H_2NH), 44.36 (C H_2N H_2), 35.34 (CH_3O).

RP-HPLC (method 1): 0.9 min (254 nm, 81.1% purity).

LRMS (ES^+): m/z (%)133.0 $[M+H]^+$ (100).

7.4.2 Synthesis of 4,4-dimethyl-2,6-dioxocyclohexyliden (Dde-OH)[14]



7.2

Acetic acid (99.92mmol, 5.72 mL) was added to DMF (500 mL) at room temperature followed by dimedone (109.9 mol, 15.41 g, Sigma-Aldrich), DCC (99.92 mol, 20.62 g, Sigma-Aldrich), and 4-dimethylaminopyridine (DMAP, 99.92 mol, 12.21 g, Sigma-Aldrich) and the reaction was stirred for 48 h. Precipitated DCU was removed by filtration and the precipitate was washed with DMF. The combined organic phases were concentrated *in vacuo* and the residue was dissolved in EtOAc (800 mL). The organic solution was washed with $KHSO_4$ (100 mL, 1 M) and the product extracted with $NaHCO_3$ (100 mL). The aqueous solution was extracted with DCM and the solution was concentrated *in vacuo* to give a crude product, which was precipitated from MeOH/H_2O 1:1 to afford Dde-OH as a yellow oily solid 11.19g (56%). Alternatively, the product could be purified by flash chromatography with eluent DCM/MeOH 97:3 to give **7.2** 11.2 g (62%).

*R*f (EtOAc) = 0.70

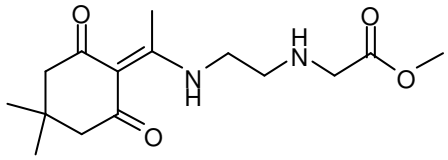
^{1}H NMR (250 MHz, $CDCl_3$): δ ppm 2.53 (s, 3H, CH_3), 2.38 (s, 4H, 2 CH_2), 1.01 (s, 6H, $C(CH_3)_2$). *OR* ^{1}H NMR (250 MHz, $CDCl_3$) : δ ppm 2.54 (s, 3H, CH_3), 2.47 (s, 2H, CH_2), 2.29 (s, 2H, CH_2), 1.01 (s, 6H, $C(CH_3)_2$).

^{13}C NMR (62.5 MHz, $CDCl_3$): δ ppm 202.74 (C=C), 197.00 (C=O), 112.70 (C=COH), 49.50 ($C(CH_3)_2$), 30.97 (C H_2), 28.82 (CH_3C=C), 28.53 (CH_3).

RP-HPLC (method 1): 3.76 min (254 nm, 73.7% purity).

LRMS (ES^+): m/z (%) 183.1 $[M+H]^+$ (100).

7.4.3 Synthesis of *N*-{2-[1-(4,4-dimethyl-2,6-dioxocyclohexyliden) ethylamino}-ethyl}glycinate (PNA backbone)[14]



7.3

7.2 (47.9 mol, 9.83 g) was added to DCM/EtOH 1:1 (200 mL) at room temperature followed by *N*,*N*-diisopropylethylamine (DIPEA, 95.87 mol, 16.20 mL, Sigma-Aldrich) and **7.1** (47.9 mol, 8.74 g) and the reaction was stirred at room temperature for 16 h. The solvent was evaporated *in vacuo* and the crude solid was dissolved in EtOAc (200 mL) and extracted with $KHSO_4$ (4 x 60 mL, 1M). The aqueous phase was brought to pH 9 with $NaHCO_3$ and the aqueous phase was extracted with EtOAc (5 x 60 mL). The combined organic phases were washed with brine (60 mL) and dried over $MgSO_4$ and concentrated *in vacuo* to yield **7.3** as a yellow oily solid 12.1 g (85%).

*R*f (MeOH/EtOAc 1:9) = 0.41.

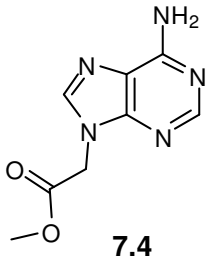
^{1}H NMR (250 MHz, DMSO): δ ppm 3.85 (s, 3H, OCH_3), 3.68 (dd, *J* = 11.44, 5.68 Hz, 2H, NHC*H*$_2$), 3.61 (s, 2H, NHC*H*$_2$CO), 2.99 (t, *J* = 5.99, 5.99 Hz, 2H, NHC*H*$_2$), 2.69 (s, 3H, CCH_3), 2.47 (s, 4H, $(CH_2)_2$-Dde), 1.16 (s, 6H, $(CH_3)_2$-Dde).

^{13}C NMR (62.5 MHz, $CDCl_3$): δ ppm 173.81 (C=O), 173.13 (C=O), 108.38 (HNC=C), 52.25 (CH_3C=C), 50.66 (HNCH_2C=O), 48.04 (HNCH_2), 43.82 (HNCH_2), 30.43 (C(CH_3)$_2$), 28.64 (CH_3O), 18.50 (CH_3).

RP-HPLC (method 1): 2.31 min (ELSD, 100% purity).

LRMS (ES$^+$): m/z (%) 297.1 [M+H]$^+$ (100).

7.4.4 Synthesis of [N^6-(4-methoxytrityl)-adenin-9-yl]-acetic acid[14]



7.4

Adenine (103.6 mmol, 14.0 g, Acros) was added to dry DMF (400 mL) at room temperature under N_2 followed by NaH (114.0 mmol, 4.56 g, 60% in mineral oil, Sigma-Aldrich) and the reaction was stirred for 1 h. The reaction was cooled to 0 °C and methyl 2-bromoacetate (103.6 mmol, 9.81 mL, Fluka) was added drop wise and the reaction was stirred for 4h. The solvent was evaporated *in vacuo*, and the crude solid was washed with H_2O, sonicated for 5 min, and filtered. The solid was dried over P_2O_5 overnight in a vacuum oven to give **7.4** as a white solid 15.01 g (69%).

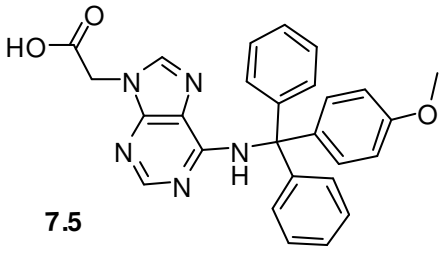
*R*f (MeOH/EtOAc 1:1) = 0.67.

^{1}H NMR (250 MHz, DMSO) δ ppm 8.14 (s, 1H, CH), 8.12 (s, 1H, CH), 7.29 (s, 2H, NH_2), 5.09 (s, 2H, CH_2), 3.71 (s, 3H, CH_3).

^{13}C NMR (62.5 MHz, DMSO): δ ppm 186.83 (H_2NC=N), 156.34 (C=O), 153.02 (=NCH=N), 150.06 ((N)2C=C), 141.57 (-NCH=N), 118.62 (-NC=C), 52.78 (CH_3), 44.17 (C H_2).

RP-HPLC (method 1): 3.69 min (ELSD, 93.1% purity).

LRMS (ES^+): m/z (%) 208.1 $[M+H]^+$ (100).



7.5

7.4 (67.57 mmol, 14.00 g) was added to dry DCM/pyridine 1:1 (500 mL) at room temperature under N_2, followed by Mmt-Cl (101.36 mmol, 31.3g, Sigma-Aldrich) and NEM (67.57 mmol, 8.55mL, Sigma-Aldrich). The reaction was heated to 40 °C and stirred for 3 h and then cooled to room temperature and stirred for 16 h. The solvent was evaporated *in vacuo* and the residue was dissolved in EtOAc (800 mL)

and washed with 1 N $KHSO_4$ (300 mL), 10% $NaHCO_3$ (300 mL), and brine (300 mL) and dried over $MgSO_4$. The solvent was evaporated *in vacuo* and the solid was washed with petroleum ether and resulting precipitate was treated with NaOH (250 mL, 1M) and refluxed for 2 h. The temperature was lowered to room temperature and the reaction was stirred for 2 h. The reaction was cooled to 0 ^oC and treated with $KHSO_4$ (350 mL, 1 N) and stirred for 30 min. The precipitate was filtered and washed with small portions of water until the filtrate was neutral and the solid was dried *in vacuo* to give **7.5** as a white solid 19.5 g (58% over two steps).

*R*f (MeOH) = 0.67.

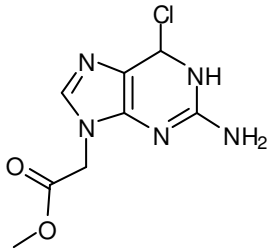
^1H NMR (250 MHz, DMSO: δ ppm 8.37 (s, 1H, CH-pur), 8.08 (s, 1H, CH-pur), 7.51-7.32 (m, 10H, CH-Mmt), 7.27 (d, *J* = 8.68 Hz, 2H, C*H*$COCH_3$-Mmt), 7.03 (d, *J* = 8.74 Hz, 2H, CH-Mmt), 5.13 (s, 2H, CH_2), 3.90 (s, 3H, OCH_3-Mmt).

RP-HPLC (method 1): 2.60 min (ELSD, 98% purity).

LRMS (ES^+): m/z (%) 466.1 $[M+H]^+$ (100).

Mp: 114-121 ^oC.

7.4.5 Synthesis of [N^2-(4-methoxytrityl)-guanin-9-yl]-acetic acid[14]



7.6

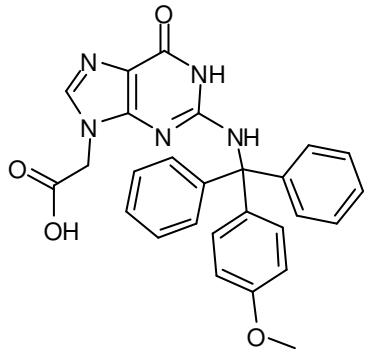
6-chloroguanine (103.6 mmol, 17.6 g, Sigma-Aldrich) was added to dry DMF (400 mL) at room temperature under N_2 followed by NaH (114.0 mmol, 4.56 g, 60% in mineral oil) and the reaction was stirred for 1 h. The reaction was cooled to 0 ^oC and methyl 2-bromoacetate (103.6 mmol, 9.81 mL) was added drop wise and the reaction was stirred for 6 h. The solvent was evaporated *in vacuo*, and the crude oil precipitated with H_2O and the precipitate was filtered and washed with H_2O to yield **7.6** as a white solid 17.85 g (71%).

*R*f (EtOAc) = 0.54.

H NMR (250 MHz, $CDCl_3$): δ ppm 7.77 (s, 1H, CH), 4.80 (s, 2H, CH_2), 3.75 (s, 3H, CH_3).

RP-HPLC (method 1): 4.73 min (ELSD, 95.8% purity).

LRMS (ES^+): m/z (%) 242.0 $[M+H]^+$ (100).



7.7

7.6 (71.82 mmol, 17.50 g) was added to dry THF (200 mL) at room temperature under N_2, followed by Mmt-Cl (107.74 mmol, 33.3g) and DIPEA (71.82 mmol, 9.28 g) and stirred for 16 h at rt. The solvent was evaporated *in vacuo* and the residue was dissolved in ether (400 mL) and washed with 1 N $KHSO_4$ (50 mL), 10% $NaHCO_3$ (50 mL), and brine (50 mL) and dried over $MgSO_4$. The solvent was evaporated *in vacuo* and the resulting crude product was precipitated from DCM/warm petroleum ether (46-60 ^oC) and the residue was treated with NaOH (400 mL, 1 M) and refluxed for 4 h. The temperature was lowered to room temperature and the reaction acidified with HCl (2 N) to pH = 2. The precipitate was filtered and washed with small portions of water until the filtrate was neutral. The solid was dried *in vacuo* for 16 h and washed with diisopropyl ether (30 mL) and dried *in vacuo* to give **7.7** as a white solid 7.33 g (49% over two steps).

*R*f (MeOH) = 0.63.

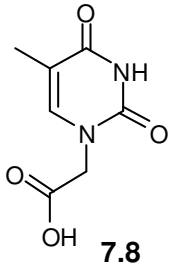
^{1}H NMR (250 MHz, DMSO): δ ppm 10.79 (s, 1H, NH-pur), 7.84 (s, 1H, NH-Mmt), 7.79 (s, 1H, CH-pur), 7.56-7.38 (m, 12H, CH-Mmt), 7.09 (d, *J* = 8.89 Hz, 2H, C*H*$COCH_3$-Mmt), 4.46 (s, 2H, CH_2), 3.96 (s, 3H, CH_3).

RP-HPLC (method 1): 2.67 min (ELSD, 100% purity).

LRMS (ES^+): m/z (%) 505.0 $[M+Na]^+$ (100).

Mp: 200-220 ^oC (decomposed).

7.4.6 Synthesis of thymine-1-yl acetic acid[14]



7.8

Thymine (103.6 mmol, 13.1 g, Acros) was added to dry DMF (400 mL) under N_2 at room temperature followed by K_2CO_3 (113.96 mmol, 26.3 g) and methyl bromoacetate (113.96 mmol, 10.8 mL) and the reaction was stirred for 16 h, which lead to formation of a crude precipitate. The precipitate was collected by filtration and dried *in vacuo*. The precipitate was cooled to 0 ^oC and treated with H_2O (125 mL) and HCl (5 mL, 4 M) and stirred for 30 min. The precipitate was collected by filtration, washed with H_2O (75 mL x 3), and treated with H_2O (125 mL) and NaOH (75 mL, 2 M) and refluxed for 10 min. The reaction was cooled to 0 ^oC and treated with HCL (110 mL, 4 M) and stirred for 30 min. The resulting precipitate collected by filtration and washed with H_2O until the filtrate was neutral and dried over P_2O_5 to yield **7.8** as a white solid 11.22 g (59% over two steps).

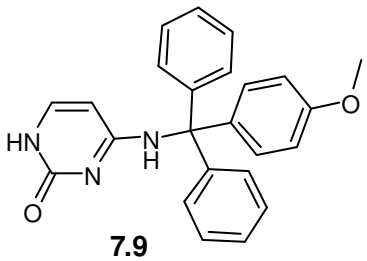
*R*f (MeOH) = 0.61

^{1}H NMR (250 MHz, DMSO): δ ppm 13.26 (s, 1H, CO_2H), 11.50 (s, 1H, NH), 7.66 (d, *J* = 1.18 Hz, 1H, CH), 4.53 (s, 2H, CH_2), 1.91 (s, 3H, CH_3).

RP-HPLC (method 1): 1.64 min (ELSD, 100% purity).

LRMS (ES$^+$): m/z (%) 185.0 $[M+H]^+$ (100).

7.4.7 Synthesis of [N^4-(4-methoxytrityl)-cytosin-9-yl]-acetic acid[14]



7.9

Cytosine (88.56 mmol, 9.84 g, Acros) was added to dry DCM/pyridine (500 mL) at room temperature under N_2, followed by Mmt-Cl (132.85 mmol, 29.7 g) and *N*-

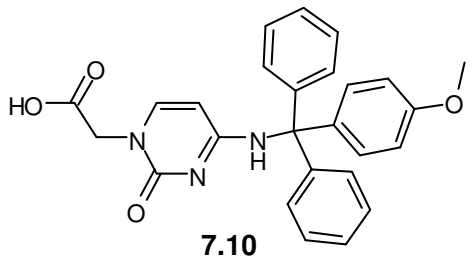
ethylmaleimide (NEM, 88.56 mmol, 10.20 mL). The reaction was heated to 50 °C and stirred for 16 h. Note some batches of cytosine were so insoluble that the reaction could not take place. In these cases the cytosine was dissolved in H_2O with 1.1 equiv. HCl and freeze dried overnight. Hereby the much more soluble cytosine·HCl salt was formed. However, in reaction of this compound 2 equiv. NEM were added to the reaction. The solvent was evaporated *in vacuo* and the residue was dissolved in EtOAc (800 mL) and washed with 1N $KHSO_4$ (300 mL), 10% $NaHCO_3$ (300 mL), and brine (300 mL) and dried over $MgSO_4$. The solvent was evaporated *in vacuo* and the solid was washed with petroleum ether to give **7.9** as a white solid 23.61 g (70%).

*R*f (EtOAc) = 0.65

^{1}H NMR (250 MHz, DMSO): δ ppm 10.17 (s, 1H, NH), 7.45 (d, *J* = 7.26 Hz, 1H, NHC*H*), 7.30-6.92 (m, 12H, CH-Mmt), 6.80 (d, *J* = 7.69 Hz, 2H, C*H*COCH$_3$-Mmt), 4.86 (d, *J* = 7.17 Hz, 1H, CH), 3.59 (s, 3H, CH_3).

RP-HPLC (method 1): 3.99 min (ELSD, 100% purity).

LRMS (ES$^+$): m/z (%) 273 [Mmt]$^+$ (100), 406.0 [M+Na]$^+$ (17).



7.10

7.9 (60.00 mmol, 23.00 g) was added to dry DMF (200 mL) at room temperature under N_2 followed by NaH (66.00mmol, 2.64g, 60% in mineral oil) and the reaction was stirred for 1 h. The reaction was cooled to 0 °C and methyl 2-bromoacetate (60.00 mmol, 5.7 mL) was added drop wise and the reaction was stirred for 4 h. The solvent was evaporated *in vacuo* and the resulting crude product was precipitated from H_2O, collected by filtration, and treated with NaOH (500 mL, 2 M) and refluxed for 2 h. The temperature was lowered to room temperature and solution was washed with DCM (100 mL x 3) and acidified with HCl (2 N) until pH = 2. The precipitate was filtered and washed with small portions of water until the filtrate was

neutral. The solid was dried *in vacuo* overnight to give **7.10** as a white solid 26.07g (98% over two steps).

*R*f (EtOAc/MeOH 1:1) = 0.47.

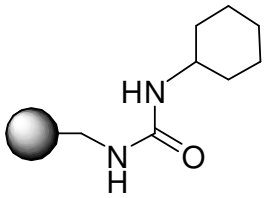
^{1}H NMR (250 MHz, DMSO): δ ppm 12.81 (s, 1H, CO_2H), 8.32 (d, *J* = 7.90 Hz, 1H, CH-Cyt), 7.46-7.08 (m, 12H, CH-Mmt), 6.81 (d, *J* = 7.90 Hz, 2H, C*H*$COCH_3$-Mmt), 6.15 (d, *J* = 7.55 Hz, 1H, CH-Cyt), 4.21 (s, 2H, CH_2), 3.70 (s, 3H, CH_3).

RP-HPLC (method 1): 4.21 min (ELSD, 100% purity).

LRMS (ES^+): m/z (%) 273 $[Mmt]^+$ (100), 442.0 $[M+H]^+$ (4).

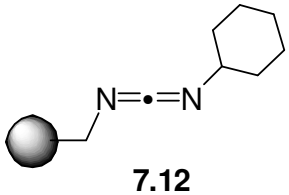
Mp: 207-209 ^oC

7.4.8 Synthesis of *N*'-methylpolystyrene-*N*'-dicyclohexylcarbodiimide (PS-DCC)



7.11

Aminomethyl PS-resin (16 mmol, 10.0 g, GLS Biochem) was swollen in dry THF (40 mL) and cyclohexyl isocyanate (80 mmol, 10.2 mL, Sigma-Aldrich) was added and the reaction was shaken overnight. The reaction was monitored by ninhydrin test. The resin was filtered, washed with DCM (x 3), DMF (x 3), MeOH (x 3) and dried in a vacuum oven overnight at 40 °C to yield **7.11** 11.85 g (99%). IR (neat): ν_{max} / (cm^{-1}) 2925s, 1627s, 1552s, 1449m, 1251m, 758m, 699s.



7.12

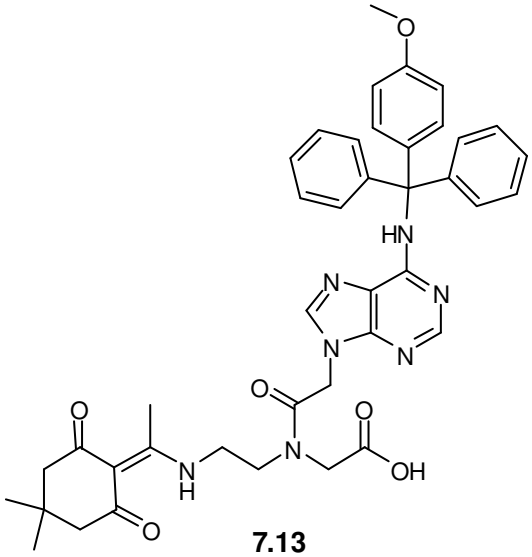
7.11 (4.6 mmol, 3.46 g) was swollen in DCM (30 mL) and triphenylphosphine (13.2 mmol, 3.50 g, Sigma-Aldrich) was added. The mixture was stirred for 10 minutes and CBr_4 (13.2 mmol, 4.42 g, Sigma-Aldrich) was added, immediately followed by drop wise addition of anhydrous TEA (43 mmol, 5.82 mL). The dark reaction mixture was shaken overnight, the resin was filtered and washed with DCM (x 3),

DMF (x 3), DCM (x 3), and dried overnight in a vacuum oven at 40 °C to give **7.12** (PS-DCC) 3.44 g (99%).

IR (neat): ν_{max} / (cm^{-1}) 2930m; 2117s; 1450m; 1338m; 1020m; 813m; 759s.

^{13}C-NMR (250 MHz, $CDCl_3$) δ ppm: 145.1, 55.6, 50.5, 34.8, 25.5, 24.5.

7.4.9 Synthesis of methyl *N*-{2-[N^6-(4-methoxytrityl)-adenin-9-yl]-acetyl}-*N*-{2-[1-(4,4-dimethyl-2,6-dioxocyclohexyliden)-ethylamino}-ethyl}-glycine (Dde-Ade(Mmt)-OH)[14]



7.13

7.5 (2.15 mmol, 1.00 g) was added to dry DMF (25 mL) under N_2 followed by HOBt (1.93 mmol, 0.30 g, Sigma-Aldrich) and the reaction was stirred for 10 min. Thereafter the reaction was added to **7.12** (3.86 mmol, 2.84 g, 1.36 mmol/g) pre-swollen in DCM and the reaction was stirred for 15 min. Then **7.3** (1.9 3mmol, 0.57 g) was added and the reaction was microwave irradiated for 30 min at 60 °C. PS-DCC/DCU was removed by filtration. The solvent was removed *in vacuo* and the residue was dissolved in DCM (100 mL) and the organic phase was washed with $KHSO_4$ (30 mL, 1 M), 10% $NaHCO_3$ (30 mL), and brine (30 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo*, which gave a brown oily solid, which was added to MeOH (20 mL) and Cs_2CO_3 (aq, 20 mL, 2 M) was added drop wise and the reaction was stirred for 1.5 h. The MeOH was evaporated *in vacuo* and the remaining solution was diluted with H_2O (100 mL). The solution was washed with EtOAc (50 mL), acidified with $KHSO_4$ (1 M), and extracted with DCM (3 x 100

mL). The combined organic phase was washed with brine (50 mL) and the solvent was evaporated *in vacuo*. The residue was precipitated from EtOAc/hexane. The solid was filtered and dried in a vacuum oven overnight. The precipitate was recrystallised from MeOH/H_2O to give **7.13** 1.07 g as a white solid (70% over two steps).

*R*f (MeOH/EtOAc 1:1) = 0.59

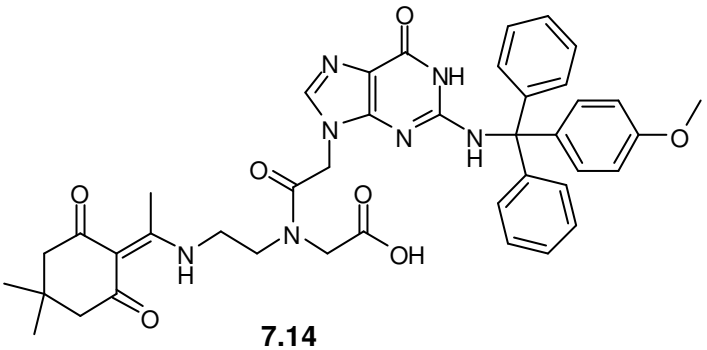
^{1}H NMR (250 MHz, $CDCl_3$, two rotamers): δ ppm 7.91 (s, 1H, CH-pur), 7.90 (s, 1H, CH-pur), 7.25-7.12 (m, 12H, CH=Mmt), 6.71 (d, *J* = 9.15 Hz, 2H, C*H*$COCH_3$-Mmt), 5.28 and 5.09 (s, 2H, CH_2COO), 4.18 and 4.03 (s, 2H, CH_2CO), 3.70 (s, 3H, OCH_3-Mmt), 3.60 (m, 4H, CH_2N), 2.47 (s, 3H, CH_3C), 2.27 (s, 4H, CH_2-Dde), 0.93 (s, 6H, $(CH_3)_2$-Dde).

RP-HPLC (method 1): 4.99 min (ELSD, 100% purity).

LRMS (ES^+): m/z (%) 273 $[Mmt]^+$ (100), 730.5 $[M+H]^+$ (16).

Mp: 170-180 °C (decomposed).

7.4.10 Synthesis of *N*-{2-[N^2-(4-methoxytrityl)-guanin-9-yl]-acetyl}-*N*-{2-[1-(4,4-dimethyl-2,6-dioxocyclohexyliden)-ethylamino}-ethyl}-glycine (Dde-Gua(Mmt)-OH)[14]



7.14

7.7 (4.15 mmol, 2.00 g) was added to dry DMF (15 mL) at 0 °C under N_2 followed by **7.3** (4.15 mmol, 1.23 g), PyBroP (4.15 mmol, 1.94 g, Merck Biosciences), and DIPEA (8.31 mmol, 1.45 mL) and the reaction was microwave irradiated for 20 min at 60 °C. The solvent was removed *in vacuo* and the residue was dissolved in DCM (150 mL) and the organic phase was washed with $KHSO_4$ (50 mL, 1 M), 10% $NaHCO_3$ (50 mL), and brine (50 mL). The organic phase was dried over $MgSO_4$ and

concentrated *in vacuo*, which gave a crude brown solid, which was added to MeOH (16.5 mL) followed by drop wise addition of Cs_2CO_3 (aq, 16.5 mL, 2 M) in the reaction was stirred for 1.5 h. MeOH was evaporated *in vacuo* and the remaining solution was diluted with water (100 mL). The solution was washed with DCM (40 mL x 3), acidified with $KHSO_4$ (1 M) to pH=2. The precipitate was collected by filtration, washed with H_2O (40 mL x 3). The precipitate was reprecipitated from MeOH/H_2O and the precipitate was filtered and dried in a vacuum oven overnight to yield **7.14** as a white solid 2.66 g (86% over two steps).

*R*f (MeOH/EtOAc 1:1) = 0.63.

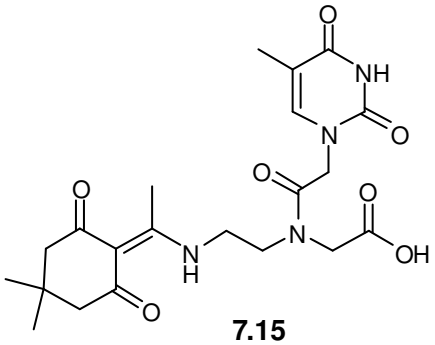
^{1}H NMR (250 MHz, DMSO, two rotamers): δ ppm 10.33 (s, 1H, NH-pur), 7.38 and 7.35 (s, 1H, NH-Mmt), 7.26 and 7.24 (s, 1H, CH-pur), 7.09-6.91 (m, 12H, CH-Mmt), 6.63 (d, *J* = 9.08 Hz, 2H, C*H*COCH$_3$-Mmt), 4.25 and 4.02 (s, 2H, CH_2COO), 3.78 and 3.76 (s, 2H, CH_2CO), 3.51 (pseudo-dd, *J* = 6.33, 2.82 Hz, 4H, CH_2N), 3.20 (s, 3H, OCH$_3$-Mmt), 2.34 (s, 3H, CH_3C), 2.11 and 2.07 (s, 4H, CH_2-Dde), 0.75 (app d, *J* = 1.84 Hz, 6H, $(CH_3)_2$-Dde).

RP-HPLC (method 1): 4.43 min (ELSD, 96.3% purity).

LRMS (ES$^+$): m/z (%) 744.2 $[M+H]^+$ (100).

Mp: 171-179 ^{o}C (decomposed).

7.4.11 Synthesis of *N*-[2-(thymin-1-yl]-acetyl]-*N*-{2-[1-(4,4-dimethyl-2,6-dioxocyclohexyliden)-ethylamino}-ethyl}-glycine (Dde-Thy-OH)[14]



7.15

7.8 (5.43 mmol, 1.00 g) was added to dry DMF (30 mL) under N_2 followed by HOBt (4.89 mmol, 0.75 g) and the reaction was stirred for 10 min. Thereafter the reaction was added to **7.12** (9.78 mmol, 7.99 g, 1.34 mmol/g) pre-swollen in DCM and stirred

for 15 min. Then **7.3** (4.89 mmol, 1.45 g) was added and the reaction was microwave irradiated for 30 min at 60 ^{o}C. DCC/DCU-resin was removed by filtration. The solvent was removed *in vacuo* and the residue was dissolved in DCM (150 mL) and the organic phase was washed with $KHSO_4$ (50 mL, 1 M), 10% $NaHCO_3$ (50 mL), and brine (50 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo*, which gave a brown solid, which was added to MeOH (20 mL) and Cs_2CO_3 (aq, 20 mL, 2 M) was added drop wise and the reaction was stirred for 1.5 h. The MeOH was evaporated *in vacuo* and the remaining solution was acidified with HCl (2 M). The solvent was evaporated *in vacuo*, and the residue was treated with hot 2-propanol and the hot suspension was filtered and the filtrate was collected and the solvent was removed *in vacuo*. The residue was sonicated with H_2O (10 mL), filtered, and dried in a vacuum oven overnight to give **7.15** as a white solid 1.58 g (65% over two steps).

*R*f (MeOH/EtOAc 1:1) = 0.58.

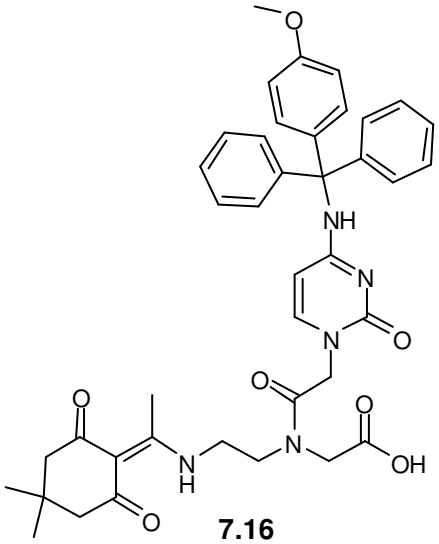
^{1}H NMR (250 MHz, DMSO, two rotamers): δ ppm 13.19 and 13.14 (s, 1H, NH-Dde), 11.31 and 11.29 (s, 1H, COOH), 7.36 and 7.33 (d, J = 1.17 Hz, 1H, CH), 4.72 and 4.54 (two d, J = 21.81 Hz, J = 8.30 Hz, 2H, CH_2COO), 4.30 and 4.03 (s, 2H, CH_2CO), 3.71 (m, 2H, CH_2N), 3.59-3.48 (m, 2H, CH_2N), 2.54 and 2.49 (s, 3H, CH_3C-Dde), 2.30 and 2.28 (s, 4H, CH_2-Dde), 1.76 (d, J = 1.00 Hz, 3H, CH_3), 0.96 and 0.95 (s, 6H, $(CH_3)_2$-Dde).

RP-HPLC (method 1): 3.30 min (ELSD, 100% purity).

LRMS (ES^{+}): m/z (%) 471.4 $[M+Na]^{+}$ (100).

Mp: 241-246 ^{o}C.

7.4.12 Synthesis of *N*-{2-[N^2-(4-methoxytrityl)-cytosin-9-yl]-acetyl}-*N*-{2-[1-(4,4-dimethyl-2,6-dioxocyclohexyliden)-ethylamino}-ethyl}-glycine (Dde-Cyt(Mmt)-OH)[14]



7.16

7.10 (4.53 mmol, 2.00 g) was added to dry DMF (30 mL) under N_2 followed by HOBt (4.08 mmol, 0.62 g) and the reaction was stirred for 10 min. Subsequently, the reaction was added to **7.12** (8.16 mmol, 6.66 g, 1.36 mmol/g) and stirred for 15 min. Then **7.3** (4.08mmol, 1.21g) was added and the reaction was microwave irradiated for 30 min at 60 ^{o}C. DCC/DCU-resin was removed by filtration. The solvent was removed *in vacuo* and the residue was dissolved in DCM (150 mL) and the organic phase was washed with $KHSO_4$ (50 mL, 1 M), 10% $NaHCO_3$ (50 mL), and brine (50 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo* to give a yellow oily solid, which was added to MeOH (20 mL) and Cs_2CO_3 (aq, 20 mL, 2 M) was added drop wise and the reaction was stirred for 1.5 h. The MeOH was evaporated *in vacuo* and the remaining solution was diluted with H_2O (100 mL). The solution was washed with EtOAc (50 mL), acidified with $KHSO_4$ (1 M), and extracted with DCM (100 mL x 3). The combined organic phase was washed with brine (50 mL) and the solvent was evaporated *in vacuo*. The residue was precipitated from EtOAc/hexane. The solid was filtered and dried a in a vacuum oven overnight. The precipitate was recrystalised from MeOH/H_2O to yield **7.16** as a white solid 2.80g (86%).

*R*f (MeOH/EtOAc 1:1) = 0.62.

^{1}H NMR (250 MHz, DMSO, two rotamers): δ ppm 9.92 (s, 1H, NH), 7.42 (d, *J* = 7.99 Hz, 1H, CH-Cyt) 7.35-6.92 (m, 12H, CH-Mmt), 6.73 (d, *J* = 7.09 Hz, 2H, CH-Mmt), 5.07 (d, *J* = 7.91 Hz, 1H, CH-Cyt), 4.67 and 4.36 (s, 2H, CH_2COO), 4.13 and 3.98 (s, 2H, CH_2CO), 3.67 (m, 7H, OCH_3-Mmt and 2 x CH_2N), 2.40 (s, 3H, CH_3C), 2.19 (s, 4H, CH_2-Dde), 0.88 (s, 6H, $(CH_3)_2$-Dde).

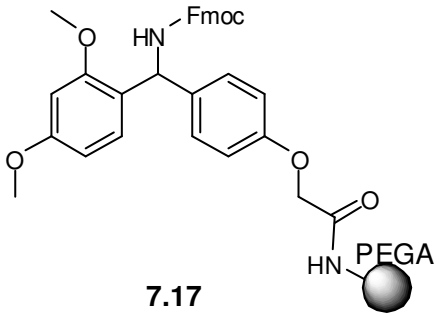
RP-HPLC (method 1): 4.55 min (ELSD, 100% purity).

LRMS (ES^+): m/z (%) 273 $[Mmt]^+$ (100), 706.2 $[M+H]^+$ (10).

Mp: 172-179 ^oC (decomposed).

7.4.13 Functionalised resins for PNA-peptide conjugate synthesis[14]

7.4.13.1 Synthesis of Fmoc-Rink-resin and down loading



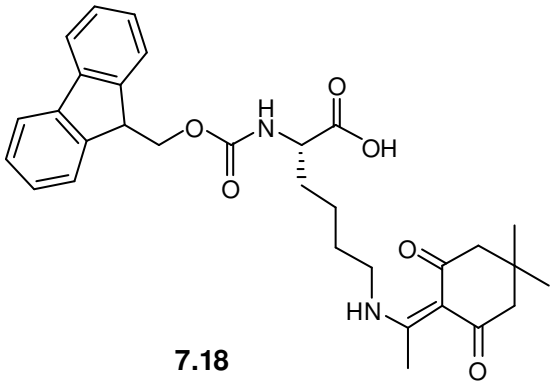
7.17

Fmoc-Rink linker (4.32 mmol, 2.34 g, GLS Biochem) was added to DMF (40 mL) at room temperature followed by HOBt (4.32 mmol, 0.660 g) and stirred for 10 min. Hereafter DIC (5.18 mmol, 0.80 mL, Sigma-Aldrich) was added and the reaction was stirred for another 10 min. The resulting solution was added to amino-PEGA resin (0.237 mmol, 0.056 mmol/g, 4.00 g, Polymer Laboratories) pre-swollen in DCM (40 mL x 3) and the reaction was microwave irradiated for 30 min at 60 ^oC. The reaction was analysed by qualitative ninhydrin test (negative). The resin was filtered and washed thoroughly with DMF (x 3), DCM (x 3), and MeOH (x 3) to give **7.17** 4.87 g (100%).

7.4.13.2 Synthesis of low loading PS-resin

Fmoc-Rink linker (0.211 mmol, 0.114 g) was added to DMF (21 mL) at room temperature followed by PyBOP (0.192 mmol, 0.100 g, GLS Biochem) and NEM (0.384 mmol, 54 μL) and the resulting solution was added to PS-resin (3.20 mmol, 3.26 g, 1.0 mmol/g, 1% DVB, 100-200 mesh, GLS Biochem) pre-swollen DCM (40 mL x 3) and the reaction was microwave irradiated at 60 ^{o}C for 10 min. The resin was filtered and washed thoroughly with DMF (x 3), DCM (x 3), and MeOH (x 3). The new loading was determined by quantitative Fmoc test. Dry resin (5-10 mg) was accurately weighed off and shaken in 20% piperidine in DMF for 10 min. The resin was filtered off and the filtrate diluted to with 20% piperidine in DMF (10 mL) and the absorbance was measured at 305 nm (ε=7800 L/mol·cm) to give a loading of 0.063 mmol/g.

7.4.13.3 Synthesis of Fmoc-Lys(Dde)-OH[14]



7.18

Fmoc-Lys-OH (7.41 mmol, 3.00 g, Iris Biotech GmBH) was added to ethanol (50 mL) at room temperature followed by Dde-OH (2.73 mmol, 2.70 g) and TFA (0.74 mmol, 0.06 mL) and the reaction was refluxed for 60 h. The solvent was removed by evaporation *in vacuo* and the orange residue was dissolved in EtOAc (150 mL). The organic phase was washed with $KHSO_4$ (1 M, 100 mL x 2), dried over Mg_2SO_4 and concentrated *in vacuo*. The crude oil was tituated with hexane to remove any unreacted Dde-OH. The product was crystallised from EtOAc/hexane to yield **7.18** as white crystals 2.62 g (66%).

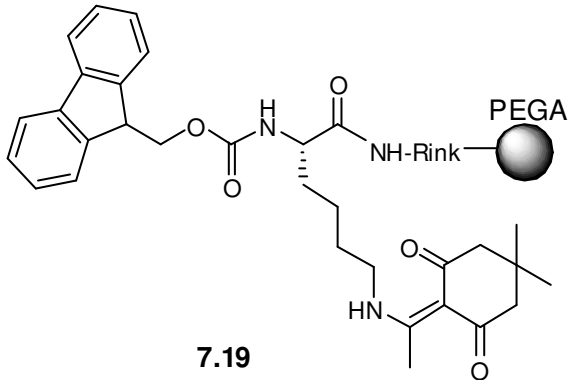
*R*f (EtOAc) = 0.45.

^{1}H NMR (250 MHz, $CDCl_3$): δ ppm 13.25 (s, 1H, COOH), 7.75-7.06 (m, 8H, CH-Fmoc), 5.70 (pseudo-d, J = 8.60 Hz, 1H, CH-Fmoc), 4.66-3.79 (m, 4H, NH, CH, CH_2-Fmoc), 3.32 (s, 2H, CH_2-NH), 2.48 and 2.47 (s, 3H, CH_3-Dde), 2.28 (s, 4H, CH_2-Dde), 2.01-1.33 (m, 6H, CH_2-Lys), 0.94 and 0.93 (s, 6H, $(CH_3)_2$-Dde).

RP-HPLC (method 1): 4.47 min (ELSD, 100% purity).

LRMS (ES^+): m/z (%) 554.5 $[M+Na]^+$ (100).

7.4.13.4 Synthesis of Fmoc-Lys(Dde)-Rink-PEGA-resin[14]



7.19

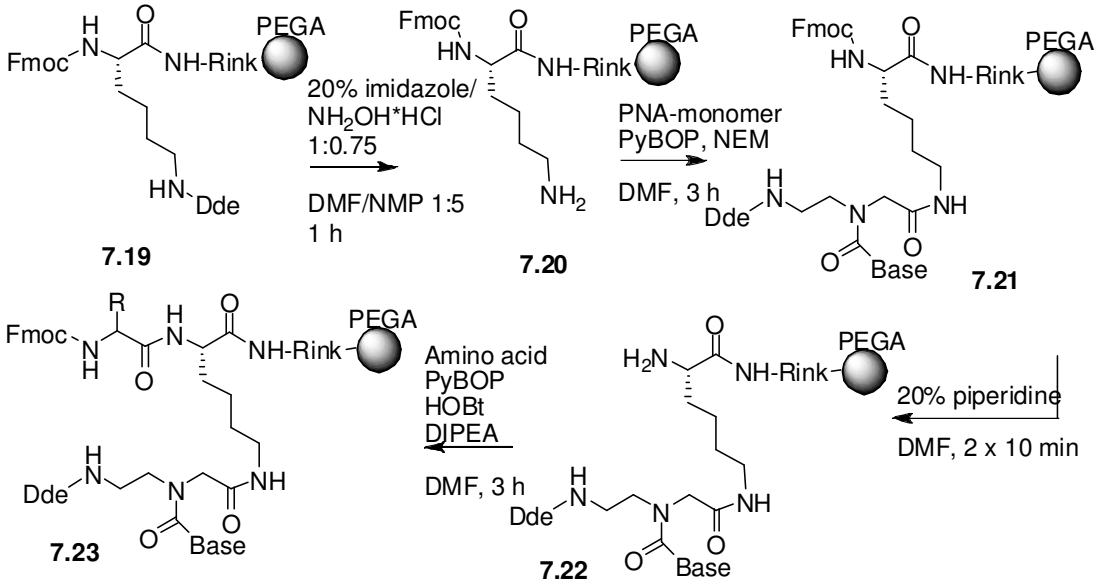
PEGA-Rink-amine resin (0.103 mmol 2.06 g, 0.050 mmol/g) was prepared by pre-swelling in DCM and Fmoc deprotection was carried out by treatment with 20% piperidine in DMF (30 mL x 2) for 10 min and washing with DMF (x 3), DCM (x 3), and MeOH (x 3). Qualitative ninhydrin test was positive.

7.18 (2.16mmol, 1.15 g) was added to DMF (21 mL) at room temperature followed by HOBt (2.16mmol, 0.292 g) and stirred for 10 min. Hereafter, DIC (2.59 mmol, 0.811 mL, Fluka) was added, the reaction was stirred for another 10 min and added to the resin and the reaction was microwave irradiated for 30 min at 60 ^oC. The reaction was monitored by a qualitative ninhydrin test (negative). The resin was filtered and washed thoroughly with DMF (x 3), DCM (x 3), and MeOH (x 3) to give **7.19** 2.45 g (0.042 mmol/g, 100%). An analytical sample was cleaved with TFA/TIS 95:5 to give Fmoc-Lys(Dde)-NH_2.

RP-HPLC (method 2): 7.76 min (ELSD, 100% purity).

LRMS (ES^+): m/z (%) 532.3 $[M+H]^+$ (100).

7.4.14 General reaction conditions for manual PNA-peptide conjugate synthesis[14]



7.4.14.1 PNA-monomer coupling

PNA-monomer (5.5 equiv.) *or* 5(6)-carboxyfluorescein (5.5 equiv., Fluka) and PyBOP (5.0 equiv.) were added to DMF (final concentrations of 0.11 M and 0.1 M respectively) followed by NEM (10 equiv., final concentration of 0.2 M) and the solution was mixed for 20 sec. The reaction mixture was added to resin (1 equiv., pre-swollen in DMF) and the reaction was shaken for 3 h (the filtrate was collected to recover the monomer). The resin was washed with DMF (x 3), DCM (x 3), and MeOH (x 3).

7.4.14.2 Recovery of PNA monomers after coupling

The unreacted monomers were recovered from the reaction mixture after use. The reaction mixture was collected after use and the resin rinsed with DMF and the filtrates were combined. Ade-, Cyt-, and Gua-monomers: DMF was removed by evaporation *in vacuo* and the monomer is precipitated by MeOH/H_2O twice yielding the respective monomer as a white solid. Thy-monomer: DMF was removed by evaporation *in vacuo* and the crude was precipitated in DCM/petroleum ether followed by EtOAc/hexane and the yellow precipitate was washed with minimum amounts of H_2O yielding the Thy-monomer as a white solid.

7.4.14.3 Amino acid and peptoid couplings

Fmoc-amino acid (5.5 equiv., GLS Biochem) *or* butyric acid (5.5 equiv., Fluka) *or* Llp (5.5 equiv., courtesy of Géraldine Escher, University of Edinburgh, UK) was added to DMF to a final concentration of 0.11 M followed PyBOP (5.0 equiv.) and HOBt (5.0 equiv.) to a final concentration of 0.1 M, hereafter DIPEA was added to a final concentration of 0.2 M (10 equiv.) and the solution was mixed for 20 sec. The reaction mixture was added to resin (1 equiv.), pre-swollen in DMF and the reaction was shaken for 3 h. The resin was washed with DMF (x 3), DCM (x 3), and MeOH (x 3).

7.4.14.4 Dde-deprotection

A stock solution of 25(wt)% imidazole/$NH_2OH{\cdot}HCl$ 1:0.75 in NMP was made up from $NH_2OH{\cdot}HCl$ (9.00mmol, 6.25g, Fluka) and imidazole (6.75mmol, 4.59g, Sigma-Aldrich) in 25 mL NMP (this was sonicated till complete dissolution). Immediately before use a solution of 20(wt%) imidazole/$NH_2OH{\cdot}HCl$ 1:0.75 in NMP/DMF 5:1 was made up from 1 mL DMF and 5 mL of the stock solution and added to the resin (pre-swollen in DMF) and the reaction was shaken for 1 h. The resin was filtered and washed with DMF (x 3), DCM (x 3), and MeOH (x 3).

Dde-deprotection was also carried out with NH_2OH/H_2O 1:1 (Sigma-Aldrich) instead of the $NH_2OH{\cdot}HCl$ salt under the same reaction conditions as described above.

7.4.14.5 Fmoc-deprotection

Fmoc deprotection was carried out by treatment of the resin with 20% piperidine (Sigma-Aldrich) in DMF for 2 x 10 min followed by washing with DMF (x 3), DCM (x 3), and MeOH (x 3).

7.4.14.6 Cleavage from resin

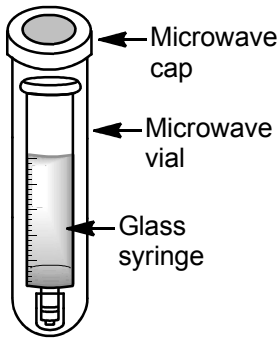
After synthesis the resins were washed twice with 20% piperidine in DMF (2 mL) for 10 min to cleave unwanted fluorescein dimmers. The PNA-peptide conjugates and peptides were cleaved from the resin by treatment with TIS (Sigma-Aldrich)/TFA 5:95 for 2 x 2 h and the products were precipitated by drop wise addition to cold ether (x 2) and collected by centrifugation.

Analytical cleavages were carried out on small aliquots of resin as described above for 1 h. Samples were analysed by Maldi-TOF or MS and RP-HPLC.

7.4.15 General reaction conditions for microwave PNA-peptide conjugate synthesis optimisation

7.4.15.1 Microwave aided PNA monomer coupling

PNA-monomer (5.5 equiv., 0.22 M) in NMP or DMF was mixed with PyBOP (5.5 equiv., 0.20 M) in NMP followed by the addition of NEM (10 equiv.) to a final concentration of 0.2 M and the reaction was stirred for 20 s and added to the resin (1 equiv.), pre-swollen in DMF in a 5 mL glass syringe with a glass filter. The entire glass syringe containing the reaction was transferred to a microwave vial, NMP or DMF was added around the syringe, and the reaction was microwave irradiated for 20 min at 60 ^{o}C. The resin was washed with DMF (x 3), DCM (x 3), and MeOH (x 3) and the filtrate was collected to recover the monomer.



7.4.15.2 Capping

After each monomer coupling the unreacted free amino-sites were capped to prevent further growth of undesirable sequences. 70% acetic anhydride and 8 (wt)% HOBt in DMF was added to the resin pre-swollen in DMF and the reaction was stirred at room temperature for 20 min. The resin was washed with DMF (x 3), DCM (x 3), and MeOH (x 3).

7.4.15.3 Microwave aided Dde-deprotection

The Dde-deprotection reaction mixtures described in section 0 were added to the resin pre-swollen in DMF in a glass syringe and the reaction was pre-stirred for 20 s. The glass syringe was transferred to a microwave vial and NMP was added around the syringe and microwave irradiated for 10 min at 60 ^{o}C. The resin was filtered and washed with DMF (x 3), DCM (x 3), and MeOH (x 3).

7.4.15.4 Microwave aided amino acid couplings and Fmoc-deprotection

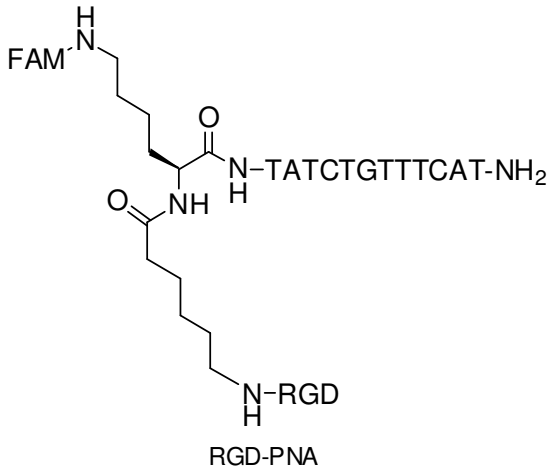
Automated peptide synthesis was performed by solid-phase Fmoc chemistry using an automated microwave peptide synthesiser for multiple peptide synthesis (Liberty, CEM). Amino acid (5.5 equiv., 0.22 M) or Llp (5.0 equiv., 0.2 M, courtesy of Géraldine Escher, University of Edinburgh, UK) in DMF were coupled using HTBU (5.0 equiv., 0.2 M, GLS Biochem)/HOBt (5.0 equiv., 0.2 M) in DMF and 20 % DIPEA in NMP (final concentration of 0.2 M) under microwave irradiation for 20 min at 60 ^{o}C. The resin was filtered and washed with DMF (x 3) and DCM (x 3). Fmoc-deprotection was carried out in the CEM synthesiser as described in section 0.

7.4.16 Synthesis of PNA-peptide conjugates and peptides

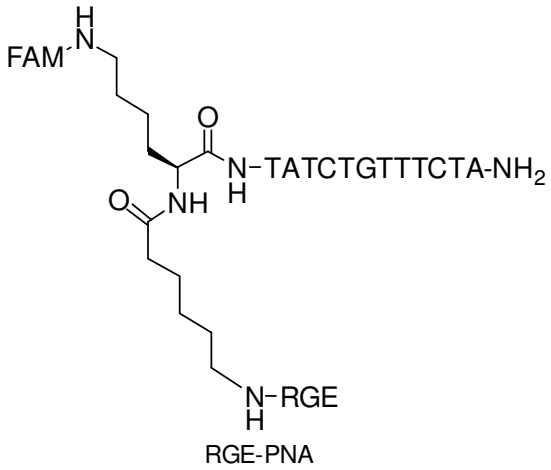
Unless otherwise stated, the PNA-peptide conjugates and peptides were purified by precipitation in cold diethyl ether followed by RP-HPLC purification: column: XTerra Prep RP18, 19 cm x 150 nm x 5µm; eluting with 100% H_2O with 0.1% TFA for 5 min 100% H_2O with 0.1% TFA to 40% H_2O with 0.1% TFA over 10 min, to 100% CH_3CN with 0.1% TFA over 4 min and 100% CH_3CN with 0.1% TFA for 2

min. Fractions were analysed by ESI-MS or MALDI-TOF and the collected fractions were freeze dried overnight.

7.4.16.1 Synthesis of RGD-PNA and RGE-PNA conjugates[14]



7.24



7.25

RGD-PNA and RGE-PNA conjugates were synthesised using microwave aided PNA and amino acid couplings as described above in an automated CEM peptide synthesiser. The conjugates were purified by RP-HPLC (method 2) and analysed by MALDI-TOF as described above. The following 4-letter codes were applied:

7.24: Arg-Gly-Asp-Ahx-Lys(FAM)-TATCTGTTTCAT-NH_2

RP-HPLC: 12.42 min. (95.2% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 4193.04 $[M]^+$ (100).

7.25: Arg-Gly-Glu-Ahx-Lys(FAM)-TATCTGTTTCTA-NH_2

RP-HPLC: 12.37 min (93.7% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 4206.33 $[M+H]^+$ (100).

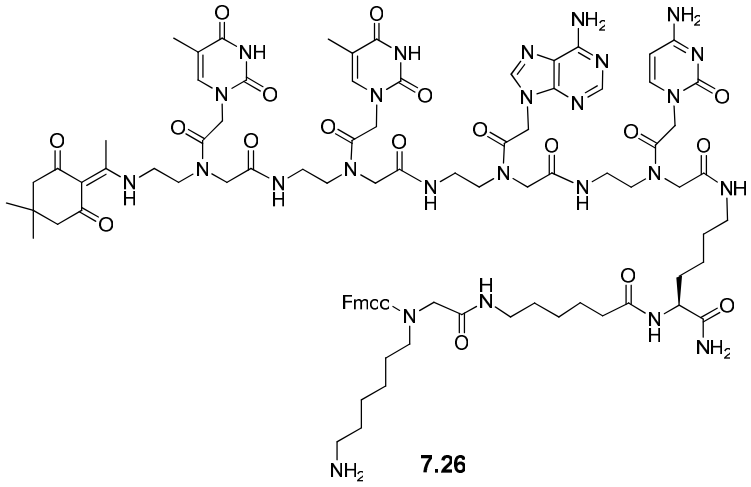
7.5 Experimental to Chapter 3

7.5.1 Synthesis of the 1296 member peptide-PNA library

The 1296 member peptide-PNA library (Library-1) was synthesised using orthogonal Dde/Fmoc-peptide chemistry on Fmoc-Lys(Dde)-Rink-PEGA resin (600 mg, 0.042 mmol/g) as described in section 0 using the following code:

Amino acid	PNA Tag (N-C)
Llp	TTAC
Pro	AAAC
Leu	TACA
Glu	ATCT
Tyr	ACAA
Lys	TCAT

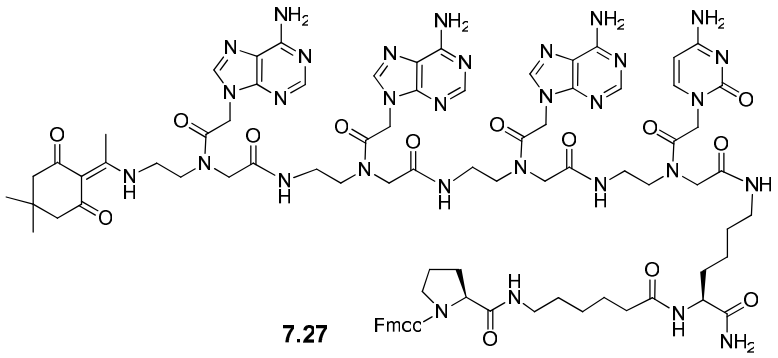
The split-and-mix synthesis was carried out by splitting the resin into 6 and coupling of four PNA monomers (Dde) and the corresponding amino acid (Fmoc) and mixing giving 1296 (6^4) different PNA encoded peptides. Before the first mix, the 6 amino acid conjugates were analysed by MALDI-TOF:



7.26

Fmoc-Llp-Ahx-Lys(Dde-TTAC)-NH_2: RP-HPLC: 2.89 min.

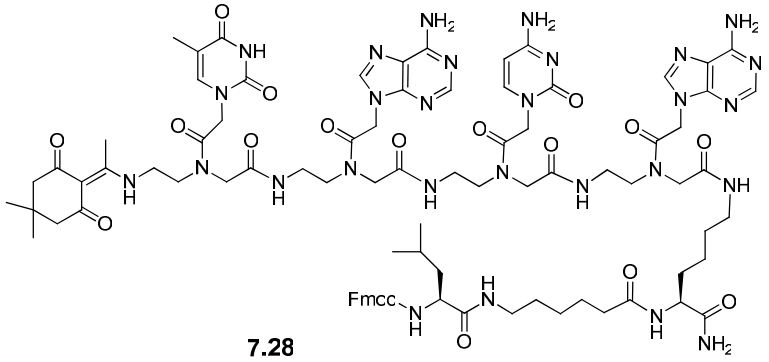
MALDI-TOF^+/MS: m/z (%) 1859.30 $[M]^+$ (100).



7.27

Fmoc-Pro-Ahx-Lys(Dde-AAAC)-NH_2: RP-HPLC: 2.82 min.

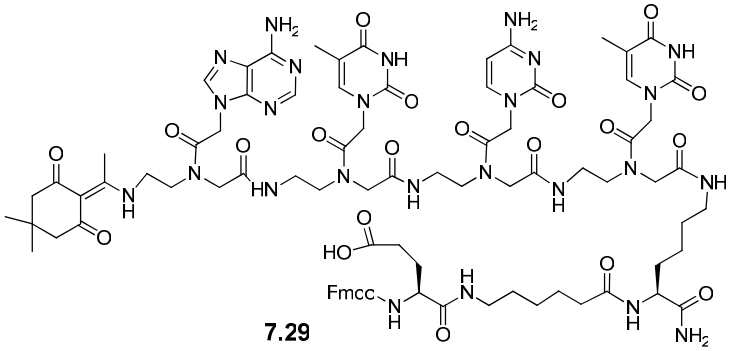
MALDI-TOF$^+$/MS: m/z (%) 1818.21 $[M]^+$ (100).



7.28

Fmoc-Leu-Ahx-Lys(Dde-TACA)-NH_2: RP-HPLC: 2.84 min.

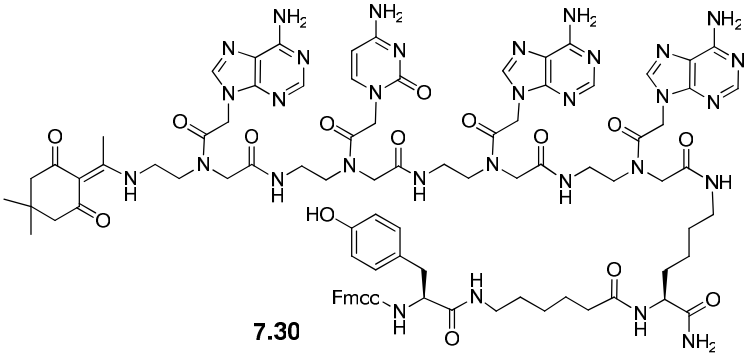
MALDI-TOF$^+$/MS: m/z (%) 1824.77 $[M]^+$ (100).



7.29

Fmoc-Glu-Ahx-Lys(Dde-ATCT)-NH_2: RP-HPLC: 2.83 min.

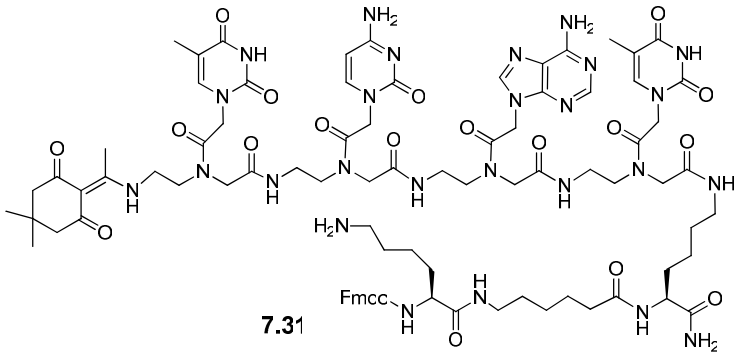
MALDI-TOF$^+$/MS: m/z (%) 1831.27 $[M]^+$ (100).



7.30

Fmoc-Tyr-Ahx-Lys(Dde-ACAA)-NH_2: RP-HPLC: 2.91 min.

MALDI-TOF$^+$/MS: m/z (%) 1883.91 $[M]^+$ (100).



7.31

Fmoc-Lys-Ahx-Lys(Dde-TCAT)-NH_2: RP-HPLC: 2.84 min.

MALDI-TOF$^+$/MS: m/z (%) 1829.64 $[M]^+$ (100).

Following split and mix synthesis (amino acid coupling and PNA synthesis) the peptide *N*-termini were capped with acetic anhydride/pyridine while the PNA amino-terminus was coupled with Fmoc-6-aminohexanoic acid followed by Fmoc cleavage and capping with 5(6)-carboxyfluorescein as described in section 0. After synthesis the peptide-PNA conjugates were cleaved from the solid support and purified as described in section 0 to yield Library-1 81 mg (51%, over 50 steps, appendix I).

7.5.2 Extraction of peptide-PNA conjugates from cells

HeLa, HEK293T, ARN8, SH-SY5Y, and E14 cells were grown to 70-90% confluence in 6-well plates and incubated with Library-1 (800 μL, 100 μM) in appropriate growth media at 37 ^oC for 2 h in 8 wells. Primary neutrophils and primary monocytes were grown to 90% confluence in 24-well plates and incubated

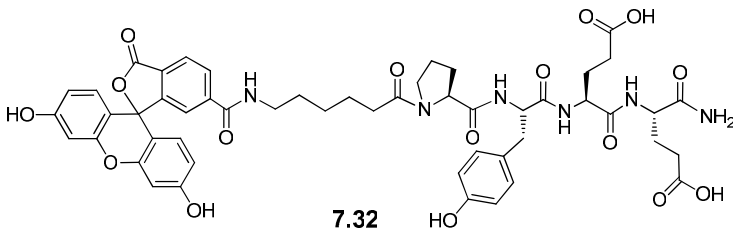
with Library-1 (300 μL, 100 μM) in appropriate growth media at 37 ^{o}C for 30 min in 12 and 24 wells respectively. Suspension cells, K562 (2x10^6 cells/mL) and primary lymphocytes (5x10^6 cells/mL) were collected by centrifugation (1100 rpm for 5 min) and incubated with Library-1 (100 μM) in RPMI CM (800 μL) and PBS (3 mL) respectively at 37 ^{o}C for 4 h. After incubation the cells were washed with PBS x 3, treated with 1 x trypsin/EDTA for 15 min, collected by centrifugation (1100 rpm for 5 min), washed with PBS, and lysed with a minimum amount of H_2O. peptide-PNA conjugates were purified by filter-centrifugation separating between 3,000 Da and 10,000 Da molecular weight filters (Nunc TM) and concentrated by speed-vac. The amount of cell penetrating PNA was estimated using a NanoVue spectrometer. Cell penetrating PNA encoded peptides extracted were HeLa (8.0 μg and 6.3 μg), HEK293T (6.9 μg), SH-SY5Y (1.9 μg), ARN8 (6.6 μg), K562 (1.5 μg), E14 (6.4 μg), primary monocytes (7.0 μg and 1.6 μg), primary neutrophils (6.2 μg, 1.8 μg), and primary lymphocytes (1.4 μg and 2.2 μg).

7.5.3 Microarray hybridisation of Library-1 extracted from cells

DNA arrays (OGT) complementary to the Library-1 were hybridised with the PNA-conjugates purified from cells in 1 x GenHyb buffer (110 μL) buffer overnight from 50 ^{o}C to 32 ^{o}C according to microarray wash protocol 2 and imaged with a Tecan LS Reloaded microarray scanner. Image analysis was carried out as described in section 7.2 (appendix II)

7.5.4 Synthesis of selective "hit"-tetramers

Peptides (X-AA_4-AA_3-AA_2-AA_1-NH_2) were synthesised in an automated CEM-peptide synthesiser on Fmoc-Rink-PS-resin using microwave aided amino acid and Llp couplings as described in section 0. After amino acid/Llp couplings the *N*-termini were coupled with 6-aminohexanoic acid followed by capping 5(6)-carboxyfluorescein as described before. After synthesis, the peptide were purified by RP-HPLC (method 1) and analysed by ES-MS (positive mode).

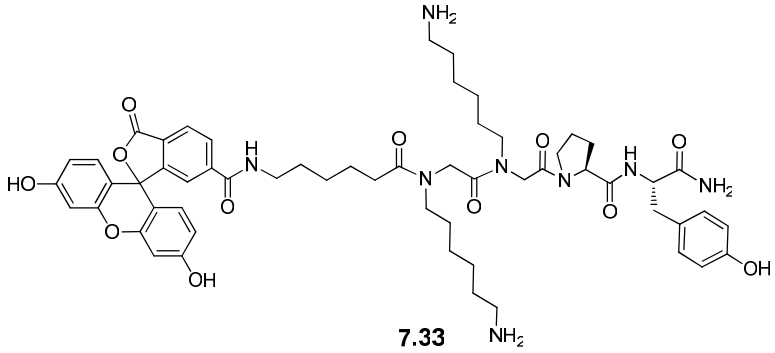


7.32

FAM-Ahx-Pro-Tyr-Glu-Glu-NH_2: 23.8 mg (39% yield).

RP-HPLC: 2.95 min. (84.3% crude purity).

LRMS (ES^+): m/z (%) 747.3 $[M\text{-}FAM]^+$ (100), 1007.3 $[M+H]^+$ (13).

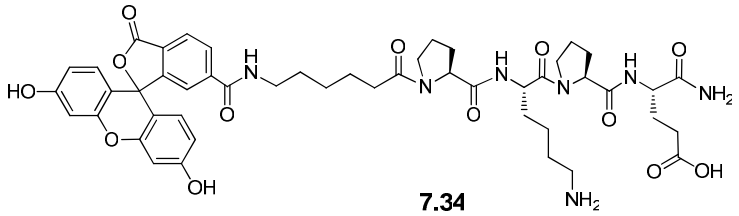


7.33

FAM-Ahx-Llp-Llp-Pro-Tyr-NH_2: 3.97 mg (42% yield)

RP-HPLC: 2.90 min. (82.6% crude purity).

LRMS (ES^+): m/z (%)703.4 $[M\text{-}FAM]^+$ (100), 1061.4 $[M+H]^+$ (7).

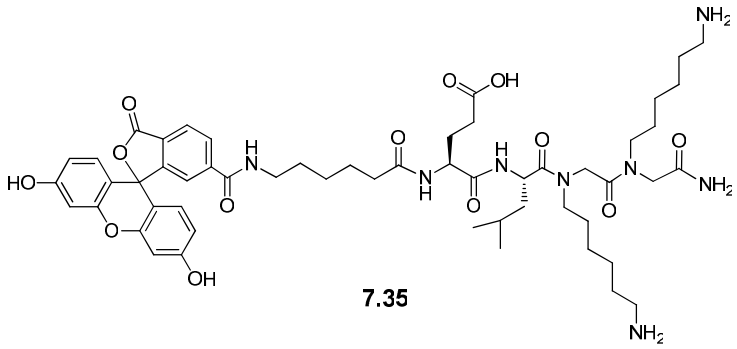


7.34

FAM-Ahx-Pro-Lys-Pro-Glu-NH_2: 25.4 mg (44% yield)

RP-HPLC: 3.35 min. (80.0% crude purity).

LRMS (ES^+): m/z (%) 470.6 $[M+2H]^{2+}$ (100), 940.4 $[M+H]^+$ 940.3 (45).

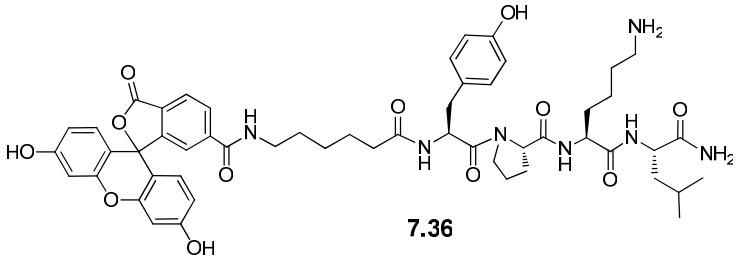


7.35

FAM-Ahx-Glu-Leu-Llp-Llp-NH_2: 12.00 mg (63% yield)

RP-HPLC: 2.90 min. (78.4% crude purity).

LRMS (ES^+): m/z (%) 521.7 $[M+2H]^{2+}$ (100), 1042.4.4 $[M]^+$ (24).

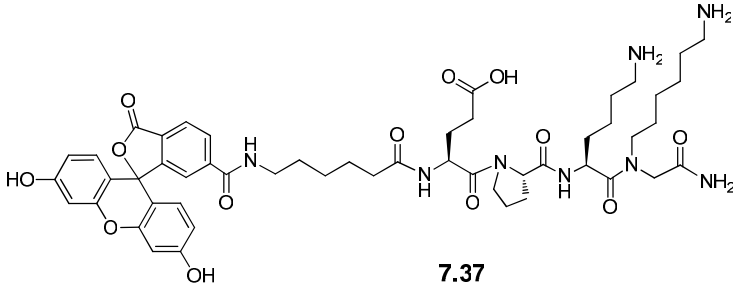


7.36

FAM-Ahx-Tyr-Pro-Lys-Leu-NH_2: 5.92 mg (19% yield)

RP-HPLC: 3.58 min. (91.5% crude purity).

LRMS (ES^+): m/z (%) 682.3 $[M\text{-}FAM]^+$ (100), 1018.3 $[M+H]^+$ (12).

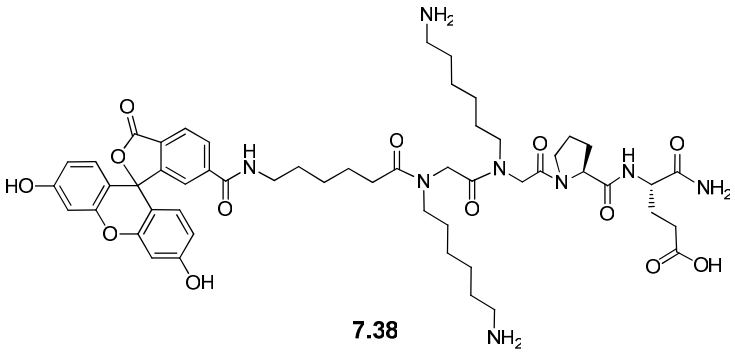


7.37

FAM-Ahx-Glu-Pro-Lys-Llp-NH_2: 10.7 mg (58% yield)

RP-HPLC: 2.99 min. (71.0% crude purity).

LRMS (ES^+): m/z (%) 500.1 $[M+2H]^{2+}$ (100), 999.3 $[M+H]^+$ (27).

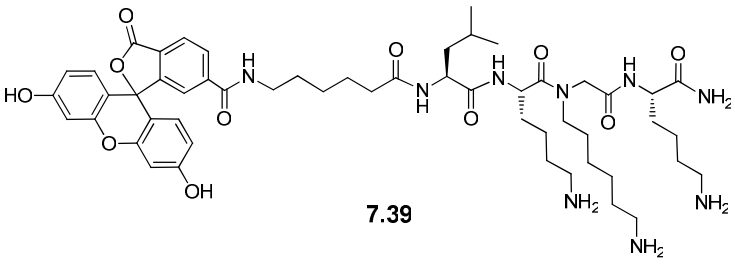


7.38

FAM-Ahx-Llp-Llp-Pro-Glu-NH_2: 10.0 mg (53% yield)

RP-HPLC: 3.00 min. (86.4% crude purity).

LRMS (ES^+): m/z (%) 514.3 $[M+2H]^{2+}$ (100), 1027.4 $[M+H]^+$ (29).

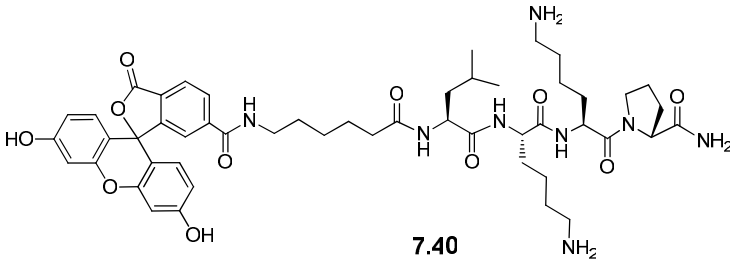


7.39

FAM-Ahx-Leu-Lys-Llp-Lys-NH_2: 11.3 mg (61% yield)

RP-HPLC: 3.26 min. (78.0% crude purity).

LRMS (ES^+): m/z (%) 502.0 $[M+2H]^{2+}$ (100).

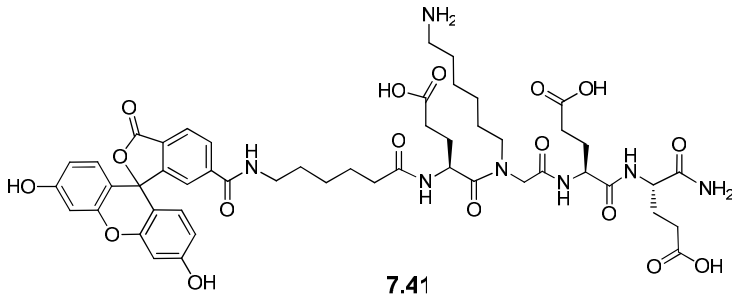


7.40

FAM-Ahx-Leu-Lys-Lys-Pro-NH_2: 41.6 mg (72% yield).

RP-HPLC: 3.18 min. (79.3% crude purity).

LRMS (ES^+): m/z (%) 478.2 $[M+2H]^{2+}$ (100), 955.3 $[M+H]^+$ (23).

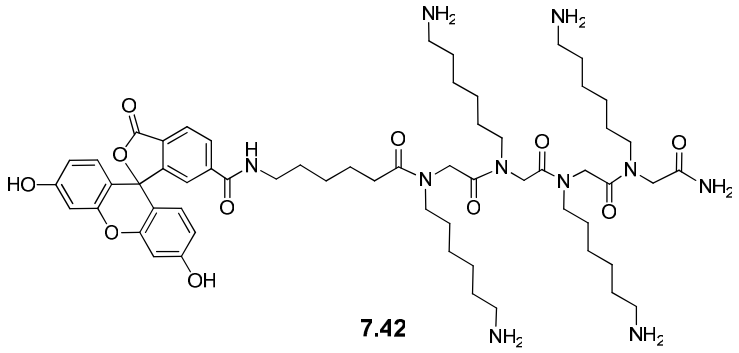


7.41

FAM-Ahx-Glu-Llp-Glu-Glu-NH_2: 7.31 mg (38% yield)

RP-HPLC: 3.11 min. (91.7% crude purity).

LRMS (ES^+): m/z (%) 516.7 $[M+2H]^{2+}$ (100), 1032.2 $[M+H]^+$ (56).



7.42

FAM-Ahx-Llp-Llp-Llp-Llp-NH_2[**]

RP-HPLC: 2.51 min.

LRMS (ES^+): m/z (%) 1113.6 $[M+H]^+$.

7.5.5 Flow cytometry of generic and selective "hit"-tetramers

Adherent cells, HeLa, HEK293T, ARN8, SH-SY5Y, and E14 were grown to 70-90% confluence in 48-well plates and incubated with the "hit"-tetramers (X-AA_4-AA_3-AA_2-AA_1-NH_2; 200 µL, 10 µM) in appropriate growth media at 37 ^oC for 4 h. Suspension cells, K562 (10^6 cells/mL) were collected by centrifugation (1100 rpm for 5 min) and incubated with "hit"-tetramers (X-AA_4-AA_3-AA_2-AA_1-NH_2; 200 µL, 10 µM) in RPMI. After incubation, the cells were washed with PBS x 4, treated with 1 x trypsin/EDTA (20 min) and suspended in tryphan blue (0.04 M, Sigma-Aldrich)

[**] Synthesised by Geraldine Escher

and propidium iodide (1.5 mM, Sigma-Aldrich) in PBS (300 μL). Flow cytometry was run analytically using a FACS Aria cytometer (selective "hit"-tetramers: X-AA_4-AA_3-AA_2-AA_1-NH_2, FITC and PE-Texas Red filters, 10.000 events) or a FACS Calibur cytometer (X-Glu-Llp-Llp-Llp-NH_2, FITC and PE/PI filters, 10.000 events).

Anticoagulated whole blood (100 μL) was incubated with "hit"-tetramers (X-AA_4-AA_3-AA_2-AA_1-NH_2;100 μL, 20 μM) in PBS at 37 ^oC for 4 h. After incubation, the red blood cells were selectively lysed by hypertonic cryohemolysis by adding 1.2 x PBS (2 mL) at 0 ^oC for 45 sek, followed by addition of 0.8 x PBS (2 mL) at 0 ^oC to re-establish isotonic concentrations. The leukocytes were collected by centrifugation (1100 rpm for 5 min), re-suspended in PBS, and analysed by flow cytometry using a FACS Calibur cytometer (FITC filter and 10.000 events). Histograms of cell count vs. FITC-fluorescence were created based on scatter plots gated for single, FITC-only positive, live cells.

In order to compare the uptake of each "hit"-tetramers (X-AA_4-AA_3-AA_2-AA_1-NH_2) across the cell types, the data set for each cell type were normalised in the same manner as the microarray data (i.e. the mean intensity of the entire data set for each cell type was arbitrarily set to 100 and 10 mean intensities in each dataset were multiplied with respective normalisation factor).

7.5.6 Confocal microscopy of selective "hit"-tetramers

Adherent cells, HeLa, HEK293, ARN8, SH-SY5Y, E14, primary neutrophils, and primary monocytes were grown on glass cover slips (primary cells) or glass cover slips coated with 0.01% poly-L-lysine [Sigma-Aldrich, washed with 1 % HCl, 70 % EtOH in H_2O and 3 x H_2O and treated with 10 % poly-lysine in H_2O for 5 min and dried at room temperature overnight] at 37 ^oC *or* 4 ^oC. Half of the wells were incubated with 10 μM "hit"-tetramer (X-AA_4-AA_3-AA_2-AA_1-NH_2) and incubated at 37 ^oC *or* 4 ^oC for 2 h in appropriate growth media. All the wells were incubated with

a lysosomal tracker dye (300 µL, 75 nM Lysotracker Red, Invitrogen) at 37°C *or* 4 °C for 2 h in appropriate growth media. Suspension cells, K562 (10^5 cells/ml) and primary lymphocytes ($6 \cdot 10^5$ cells/ml) were collected by centrifugation (1100 rpm for 5 min) and incubated with 10 µM "hit"-tetramer (X-AA_4-AA_3-AA_2-AA_1-NH_2) in RPMI (K562) and PBS (lymphocytes) at 37 °C *or* at 4 °C for 2 h, washed with PBS x 3 and incubated with LysoTracker Red as described above. Subsequently, the cells were spun onto glass cover slips by centrifugation (1100 rpm for 5 min). Cells were analysed by confocal microscopy using a Deltavision Tollervey microscope with a Photometrics Coolsnap HQ camera, SoftWoRx (Applied Precision) software using FITC (Ex = 490/20 nm, Em = 528/38 nm) and TRITC (Ex = 545/30 nm, Em = 610/75 nm) filters (100 x magnification).

7.5.7 Haemolysis assays

Erythrocytes were isolated from freshly drawn, anticoagulated, human blood and resuspended to 20% in PBS (pH 7.4). In a 96-well microtiter plate, 100 µl of erythrocyte suspension was added to 100 µl of the "hit"-tetramers (final concentration 10 µM) in PBS (prepared by 1:2 serial dilutions) or 100 µl of PBS in the case of negative controls. One-hundred percent haemolysis wells contained 100 µl of red cell suspension with 100µl of 0.2% Triton 100. The plate was incubated for 1 h at 37 °C, and then each well was diluted with 150 µl of PBS. The plate was then centrifuged at 1,200 *g* for 15 min, and 100 µl of the supernatant from each well was transferred to a fresh microtiter plate, and A570 was measured. Percentage of haemolysis was determined as (*A*-*A*0)/(*A*total-*A*0) X 100, where *A* is the absorbance of the test well, *A*0 the absorbance of the negative controls, and *A*total the absorbance of 100% haemolysis wells, all at 570 nm on a Biotek plate reader.

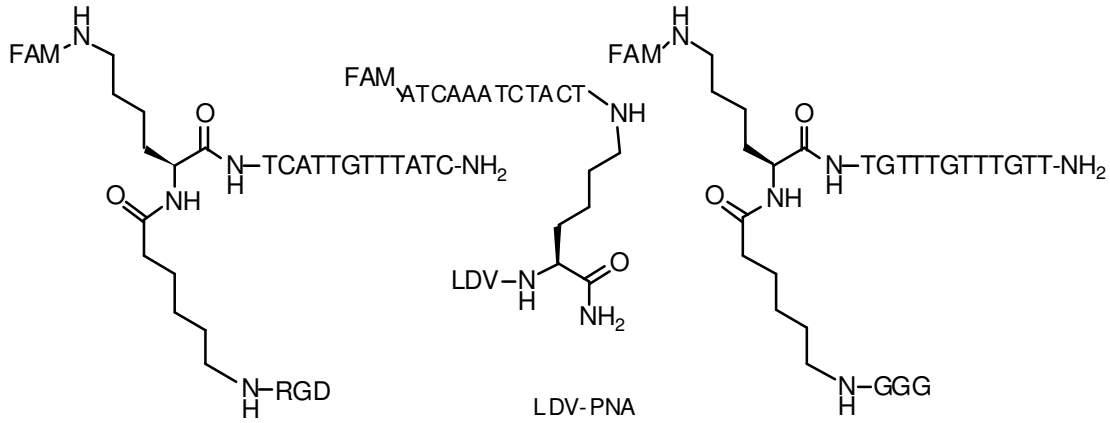
7.6 Experimental to Chapter 4

7.6.1 Synthesis of PNA-peptide conjugates and peptides

Unless otherwise stated the PNA-peptide conjugates and peptides were purified by precipitation in cold diethyl ether followed by RP-HPLC purification (column:

XTerra Prep RP18, 19 cm x 150 nm x 5µm) eluting with 100% H_2O with 0.1% TFA for 5 min 100% H_2O with 0.1% TFA to 40% H_2O with 0.1% TFA over 10 min; to 100% CH_3CN with 0.1% TFA over 4 min and 100% CH_3CN with 0.1% TFA for 2 min. Fractions were analysed by ESI-MS or MALDI-TOF and the collected fractions were freeze dried overnight.

7.6.1.1 Synthesis of RGD-, LDV-, and GGG-PNA conjugates[14]



RGD-, LDV-, and GGG-PNA conjugates were synthesised using microwave aided PNA and amino acid couplings as described above in an automated CEM peptide synthesiser. The conjugates were purified by RP-HPLC (method 2) and analysed by MALDI-TOF as described above. The following 4-letter codes were applied:

7.43: Arg-Gly-Glu-Ahx-Lys(FAM)-TCATTGTTTATC-NH_2

RP-HPLC: 12.42 min. (95.2% crude purity).

MALDI-TOF: m/z (%) 4164.75 $[M+H]^+$ (100).

7.44: Leu-Glu-Val-Lys(FAM-ATCAAATCTACT)-NH_2

RP-HPLC: 12.37 min. (75.4% crude purity).

MALDI-TOF: m/z (%) 4098.25 $[M+H]^+$ (100).

7.45: Gly-Gly-Gly-Ahx-Lys(FAM)-TGTTTGTTTGTT-NH_2

RP-HPLC: 11.80 min. (89.7% crude purity).

MALDI-TOF: m/z (%) 4100.71 $[M+H]^+$ (100).

7.6.2 Fabrication of RGD, LDV, and GGG-displaying microarray

DNA oligonucleotides (200 μM, Sigma-Aldrich) complementary to RGD-PNA, LDV-PNA, GGG-PNA, TCTTA (PNA), and marker DNA in 1 x Printing buffer (Agilent) were printed by contact printing using 4 solid pins, 1 stamp per ink, 3 stamps per spot, and 10 ms stamp- and inking time under 70% humidity in 9 spots/feature (squares) in a 3 x 3 grid x 4 subarrays. After printing the DNA was immobilised by keeping the microarray overnight in a box with filter paper soaked with NaCl (3 M). Thereafter the array was blocked with perfluorooctyl propylamine (2 M) and DIPEA (2 M) in acetonitrile for 30 min. The microarray was briefly washed with acetonitrile (3 x 20 mL), H_2O (3 x 20 mL), and Tris pH 8 (20 mL, 10 mM).

7.6.3 Microarray hybridisation of RGD-PNA, LDV-PNA, and GGG-PNA conjugates

Before hybridisation the PNA-peptide conjugate solution was denatured at 40 ^oC for 1 min. RGD-PNA, LDV-PNA, GGG-PNA, and TCTTA (1 μM) were hybridised in 1 x GenHyb buffer from 50-22 ^oC overnight in a GeneMachines Hyb4 hybridizer and the array was washed according to the Codelink microarray wash protocol.

7.6.4 HeLa growth on RGD, LDV, and GGG-displaying microarray

HeLa cells (5 mL, 20,000 cells/mL) detached with partly inactivated trypsin were incubated in RPMI CM enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) for 72 h. The microarray was washed with PBS and attached cells were fixed and stained with

phalloidin and DAPI as described in section 0. The microarray was imaged with a BioTEc LaVision BioAnalyser and a microscope (Leica DM IRB).

7.6.5 Flow cytometry of RGD-PNA, LDV-PNA, and GGG-PNA treated HeLa cells

HeLa cells were grown to 80% confluence in a 48-well plate and a solution of RGD-, GGG-, or LDV-PNA (10 μM) in DMEM enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) was added and incubated for 30 min to 2 h. The media was removed and the cells gently washed with PBS x 3. Cells were detached by scraping or by standard trypsination (15 min), suspended in PBS enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM, 300 μL), and analysed on a Calibur flow cytometer (FITC filter with 10,000 events).

7.6.6 Microarray hybridisation of Library-2

The DNA microarrays were hybridised with Library-2 (5 μM) in 1 x GenHyb buffer (150 μL, Genetix) overnight from 65 ^{o}C to 37 ^{o}C. The arrays were washed according to microarray wash protocol 1 and imaged with a Tecan LS Reloaded microarray scanner producing 4,000 x 12,000 pixel images.

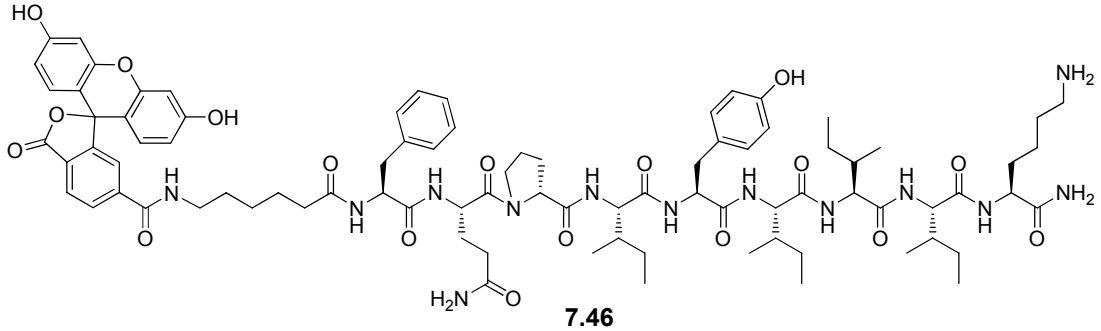
7.6.7 Cell growth on Library-2 displaying microarrays

D54 and HEK293T-CCR6 cells were grown to 80% confluence in DMEM CM enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) and detached by trypsination (above). Cells were allowed to re-attach for 8 h to allow regeneration of surface receptors. The cells were then detached by gently tapping the flask and incubated (5 mL, 50,000 cells/mL) with the Library-2 displaying microarray (section 0) in DMEM enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) for 2 h at 37 ^{o}C. The media was removed and fresh media was added and the incubation continued for 16 h. The microarray was washed briefly in PBS enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) and attached cells were fixed and stained with phalloidin and DAPI as described in section 0. The microarrays were imaged using a Nikon Eclipse 50i microscope producing 5,000 x 13,000 pixel images. The DAPI-filtered image and the

FITC-filtered image acquired prior to cellular incubation (section 0) were aligned in the following manner: the two FITC-filtered images acquired *pre* and *post* cellular incubation were superimposed and aligned using ImageViewer 1.0 software. Hereafter, the resulting realigned FITC-filtered image acquired prior to cellular incubation and the DAPI image were analysed as described in section 7.2 (appendix III).

7.6.8 Synthesis of integrin and GPCR ligands and non-binding peptides

Peptides were synthesised using microwave aided amino acid couplings and 5(6)-carboxyfluorescein couplings as described in section 0 and purified by RP-HPLC method 3 and ES-MS (positive mode).

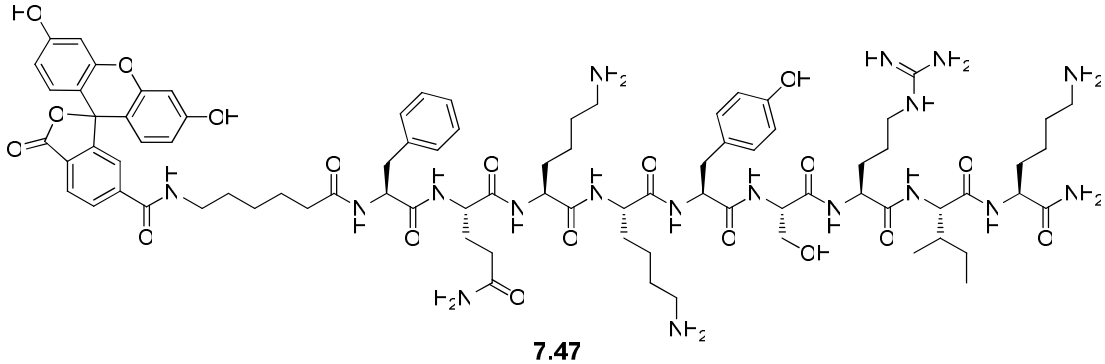


7.46

FAM-Ahx-FQ*p*IYIIIK-NH_2: 7.35 mg (41%).

RP-HPLC: 4.195 min. (46.4% crude purity).

MALDI-TOF^+/MS: m/z (%) 1577.5 $[M+H]^+$ (100).

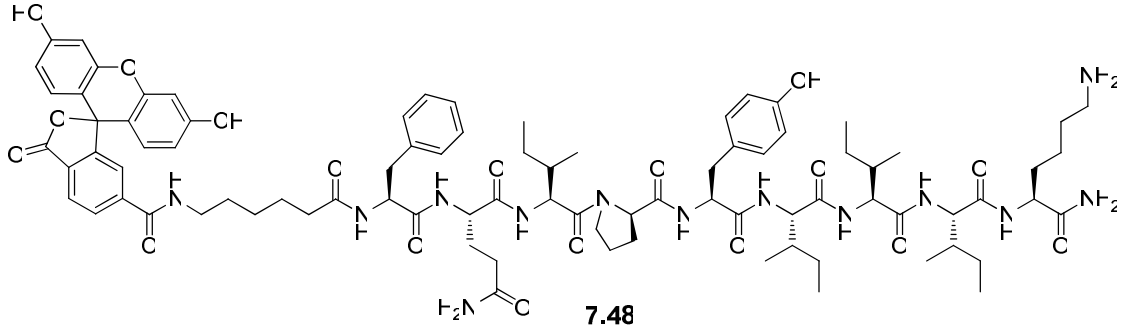


7.47

FAM-Ahx-FQKKYSRIK-NH_2: 13.5 mg (72%).

RP-HPLC: 2.818 min. (65.0% crude purity).

MALDI-TOF^+/MS: m/z (%) 1652 $[M+H]^+$ (100).

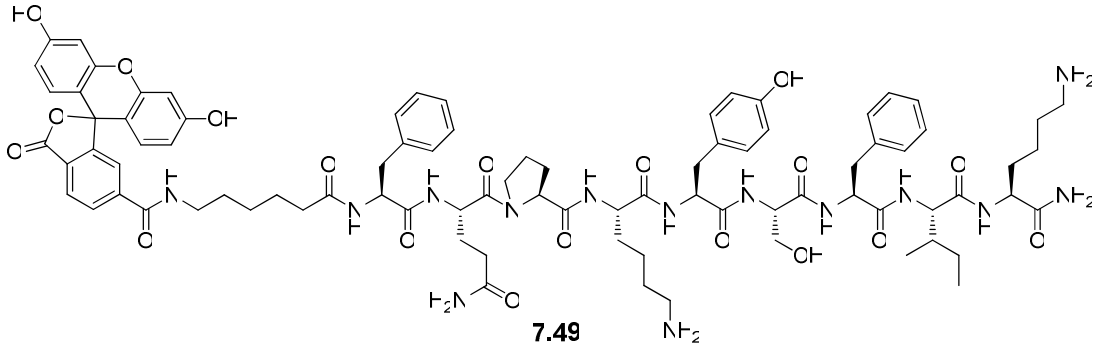


7.48

FAM-Ahx-FQIPYIIIK-NH_2: 6.53 mg (36%).

RP-HPLC: 4.190 min. (55.2% crude purity).

MALDI-TOF$^+$/MS: m/z (%)1576.6 $[M+H]^+$ (100).

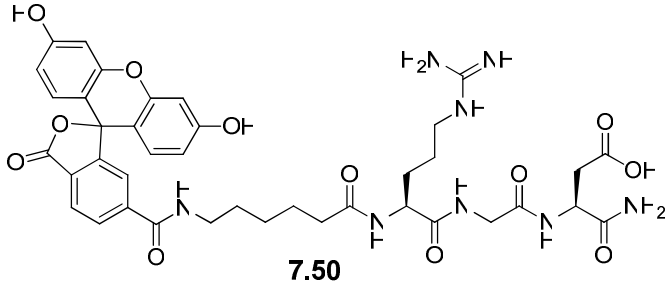


7.49

FAM-Ahx-FQPKYSFIK-NH_2: 14.4 mg (79%).

RP-HPLC: 3.457 min. (79.5% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 1600.4 $[M+H]^+$ (100).



7.50

FAM-Ahx-RGD-NH_2: 6.65 mg (74%).

RP-HPLC: 3.030 min. (95.4% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 789.2 $[M+H]^+$ (100).

7.6.9 Fabrication of "hit"-displaying microarrays

FAM-Ahx-FQ*p*IYIIIK-NH_2 (200 nM), FAM-Ahx-FQKKSRIK-NH_2 (200 nM), FAM-Ahx-FQIPYIIIK-NH_2 (200 nM), and FAM-Ahx-FQPKYSFIK-NH_2 (200 nM)

in sodium phosphate (50 mM, pH 8.5) were contact printed on Codelink slides (GE Healthcare) in 45% humidity using 4 solid pins, 1 stamp per ink, 3 stamps per spot, and 10 ms stamp- and inking time in 9 spots/feature (squares) in a 3 x 3 grid in 3 subarrays using a Genetix QArray Mini Microarrayer. The peptides were immobilised by incubating the microarrays in a box with filter paper soaked with NaCl (3 M) overnight. Unreacted sites were blocked with 2 M perfluorooctyl propylamine and DIPEA (2 M) in acetonitrile for 30 min. The microarrays were washed with acetonitrile (2 x 20 mL) briefly, H_2O (2 x 20 mL) briefly, 0.1% SDS in 4 x SSC (20 mL) at 50 ^oC for 30 min, H_2O (2 x 20 mL) briefly, and Tris pH 8 (20 mL, 10 mM) briefly. The array was imaged with a BioTEc LaVision BioAnalyser.

7.6.10 Cell growth on "hit"-displaying microarrays

Cell growth on the FAM-Ahx-FQpIYIIIK-NH_2/FAM-Ahx-FQKKSRIK-NH_2 displaying microarray (section 0) was carried out as described in section 0 with D54 (3 mL, 100,000 cells/mL) in DMEM enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM). In addition, D54 (3 mL, 20,000 cells/mL) were pre-incubated with FAM-Ahx-RGD-NH_2 (100 μM) in DMEM enriched $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) for 20 min before incubation on the microarray. Likewise, HEK293TCCR6 or HEK293T cells (3 mL, 100,000 cells/mL) were incubated on FAM-Ahx-FQIPYIIIK-NH_2/FAM-Ahx-FQPKYSFIK-NH_2 displaying microarrays (section 0) in DMEM enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM). In addition, HEK293TCCR6 cells (3 mL, 100,000 cells/mL) were pre-incubated with PerCP-Anti human CCR6 (5 μg/mL, R&D systems) in DMEM enriched $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) for 20 min before incubation on the microarray. Cells were fixed and stained as described above and imaged with a BioTEc LaVision BioAnalyser and a Nikon Eclipse 50i microscope.

7.7 Experimental to Chapter 5

7.7.1 Fabrication of single oligonucleotide microarray

The microarray containing a single DNA oligonucleotide was generated by contact printing a 3'-amino modified DNA oligonucleotide (Microsynth) on a Codelink slide in a 10 x 10 pattern. After printing, the unreacted sites on the slide were blocked with ethanolamine and the array was washed according to the Codelink protocol.

All other DNA microarrays were custom made by OGT.

7.7.2 PCR over microarrays

Extension reaction mix (200 µL) without primers was prepared according to a Promega standard protocol using a PCR master mix (Promega; 25 U/mL *Taq* Polymerase, 200 µM dNTP, 1.5 mM $MgCl_2$), which was loaded onto the microarray using an Agilent hybridisation cover slide. The first extension step was carried out at 50 ^oC (primer-2) or 55 ^oC (Solexa-primer-2) for 16 h (overnight). Thereafter, the reaction mixture was removed using a pipette and fresh PCR reaction mix (200 µL) with primer-1 and 2 (0.1 µM) *or* Solexa-primer-1 and 2 (0.1 µM) were loaded onto the microarray and a standard PCR cycle was set up in a GeneMachines Hyb4 automated hybridiser [40 cycles; denaturation at 94 ^oC for 30 s for 10 cycles and 88 ^oC for 30 s for 30 cycles, annealing at 49 ^oC for 1 min, extension at 50 ^oC for 5 min (primer-1 and 2); *or* denaturation at 94 ^oC for 30 s, annealing at 65 ^oC for 1 min, extension at 70 ^oC for 1 min (Solexa-primers-1 and 2)]. In addition, an initial 3 min denaturation step at 94 ^oC and a final 15 min extension step at 50 ^oC (primer-2) or 70 ^oC (Solexa-primer-2) were carried out. Immediately after the PCR had finished the reaction mix was collected using a pipette and the microarray washed with H_2O (3 x 50 µL). The aqueous fractions were pooled together and concentrated by speed vac followed by purification by preparative DNA gel electrophoresis (dsDNA-3875: 1.75

µg, 40% isolated yield; dsDNA-Library-2: 1.65 µg, 40% isolated yield; equation 7.2)[††].

$$yield\ per\ cycle = \left(\frac{n_{product}}{n_{theoretical}} \right)^{\frac{2}{cycles}} \cdot 100\% = \left(\frac{n_{product}}{n_{template} \cdot 2^{cycles}} \right)^{\frac{2}{cycles}} \cdot 100\% \quad \text{(Equation 7.2)}$$

7.7.3 PCR in solution

The purified products (250-875 ng) from each of the PCR "read-off" microarrays were used as templates in another round of PCR with Primer 1 and 2 (1 µM) *or* Solexa-primer-1 and 2 (1 µM) in a 1 x PCR Master Mix (200 µL) in a Techne TC-312 PCR cycler with the same cycle as on microarray. After PCR the samples were concentrated by speed vac followed by purification by preparative DNA gel electrophoresis (dsDNA-3875-2: 1.11 µg, 27% isolated yield; dsDNA-Library-2-2: 16.8 µg, 30% isolated yield).

The PCR products were amplified in another round of PCR carried out with DNA (250-875 ng), primer-1, and primer-2-FAM (2.5 µM), *or* primer-3, and primer-4-FAM [2.5 µM, Sigma-Aldrich; 40 cycles; denaturation at 94 ^{o}C for 30 s for 10 cycles and 88 ^{o}C for 30 s for 30 cycles, annealing at 49 ^{o}C for 1 min, extension at 50 ^{o}C for 5 min (primer-1 and primer-2-FAM); *or* denaturation at 94 ^{o}C for 30 s, annealing at 58 ^{o}C for 1 min, extension at 66 ^{o}C for 1 min (primers-3 and primer-4-FAM)]. In addition, an initial 3 min denaturation step at 94 ^{o}C and a final 15 min extension step at 50 ^{o}C (primer-1 and primer-2-FAM) or 66 ^{o}C (primers-3 and primer-4-FAM) were carried out. After PCR, the samples were concentrated by speed vac followed by purification by preparative DNA gel electrophoresis (dsDNA-Library-2-FAM: 20.5 µg, 29% isolated yield; dsDNA-3875-FAM: 28.3 µg, 30% isolated yield).

[††] Yield calculations were based on an assumed 25.000 DNA oligos per spot on the microarray. The actual amount of DNA per feature could not be disclosed by the manufacturer.

dsDNA_Library-2_FAM and dsDNA3,875_FAM were used as templates in ssDNA PCR amplification with Microarray Primer FAM (10 μM) in 1 x PCR Master Mix (600 μL) with the same cycle as described before for this primer. After PCR the samples were concentrated by speed vac followed by purification by preparative DNA gel electrophoresis (7.60 μg ssDNA-3875-FAM; 6.04 μg ssDNA-Library-2-FAM).

7.7.4 Digestion enzyme assay

dsDNA-Library-2-2 (200 ng) *or* dsDNA-3875-2 (200 ng) were digested with 0.25 units/μL EcoICRI (Promega) in 1 x RE buffer (20 μL, Promega) containing 0.1 μg/μL Acetylated BSA (Promega) at 37 ^{o}C for 4 h followed by analytical DNA gel electrophoresis.

7.7.5 Microarray hybridisation of the PCR products

The purified fluorescent PCR constructs were dissolved in 0.1% SDS in 4 x SSPE buffer (110 μL) and denatured at 65 ^{o}C for minimum 5 min. This solution was hybridised on customized DNA arrays (OGT) in an Agilent hybridisation chamber from 65-27 ^{o}C over 24 h. The arrays were washed according to microarray wash protocol 3 and the microarray was imaged with a Tecan LS Reloaded microarray scanner. Image analysis was carried out as described in section 7.2 (appendix IV).

7.7.6 Solexa Illumina Sequencing

dsDNA-Library-2 (200 nmol) was sequenced by Illumina Solexa DNA sequencing of 36-base reads off the Solexa Primer 1 domain at the end of each oligonucleotide (The GenePool, The University of Edinburgh). The resulting reads were clustered against a list of the 10,000 oligonucleotides in the 10,000 library and a list of the identified sequences was generated, including the number of times each oligonucleotide was seen. Another list of the sequences not seen by Illumina Solexa sequencing was generated (appendix V).

7.8 Experimental to Chapter 6

7.8.1 Library-2 extraction from cells

D54 and HEK293TCCR6 grown to 80-90% confluence in a T-25 flasks were incubated with Library-229 (4 mL, 10 μM) in DMEM enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) for 30 min at 37 ^oC. The media was removed and the cells were washed with PBS (x 4) enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM) and treated with 1 x pronase/EDTA (Merck Chemicals) in PBS for 20 min at 37 ^oC. The cells were pelleted by centrifugation (1100 rpm for 5 min) and the supernatant was recovered. PNA-peptide conjugates were purified by filter-centrifugation (3,000 Da and 10,000 Da molecular weight, Nunc TM) and the amount of PNA was quantified by measuring the emission of 5(6)-carboxyfluorescein at 520 nm. This afforded surface bound PNA encoded peptides extracted from D54 (4.4 pmol), and HEK293TCCR6 (2.4 pmol).

7.8.2 Production of ssDNA-Library-2

PCR was carried out with dsDNA-Library-2 (~750 ng, section 0) in 1 x PCR Master Mix (600 μL, Promega) with primer-3 (10 μM, Sigma-Aldrich) via: 40 cycles; denaturation at 94 ^oC for 30 s; annealing at 62 ^oC for 2 min; and extension at 68 ^oC for 2 min in a Techne TC-312 PCR cycler. In addition, an initial 3 min denaturation step at 94 ^oC and a final 15 min extension step at 68 ^oC were carried out. After PCR, the samples were concentrated by speed vac followed by purification by preparative DNA gel electrophoresis.

7.8.3 Amplification of PNA

7.8.3.1 Single oligonucleotide

Oligonucleotide-1 (0.30 nmol, Sigma, Table 6.1) was hybridised with peptide-PNA (Gly-Gly-Gly-TGTTTGTTTGTT, 0.3 nmol) and oligonucleotide-2 and 3 (0.60 nmol, Sigma, Table 6.1) in 4 x SSPE buffer (30 μL, Invitrogen) from 65 ^oC to 37 ^oC over 24 h. Half of the warm solutions were kept as a positive control for PNA/DNA hybridisation and the remaining were diluted 10 times with 37 ^oC 1.1 x S1 Nuclease Reaction buffer (Promega) and the sample was treated with DNA S1 Nuclease (100

U/1 μg, Promega) for 15 min at 37 °C. Similarly, oligonucleotide-1 (0.15 nmol) was hybridised with oligonucleotide-2 and 3 (0.30 nmol) as a positive control for primer hybridisation. The hybridisation mixtures were analysed by DNA gel electrophoresis.

7.8.3.2 10,000 oligonucleotides

ssDNA-Library-2 (0.20 nmol, section 0) was hybridised with PNA-Library-229 (0.10 nmol) and oligonucleotide-2 and 3 (0.40 nmol, Table 6.1) as described above. Similarly, ssDNA-Library-2 (5 equiv.) was hybridised with oligonucleotide-2 and 3 (6 equiv., Table 6.1) as well as PNA-D54 (4.4 pmol, 1 equiv.) *or* PNA-HEK293T-CCR6 (2.44 pmol, 1 equiv.). The warm solutions were diluted 10 times with 37 °C 1.1 x S1 Nuclease Reaction buffer (Promega) and the sample was treated with 100 U/1 μg DNA S1 Nuclease (Promega) for 15 min at 37 °C. The dsDNA/PNA-Library-2 complexes were purified by Sephadex G-25 gel column filtration (GE Healthcare) according to manufacturer's protocol and the eluates concentrated by speed vac.

The purified dsDNA/PNA complexes were amplified by PCR in 1 x PCR Master Mix (50 μL) with primer-3 (2 μM) and primer-4-FAM (2 μM) in a Techne TC-312 PCR cycler with the same cycle as described before. The PCR products were concentrated by speed vac purified by preparative DNA gel electrophoresis. The product was amplified by PCR again in new PCR mix (600 μL) according to the same PCR cycle and purification procedure. The dsDNA-FAM were used as templates in ssDNA PCR amplification in 1 x PCR Master Mix (600 μL) with primer-4-FAM (10 μM) carried out with the same cycle as described above. After PCR, the samples were concentrated by speed vac followed by purification by preparative DNA gel electrophoresis (ssDNA-Library-2-FAM: 6.75 μg, ssDNA-D54-FAM: 4.73 μg, ssDNA-HEK293TCCR6-FAM: 3.21 μg).

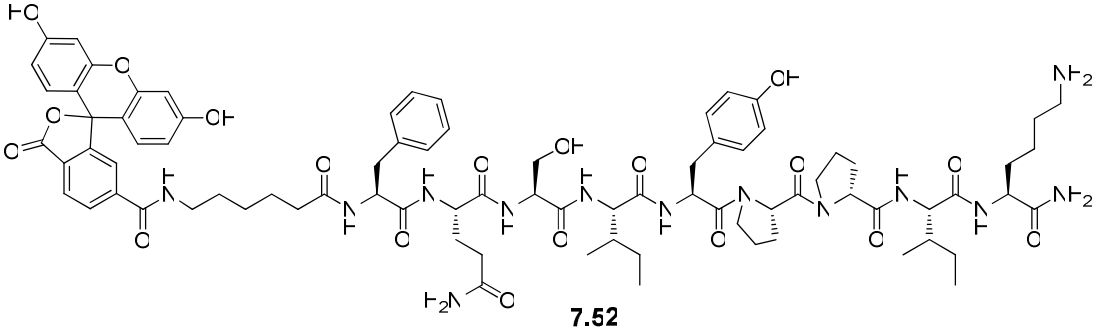
7.8.4 Microarray hybridisation of ssDNA-Library-2

Library-2 complementary DNA microarrays (OGT) were hybridised with the purified ssDNA-FAM in 0.1% SDS in 4 x SSPE buffer (110 μL) overnight from 65 °C to 27 °C and the arrays were washed according to microarray wash protocol 3 and

imaged with a Tecan LS Reloaded microarray scanner. Image analysis was carried out as described in section 7.2 (appendix VI).

7.8.5 Peptide synthesis of integrin and GRCP ligands and non-binding peptides

Peptides were synthesised using microwave aided amino acid couplings and 5(6)-5(6)-carboxyfluorescein couplings as described in section 0 and purified by RP-HPLC method 3 and ES-MS (positive mode).

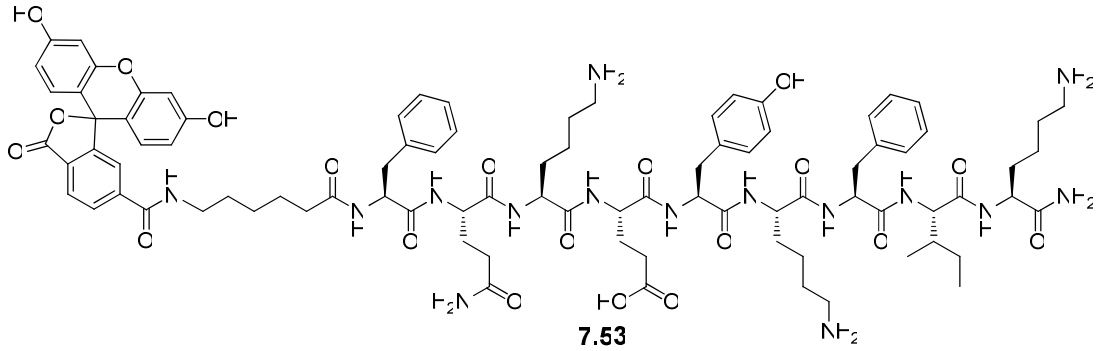


7.52

FAM-Ahx-FQSIYP*p*IK-NH$_2$: 12.8 mg (73%).

RP-HPLC: 3.943 min. (81.1% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 1633.4 $[M]^+$ (100).

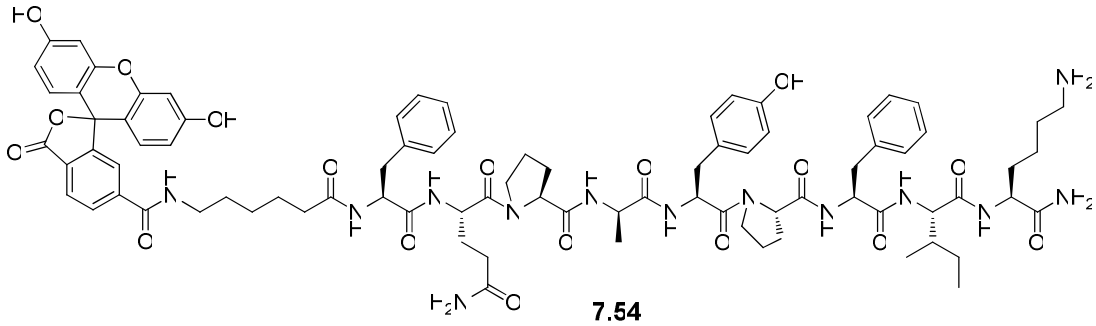


7.53

FAM-Ahx-FQKEYKFIK-NH$_2$: 13.0 mg (68%).

RP-HPLC: 3.197 min. (72.8% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 1672.7 $[M+H]^+$ (100).

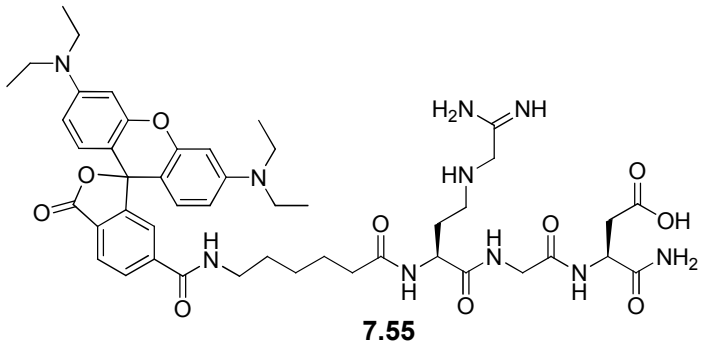


7.54

FAM-Ahx-FQP*a*YPFIK-NH$_2$: 11.2 mg (63%).

RP-HPLC: 3.868 min. (48.8% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 1553.4 $[M+H]^+$ (100).



7.55

Rho-Ahx-RGD-NH$_2$: 10.2 mg (100%).

RP-HPLC: 3.269 min. (71.5% crude purity).

MALDI-TOF$^+$/MS: m/z (%) 899.3 $[M+1H]^+$ (100).

7.8.6 Flow cytometry of integrin and GRCP ligands and non-binding peptides

D54, HEK293T, and HEK293TCCR6 cells were grown to 80-90% confluence in 24-well plates. HEK293T and HEK293TCCR6 cells were incubated with PerCP-anti human CCR6 (10 µg/mL), FITC-anti human CCR6 (10 µg/mL, R&D systems), FAM-Ahx-FQSIYP*p*IK-NH$_2$ (10 µM), FAM-Ahx-FQP*a*YPFIK-NH$_2$ (10 µM), PerCP-anti human CCR6 (10 µg/mL) *and* FAM-Ahx-FQSIYP*p*IK-NH$_2$ (10-100 µM) *or* FAM-Ahx-FQP*a*YPFIK-NH$_2$ (10 µM). D54 incubation with: FAM-Ahx-RGD-NH$_2$ (10 µM), Rho-Ahx-RGD-NH$_2$ (10 µM), FAM-Ahx-FQIPYIIIK-NH$_2$ (10 µM), FAM-Ahx-FQKDYKFIK-NH$_2$, (10 µM) Rho-Ahx-RGD-NH$_2$ (10 µM) *and* FAM-Ahx-FQIPYIIIK-NH$_2$ (10-100 µM) *or* FAM-Ahx-FQKDYKFIK-NH$_2$ (10 µM) in DMEM enriched with $CaCl_2$ (50 µM) and $MgCl_2$ (50 µM) for 30 min at 37 ^oC. The

media was removed and the cells were washed with PBS x 3, detached by scraping or by treatment with 1 x pronase/EDTA for 15 min at 37 ^{o}C, and suspended in PBS enriched with $CaCl_2$ (50 μM) and $MgCl_2$ (50 μM). The samples were analysed by flow cytometry using a BD FACSAria cytometer (D54 cells, FITC and PE-Texas Red filters and 10,000 events) or a BD Calibur cytometer (HEK293T and HEK293TCCR6 cells, FITC and PE/PI filters and 10,000 events).

References

1. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991,** *254*, 1497-1500.
2. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D.A.; Berg, R. H.; Kim, S. K.; Nordén, B.; Nielsen, P. E. *Nature* **1993**, *365*, 566-568.
3. Albericio, F.; Carpino L. A. *Methods in Enzymology, Solid-phase peptide synthesis; Academic Press: Orlando*, **1997**.
4. Chan W. C.; White P. D. *Fmoc solid phase peptide synthesis: a practical approach (2nd ed.); Oxford University Press* **2004**.
5. Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.
6. Nielsen P. E. *Peptide nucleic acids: protocols and applications (2nd ed.)*; *Horizon Bioscience,* **2004**.
7. Casale, R.; Jensen, I. S.; Egholm, M. *Peptide Nucleic Acids (2nd ed.); Horizon Bioscience* **2004**, 61-76.
8. Dueholm, K. L.; Egholm, M.; Behrens, C.; Christensen, L.; Hansen, H. F.; Vulpius, T.; Petersen, K. H.; Berg, R. H.; Nielsen, P. E.; Buchardt, O. *J. Org. Chem.* **1994**, *59*, 5767-5773.
9. Pothukanuri, S.; Pianowski, Z.; Winssinger, N. *Eur. J. Org. Chem.* **2008**, 3141-3148.
10. Wojciechowski, F.; Hudson, R. H. E. *J. Org. Chem.* **2008,** *73*, 3807-3816.
11. Porcheddu, A.; Giacomelli, G.; Piredda, I.; Carta, M.; Nieddu, G. *Eur. J. Org. Chem.* **2008**, *34*, 5786-5797.
12. Breipohl, G.; Knolle, J.; Langner, D.; O'Malley, G.; Uhlmann, E. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 665-670.
13. Thomson, S. A.; Josey, J. A.; Cadilla, R.; Gaul, M. D.; Hassman, C. F.; Luzzio, M. J.; Pipe, A. J.; Reed, K. L.; Ricca, D. J.; Wiethe, R. W.; Noble, S. A. *Tetrahedron* **1995**, *51*, 6179-6194.
14. Bialy L. Díaz-Mochón J.J Specker E. Keinicke L. Bradley M., *Tetraheron* **2005**, *61*, 8295-8305.
15. Diaz-Mochon, J. J.; Bialy, L.; Bradley, M. *Org. Lett.* **2004**, *6*, 1127-1129.
16. Bycroft, B. W.; Chan, W. C.; Chhabra, S. R.; Teesdale-Spittle, P. H.; Hardy, P. M. *J. Chem. Soc., Chem. Commun.* **1993**, *39*, 776-777.
17. Svensen, N.; Díaz-Mochón, J.J.; Bradley, M. *Tetrahedron Lett.* **2008**, *49*, 6498-6500.
18. Fukuda, H.; Takao, T.; Coull, J.; Shimonishi, Y. *Pept. Chem.* **1993**, 45-48.
19. Christensen, L.; Fitzpatrick, R.; Gildea, B.; Petersen, K.H.; Hansen, H.F.; Koch, T.; Egholm, M.; Buchardt, O.; Nielsen, P.E.; Coull, J.; Berg, R.H. *J. Pept. Sci.* **1995**, *1*, 175-183.
20. Christensen, L.; Fitzpatrick, R.; Gildea, B.; Warren, B.; Coull, J. *Improved synthesis, purification, and characterisation of PNA Oligomers. In: Solid phase synthesis; Mayflower World-Wide Limited* **1994**, 149-156.
21. Uhlmann, E.; Peyman, A.; Breipohl, G.; Will, D. W. *Angew. Chem., Int. Ed.* **1998**, *37*, 2796-2823.
22. Nielsen, P. E.; Egholm, M. *Current Issues Molec. Biol.* **1999**, *1*, 89-104.

23. Lohse, J.; Dahl, O.; Nielsen, P. E. *Proc. Nat. Acad. Sci. U.S.A.* **1999**, *96*, 11804-11808.
24. Wittung, P.; Nielsen, P. E.; Norden, B. *Biochemistry* **1997**, *36*, 7973-7979.
25. Demidov, V. V.; Yavnilovich, M. V.; Belotserkovskii, B. P.; Frank-Kamenetskii, M. D.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 2637-2641.
26. Veselkov, A. G.; Demidov, V. V.; Frank-Kamenetskii, M. D.; Nielsen, P. E. *Nature* **1996**, *379*, 214-215.
27. Kuhn, H.; Hu, Y.; Frank-Kamenetskii, M. D.; Demidov, V. V. *Biochemistry* **2003**, *42*, 4985-4992.
28. Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchard, O.; Sonnichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994**, *48*, 1310-1313.
29. Pouchain D. Diaz-Mochon J.J. Bialy L. and Bradley M., *ACS Chemical Biol.* **2007**, *2*, 810-818.
30. Ray, A.; Nordén, B. *FASEB J.* **2000**, *14*, 1041-1060.
31. Meldal, M.; Svendsen, I.; Breddamand, K.; Auzanneau, F. I. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3314-3318.
32. Tornoe, C. W.; Sanderson, S. J.; Mottram, J. C.; Coombs, G. H.; Meldal, M. *J. Comb. Chem.* **2004**, *6*, 312-324.
33. Diaz-Mochon, J. J.; Bialy, L.; Bradley, M. *Chem. Commun.* **2006**, *38*, 3984-3986.
34. Seitz, O. *Angew. Chem., Int. Ed.* **2000**, *39*, 3249-3252.
35. Socher, E.; Bethge, L.; Knoll, A.; Jungnick, N.; Herrmann, A.; Seitz, O. *Angew. Chem., Int. Ed.* **2008**, *47*, 9555-9559.
36. Stender, H. *Expert Rev. Mol. Diagn.* **2003**, *3*, 649-655.
37. Lansdorp, P. M.; Verwoerd, N. P.; van de Rijke, F. M.; Dragowska, V.; Little, M.-T.; Dirks, R. W.; Raap, A. K.; Tanke, H. J. *Hum. Mol. Genet.* **1996**, *5*, 685-691.
38. Igloi, G. L. *Expert Rev. Mol. Diagn.* **2003**, *3*, 17-26.
39. Thiede, C.; Bayerdoerffer, E.; Blasczyk, R.; Wittig, B.; Neubauer, A. *Nucleic Acids Res.* **1996**, *24*, 983-984.
40. Perry-O'Keefe, H.; Yao, X.-W.; Coull, J. M.; Fuchs, M.; Egholm, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14670-14675.
41. Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G. *Science* **1992**, *258*, 1481-1485.
42. Boutimah-Hamoudi, F.; Leforestier, E.; Senamaud-Beaufort, C.; Nielsen, P. E.; Giovannangeli, C.; Saison-Behmoaras, T. E. *Nucleic Acids Res.* **2007**, *35*, 3907-3917.
43. Boffa, L. C.; Scarfi, S.; Mariani, M. R.; Damonte, G.; Allfrey, V. G.; Benatti, U.; Morris, P. L. *Cancer Res.* **2000**, *60*, 2258-2262.
44. Tan, X.-X.; Actor, J. K.; Chen, Y. *Antimicrob. Agents Chemother.* **2005**, *49*, 3203-3207.
45. Tripathi, S.; Chaubey, B.; Barton, B. E.; Pandey, V. N. *Virology* **2007**, *363*, 91-103.

46. Turner, J. J.; Ivanova, G. D.; Verbeure, B.; Williams, D.; Arzumanova, A. A.; Abes, S.; Lebleu, B.; Gait, M. J. *Nucleic Acids Res.* **2005**, *33*, 6837-6849.
47. Lechardeur, D.; Sohn, K-J.; Haardt, M.; Joshi, P. B.; Monck, M.; Graham, R. W.; Beatty, B.; Squire, J.; O'Brodovich, H.; Lukacs, G. L . *Gene Therapy* **1999**, *6*, 482-497.
48. Larsen, H. J.; Nielsen, P. E. *Nucleic Acids Res.* **1996**, *24*, 458-463.
49. Janowski, B. A.; Kaihatsu, K.; Huffman, K. E.; Schwartz, J. C.; Ram, R.; Hardy, D.; Mendelson, C. R.; Corey, D. R. *Nat. Chem. Biol.* **2005**, *1*, 210-215.
50. Liebling, M. R.; Jou, N.-T.; Fang, W.; Louie, J. S. *Mol. Biotechnol.* **2003**, *25*, 229-240.
51. Diviacco, S.; Rapozzi, V.; Xodo, L.; Helene, C.; Quadrifoglio, F.; Giovannangeli, C. *FASEB J.* **2001**, *15*, 2660-2668.
52. Fabani, M. M.; Gait, M. J. *RNA* **2008**, *14*, 336-346.
53. Koppelhus, U.; Nielsen, P. E. *Adv. Drug Deliv. Rev.* **2003**, *55*, 267-280.
54. Rebuffat, A. G.; Nawrocki, A. R.; Nielsen, P. E.; Bernasconi, A. G.; Bernal-Mendez, E.; Frey, B. M.; Frey, F. J. *FASEB J.* **2002**, *16*, 1426-1428.
55. Diaz-Mochon, J. J.; Bialy, L.; Watson, J.; Sanchez-Martin, R. M.; Bradley, M. *Chem. Commun.* **2005**, *26*, 3316-3318.
56. Koppelhus, U.; Awasthi, S. K.; Zachar, V.; Holst, H. U.; Ebbesen, P.; Nielsen, P. E. *Antisense Nucleic Acid Drug Dev.* **2002**, *12*, 51-63.
57. Dragulescu-Andrasi, A.; Rapireddy, S.; He, G.; Bhattacharya, B.; Hyldig-Nielsen, J. J.; Zon, G.; Ly, D. H. *J. Am. Chem. Soc.* **2006**, *128*, 16104-16112.
58. Mehiri, M.; Upert, G.; Tripathi, S.; Di Giorgio, A.; Condom, R.; Pandey, V. N.; Patino, N. *Oligonucleotides* **2008**, *18*, 245-256.
59. Hamilton, S. E.; Simmons, C. G.; Kathiriya, I. S.; Corey, D. R. *Chem. Biol.* **1999**, *6*, 343-351.
60. Shammas, M. A.; Simmons, C. G.; Corey, D. R.; Shmookler Reis, R. J. *Oncogene* **1999**, *18*, 6191-6200.
61. Orum, H.; Nielsen, P. E.; Jorgensen, M.; Larsson, C.; Stanley, C.; Koch, T. *Biotechniques* **1995**, *19*, 472-480.
62. Demers, D. B.; Curry, E. T.; Egholm, M.; Sozer, A. C. *Nucleic Acids Res.* **1995**, *23*, 3050-3055.
63. Shepard, J. R.; Addison, R. M.; Alexander, B. D.; Della-Latta, P.; Gherna, M.; Haase, G.; Hall, G.; Johnson, J. K.; Merz, W. G.; Peltroche-Llacsahuanga, H.; Stender, H.; Venezia, R. A.; Wilson, D.; Procop, G. W.; Wu, F.; Fiandaca, M. J. J., *Clin. Microbiol.* **2008**, *46*, 50-55.
64. Geysen, H. M.; Meloen, R. H.; Barteling, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998-4002.
65. Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131-5135.
66. Nicolaou, K. C.; Pfefferkorn, J. A.; Mitchell, H. J.; Roecker, A. J.; Barluenga, S.; Cao, G. Q.; Affleck, R. L.; Lillig, J. E. *J. Am. Chem. Soc.* **2000**, *122*, 9954-9967.
67. Nicolaou, K. C.; Xiao, X.-Y.; Parandoosh, Z.; Senyei, A.; Nova, M. P. *Angew. Chem., Int. Ed.* **1995**, *34*, 2289-2991.
68. Frank, R. *Tetrahedron* **1992**, *48*, 9217-9232.
69. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Abstr. 14th Int. Congr. Biochem.*, *Prague, Czechoslovakia.* **1988**, *5*, 47.

70. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487-493.
71. Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82-84.
72. Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381-5383.
73. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700-10704.
74. Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc.* **1993**, *115*, 9812-9813.
75. Ben-Dor, A.; Karp, R.; Schwikowski, B.; Yakhini, Z. *J. Comput. Biol.* **2000**, *7*, 503-519.
76. Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Neri, D. *J. Biotechnol.* **2006**, *126*, 568-581.
77. Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. *Nat. Biotechnol.* **2004**, *22*, 568-574.
78. Franch, T. *et al. Patent number WO2007062664*, **2007**.
79. Morgan, B. *et al. Patent number WO2005058479*, **2005**.
80. Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K. Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. *Nature Chem. Biol.* **2009**, *5*, 647-654.
81. Ladoukakis, E. D.; Zouros, E. *Mol. Biol. Evol.* **2001**, *18*, 1168-75.
82. Webb, C. O.; Ackerly, D. D.; McPeek, M. A.; Donoghue, M. J. *Annu. Rev. Ecol. Syst.* **2002**, *33*, 475-505.
83. Kress, W. J.; Erickson, D. L. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2761-2762.
84. Pianowski, Z. L.; Winssinger, N. *Chem. Soc. Rev.* **2008**, *37*, 1330-1336.
85. Winssinger, N.; Harris, J. L.; Backes, B. J.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2001**, *40*, 3152-3155.
86. Winssinger, N.; Ficarro, S.; Schultz, P. G.; Harris, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11139-11144.
87. Diaz-Mochon, J. J.; Bialy, L.; Keinicke, L.; Bradley, M. *Chem. Commun.* **2005**, *11*, 1384-1386.
88. Urbina, H. D.; Debaene, F.; Jost, B.; Bole-Feysot, C.; Mason, D. E.; Kuzmic, P.; Harris, J. L.; Winssinger, N. *Chem. Bio. Chem.* **2006**, *7*, 1790-1797.
89. Kulesh, D. A.; Clive, D. R.; Zarlenga D. S.; Greene J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8453-8457.
90. Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995**, *270*, 467-470.
91. Lashkari, D. A.; DeRisi, J. L.; McCusker, J. H.; Namath, A. F.; Gentile, C.; Hwang, S. Y.; Brown, P. O.; Davis, R. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13057-13062.
92. Muller, U. R.; Niolau, D. V. *Microarray technology and its applications (1st ed.)*; *Springer-Verlag* **2005**.

93. Houseman, B. T.; Huh, J. H.; Kron, S. J.; Mrksich, M. *Nat. Biotechnol.* **2002**, *20*, 270-274.
94. MacBeath, G.; Schreiber, S. L. *Science* **2000**, *289*, 1760-1763.
95. Zlauddin, J.; Sabatini, D. M. *Nature* **2001**, *411*, 107-110.
96. Kononen, J.; Bubendorf, L.; Kallioniemi, A.; Barlund, M.; Schraml, P.; Leighton, S.; Torhorst, J.; Mihatsch, M. J.; Sauter, G.; Kallioniemi, O. P. *Nat. Med.* **1998**, *4*, 844-847.
97. Wang, D.; Liu, S.; Trummer, B. J.; Deng, C.; Wang, A. *Nat. Biotechnol.* **2002**, *20*, 275-281.
98. Mant, A.; Tourniaire, G.; Diaz-Mochon, J. J.; Elliott, T. J.; Williams, A. P.; Bradley, M. *Biomaterials* **2006**, *27*, 5299-5306.
99. MacBeath, G.; Koehler, A. N.; Schreiber, S. L. *J. Am. Chem. Soc.* **1999**, *121*, 7967-7968.
100. Stears, R. L.; Martinsky, T.; Schena, M. *Nat. Med.* **2003**, *9*, 140-145.
101. Brown, P. O.; Botstein, D. *Nat. Genet.* **1999**, *21*, 33-37.
102. Gresham, D.; Dunham, M. J.; Botstein, D. *Nat. Rev. Genet.* **2008**, *9*, 291-302.
103. Pollack, J. R.; Perou, C. M.; Alizadeh, A. A.; Eisen, M. B.; Pergamenschikov, A.; Williams, C. F.; Jeffrey, S. S.; Botstein, D.; Brown, P. O. *Nat. Genet.* **1999**, *23*, 41-46.
104. Pease, A. C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. P.; Fodor, S. P. A. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5022-5026.
105. Pirrung, M. C. *Angew. Chem., Int. Ed.* **2002**, *41*, 1276-1289.
106. Singh-Gasson, S.; Green, R. D.; Yue, Y.; Nelson, C.; Blattner, F.; Sussman, M. R.; Cerrina, F. *Nat. Biotechnol.* **1999**, *17*, 974-978.
107. Stengele, K.; Bühler, J.; Bühler, S.; Kvassiouk, E.; Green, R.; Prykota, T.; Pfleiderer, W. *Nucleos. Nucleot. Nucl.* **2005**, *24*, 891-896.
108. Hughes, T. R.; Mao, M.; Jones, A.R.; Burchard, J.; Marton, M.J.; Shannon, K.W.; Lefkowitz, S.M.; Ziman, M.; Schelter, J.M.; Meyer, M.R. *Nat. Biotechnol.* **2001**, *19*, 342-347.
109. Lausted, C.; Dahl, T.; Warren, C.; King, K.; Smith, K.; Johnson, M.; Saleem, R.; Aitchison, J.; Hood, L.; Lasky, S.R. *Genome Biol.* **2004**, *5*, R58.1-R58.16.
110. Liu, R. H.; Dill, K.; Fuji, H. S.; McShea, A. *Expert Rev. of Mol. Diag*. **2006**, *6*, 253-261.
111. Mamanova, L.; Coffey, A. J.; Scott, C. E.; Kozarewa, I.; Turner, E. H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D. J. *Nature methods* **2010**, *7*, 111-8.
112. Schena, M. *Microarray Analysis; New York: Wiley-Liss, John Wiley & Sons, Inc.* **2003**, 95-120.
113. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell (4th ed.)*; *Garland Science* **2002**.
114. Humphries, M.J. *Biochem. Soc. Trans*. **2000**, *28*, 311-339.
115. Filmore, D. *Mod. Drug Discov.* **2004**, *7*, 24-28.
116. Kroeze, W. K.; Sheffler, D. J.; Roth, B. L. *J. Cell Sci.* **2003**, *116*, 4867-4869.
117. Hynes R. *Cell* **2002**, *110*, 673-687.
118. Humphries, J. D.; Byron, A.; Humphries, M. J. *J. Cell Sci.* **2006**, *119*, 3901-3903.
119. Morgan, M. R.; Humphries, M. J.; Bass, M. D. *Nat. Rev. Mol. Biol.* **2007**, *8*, 957-969.

120. Hermann, P.; Armant, M.; Brown, E.; Rubio, M.; Ishihara, H.; Ulrich, D.; Caspary, R. G.; Lindberg F. P.; Armitage, R.; Maliszewski, C.; Delespesse, G.; Sarfati, M. *J. Cell Biol.* **1999**, *144*, 767-75.
121. O'Toole, T. E.; Katagiri, Y; Faull, R. J.; Peter, K.; Tamura, R.; Quaranta, V.; Loftus, J. C.; Shattil, S. J.; Ginsberg, M. H. *J. Cell Biol.* **1994**, *124*, 1047-1059.
122. Gupta, V.; Gylling, A.; Alonso, J.-L.; Sugimori, T.; Ianakiev, P.; Xiong, J.-P.; Arnaout, M. A. *Blood* **2007**, *109*, 3513-20.
123. Calderwood, D. A.; Shattil, S. J.; Ginsberg, M. H. *J. Biol. Chem.* **2000**, *275*, 22607-22610.
124. Carey, D. J. *Biochem. J.* **1997**, *327,* 1-16.
125. Bernfield, M.; Kokenyesi, R.; Kato, M.; Hinkes, M. T.; Spring, J.; Gallo, R. L.; Lose, E. J. *Annu. Rev. Cell. Biol.* **1992**, *8,* 365–393.
126. Beauvais, D. M.; Burbach, B. J.; Rapraeger, A. C. *J. Cell Biol.* **2004**, *167*, 171-181.
127. Morgan, M. R.; Humphries, M. J.; Bass, M. D. *Nat. Mol. Cell Biol.* **2007**, *8*, 957-969.
128. Ng, T.; Shima, D.; Squire, A.; Bastiaens, P. I. H.; Gschmeissner, S.; Humphries, M. J.; Parker, P. J. *EMBO J.* **1999**, *18*, 3909-3923.
129. Han, J.; Lim, C.J.; Watanabe, N.; Soriani, A.; Ratnikov, B.; Calderwood, D.A.; Puzon-McLaughlin, W.; Lafuente, E.M.; Boussiotis, V.A.; Shattil, S.J.; Ginsberg, M.H. *Curr. Biol.* **2006**, *16*, 1796-1806.
130. Koo, B. K.; Jung, Y. S.; Shin, J.l Han, I.; Mortier, E.; Zimmermann, P.; Whiteford, J. R.; Couchman, J. R.; Oh, E. S.; Lee, W. *J. Mol. Biol.* **2006**, *355*, 651-663.
131. Zimmermann, P.; Zhang, Z.; Degeest, G.; Mortier, E.; Leenaerts, I.; Coomans, C.; Schulz, J.; N'Kuli, F.; Courtoy, P. J.; David, G. *Dev. Cell* **2005**, *9*, 377–388.
132. Yoo, J.; Jeong, M. J. Cho, H. J.; Oh, E. S.; Han, M. Y. *Biochem. Biophys. Res. Commun.* **2005**, *328*, 424-431.
133. Gilman A.G. *Annu. Rev. Biochem.* **1987**, *56*, 615–649.
134. Zaballos, A.; Varona, R.; Gutiérrez, J.; Lind, P.; Márquez, G. *Biochem. Biophys. Res. Commun.* **1996**, *227*, 846-853.
135. Baba, M.; Imai, T.; Nishimura, M.; Kakizaki, M.; Takagi, S.; Hieshima, K.; Nomiyama, H.; Yoshie, O. *J. Biol. Chem.* **1997**, 272, 14893-14898.
136. Ai, L. S.; Liao, F. *Biochemistry* **2002**, *41*, 8332-8341.
137. Insel, P. A.; Tang, C-M.; Hahntow, I.; Michel, M. C. *Biochim. Biophys. Acta.* **2007**, *1768*, 994-1005.
138. Leskov, I. B.; Klenchin, V. A.; Handy, J. W.; Whitlock, G. G.; Govardovskii, V. I.; Bownds, M. D.; Lamb, T. D.; Pugh, E. N.; Arshavsky, V. Y. *Neuron* **2000**, *27*, 525-537.
139. Mukherjee, S.; Ghosh, R. N.; Maxfield, F. R. *Physiol. Rev.* **1997**, *77*, 759-803.
140. Miaczynska, M.; Stenmark, H. *J. Cell Biol.* **2008**, *180*, 7-11.
141. Soldati, T.; Schliwa, M. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 897-908.
142. Aderem, A.; Underhill, D. M. *Annu. Rev. Immunol.* **1999**, *17,* 593-623.
143. Falcone, S.; Cocucci, E.; Podini, P.; Kirchhausen, T.; Clementi, E.; Meldolesi, J. *J. Cell Sci.* **2006**, *119*, 4758-4769.
144. Rappoport, J. Z. *Biochemical J.* **2008**, *412*, 415-423.

145. Luzio, J. P.; Rous, B. A.; Bright, N. A.; Pryor, P. R.; Mullock, B. M.; Piper, R. C. *J. Cell Sci.* **2000**, *113*, 1515-1524.
146. Mitchell, D. J.; Kim, D. T.; Steinman, L.; Fathman, C. G.; Rothbard, J. B. *J. Pept. Res.* **2000**, *56*, 318-325.
147. Green M., Loewenstein P. M. *Cell* **1988**, *55*, 1179-1188.
148. Frankel, A. D.; Pabo, C. O. *Cell* **1988**, *55*, 1189-1193.
149. Astriab-Fisher, A.; Sergueev, D. S.; Fisher, M.; Shaw, B. R.; Juliano, R. L. *Biochem. Pharmac.* **2000**, *60*, 83-90.
150. Fawell, S.; Seery, J.; Daikh, Y.; Moore, C.; Chen, L. L.; Pepinsky, B.; Barsoum, J. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 664-668.
151. Eguchi, A.; Akuta, T.; Okuyama, H.; Senda, T.; Yokoi, H.; Inokuchi, H.; Fujita, S.; Hayakawa, T.; Takeda, K.; Hasegawa, M.; Nakanishi, M., J. *Biol. Chem.* **2001**, *276*, 26204-26210.
152. Torchilin, V. P.; Rammohan, R.; Weissig, V.; Levchenko, T.S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8786-8791.
153. Lewin, M.; Carlesso, N.; Tung, C. H.; Tang, X. W.; Cory, D.; Scadden, D.T.; Weissleder, R. *Nat. Biotech.* **2000**, *18*, 410-414.
154. Schwarze, S. R.; Ho, A.; Vocero-Akbani, A.; Dowdy, S. F. *Science* **1999**, *285*, 1569-1572.
155. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell (4th ed.)*; *Garland Science* **2002**.
156. Murphy, J. E.; Uno, T.; Hamer, J. D.; Cohen, F. E.; Dwarki, V.; Zuckermann, R. N. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1517-1522.
157. Luo, D.; Saltzman, W. M. *Nat. Biotechnol.* **2000**, *18*, 33-37.
158. Herce, H. D.; Garcia, A. E. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20805-20810.
159. Herce, H. D.; Garcia, A. E. *J. Biological Physics* **2008**, *33*, 345-356.
160. Herce, H. D.; Garcia, A.E.; Litt, J.; Kane, R.S.; Martin, P.; Enrique, N.; Rebolledo, A.; Milesi, V. *Biophysical J.* **2009**, *97*, 1917-1925.
161. Vives, E.; Brodin, P.; Lebleu, B. *J. Biol. Chem.* **1997**, *272*, 16010-16017.
162. Suzuki, T.; Futaki, S.; Niwa, M.; Tanaka, S.; Ueda, K.; Sugiura, Y. *J. Biol. Chem.* **2002**, *277*, 2437-2443.
163. Wadia, J. S.; Dowdy. S. F. *Curr. Opin. Biotech.* **2002**, *13*, 52-56.
164. Richard, J. P.; Melikov, K.; Vives, E.; Ramos, C.; Verbeure, B.; Gait, M. J.; Chernomordik, L. V.; Lebleu, B. *J. Biol. Chem.* **2003**, *278*, 585-590.
165. Tyagi, M.; Rusnati, M.; Presta, M.; Giacca, M. *J. Biol. Chem.* **2001**, *276*, 3254-3261.
166. Silhol, M.; Tyagi, M.; Giacca, M.; Lebleu, B.; Vives, E. *Eur. J. Biochem.* **2002**, *269*, 494-501.
167. Bacchetti, S.; Graham, F. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1590-1594.
168. Aird, W. C. *Circ. Res.* **2007**, *100*, 158–173.
169. Aird, W. C. *Circ. Res.* **2007**, *100*, 174–190.
170. Mae, M.; Myrberg, H.; El-Andaloussi, S.; Langel, U. *Int. J. Pept. Res. Ther.* **2009**, *15*, 11-15.
171. Gregorc, V.; Santoro, A.; Bennicelli, E.; Punt, C. J.; Citterio, G.; Timmer-Bonte, J. N.; Caligaris Cappio, F.; Lambiase, A.; Bordignon, C.; van Herpen, C. M. *Brit. J. Cancer* **2009**, *101*, 219-224.

172. Eskens, F. A.; Dumez, H.; Hoekstra, R.; Perschl, A.; Brindley, C.; Bottcher, S.; Wynendaele, W.; Drevs, J.; Verweij, J.; van Oosterom, A. T. *Eur. J. Cancer* **2003**, *39*, 917-926.
173. Stupp, R.; Hegi, M. E. *J. Clin. Oncol.* **2007**, *25*, 1459-1460.
174. Laakkonen, P.; Vuorinena, K. *Integr. Biol.* **2010**, *2*, 326-337.
175. Weinshenker, N. M.; Shen, C. *Tet. Lett.* **1972**, *13*, 3281-3284.
176. Lange, U. E. W. *Tetrahedron Lett.* **2002**, *43*, 6857-6860.
177. Jones SW, Christison R, Bundell K, Voyce CJ, Brockbank SMV, Newham P, Lindsay M.A. *Brit. J. Pharmacol.* **2005**, *145*, 1093-1102.
178. Bernatowicz, M. S.; Daniels, S. B.; Köster, H. *Tet. Lett.* **1989**, *30*, 4645-4648.
179. Li, C.; Tseng, G. C.; Wong, W. H. *Statistical analysis of gene expression microarray data; CRS Press* **2003**, 1-34.
180. Keiding, N.; Morgan, B.; Speed, T.; van der Heijden, P. *Interdisciplinary Statistics Series; CRC Press* **2008**.
181. Mosiman, V. L.; Patterson, B. K.; Canterero, L.; Goolsby, C. L. *Cytometry (Communications in Clinical Cytometry)* **1997**, *30*, 151-156.
182. Lecoeur, H. *Exp. Cell Res.* **2002**, *277*, 1-14.
183. Segal, A. W. *Annu. Rev. Immunol.* **2005**, *9*, 197-223.
184. Klein, E.; Ben-Bassat, H.; Neumann, H.; Ralph, P.; Zeuthen, J.; Polliack, A.; Vánky, F. *Int. J. Cancer* **1976**, *18*, 421-431.
185. Lozzio, B.B.; Lozzio, C.B.; Bamberger, E.G.; Feliu, A.S. *Proc. Soc. Exp. Biol. Med.* **1981**, *166*, 546-550.
186. Mosmann T. *J. Immunol. Methods* **1983**, *65*, 55-63.
187. Miller, W. H.; Keenan, R. M.; Willette, R. N.; Lark, M. W., *Drug. Discov. Today* **2000**, *5*, 397-408.
188. Insel, P. A.; Tang, C-M.; Hahntow, I.; Michel, M. C. *Biochim. Biophys. Acta.* **2007**, *1768*, 994-1005.
189. Schroeder, H.; Ellinger, B.; Becker, C. F. W.; Waldmann, H.; Niemeyer, C. M. *Angew. Chem., Int. Ed.* **2007**, *46*, 4180-4183.
190. Derda, R.; Li L.; Orner, B.P.; Lewis, R. L.; Thomson, J. A.; Kiessling, L. L. *ACS Chem. Biol.* **2007**, *2*, 347-355.
191. Elefteriou, F.; Exposito, J.Y.; Garrone, R.; Lethias, C. *Eur. J. Biochem.* **1999**, *263*, 840-848.
192. Lau, D. H.; Guo, L.; Liu, R.; Song, A.; Shao, C.; Lam, K. S. *Biotechnol. Lett.* **2002**, *24*, 497-500.
193. Kuratomi, Y.; Nomizu, M.; Nielsen, P.K.; Tanaka, K.; Song, S-Y.; Kleinman, H.K.; Yamada, Y. *Exp. Cell Res.* **1999**, *249*, 386-395.
194. Makino, M.; Okazaki, I.; Kasai, S.; Nishi, N.; Bougaeva, M.; Weeks, B.S.; Otaka, A.; Nielsen, P.K.; Yamada, Y.; Nomizu, M. *Exp. Cell Res.* **2002**, *277*, 95-106.
195. Irvine, D. J.; Hue, K.A.; Meyers, A. M.; Griffith, L. G. *Biophys. J.* **2002**, *82*, 120-132.
196. Fueyo, J.; Gomez-Manzano, C.; Puduvalli, V. K.; Martin-Duque, P.; Perez-Soler, R.; Levin, V. A.; Yung, W. K.; Kyritsis, A. P. *Int. J. Oncol.* **1998**, *12*, 665-669.

197. Gomez-Manzano, C.; Fueyo, J.; Kyritsis, A. P.; Steck, P. A.; Roth, J. A.; McDonnell, T. J.; Steck, K. D.; Levin, V. A.; Yung, W. K. *Cancer Res.* **1996**, *56*, 694-699.
198. Coughlan, L.; Vallath, S.; Saha, A.; Flak, M.; McNeish, I. A.; Vassaux, G.; Marshall, J. F.; Hart, I. R.; Thomas, G. J. *J. Virol.* **2009**, *83*, 6416-6428.
199. Ridley, A. J.; Schwartz, M. A.; Burridge, K.; Firtel, R. A.; Ginsberg, M. H.; Borisy, G.; Parsons, J. T.; Horwitz, A. R. *Science* **2003**, *302*, 1704-1709.
200. Adomas, A.; Heller, G.; Olson, A.; Osborne, J.; Karlsson, M.; Nahalkova, J.; Van Zyl, L.; Sederoff, R.; Stenlid, J.; Finlay, R.; Asiegbu, F. O. *Tree Physiol.* **2008**, *28*, 885-897.
201. Baldi, P.; Hatfield, G. W. *DNA Microarrays and Gene Expression, From Experiments to Data Analysis and Modeling*, *(1st ed.); Cambridge University Press* **2004**.
202. Adessi, C.; Matton, G.; Ayala, G.; Turcatti, G.; Mermod, J-J.; Mayer, P.; Kawashima, E. *Nucleic Acids Res.* **2000**, *28*, e87.
203. Fedurco, M.; Romieu, A.; Wiiams, S.; Lawrence, I.; Turcatti, G., *Nucleic Acids Res.* **2006**, *34*, e22.
204. Khan, Z.; Poetter, K.; Park, D. J. *Anal. Biochem.* **2008**, *375*, 391-393.
205. Shapero, M. H.; Leuther, K. K.; Nguyen, A.; Scott, M.; Jones, K. W. *Genome Res.* **2001**, *11*, 1926-1934.
206. Gnirke, A.; Melnikov, A.; Maguire, J.; Rogov, P.; LeProust, E. M.; Brockman, W.; Fennell, T.; Giannoukos, G.; Fisher, S.; Russ, C.; Gabriel, S.; Jaffe, D. B.; Lander, E. S.; Nusbaum, C. *Nat. Biotech.* **2009**, *27*, 182-189.
207. Owczarzy, R.; Vallone, P. M.; Gallo, F. J.; Paner, T. M.; Lane, M. J.; Benight, A. S.; *Biopolymers* **1997**, *44*, 217-239.
208. Fan, J. B.; Chee, M. S.; Gunderson, K. L. *Nat. Rev. Genet.* **2006**, *7*, 632-644.
209. McCalla, S. E.; Luryi, A. L.; Tripathi, A. *Langmuir* **2009**, *25*, 6168-6175.
210. Hamilton, A.; Baulcombe, D. *Science* **1999**, *286*, 950-952.
211. Elbashir, S.; Harborth, J.; Lendeckel, W.; Yalcin, A.; Weber, K.; Tuschl, T. *Nature* **2001**, *411*, 494-988.
212. Hannon, G.; Rossi, J. *Nature* **2004**, *431*, 371-378.
213. Fire, A.; Xu, S.; Montgomery, M.; Kostas, S.; Driver, S.; Mello, C. *Nature* **1998**, *391*, 806-811.
214. Gartel, A.L.; Kandel, E.S. *Biomol. Eng.* **2006**, *23*, 17-34.
215. Still, W. C.; Kahn, M.; Mintra, A. *J. Org. Chem.* **1978**, *43*, 2923-2925.
216. Taylor, E. L.; Rossi, A.G.; Dransfield, I.; Hart, S. P. *Method. Mol. Biol.* **2007**, 177-200.