Design and recombinant production of combinatorial peptide libraries for gene delivery

Markus de Raad

The printing of this thesis was financially supported by:

STW Enabling new technology

UIPS Utrecht Institute for Pharmaceutical Sciences

Design and recombinant production of combinatorial peptide libraries for gene delivery

Markus de Raad

Ph.D. Thesis

Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences (UIPS),
Faculty of Science, Utrecht University, The Netherlands
November 2013

ISBN: 978-90-393-6059-0

Cover
Promotie In Zicht en Markus de Raad

Layout
Promotie In Zicht, Arnhem

Print
CPI Wöhrmann Print Service, Zutphen

Design and recombinant production of combinatorial peptide libraries for gene delivery

Ontwerp en recombinante productie van een verzameling combinatoriële peptiden voor genafgifte

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op woensdag 20 november 2013
des middags te 12.45 uur

door

Markus de Raad
geboren op 13 maart 1984
te Gouda

Promotoren

Prof.dr. D.J.A. Crommelin
Prof.dr. P.J.M. Rottier

Copromotor

Dr. E. Mastrobattista

“Here comes a delivery, straight from the heart to you”
Delivery - Babyshambles

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs (project number 10243).

Table of contents

1

Chapter

General Introduction

Introduction

Gene therapy can be defined as the introduction of exogenous nucleic acids into cells with the intention of altering gene expression to prevent, halt or reverse a pathological process. It forms an attractive approach for therapeutic intervention of a wide range of diseases, including genetic diseases, metabolic disorders, infectious diseases, chronic illnesses, cardiovascular diseases and cancer (1,2). Gene therapy can be carried out by three routes: gene addition/replacement, gene modulation/knockdown and gene correction/repair (1,3).

Gene addition/replacement therapy is used to provide therapeutic benefit or to supply a protein that is missing and is mainly applied to correct monogenic loss-of-function mutations (1,3). Other applications are suicide therapy, where the expression of a suicide gene leads to induced cell death, or the expression of ligands for surface display to activate cytotoxic T cells (1). For gene addition/replacement therapy, plasmid DNA (pDNA) or messenger RNA (mRNA) can be used.

Gene modulation/knockdown refers to the downregulation or complete inhibition of gene expression by intervening at either the transcriptional or translational level. Gene modulation/knockdown can be applied to reverse the deleterious effects caused by the abnormal expression of a mutated protein, an oncogene or a virulence factor (3). Gene modulation/knockdown can be mediated through several mechanisms. Triple-helix forming oligodeoxynucleotides (TFOs), DNA- and RNA decoys can be used for physically blocking transcription or translation (4). RNA interference (RNAi) using small interfering-RNA (siRNA) or short hairpin-RNA (shRNA)-expressing pDNA, microRNA (miRNA), ribozymes, DNAzymes and antisense RNA or DNA can be used to destabilize and destruct mRNA, thereby inhibiting translation (3-5).

Gene correction/repair is a method to restore wild type functions in dominant negative mutations and can be implemented at the mRNA level or at the genome level (3). Oligonucleotides or small DNA fragments can be used to promote mismatch repair, site-specific recombination or splice site modulation (3,5).

In order for gene-based therapeutics to become effective, the therapeutic nucleic acids must be delivered into target cells and have to reach their site of action within the cell. However, due to the high charge density and large molecular weight, nucleic acids are generally impermeable to cellular membranes and require assistance in order to reach their target site (6). To facilitate the uptake by target cells and delivery of nucleic acids at their target site, a sophisticated delivery system is required which must be capable of targeting the diseased cell, facilitate uptake and intracellular trafficking of the nucleic acid cargo to their site-of-action.

Viral and non-viral gene delivery

Gene delivery systems can be divided into two broad classes: viral vectors and non-viral (synthetic) vectors. Viral delivery systems are derived from viruses, whereas non-viral systems are based on macromolecular complexes. About 70% of all approved gene therapy clinical trials worldwide used viral vectors (7). In 2003, China became the first country to approve a viral based gene therapy product (Gendicine™) for clinical use and just recently, the EMA recommended for the first time a virus based gene therapy product for approval in the European Union (8). Viruses are highly complex and are adapted to infect cells and deliver their RNA/DNA cargo (9). The transfection efficiencies of viral vectors remain unprecedented and outperform their non-viral counterparts. However, viral vectors have several drawbacks, including their immunotoxicity and the chances for insertional mutagenesis (10). Subsequently, non-viral delivery systems have emerged as potential alternatives to viral vectors.

Most non-viral delivery systems are based on self-assembling complexes of nucleic acids with positively charged molecules, such as polymers, lipids and peptides, through electrostatic interactions (11). In general, non-viral vectors have a low immunogenicity, are nonpathogenic, and thus lack the major safety issues associated with their viral counterparts (12). Also, non-viral vectors can be produced at relatively low costs, and have no limitation on the DNA-size and carrying capacity (13). However, compared to viral vectors, the gene delivery efficiency of non-viral vectors is poor.

Peptide-based gene delivery systems

Peptide-based gene delivery systems offer advantages compared to polymer- or lipid based non-viral gene delivery systems. Both cationic polymers and lipids are known to be cytotoxic (14). Also, functionalization of polymers and lipids with targeting ligands, endosomal escape agents or polyethylene glycol (PEG) is often required in order to mediate efficient gene transfer. Reproducible incorporation and/or attachment of functional components is a significant challenge and often results in compositional variations, which is unfavorable from a pharmaceutical point of view.

Peptide-based gene delivery systems, on the other hand, may offer a versatile platform for efficient gene delivery. Peptides are biodegradable, biocompatible and various peptides have been identified that can perform several basic functions for gene delivery, such as DNA condensation or membrane disruption. By assembling different functional peptides required for effective gene delivery into a single-chain, the ideal gene delivery system can be created, thereby eliminating compositional variations, facilitate pharmaceutical formulation, and achieve reproducibility at the molecular level.

Aim

The aim of this thesis is to set up a high-throughput screening method to select out of a large library of multimodular peptides those candidates that are able to efficiently deliver therapeutic nucleic acids into target cells at their site of action. To achieve this, we propose a design strategy that follows a random, integrative approach selecting multimodular peptides containing combinations of functional traits that are optimal for efficient gene transfer (Fig.1). By randomly combining peptides with properties needed for gene delivery (e.g. DNA condensing peptides and membrane disrupting peptides), a combinatorial library encoding multimodular peptides will be generated. This library will be used for the recombinant production and screening of these multimodular peptides for their transfection efficiency. Several rounds of screening and selection will be performed to obtain multimodular peptides for effective gene delivery.

We opt for recombinant production of such multimodular peptides for several reasons. First, synthesis of long synthetic peptides of over >50 amino acids is difficult and often results in low yields and high heterogeneity (15,16). Second, by using recombinant protein synthesis technology, the precision of the cellular machinery can be exploited to reproducibly create an exactly specified protein-based material (17).

Outline

Chapter 2 provides an overview of functional peptides used for gene delivery, either individually or in multifunctional/multi-component vectors. The chapter is concluded with a discussion on why an integrative approach is beneficial for the generation of new successful non-viral gene delivery systems.

The development of a method for “scarless” ligation of multipart gene segments in a truly sequence-independent fashion is described in **Chapter 3**. This method is based on the ligation of single-stranded or double-stranded oligodeoxynucleotides (ODN) and PCR products immobilized on a solid support. Different settings were tested to optimize the solid-support ligation.

In **Chapter 4**, the development of a high-content screening (HCS) assay for rapid screening of non-viral gene delivery systems is described. For a proof-of-principle, a small library of peptide-based transfectants was simultaneously screened for transfection efficiency, cytotoxicity, induction of cell permeability and the capacity to transfect non-dividing cells.

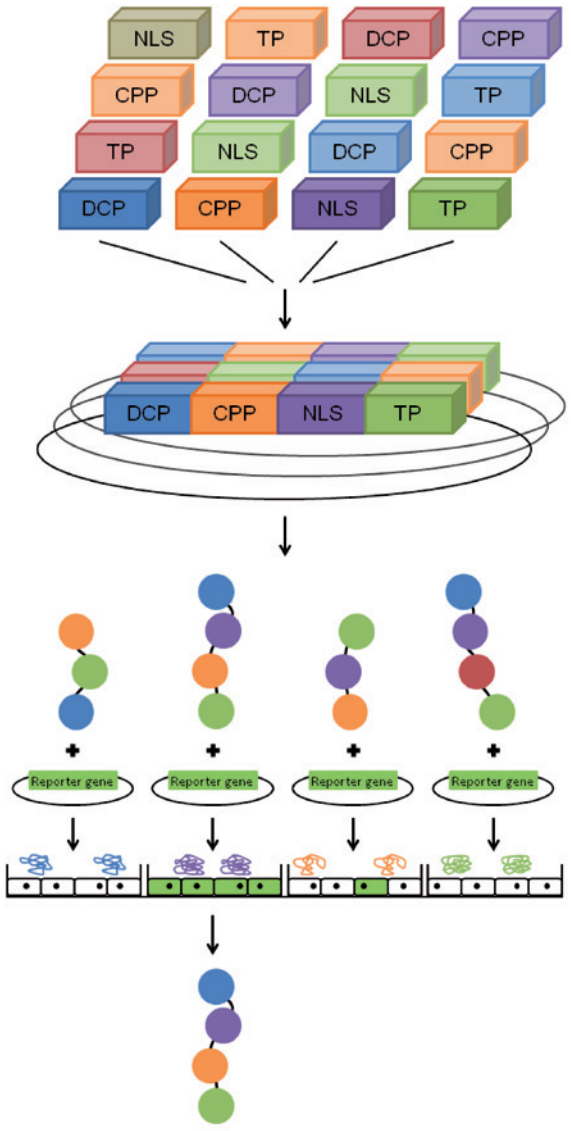


Figure 1 The proposed random, integrative approach that selects for optimal combinations of functional peptides for efficient gene transfer using combinatorial protein engineering. Functional domains (DNA condensing peptides (DCP), cell penetrating peptides (CPP), NLS peptides (NLS) and targeting peptides (NP)) can be randomly combined on a genetic level to generate a combinatorial gene library encoding for multimodular peptide vectors. The gene library can then be inserted into an expression vector and expressed in a suitable host organism. After subsequent high-throughput expression and purification, the multimodular peptide vectors can directly be screened and most efficient candidates selected.

The generation of a genetic library encoding multimodular peptides is reported in **Chapter 5**. Individual multimodular peptides were obtained after recombinant expression and purification of the generated library.

In **Chapters 6 and 7**, the multimodular peptide libraries were fused to a fusion protein in order to increase the recombinant protein yield of the individual constructs. **Chapter 6** describes the generation of targeted multimodular peptide libraries for gene delivery by fusing of the anti epidermal growth factor receptor (EGFR) biparatopic nanobody to the multimodular peptide library. **Chapter 7** describes the generation of enhanced green fluorescent protein (EGFP) fused multimodular peptide libraries for gene delivery.

Chapter 8 summarizes this thesis and discusses the findings and conclusions.

References

(1) Kay MA. State-of-the-art gene-based therapies: The road ahead. Nature Reviews Genetics 2011;12(5): 316-328.

(2) Kaiser J. Gene therapists celebrate a decade of progress. Science 2011;334(6052):29-30.

(3) Hsu CYM, Uludag H. Nucleic-acid based gene therapeutics: Delivery challenges and modular design of nonviral gene carriers and expression cassettes to overcome intracellular barriers for sustained targeted expression. J Drug Target 2012;20(4):301-328.

(4) Opalinska JB, Gewirtz AM. Nucleic-acid therapeutics: Basic principles and recent applications. Nature Reviews Drug Discovery 2002;1(7):503-514.

(5) Eckstein F. The versatility of oligonucleotides as potential therapeutics. Expert Opinion on Biological Therapy 2007;7(7):1021-1034.

(6) Mann A, Thakur G, Shukla V, Ganguli M. Peptides in DNA delivery: current insights and future directions. Drug Discov Today 2008 Feb;13(3-4):152-60.

(7) Gene Therapy Clinical Trials Worldwide. Available at: http://www.wiley.com//legacy/wileychi/genmed/clinical/.

(8) Wirth T, Parker N, Ylä-Herttuala S. History of gene therapy. Gene 2013;525(2):162-169.

(9) Jang JH, Lim KI, Schaffer DV. Library selection and directed evolution approaches to engineering targeted viral vectors. Biotechnol Bioeng 2007 Oct 15;98(3):515-24.

(10) Mingozzi F, High KA. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 2011 May;12(5):341-355.

(11) Glover DJ, Lipps HJ, Jans DA. Towards safe, non-viral therapeutic gene expression in humans. Nat Rev Genet 2005 Apr;6(4):299-310.

(12) Douglas KL. Toward development of artificial viruses for gene therapy: a comparative evaluation of viral and non-viral transfection. Biotechnol Prog 2008 Jul-Aug;24(4):871-83.

(13) De Laporte L, Cruz Rea J, Shea LD. Design of modular non-viral gene therapy vectors. Biomaterials 2006 Mar;27(7):947-54.

(14) Canine BF, Hatefi A. Development of recombinant cationic polymers for gene therapy research. Adv Drug Deliv Rev 2010 Dec 30;62(15):1524-1529.

(15) Corradin G, Kajava AV, Verdini A. Long synthetic peptides for the production of vaccines and drugs: A technological platform coming of age. Science Translational Medicine 2010;2(50).

(16) Raibaut L, Ollivier N, Melnyk O. Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 2012;41(21):7001-7015.

(17) Dimarco RL, Heilshorn SC. Multifunctional materials through modular protein engineering. Adv Mater 2012;24(29):3923-3940.

2

Chapter

Peptide Vectors for Gene Delivery: From Single Peptides to Multifunctional Peptide Nanocarriers

Markus de Raad[a], Erik A. Teunissen[a] and Enrico Mastrobattista[a]

[a] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

Submitted for publication

Abstract

In order to deliver nucleic acids into target cells for therapeutic benefit, the gene delivery system must overcome several extracellular and intracellular barriers. It should thus possess several essential characteristics, including DNA protection in the extracellular environment, capacity to bind to and be internalized by cells, endosomal escape and release of nucleic acids within the right intracellular compartment (i.e. cytosol for mRNA or siRNA or nucleus for pDNA). Naturally derived or synthetic peptides offer an attractive platform for non-viral gene delivery, as several functional peptide classes exist capable of performing these tasks. However, none of these functional peptide classes contain all the essential characteristics required to overcome all of the cellular barriers associated with successful gene delivery.

Combining functional peptides into novel multifunctional peptide vectors will be pivotal for improving peptide based gene delivery systems. By using combinatorial strategies and high-throughput screening, the identification of novel multifunctional peptide vectors will without doubt accelerate the optimization of peptide-based gene delivery systems.

Barriers in gene delivery

The majority of human gene therapy trials currently under investigation rely on viral vectors. Viruses are highly complex systems and are adepts at infecting host cells (1). They evolved numerous strategies to survive the extracellular matrix, cross cellular membranes and deliver their genetic material into the appropriate cellular compartment (2). Through these mechanisms, viral vectors are very effective in gene delivery and transient gene expression.

However, viral vectors have several drawbacks, including immunogenicity, limited nucleic acid carrying capacity, broad tropism and, in case of retroviral and lentiviral vectors, the risk of insertional mutagenesis (3,4). These limitations have warranted the parallel development of non-viral gene delivery systems.

In order to deliver therapeutic nucleic acids into the target cell, the gene delivery system must overcome several extracellular and intracellular barriers (Fig. 1). Directly after either local or intravenous administration, the gene delivery system is exposed to a harsh environment. Serum nucleases rapidly attack and degrade exogenous nucleic acids. Blood components, serum proteins, antibodies, extracellular matrix and cells may interact with the vector, which can affect its functional integrity. Also, these conditions can result in aggregation and/or opsonization of the vector, which can alter its biodistribution and can lead to the clearance by the mononuclear phagocyte system (MPS) (5).

For intravenous administration, the gene delivery system has to extravasate from the bloodstream in order to reach the specific target cells. Extravasation from the vasculature is dependent on the size of the vector and permeability of the endothelium (6). After local administration or extravasation from the vasculature, the vector has to penetrate into the tissue to reach its target cells. Movement through the extracellular matrix is mediated either by convection or by passive diffusion and can be slowed down by non-specific interactions with the extracellular matrix or high interstitial fluid pressure (7).

In order to allow cellular uptake, the gene delivery system has to bind to the cell surface. Many gene delivery systems have a net positive surface charge, which allows for the adsorption onto negatively charged cellular membranes. This type of binding is, however, nonspecific and cell-specific targeting is prohibited.

After reaching and binding to the target cell membrane, the vector should mediate direct cellular entry or has to be internalized. As the plasma membrane of mammalian cells restricts the direct passage of large, hydrophilic or charged particles, endocytosis is the main route of cellular entry (8). Thereafter, the internalized vector is entrapped in endocytic compartments, from which it must escape in order to deliver its nucleic acid cargo at the intended intracellular site of action. Otherwise, the vector stays entrapped in the endocytic compartments or can be degraded

through lysosomal degradation (6,9). After managing the escape from lysosomal degradation, the gene delivery system is released in the cytoplasm.

If the nucleic acid cargo requires nuclear delivery, the delivery system has to cross the cytoplasm in order to reach the nucleus. The mobility rate of large particles in the cytosol is dramatically reduced due to the high viscosity, through interactions with cytosolic components and by the sieve-like effect of the filamentous structures (10). Dissociation of the delivery system at this stage might increase the mobility and allows further transport of the released nucleic acids. If the target of the nucleic acid cargo is located in the cytoplasm, dissociation from the vector is also required to access their targets. However, nucleic acids are largely immobile in the cytosol and have a half live of 50-90 minutes in this compartment, primarily due to cytosolic nucleases (11).

For nuclear delivery, the free or vector-associated nucleic acids have to gain entry into the nuclear compartment. The nucleus is surrounded by the nuclear envelope, which consists of a double membrane interrupted by large integral protein structures, the nuclear pore complexes (NPC) (12,13). Molecules up to 60kDa and DNA fragments <250bp can freely diffuse into the nucleus via the NPC (14). However, larger molecules or particles (up to 40nm in size), can only enter the nucleus via shuttle molecules, which requires specific interaction with the NPC and is energy-dependent (15). Alternatively, during mitosis the nuclear membrane disassembles and the nucleus becomes easier accessible. Since most cells in our body only slowly divide or have reached a differentiation state beyond which cell division does not take place, active transport is often required for the nuclear entry of nucleic acids.

Besides cellular barriers, immunogenicity of non-viral vectors can form an obstacle. Initially, non-viral vectors were thought to be non-immunogenic compared to viral vectors. Recently, studies show that systemic injection of certain non-viral vectors can induce innate immune response and can under specific conditions cause tissue damage (16).

To overcome the barriers faced in gene delivery, non-viral vectors are required that have several essential characteristics: (i) they should condense or entrap the nucleic acid cargo into nanoparticles in the size range of 10-250nm to provide stable transport and protect it against extracellular and intracellular nucleases, (ii) they should be able to bind specifically or nonspecifically to the intended target cells, (iii) they should mediate cellular entry and/or endosomal escape, (iv) they should release the nucleic acid cargo in the targeted subcellular domain, and (v) they should be poorly immunogenic and should minimize systemic and cellular toxicity (16-19).

Naturally derived or synthetic peptides offer an attractive platform for non-viral gene delivery. Peptides are relatively stable, easy to produce and display a low cytotoxicity

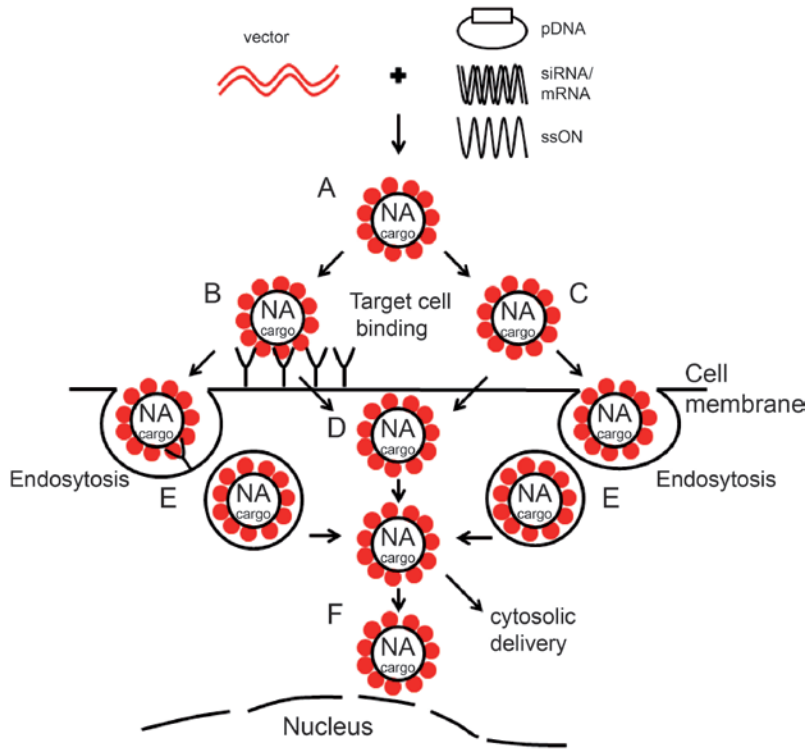


Figure 1 A schematic overview of the cellular barriers to gene delivery. The gene delivery system must be able to (A) tightly condense the nucleic acid cargo into nano-sized particles and protect it against degradation; bind (B) specifically or (C) nonspecifically to the intended target cells; either (D) facilitate direct cellular entry or (E) be internalized and induce escape from endosomolytic vesicles; and in case of nuclear delivery (F) carry the nucleic acid cargo into the nucleus.

and immunogenicity profile (20-22). Cationic peptides rich in basic residues such as lysine and/or arginine (e.g. P-1 derived from salmon protamine) are able to efficiently bind and condense nuclei acids into small, nano-sized particles (23). Peptide ligands (e.g. SP5.2 binds selectively to the VEGF receptor) can be used for targeting purposes (24). Fusogenic peptides (e.g. HA2 derived from influenza virus) are able to cross cellular membranes or provide endosomal escape and peptides containing nuclear localization signals (e.g. SV40 derived from simian virus 40) can be utilized for transport of the nucleic acid cargo to the nucleus (25,26). In this context, peptides are defined as polypeptide chains of less than 50 amino acids or 5 kDa molecular weight; larger polypeptides will be referred to as proteins.

Furthermore, peptides can be modified to enhance their gene delivery capacities. Incorporating or replacing natural amino acids by non-natural amino acids (e.g. D-amino acids or N-methylated amino acids) improves the metabolic stability of peptides by decreasing their susceptibility towards proteases (27,28). Branching of the peptide backbone can introduce extra flexibility in the molecule for e.g. improved peptide-nucleic acid cargo interactions (29,30). Several N- or C-terminal modifications have shown to improve peptide characteristics. C-terminal cysteamidation improves stability by decreasing the susceptibility towards proteases, N-terminal acetylation and the addition of stearic acid or cholesterol improves stability/formation of peptide-nucleic acid complexes (31). And in case of targeting peptides, cyclization can improve the affinity of the peptide towards the target.

DNA condensing peptides

The formation of stable nucleic acid/peptide complexes is vital for successful and efficient gene delivery, as condensation is essential for nuclease protection and stabilization in the extracellular media.

Peptides comprising multiple basic residues are able to interact with the negatively charged phosphate backbone of nucleic acids through electrostatic interactions. Also hydrophobic interactions are involved in the condensation and stabilization process. Through these electrostatic and hydrophobic interactions, pDNA and mRNA can be condensed into small particles. A minimum of 6-8 positive charges per peptide are required to condense pDNA, whereas peptides with 13 or more positive charges are able to condense pDNA tightly into nano-sized particles, which are more stable in the presence of serum and protect the pDNA cargo against degradation by endonucleases (32). However, smaller nucleic acids (>300 bp's), such as siRNAs and single-stranded oligonucleotides (ssONs), interact more disorderly with cationic peptides and can only complex the nucleic acids, which could result in excessive size and poor stability of the formed complexes (33,34). Besides condensing nucleic acids into small particles, the net positive charge exhibited by the nucleic acid/peptide complexes enables interactions with cellular membranes and internalization into the cell (18).

Many different cationic peptides have been studied for nucleic acid condensation, all involving variations in the use and number of lysines and arginines. The simplest, non-natural DNA condensing peptides are polylysines and polyarginines, usually with a length between 6 and 20 residues. Other non-natural DNA binding peptides are mostly variations on the polylysine and polyarginine peptides.

Examples are the lysine-based peptides with the sequence YKAK$_n$WK, where the central lysine cluster varies between 4-12 residues or branched peptides with different number and type of cationic amino acids ((R$_n$)$_2$KGGC or (K$_n$)$_2$KGGC) (29,35).

In living organisms, DNA is usually found associated with cationic proteins, as thermodynamically stable, tight complexes (36). Besides synthetic cationic peptides, peptides derived from cationic proteins are used for nucleic acid condensation. Notably, peptides derived from protamines, histones and cationic domains of viral capsid proteins are able to condense DNA (23,37,38).

On the other hand, strong condensation can lead to incomplete dissociation of the complexes, which hinders the nucleic acid cargo in its therapeutic function. A fine balance between complex formation and release is required. Strong condensation is favorable for stability but could be irreversible, whereas weak condensation is favorable for reversible dissociation but could cause instability and premature disintegration of the complexes. By introducing controlled reversibility via cross-links, these problems can be circumvented. For instance, McKenzie et al. synthesized a series of lysine-containing peptides with 1-4 cysteine residues (39,40). After pDNA condensation, the cysteine residues can form interpeptide disulfide bonds, which prevents the pDNA from dissociating and leads to smaller nanoparticles (39,40). Once internalized, the disulfide bonds can be reversed by the reducing environment within the cell cytoplasm and allows the pDNA to dissociate.

Due to the net positive charge exhibited by the complexes, they lack the ability to target specific cells. The net positive charge can also induce toxicity by nonspecific interactions with anionic proteins and molecules in the bloodstream and/or in the cell. Shielding of the positive charge of the complex by the addition of flexible, hydrophilic polymers such as polyethylene glycols (PEGs) can reduce nonspecific interactions, (21). However, decoration of the DNA condensing peptides with a steric hydrophilic layer requires chemical modifications and can hinder nucleic acid condensation.

There are many DNA condensing peptides that are able to condense nucleic acids into compact, stable particles and protect its cargo from physical or enzymatic degradation. However, they are unable to facilitate effective gene delivery *in vitro* and *in vivo*. Only chemically modified polyarginines (stearyl-R$_8$, stearyl-(RxR)$_4$ and chol-R$_9$) were able to mediate pDNA, splice-correcting ONs (scONs) and siRNA delivery *in vitro* and *in vivo* (41-43). Although positive charges in peptides are required for nucleic acid condensation into stable particles, these charges alone are not sufficient for efficient gene delivery and transfection.

Cell Penetrating peptides

In order to deliver their nucleic acid cargo inside the target cell, peptides should mediate direct cellular entry via membrane translocation or facilitate endosomal escape after endocytosis. Cell-penetrating peptides (CPP) are a class of peptides with

membrane translocating properties. In the past twenty years many different CPP have been studied, including virus-derived peptides (e.g. INF-7 & E^{rns} peptide), anti-microbial peptides (e.g. Bactenecin 7 & Buforin 2) and synthetic mimics (e.g. LAH4 & ppTG20) (44-49).

There is still no unambiguous definition of the family of CPP and they can be divided in several different sub-classes, based on their functioning, origin or structure. For gene delivery purposes, CPP can be divided into 2 classes, endosomolytic peptides and membrane-disrupting peptides. Endosomolytic peptides are activated by the acidic environment of early and late endosomes. Once inside the endosome, these peptides can either buffer against the proton pump to cause endosomal burst or interact with the endosomal membrane to cause membrane disruption. Histidine rich peptides (e.g. H5WYG & $CH_6K_3H_6C$) are thought to be able to buffer against the influx of protons through the 'proton sponge effect' (50,51). The general idea is that the histidine imidazole groups (pKa 6.5) absorb protons, which leads to additional influx of protons into the endosome. This is accompanied by the influx of chloride ions to maintain charge neutrality and results in osmotic swelling and eventually in the rupture of the endosome, allowing the release of the peptide complex. Another strategy is the use membrane-disrupting amphipathic peptides (e.g. KALA & HA2 synthetic derivatives), which adapt an α-helical configuration at endosomal pH that can interact with the endosomal membrane, causing membrane disruption and/or pore formation (52,53).

Membrane-disrupting peptides (e.g. TAT & Penetratin) can interact with and efficiently cross biological membranes via membrane destabilization and/or pore formation (54,55). Many membrane-disrupting peptides have also been used to promote endosomal escape.

The uptake mechanism of CPP by cells is still unclear and controversial. Although cell entry via various direct penetration pathways (bypassing endocytic routes) could be possible, it is accepted that most CPP utilize endocytic pathways to gain cellular access. Especially CPP associated with cargo or nanoparticles are dependent on endocytic pathways. As a consequence, endosomal escape is required.

In order to facilitate gene delivery, the CPP and the nucleic acid cargo should be tightly associated. This can be achieved either by covalent conjugation or by non-covalent particle formation. Covalent conjugation of the CPP and its cargo yields well defined chemical entities with a known stoichiometry (56). However, the production of these conjugates is laborious, expensive and is not compatible with pDNA (56). Non-covalent complexation of nucleic acids is mediated by electrostatic and hydrophobic interactions. CPP-nucleic acid complexes suffer from instability problems, especially *in vivo*. Particle formation and stability can be improved by N-terminal acetylation, introduction of hydrophobic residues, such as stearic acid or cholesterol, and by cross-linking the CPP after complex formation (57). Complexation

also influences cellular uptake, as complex size and cargo type determine entry pathways, although CPP complexes are mostly dependent on endocytic pathways. And in order to efficiently escape endosomes, dissociation of CPP and cargo may be required for configurational adaptation.

Similar to DNA condensing peptides, CPP lack the ability to target specific cells and can induce toxicity by nonspecific interactions. Also, CPP can induce toxicity through their mechanism of action, by destabilization of and/or pore formation in the membranes of non-targeted cells.

Although the mechanism of action is still unclear, CPP display effective endosomal disruption and several different CPP have been successfully used in the delivery of pDNA, ssONs and siRNA *in vitro* and *in vivo* (31,49,58-60). However, most successful CPP have been chemically modified, mainly because most unmodified CPP do not enable efficient nucleic acid condensation (56).

Peptides for active cytosolic transport and Nuclear Localization Signal peptides

Upon escaping the endosome, the peptide-nucleic acid complexes can reach the cytosol. Depending on the site-of-action, the nucleic acids have to either reach the nucleus and cross the nuclear membrane (pDNA and ssONs) or find their cytosolic target (mRNAs, siRNAs and antisense-based ONs). The cytoplasm is a molecularly crowded environment containing organelles and soluble macromolecules with a highly structured, mesh-like cytoskeleton (12).Therefore, large macromolecules have limited diffusional mobility in the cytoplasm. Where pDNA smaller than 500bp can freely diffuse through the cytosol, particles greater than 20nm and pDNA greater than 2kb are largely immobile in the cytoplasm (12,61).

The cell uses different kinds of molecular motors to transport intracellular vesicles, organelles and large macromolecules through the cytoplasm (2,5). For transport of cargos inwards to the perinuclear region, the molecular motors (dyneins) are used. It is believed that peptide/nucleic acid complexes can be actively transported towards the nucleus by incorporation of dynein light chain association sequences (DLCAS) into the complexes, which specifically interact with the dynein molecular motor complex (62). These sequences include the consensus motifs KSTQT, GIQVD and SKCSR (63). The use of DLCAS enhanced the nuclear accumulation of the rabies virus P-protein and can promote the rate and amount of gene expression of pDNA delivered by a nanocarrier (63,64).

After reaching the nucleus, the nucleic acids have to cross the nuclear envelope (NE) that acts as a very efficient barrier to macromolecular entry into the nucleus. Transport into the nucleus occurs either by passive diffusion or by active transport

through nuclear pore complexes (NPC). Only molecules <45kDa or DNA molecules <250bp can passively diffuse through NPC channels (2). The transport of larger molecules, like pDNA and particles, across the NPC is mediated through a process that generally involves the recognition of specific signal peptides, called nuclear localization signals (NLS) (16).

To facilitate transport through the cytosol and nuclear targeting, many NLS peptides have been studied. NLS peptides can be recognized by and bind to cytoplasmic receptors called importins and via binding of the importins to the NPC, the NLS can be translocated into the nucleus through the NPC in a GTP-dependent manner (12).

Different classes of NLS sequences exist and can be subdivided into classical and non-classical sequences (13). Classical NLS peptides are characterized by a short stretch of basic, charged amino acids. Classical NLS sequences can generally be divided into monopartite peptides, which have one cluster of basic amino acids, and bipartite peptides, which contain two clusters of basic amino acids separated by 10-12 neutral amino acids (30). The best understood and most frequently used NLS peptide is derived from the large tumor antigen of the simian virus 40 (SV40). This is a monopartite NLS comprising a single stretch of basic amino acids with the sequence PKKKRKV and is able to increase the speed of nuclear uptake of pDNA by up to 100-fold (65,66). Examples of bipartite NLS sequences are KRPAATKKAGQAKKKK, a NLS derived from the *Xenopus* protein nucleoplasmin and the nucleolin NLS, KRKKE-MANKSAPEAKKKK (30).

Non-classical NLS sequences lack the stretch of basic amino acids and consists of charged/polar residues that are interspersed with non-polar residues, like the 38-residue M9 NLS sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY derived from the human mRNA-binding protein hnRNP A1 and the yeast MATα2 homeodomain-containing derived NLS sequences NKIPIKDLLNPQ and VRILESWFAKNI (12,30).

Recruitment of the nuclear import machinery to facilitate efficient nuclear uptake of nucleic acids or complexes into the nucleus requires direct or indirect coupling of NLS peptides to the nucleic acids (8,13). Coupling of NLS peptides can be achieved through electrostatic interactions through basic residues of the NLS peptides. NLS peptides can also be covalently attached to the nucleic acids via chemical linkage, although the chemical modifications of the nucleic acids can lead to loss of therapeutic activity, due to transcriptional inactivation.

However, NLS-nucleic acid complexes formed through electrostatic interactions could only modestly enhance transfection (12,67). This could be caused by the interactions of the NLS peptides with the nucleic acid cargo, which makes the NLS unavailable for binding to importins (12). Also, covalent coupling of NLS peptides to nucleic acids resulted in modest or no enhancement of the transfection efficiency (12,67,68).

Besides active transport mediated by NLS peptides, direct entry of the nucleus by complexes has been reported through membrane fusion or pore formation (16). Also, nuclear import can take place indirectly through processes associated with mitotic events, whereby transient breakdown of nuclear membrane temporarily removes the physical barrier, allowing complexes or nucleic acids into the nucleoplasm (16).

Although NLS peptides can enhance nuclear delivery of nucleic acids, the NLS sequence alone is not sufficient for efficient gene delivery and transfection.

Targeting peptides for specific cell binding

The specificity of delivery systems for the target site can be increased by targeted delivery.

In the field of gene delivery, 'active targeting' is defined as any strategy that increases the accumulation at sites of diseases through a modulation of the biodistribution of the nucleic acid cargo via the attachment of ligands to the surface of nanocarriers (69,70). Generally, the active targeting functionality is based on ligand-receptor interactions, where ligands attached to the surface of the delivery system bind to proteins or cell surface receptors overexpressed on the plasma membrane of diseased cells (71). Normally, receptor-ligand interactions are involved in various processes, including viral entry and cellular signaling.

Besides improving target specificity and increasing accumulation at sites of disease, receptor-ligand interactions can be utilized for receptor mediated uptake and internalization of nanocarriers (2).

Many peptides can function as targeting ligands to specific cell and/or tissue types, such as cancer cells, tumor vasculature and neuronal cells (72-75). Targeting peptides can be short sequences, which are the essential amino acids needed for receptor recognition known from nature, or *de novo* synthetic peptide ligands, which have been identified by phage display or combinatorial peptide libraries (76). The most studied targeting peptides are peptides containing the RGD motif, which binds to αvβ3 integrins which play a role in the angiogenesis, invasion and metastatic activity of solid tumors. Numerous other targeting peptides have been identified, including GE7 (NPVVGYIGERPQYRDL) which targets the epidermal growth factor (EGF), LTVSPWY which targets the Human Epidermal Growth Factor Receptor 2 (HER2) and CGNKRTRGC which targets tumor lymphatics (30,71,77).

Cyclization can improve the affinity of the peptide towards the target receptor in comparison to the linear peptide. For example, cyclic RGD has a higher affinity for the binding pocket of the integrin receptor compared to the linear RGD, although cyclization requires chemical modification (78). Another strategy to increase the binding strength of targeting peptide is multimerization. A dimeric ligand can have a higher binding avidity compared to the monomeric ligand. This is the result of a

higher local concentration of ligands because of the direct linkage between two moieties and/or by binding two adjacent receptors, increasing the avidity (77). The use of multimeric RGD peptides increased the avidity for cells through multiple interactions with the integrin receptor on the cell surface compared to the monomeric peptide, where the tetra-RGD moieties had the highest avidity (79).

Recently, a new class of targeting peptides has been discovered, namely tumor-penetrating peptides. This class of peptides combines the homing ability of targeting peptides with cellular penetration capacity of CPP. Examples are the iNGR and the F3 peptide which targets tumor vasculature and the iRGD peptide which binds to αvβ3 integrin (80-82). The peptides contain the cell penetration sequence CendR (R/KXXR/K) which increases the penetration of peptides and nanoparticles into tissues through the engagement of neuropilin- 1 (NRP) (82). First, the peptides bind to their specific target. Then, after proteolytic cleavage, the CendR motif is exposed and enables internalization via NRP-1.

For gene delivery purposes, targeting peptides need to be associated to the nucleic acid cargo. Condensation of the nucleic acids through electrostatic interactions is only possible if the peptides contain basic residues. Otherwise covalent conjugation of the nucleic acid cargo to the targeting peptide is required. Also, the complexes or conjugates formed should protect the nucleic acid cargo against degradation, which can be problematic with low affinity between peptides and nucleic acids.

After binding to the target cell, the complexes require internalization and escape from endocytic compartments in order to deliver their nucleic acid payload inside the cell. Internalization can be accomplished by selecting a target which facilitates receptor-mediated internalization upon binding. However, receptor-mediated uptake is limited by the amount of receptors present on cell surface (16).

Although tumor-penetrating peptides are able to deliver nanoparticles inside cells *in vitro* and *in vivo*, most targeting peptides lack these abilities (80,83).

Improving peptide mediated gene delivery: towards multifunctional peptide constructs

So far, the utilization of individual functional peptides for gene delivery has resulted in limited clinical success. The cationic peptide CL22 has been tested in a phase I/II trial for transfection of human dendritic cells with pDNA encoding melanocyte differentiation antigens, as a therapeutic vaccine for patients with metastatic melanoma (84). However, chloroquine was added during incubation of the complexes with cells to boost transfection (84).

None of these functional peptide classes contain all the essential features required to overcome all of the cellular barriers associated with successful gene delivery.

In order to improve the gene delivery efficiency of functional peptides, multiple functionalities should be combined into a single peptide vector. Several different approaches have been used to create multifunctional peptides, from simply fusing 2 peptides together to the creation of recombinant multifunctional peptide constructs. Each individual peptide domain in the bi-, tri- or multifunctional vector is responsible for overcoming a particular cellular barrier (e.g. endosomal escape or nuclear trafficking), but together, the domains function synergistically to direct the efficient delivery of the nucleic acid cargo to the target cell.

The most straightforward multifunctional constructs are bi-functional peptide vectors, where 2 peptide domains are covalently linked together (Table 1). For example, in the MPG peptide, a CPP peptide derived from the glycine-rich region of the fusion sequence of viral gp41 (GALFLGFLGAAGSTMGA) and a NLS peptide derived from the SV40 large T antigen (PKSKRKV) were linked via a peptide linker (WSQ) (85).

Table 1 Bi- and tri-functional peptide constructs

Bi-functional peptide constructs					
Construct name	**DNA condensing**	**CPP**	**NLS**	**Targeting**	**ref**
MPG peptide	-	gp41	SV40	-	(58,85)
TAT-Mu	Mu	TAT	-	-	(87)
TM	Mu	TAT	-	-	(88)
Mu-Mu	Mu_2	-	-	-	(87)
RVG-9R	R_9	-	-	RVG	(73)
PKKKRKV-R8	R_8	-	SV40	-	(89)
KC20	K_4H	KALA (cross-linkable)	-	-	(90)
KC23	K_6H	KALA (cross-linkable)	-	-	(90)
Tri-functional peptide constructs					
Construct name	**DNA condensing**	**CPP**	**NLS**	**Targeting**	**ref**
RC26	K_4H	KALA (cross-linkable)	-	RGD	(90)
RC29	K_6H	KALA (cross-linkable)	-	RGD	(90)
DBV	4HP	GALA	-	HER2 targeting motif	(86)
TNM	Mu	TAT	SV40 (3x)	-	(91)

A tri-functional peptide vector was constructed by Wang et al., containing the fusogenic peptide GALA, a DNA condensing motif built from 4 repeating histone H2A peptides and a HER2 targeting peptide (Table 1, (86)).

Besides linking 2 or 3 peptide together, bi- and tri-functional protein-peptide fusion constructs have been created, mostly containing a protein targeting motif (Table 2). For example, $(SPKR)_4$INV, in which INV (C-terminal domain of the *Yersinia pseudotuberculosis* invasin protein) targets β1 integrins and $SPKR_4$ is used for pDNA condensation. Another example is TMAF, a trifunctional protein-peptide fusion construct containing HER2-binding affibody for targeting HER2 receptor overexpressing cells, the Mu peptide for nucleic acid condensation and the TAT peptide for membrane translocation (88).

Table 2 Bi- and tri-functional protein-peptide fusion constructs

Bi-functional protein-peptide fusion constructs					
Construct name	**DNA condensing**	**CPP**	**NLS**	**Targeting**	**ref**
protein 249AL[a]	K_{10}	-	-	FMDV	(65)
GH_c	GAL4	-	-	tetanus toxin HC	(92)
TH_c	-	Diphtheria Toxin TD	-	tetanus toxin HC	(92)
cKH-FGF2	$(KKKHHHHKKK)_6$	-	-	FGF2	(93)
dKH-FGF2	$(KHKHKHKHK)_6$	-	-	FGF2	(93)
TG	GAL4	TAT	-	-	(94)
$(SPKR)_4$ inv protein	$(SPKR)_4$	-	-	Inv	(95)
EG	-	Pseudomonas exotoxin A	-	scFv against ErbB2	(96)
5G	GAL4	-	-	scFv against ErbB2	(96)
MAF	Mu	-	-	HER2 affibody	(88)
LF254–GAL4	GAL4	-	-	LF254	(97)
PBK10	K10	Ad5 penton protein	-	-	(98)
Hph-1-GAL4	GAL4	Hph-1	-	-	(99)
F105-P	protamine	-	-	Fab against HIV-1	(100)
C9LZ-P	protamine-1 peptide	-	-	EpCAM-binding DARPin	(101)
TAT-DRBD	(ds)RNA-binding domain	TAT	-	-	(102)
NLS-HI	Human histone HI	-	SV40	-	(25)
scFv-tP	truncated protamine	-	-	scFv against avian influenza virus	(103)

Table 2 Continued

Tri-functional protein-peptide fusion constructs					
Construct name	**DNA condensing**	**CPP**	**NLS**	**Targeting**	**ref**
NLSCt[a]	K_{10}	-	SV40	FMDV	(65)
TMAF	Mu	TAT	-	HER2 affibody	(88)
5EG	GAL4	Pseudomonas exotoxin A	-	scFv against ErbB2	(96)
TEG	GAL4	Pseudomonas exotoxin A	-	TGF-a (1-50)	(104)
HerPBK10	K10	Ad5 penton protein	-	heregulin-α	(98)
GD5	GAL4	DT-T	-	scFv against ErbB2	(105)
GTH_c	GAL4	Diphtheria Toxin TD	-	tetanus toxin HC	(92)
P-DT-NLS	Human protamine 8-29	DT 198–384	opT-NLS	-	(106)
P-NLS-MSH	Human protamine 8-29	-	opT-NLS	α-MSH	(106)
P-DT-MSH	Human protamine 8-29	DT 198–384	-	α-MSH	(106)

[a]Functional peptides fused to E. coli β-galactosidase

Multifunctional peptide or peptide-protein fusion vectors containing 4 or more different functionalities to facilitate gene delivery, have also been created (Table 3). For example, Canine et al. created a multifunctional peptide vector, containing the fusogenic peptide GALA, the DNA condensing and endosomolytic peptide DCE, the NLS M9 and a HER-2 targeting peptide (107).

Table 3 Multifunctional peptide/peptide-protein constructs

Multifunctional peptide/peptide-protein constructs					
Construct name	**DNA condensing**	**CPP**	**NLS**	**Targeting**	**ref**
P-DT-NLS-MSH	Human protamine 8-29	DT 198–384	opT-NLS	α-MSH	(106)
multi-domain biopolymer	DCE $(RRXRRXHHXHHX)_3$	GALA	M9	HER2 targeting motif	(107)
DBV	Mu	GALA	NLS HIV-1	cyclic targeting peptide	(108)

Although almost all bi-, tri- and multifunction peptide and protein-peptide fusion constructs were able to transfect cells *in vitro*, only 5 constructs demonstrated *in vivo* delivery of nucleic acids, namely Hph-1-GAL4, F105-P, TAT-DRBD, TMAF and RVG-9R (73,88,99,100,102). Also, a direct comparison of the *in vitro* transfection efficiencies of the different constructs is difficult since these experiments were performed on different cell types, using different controls, and under different experimental conditions.

Besides determining the transfection efficiency, the functionality of the separate peptide/protein modules should also be evaluated. For bi-functional constructs, especially the DNA condensing/targeting constructs, this is relatively straight forward and many investigators demonstrated the functionality of both modules. However, for tri- and multifunctional constructs the evaluation of the individual modules is more difficult. In general, most investigators evaluate the functionality of the individual modules by looking at its contribution to the total gene delivery process.

Multimodular Peptide Libraries for Gene Delivery

Results obtained with multifunctional peptide or peptide-protein constructs have demonstrated that the individual peptides can remain functional and contribute to the delivery process. However, compared to 'simple' monofunctional peptide vectors, their gene delivery capacity did not improve and most promising peptide vectors are still based on single peptides. By fusing multiple peptides into a single vector, different and new challenges arise.

Some functionalities within the vector can be conflicting. Positive charged peptides are required for nucleic acid condensation. However, the positive charges can hinder the specificity of the targeting peptide via nonspecific cell binding. Also, negatively charged CPP can influence the interactions between the nucleic acid cargo and the DNA condensing peptides.

Another difficulty is the configuration of the formed complexes. A functional peptide can be buried inside the core of the complex which could be suboptimal for its activity. For example, if the targeting peptide is buried inside the complex, binding efficiency to the target cell can strongly be reduced.

Due the lack of knowledge on the structure/activity relationship of gene delivery systems, it is difficult to rationalize beforehand which combination of the functional peptides will be optimal. Also, the optimal position (N-terminal, C-terminal or in between) of the different functional peptides in the multi-peptide chain is unknown. Rational design has been the main strategy to optimize multifunctional peptide vectors, in which often a single step is optimized without considering the effect of this optimization on other steps of this process. However, with the above mentioned challenges in mind, we feel that the design strategy should follow a more random,

integrative approach that selects for combinations of functional peptides that are optimal for efficient gene transfer. It is expected that by randomly combining functional peptides with properties needed for gene delivery, a combinatorial library of multifunctional peptide vectors can be created and screened for gene delivery efficiency. Using several screening and selection rounds, the most effective multi-functional peptide vectors can be identified.

The advantage of generating and screening large random multifunctional peptide vector libraries as opposed to a rational design strategy, is that problems with conflicting/opposing functionalities and particle configuration are circumvented and can accelerate the identification of the structure/activity relationships. Also, screening of more candidates increases the chance of identifying promising gene delivery vectors.

The random, integrative approach requires a combinatorial, modular design strategy for the generation of all possible combinations of functional peptides, and complementary high-throughput characterization for the subsequent screening (Fig.1 in Chapter 1). Genetic modular assembly strategies allow for generation of gene libraries that encode for multifunctional peptide vectors in a fast, simple, and a combinatorial way (109-111). The gene library can then be inserted into an expression vector and transformed into suitable host organism. After subsequent high-throughput expression and purification, the multifunctional peptide vectors can directly be screened, for example, on nucleic acid delivery *in vitro* using high-throughput transfection assays (112,113).

An alternative to recombinant production of multimodular peptide constructs, is peptide synthesis. Using solid phase peptide synthesis, long synthetic peptides over 120 amino acids residues in length, have been created (114). However, synthesis of long synthetic peptides over >50 amino acids is not possible for "difficult sequences", and often results in low yields and high heterogeneity (114,115). Long peptides and small proteins (up to ± 200 amino acids) can be generated by native chemical ligation (NCL), where short synthetic peptides are assembled into larger polypeptides or small proteins (115,116). Also, combinatorial assembly approaches have been designed for the preparation of protein variants (117). For the generation of multimodular peptide libraries, however, further advances in peptide synthesis and in chemical ligation are required. For example, aggregation of intermediate products, "difficult peptide sequences", low yields and large time consumption are challenges which have to be resolved in order to produce synthetic multimodular peptide libraries (114,115,117).

Conclusion

Many individual functional peptides have been utilized for gene delivery purposes. Despite the progress which has been made towards the generation of more efficient peptide vectors, several challenges remain. The development of novel multifunctional peptide vectors will be pivotal for improving peptide based gene delivery systems. By using combinatorial strategies and high throughput screening, the identification of novel multifunctional peptide vectors, optimal for efficient gene transfer will without doubt accelerate the optimization of peptide-based gene delivery systems needed for clinical use.

Abbreviations

bp: base pairs
CPP: cell penetrating peptides
mRNA: messenger RNA
NCL: natice chemical ligation
NLS: nuclear localization signal
NPC: nuclear pore complexes
ON: oligonucleotide
pDNA: plasmid DNA
PEG: polyethylene glycol
siRNA: small interfering-RNA
ss: single-stranded

References

(1) Jang JH, Lim KI, Schaffer DV. Library selection and directed evolution approaches to engineering targeted viral vectors. Biotechnol Bioeng 2007 Oct 15;98(3):515-24.
(2) Glover DJ. Artificial viruses: exploiting viral trafficking for therapeutics. Infect Disord Drug Targets 2012 Feb;12(1):68-80.
(3) Mingozzi F, High KA. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 2011 May;12(5):341-355.
(4) Thomas CE, Ehrhardt A, Kay MA. Progress and problems with the use of viral vectors for gene therapy. Nature Reviews Genetics 2003;4(5):346-358.
(5) Mastrobattista E, van der Aa MA, Hennink WE, Crommelin DJ. Artificial viruses: a nanotechnological approach to gene delivery. Nat Rev Drug Discov 2006 Feb;5(2):115-21.
(6) Read ML, Logan A, Seymour LW. Barriers to gene delivery using synthetic vectors. Adv Genet 2005;53:19-46.
(7) Burke RS, Pun SH. Extracellular barriers to in Vivo PEI and PEGylated PEI polyplex-mediated gene delivery to the liver. Bioconjug Chem 2008 Mar;19(3):693-704.
(8) Wagstaff KM, Jans DA. Nucleocytoplasmic transport of DNA: enhancing non-viral gene transfer. Biochem J 2007 Sep 1;406(2):185-202.
(9) Douglas KL. Toward development of artificial viruses for gene therapy: a comparative evaluation of viral and non-viral transfection. Biotechnol Prog 2008 Jul-Aug;24(4):871-83.
(10) Dauty E, Verkman AS. Actin cytoskeleton as the principal determinant of size-dependent DNA mobility in cytoplasm: a new barrier for non-viral gene delivery. J Biol Chem 2005 Mar 4;280(9):7823-8.
(11) Merdan T, Kopecek J, Kissel T. Prospects for cationic polymers in gene and oligonucleotide therapy against cancer. Adv Drug Deliv Rev 2002 Sep 13;54(5):715-58.
(12) Pouton CW, Wagstaff KM, Roth DM, Moseley GW, Jans DA. Targeted delivery to the nucleus. Adv Drug Deliv Rev 2007 Aug 10;59(8):698-717.
(13) van der Aa MA, Mastrobattista E, Oosting RS, Hennink WE, Koning GA, Crommelin DJ. The nuclear pore complex: the gateway to successful nonviral gene delivery. Pharm Res 2006 Mar;23(3):447-59.
(14) Talcott B, Moore MS. Getting across the nuclear pore complex. Trends Cell Biol 1999 Aug;9(8):312-8.
(15) Pante N, Kann M. Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. Mol Biol Cell 2002 Feb;13(2):425-34.
(16) Hsu CYM, Uludag H. Nucleic-acid based gene therapeutics: Delivery challenges and modular design of nonviral gene carriers and expression cassettes to overcome intracellular barriers for sustained targeted expression. J Drug Target 2012;20(4):301-328.
(17) Miyata K, Nishiyama N, Kataoka K. Rational design of smart supramolecular assemblies for gene delivery: Chemical challenges in the creation of artificial viruses. Chem Soc Rev 2012;41(7):2562-2574.
(18) Saccardo P, Villaverde A, Gonzalez-Montalban N. Peptide-mediated DNA condensation for non-viral gene therapy. Biotechnol Adv 2009 Jul-Aug;27(4):432-8.
(19) Pezzoli D, Candiani G. Non-viral gene delivery strategies for gene therapy: A “ménage à trois” among nucleic acids, materials, and the biological environment: Stimuli-responsive gene delivery vectors. Journal of Nanoparticle Research 2013;15(3).
(20) Mann A, Thakur G, Shukla V, Ganguli M. Peptides in DNA delivery: current insights and future directions. Drug Discov Today 2008 Feb;13(3-4):152-60.
(21) Sato AK, Viswanathan M, Kent RB, Wood CR. Therapeutic peptides: technological advances driving peptides into development. Curr Opin Biotechnol 2006;17(6):638-642.
(22) Hoyer J, Neundorf I. Peptide vectors for the nonviral delivery of nucleic acids. Acc Chem Res 2012;45(7):1048-1056.
(23) Kharidia R, Friedman KA, Liang JF. Improved gene expression using low molecular weight peptides produced from protamine sulfate. Biochemistry (Mosc) 2008 Oct;73(10):1162-8.
(24) El-Mousawi M, Tchistiakova L, Yurchenko L, Pietrzynski G, Moreno M, Stanimirovic D, et al. A Vascular Endothelial Growth Factor High Affinity Receptor 1-specific Peptide with Antiangiogenic Activity Identified Using a Phage Display Peptide Library. J Biol Chem 2003;278(47):46681-46691.

(25) Fritz JD, Herweijer H, Zhang G, Wolff JA. Gene transfer into mammalian cells using histone-condensed plasmid DNA. Hum Gene Ther 1996;7(12):1395-1404.
(26) Durrer P, Galli C, Hoenke S, Corti C, Glück R, Vorherr T, et al. H+-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region. J Biol Chem 1996;271(23):13417-13421.
(27) Rennert R, Neundorf I, Beck-Sickinger AG. Calcitonin-derived peptide carriers: Mechanisms and application. Adv Drug Deliv Rev 2008;60(4-5):485-498.
(28) Veldhoen S, Laufer SD, Restle T. Recent developments in peptide-based nucleic acid delivery. International Journal of Molecular Sciences 2008;9(7):1276-1320.
(29) Plank C, Tang MX, Wolfe AR, C. SF,Jr. Branched cationic peptides for gene delivery: role of type and number of cationic residues in formation and in vitro activity of DNA polyplexes. Hum Gene Ther 1999 Jan 20;10(2):319-32.
(30) Martin ME, Rice KG. Peptide-guided gene delivery. Aaps J 2007;9(1):E18-29.
(31) Lehto T, Simonson OE, Mäger I, Ezzat K, Sork H, Copolovici D-, et al. A peptide-based vector for efficient gene transfer in vitro and in vivo. Molecular Therapy 2011;19(8):1457-1467.
(32) Adami RC, Collard WT, Gupta SA, Kwok KY, Bonadio J, Rice KG. Stability of peptide-condensed plasmid DNA formulations. J Pharm Sci 1998 Jun;87(6):678-83.
(33) Gary DJ, Puri N, Won Y-. Polymer-based siRNA delivery: Perspectives on the fundamental and phenomenological distinctions from polymer-based DNA delivery. J Controlled Release 2007;121(1-2):64-73.
(34) Endoh T, Ohtsuki T. Cellular siRNA delivery using cell-penetrating peptides modified for endosomal escape. Adv Drug Deliv Rev 2009 Jul 25;61(9):704-9.
(35) Gottschalk S, Sparrow JT, Hauer J, Mims MP, Leland FE, Woo SL, et al. A novel DNA-peptide complex for efficient gene transfer and expression in mammalian cells. Gene Ther 1996 May;3(5):448-57.
(36) Aris A, Villaverde A. Modular protein engineering for non-viral gene therapy. Trends Biotechnol 2004 Jul;22(7):371-7.
(37) Schwartz B, Ivanov MA, Pitard B, Escriou V, Rangara R, Byk G, et al. Synthetic DNA-compacting peptides derived from human sequence enhance cationic lipid-mediated gene transfer in vitro and in vivo. Gene Ther 1999 Feb;6(2):282-92.
(38) Tecle M, Preuss M, Miller AD. Kinetic study of DNA condensation by cationic peptides used in nonviral gene therapy: analogy of DNA condensation to protein folding. Biochemistry 2003 Sep 9;42(35):10343-7.
(39) McKenzie DL, Kwok KY, Rice KG. A potent new class of reductively activated peptide gene delivery agents. J Biol Chem 2000 Apr 7;275(14):9970-7.
(40) McKenzie DL, Smiley E, Kwok KY, Rice KG. Low molecular weight disulfide cross-linking peptides as nonviral gene delivery carriers. Bioconjug Chem 2000 Nov-Dec;11(6):901-9.
(41) Khalil IA, Futaki S, Niwa M, Baba Y, Kaji N, Kamiya H, et al. Mechanism of improved gene transfer by the N-terminal stearylation of octaarginine: Enhanced cellular association by hydrophobic core formation. Gene Ther 2004;11(7):636-644.
(42) Lehto T, Abes R, Oskolkov N, Suhorutšenko J, Copolovici D-, Mäger I, et al. Delivery of nucleic acids with a stearylated (RxR)4 peptide using a non-covalent co-incubation strategy. J Controlled Release 2010;141(1):42-51.
(43) Yin H, Moulton HM, Seow Y, Boyd C, Boutilier J, Iverson P, et al. Cell-penetrating peptide-conjugated antisense oligonucleotides restore systemic muscle and cardiac dystrophin expression and function. Hum Mol Genet 2008;17(24):3909-3918.
(44) Sadler K, Eom KD, Yang JL, Dimitrova Y, Tam JP. Translocating proline-rich peptides from the antimicrobial peptide bactenecin 7. Biochemistry 2002 Dec 3;41(48):14150-7.
(45) Takeshima K, Chikushi A, Lee KK, Yonehara S, Matsuzaki K. Translocation of analogues of the antimicrobial peptides magainin and buforin across human cell membranes. J Biol Chem 2003 Jan 10;278(2):1310-5.
(46) Langedijk JP. Translocation activity of C-terminal domain of pestivirus Erns and ribotoxin L3 loop. J Biol Chem 2002 Feb 15;277(7):5308-14.
(47) Plank C, Oberhauser B, Mechtler K, Koch C, Wagner E. The influence of endosome-disruptive peptides on gene transfer using synthetic virus-like gene transfer systems. J Biol Chem 1994 Apr 29;269(17):12918-24.
(48) Kichler A, Leborgne C, Marz J, Danos O, Bechinger B. Histidine-rich amphipathic peptide antibiotics promote efficient delivery of DNA into mammalian cells. Proc Natl Acad Sci U S A 2003 Feb 18;100(4):1564-8.

(49) Rittner K, Benavente A, Bompard-Sorlet A, Heitz F, Divita G, Brasseur R, et al. New basic membrane-destabilizing peptides for plasmid-based gene delivery in vitro and in vivo. Mol Ther 2002 Feb;5(2):104-14.
(50) Midoux P, Kichler A, Boutin V, Maurizot JC, Monsigny M. Membrane permeabilization and efficient gene transfer by a peptide containing several histidines. Bioconjug Chem 1998 Mar-Apr;9(2):260-7.
(51) Read ML, Singh S, Ahmed Z, Stevenson M, Briggs SS, Oupicky D, et al. A versatile reducible polycation-based system for efficient delivery of a broad range of nucleic acids. Nucleic Acids Res 2005;33(9):1-16.
(52) Baru M, Nahum O, Jaaro H, Sha'anani J, Nur I. Lysosome-disrupting peptide increases the efficiency of in-vivo gene transfer by liposome-encapsulated DNA. J Drug Target 1998;6(3):191-9.
(53) Wyman TB, Nicol F, Zelphati O, Scaria PV, Plank C, C. SF,Jr. Design, synthesis, and characterization of a cationic peptide that binds to nucleic acids and permeabilizes bilayers. Biochemistry 1997 Mar 11;36(10):3008-17.
(54) Fischer PM, Zhelev NZ, Wang S, Melville JE, Fahraeus R, Lane DP. Structure-activity relationship of truncated and substituted analogues of the intracellular delivery vector Penetratin. J Pept Res 2000 Feb;55(2):163-72.
(55) Ignatovich IA, Dizhe EB, Pavlotskaya AV, Akifiev BN, Burov SV, Orlov SV, et al. Complexes of plasmid DNA with basic domain 47-57 of the HIV-1 Tat protein are transferred to mammalian cells by endocytosis-mediated pathways. J Biol Chem 2003 Oct 24;278(43):42625-36.
(56) Lehto T, Kurrikoff K, Langel Ü. Cell-penetrating peptides for the delivery of nucleic acids. Expert Opinion on Drug Delivery 2012;9(7):823-836.
(57) Lehto T, Ezzat K, Langel U. Peptide nanoparticles for oligonucleotide delivery. Progress in Molecular Biology and Translational Science 2011;104:397-426.
(58) Simeoni F, Morris MC, Heitz F, Divita G. Insight into the mechanism of the peptide-based gene delivery system MPG: implications for delivery of siRNA into mammalian cells. Nucleic Acids Res 2003 Jun 1;31(11):2717-24.
(59) El Andaloussi S, Lehto T, Mäger I, Rosenthal-Aizman K, Oprea II, Simonson OE, et al. Design of a peptide-based vector, PepFect6, for efficient delivery of siRNA in cell culture and systemically in vivo. Nucleic Acids Res 2011;39(9):3972-3987.
(60) Mäe M, EL Andaloussi S, Lundin P, Oskolkov N, Johansson HJ, Guterstam P, et al. A stearylated CPP for delivery of splice correcting oligonucleotides using a non-covalent co-incubation strategy. J Controlled Release 2009;134(3):221-227.
(61) Vázquez E, Ferrer-Miralles N, Villaverde A. Peptide-assisted traffic engineering for nonviral gene therapy. Drug Discov Today 2008;13(23-24):1067-1074.
(62) Cohen RN, Rashkin MJ, Wen X, Szoka Jr. FC. Molecular motors as drug delivery vehicles. Drug Discovery Today: Technologies 2005;2(1):111-118.
(63) Moseley GW, Roth DM, DeJesus MA, Leyton DL, Filmer RP, Pouton CW, et al. Dynein light chain association sequences can facilitate nuclear protein import. Mol Biol Cell 2007;18(8):3204-3213.
(64) Tanaka K, Kanazawa T, Sugawara K, Horiuchi S, Takashima Y, Okada H. A cytoplasm-sensitive peptide vector cross-linked with dynein light chain association sequence (DLCAS) enhances gene expression. Int J Pharm 2011 10/31;419(1–2):231-234.
(65) Aris A, Villaverde A. Engineering nuclear localization signals in modular protein vehicles for gene therapy. Biochem Biophys Res Commun 2003 May 16;304(4):625-31.
(66) Belting M, Sandgren S, Wittrup A. Nuclear delivery of macromolecules: barriers and carriers. Adv Drug Deliv Rev 2005 Feb 28;57(4):505-27.
(67) Cartier R, Reszka R. Utilization of synthetic peptides containing nuclear localization signals for nonviral gene transfer systems. Gene Ther 2002;9(3):157-167.
(68) van der Aa MAEM, Koning GA, d'Oliveira C, Oosting RS, Wilschut KJ, Hennink WE, et al. An NLS peptide covalently linked to linear DNA does not enhance transfection efficiency of cationic polymer based gene delivery systems. J Gene Med 2005;7(2):208-217.
(69) Schatzlein AG. Targeting of Synthetic Gene Delivery Systems. J Biomed Biotechnol 2003;2003(2):149-158.
(70) Petros RA, Desimone JM. Strategies in the design of nanoparticles for therapeutic applications. Nature Reviews Drug Discovery 2010;9(8):615-627.
(71) Kamaly N, Xiao Z, Valencia PM, Radovic-Moreno AF, Farokhzad OC. Targeted polymeric therapeutic nanoparticles: Design, development and clinical translation. Chem Soc Rev 2012;41(7):2971-3010.

(72) Balestrieri ML, Napoli C. Novel challenges in exploring peptide ligands and corresponding tissue-specific endothelial receptors. Eur J Cancer 2007 May;43(8):1242-50.
(73) Kumar P, Wu H, McBride JL, Jung KE, Kim MH, Davidson BL, et al. Transvascular delivery of small interfering RNA to the central nervous system. Nature 2007 Jul 5;448(7149):39-43.
(74) Shadidi M, Sioud M. Selective targeting of cancer cells using synthetic peptides. Drug Resist Updat 2003 Dec;6(6):363-71.
(75) Levine RM, Scott CM, Kokkoli E. Peptide functionalized nanoparticles for nonviral gene delivery. Soft Matter 2013;9(4):985-1004.
(76) Aina OH, Liu R, Sutcliffe JL, Marik J, Pan CX, Lam KS. From combinatorial chemistry to cancer-targeting peptides. Mol Pharm 2007 Sep-Oct;4(5):631-51.
(77) Vives E, Schmidt J, Pelegrin A. Cell-penetrating and cell-targeting peptides in drug delivery. Biochim Biophys Acta 2008 Dec;1786(2):126-38.
(78) Pierschbacher MD, Ruoslahti E. Influence of stereochemistry of the sequence Arg-Gly-Asp-Xaa on binding specificity in cell adhesion. J Biol Chem 1987;262(36):17294-17298.
(79) Dijkgraaf I, Kruijtzer JAW, Liu S, Soede AC, Oyen WJG, Corstens FHM, et al. Improved targeting of the avß3 integrin by multimerisation of RGD peptides. European Journal of Nuclear Medicine and Molecular Imaging 2007;34(2):267-273.
(80) Sugahara KN, Teesalu T, Karmali PP, Kotamraju VR, Agemy L, Girard OM, et al. Tissue-Penetrating Delivery of Compounds and Nanoparticles into Tumors. Cancer Cell 2009;16(6):510-520.
(81) Porkka K, Laakkonen P, Hoffman JA, Bernasconi M, Ruoslahti E. A fragment of the HMGN2 protein homes to the nuclei of tumor cells and tumor endothelial cells in vivo. Proc Natl Acad Sci U S A 2002;99(11):7444-7449.
(82) Alberici L, Roth L, Sugahara KN, Agemy L, Kotamraju VR, Teesalu T, et al. De Novo design of a tumor-penetrating peptide. Cancer Res 2013;73(2):804-812.
(83) Sugahara KN, Teesalu T, Prakash Karmali P, Ramana Kotamraju V, Agemy L, Greenwald DR, et al. Coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. Science 2010;328(5981):1031-1035.
(84) Steele JC, Rao A, Marsden JR, Armstrong CJ, Berhane S, Billingham LJ, et al. Phase I/II trial of a dendritic cell vaccine transfected with DNA encoding melan A and gp100 for patients with metastatic melanoma. Gene Ther 2011;18(6):584-593.
(85) Morris MC, Vidal P, Chaloin L, Heitz F, Divita G. A new peptide vector for efficient delivery of oligonucleotides into mammalian cells. Nucleic Acids Res 1997 Jul 15;25(14):2730-6.
(86) Wang Y, Mangipudi SS, Canine BF, Hatefi A. A designer biomimetic vector with a chimeric architecture for targeted gene transfer. J Controlled Release 2009;137(1):46-53.
(87) Rajagopalan R, Xavier J, Rangaraj N, Rao NM, Gopal V. Recombinant fusion proteins TAT-Mu, Mu and Mu-Mu mediate efficient non-viral gene delivery. J Gene Med 2007;9(4):275-286.
(88) Govindarajan S, Sivakumar J, Garimidi P, Rangaraj N, Kumar JM, Rao NM, et al. Targeting human epidermal growth factor receptor 2 by a cell-penetrating peptide-affibody bioconjugate. Biomaterials 2012;33(8):2570-2582.
(89) Wang H-, Chen J-, Sun Y-, Deng J-, Li C, Zhang X-, et al. Construction of cell penetrating peptide vectors with N-terminal stearylated nuclear localization signal for targeted delivery of DNA into the cell nuclei. J Controlled Release 2011;155(1):26-33.
(90) Guo XD, Wiradharma N, Liu SQ, Zhang LJ, Khan M, Qian Y, et al. Oligomerized alpha-helical KALA peptides with pendant arms bearing cell-adhesion, DNA-binding and endosome-buffering domains as efficient gene transfection vectors. Biomaterials 2012;33(26):6284-6291.
(91) Xavier J, Singh S, Dean DA, Rao NM, Gopal V. Designed multi-domain protein as a carrier of nucleic acids into cells. J Controlled Release 2009;133(2):154-160.
(92) Box M, Parks DA, Knight A, Hale C, Fishman PS, Fairweather NF. A multi-domain protein system based on the HC fragment of tetanus toxin for targeting DNA to neuronal cells. J Drug Target 2003;11(6):333-343.
(93) Canine BF, Wang Y, Hatefi A. Evaluation of the effect of vector architecture on DNA condensation and gene transfer efficiency. J Controlled Release 2008;129(2):117-123.

(94) Gao P, Li X, Liu Y, Liu Y, Kan S, Jin J, et al. Construction, expression and characterization of a chimeric multi-domain protein mediating specific DNA transfer. Protein Expr Purif 2010;74(2):189-195.
(95) Fortunati E, Ehlert E, van Loo ND, Wyman C, Eble JA, Grosveld F, et al. A multi-domain protein for beta1 integrin-targeted DNA delivery. Gene Ther 2000 Sep;7(17):1505-15.
(96) Fominaya J, Wels W. Target cell-specific DNA transfer mediated by a chimeric multidomain protein: Novel non-viral gene delivery system. J Biol Chem 1996;271(18):10560-10568.
(97) Gaur R, Gupta PK, Goyal A, Wels W, Singh Y. Delivery of nucleic acid into mammalian cells by anthrax toxin. Biochem Biophys Res Commun 2002;297(5):1121-1127.
(98) Medina-Kauwe LK, Maguire M, Kasahara N, Kedes L. Nonviral gene delivery to human breast cancer cells by targeted Ad5 penton proteins. Gene Ther 2001;8(23):1753-1761.
(99) Kim E-, Yang S-, Hong D-, Kim W-, Kim H-, Lee S-. Cell-penetrating DNA-binding protein as a safe and efficient naked DNA delivery carrier in vitro and in vivo. Biochem Biophys Res Commun 2010;392(1):9-15.
(100) Song E, Zhu P, Lee S-, Chowdhury D, Kussman S, Dykxhoorn DM, et al. Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. Nat Biotechnol 2005;23(6):709-717.
(101) Winkler J, Martin-Killias P, Plückthun A, Zangemeister-Wittke U. EpCAM-targeted delivery of nanocomplexed siRNA to tumor cells with designed ankyrin repeat proteins. Molecular Cancer Therapeutics 2009;8(9):2674-2683.
(102) Eguchi A, Meade BR, Chang Y-, Fredrickson CT, Willert K, Puri N, et al. Efficient siRNA delivery into primary cells by a peptide transduction domain-dsRNA binding domain fusion protein. Nat Biotechnol 2009;27(6):567-571.
(103) Zhang T, Wang C-, Zhang W, Gao Y-, Yang S-, Wang T-, et al. Generation and characterization of a fusion protein of single-chain fragment variable antibody against hemagglutinin antigen of avian influenza virus and truncated protamine. Vaccine 2010;28(23):3949-3955.
(104) Fominaya J, Uherek C, Wels W. A chimeric fusion protein containing transforming growth factor-a mediates gene transfer via binding to the EGF receptor. Gene Ther 1998;5(4):521-530.
(105) Uherek C, Fominaya J, Wels W. A Modular DNA Carrier Protein Based on the Structure of Diphtheria Toxin Mediates Target Cell-specific Gene Delivery. J Biol Chem 1998;273(15):8835-8841.
(106) Glover DJ, Su MN, Mechler A, Martin LL, Jans DA. Multifunctional protein nanocarriers for targeted nuclear gene delivery in nondividing cells. FASEB Journal 2009;23(9):2996-3006.
(107) Canine BF, Wang Y, Hatefi A. Biosynthesis and characterization of a novel genetically engineered polymer for targeted gene transfer to cancer cells. J Controlled Release 2009;138(3):188-196.
(108) Mangipudi SS, Canine BF, Wang Y, Hatefi A. Development of a genetically engineered biomimetic vector for targeted gene transfer to breast cancer cells. Molecular Pharmaceutics 2009;6(4):1100-1109.
(109) Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, et al. BglBricks: A flexible standard for biological part assembly. J Biol Eng 2010 Jan 20;4(1):1.
(110) de Raad M, Kooijmans SA, Teunissen EA, Mastrobattista E. A Solid-Phase Platform for Combinatorial and Scarless Multipart Gene Assembly. ACS Synth Biol 2013 Mar 15.
(111) Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A modular cloning system for standardized assembly of multigene constructs. PLoS One 2011 Feb 18;6(2):e16765.
(112) Lin C-, Moore PA, Auberry DL, Landorf EV, Peppler T, Victry KD, et al. Automated purification of recombinant proteins: Combining high-throughput with high yield. Protein Expr Purif 2006;47(1):16-24.
(113) de Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E. High-content screening of peptide-based non-viral gene delivery systems. J Control Release 2012 Mar 28;158(3):433-442.
(114) Corradin G, Kajava AV, Verdini A. Long synthetic peptides for the production of vaccines and drugs: A technological platform coming of age. Science Translational Medicine 2010;2(50).
(115) Raibaut L, Ollivier N, Melnyk O. Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 2012;41(21):7001-7015.
(116) Chandrudu S, Simerska P, Toth I. Chemical methods for peptide and protein production. Molecules 2013;18(4):4373-4388.
(117) Lindgren J, Ekblad C, Abrahmsén L, Eriksson Karlström A. A Native Chemical Ligation Approach for Combinatorial Assembly of Affibody Molecules. ChemBioChem 2012;13(7):1024-1031.

3

Chapter

A Solid-Phase Platform for Combinatorial and Scarless Multipart Gene Assembly

Markus de Raad[a], Sander A. A. Kooijmans[b], Erik A. Teunissen[a] and Enrico Mastrobattista[a]

[a] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands
[b] Department of Clinical Chemistry and Haematology, University Medical Center Utrecht, Heidelberglaan 100, 3584CX Utrecht, The Netherlands

ACS Synthetic Biology, 2 (6): 316-326 (2013)

Abstract

With the emergence of standardized genetic modules as part of the synthetic biology toolbox the need for universal and automatable assembly protocols increases. Although several methods and standards have been developed, these all suffer from drawbacks such as the introduction of scar sequences during ligation or the need for specific flanking sequences or *a priori* knowledge of the final sequence to be obtained. We have developed a method for scarless ligation of multipart gene segments in a truly sequence-independent fashion. The big advantage of this approach is that it is combinatorial, allowing the generation of all combinations of several variants of two or more modules to be ligated in less than a day. This method is based on the ligation of single-stranded or double-stranded oligodeoxynucleotides (ODN) and PCR products immobilized on a solid support. Different settings were tested to optimize the solid-support ligation. Finally, to show proof of concept for this novel multipart gene assembly platform, a small library of all possible combinations of 4 BioBrick modules was generated and tested.

Introduction

The field of synthetic biology heavily relies on robust methods to enable the construction of genes, pathways and genomes (1-5). All these methods differ in both mechanism and scale. This offers the user methods for self-assembly of many parts in a single reaction (parallel assembly), assembly of constructs with a pre-defined physical arrangement (ordered assembly) and allows the use of multiple parts simultaneously (combinatorial assembly) (6).

The ideal assembly method should be based on a modular assembly strategy, allowing seamless and sequence-independent fusion of modules in a fast, simple and combinatorial way. A modular strategy enables the use of standardized parts and simplifies the generation of combinatorial gene libraries. Fusion of the modules must be seamless to avoid unwanted sequences which could have undesired effects and should be sequence-independent in order to allow for the creation of combinatorial gene libraries. Furthermore, the fusion method should be fast and simple to increase the number of generated constructs. However, current methods do not possess all these characteristics.

Almost all currently used modular strategies (including BioBrick, BglBrick and Golden Gate) that are often based on BioBrick assembly standards, require restriction endonuclease digestion which results in scar formation and is sequence dependent (4,5,7-10). Also, most modular strategies require multiple ligation, transformation, overnight growth and plasmid extraction steps in order to align multiple modules in a construct.

Methods for seamless fusion, including SLIC, Gibson and CPEC, rely on synthetic oligonucleotides or PCR generated modules with complementary regions, which are annealed and then fused by using either ligases or polymerases (5,8,11-15). Using these methods, large constructs can be created in single-step assembly reactions which improves construction speed (13,16). However, due to the use of complementary regions, fusion is sequence dependent and hinders simultaneous assembly of different modules for the construction of combinatorial gene libraries. Also, repetitive and GC-rich sequences are problematic (6).

Here we present a modular assembly strategy for seamless, sequence-independent and fast combinatorial gene construction at the gene/pathway level. The assembly strategy is based on the ligation of single- or double-stranded oligonucleotides (ssODN or dsODN) using a solid-phase platform, as initially described by Dietrich et al. (17). Their method is further optimized by a decreased ligation time, increased amount of ligated dsODNs and expanded by including the assembly of ssODNs as well as using PCR generated dsODN. The use of a solid-phase system facilitates purification of the obtained products and enables automation for high-throughput gene construction.

Our strategy includes three different assembly methods (I) solid-phase assembly of single-stranded oligonucleotides, (II) solid-phase assembly of single-stranded oligonucleotides using linkers and (III) solid-phase assembly of double-stranded oligonucleotides, either obtained by PCR or by annealing complementary ssODN. To demonstrate the feasibility of our methods, a protein expression device containing a small RBS library taken from the BioBrick parts registry and a combinatorial library encoding multimodular peptides for gene delivery were created.

Materials and Methods

Materials

All chemicals and reagents were purchased from Sigma-Aldrich Chemie B.V. (Zwijndrecht, The Netherlands), unless stated otherwise. Dynabeads M-280 Streptavidin were purchased from Dynal (Oslo, Norway). Restriction enzymes, FastAP™ (Thermosensitive Alkaline Phosphatase), Generuler™ DNA and Riboruler™ RNA ladders, *Pfu* polymerase, Rapid DNA Ligation Kit, T4 RNA ligase, and GeneJET™ Plasmid Miniprep Kit were purchased from Thermo Fisher Scientific (St. Leon-Rot, Germany). Thermostable RNA ligase was purchased from Epicentre Biotechnologies (Madison, WI, USA). Precast denaturing polyacrylamide TBE-urea gels (6% and 10%) were purchased from Invitrogen (Breda, The Netherlands).

Synthetic oligodeoxynucleotides

All oligodeoxynucleotides were synthesized by Eurogentec S.A. (Seraing, Belgium). ssODN were synthesized with 5'- and 3'- phosphates and PAGE purified at a 40nmol scale or with 5'biotin-TEG and PAGE purified at the 200nmol scale. ssODN linkers, primers and complementary strands were synthesized unmodified and SePOP desalted at a 40nmol scale, except for the primer fw_EGFP, which was 5'-phosphorylated and PAGE purified. Oligodeoxynucleotides were dissolved in nuclease-free water and stored at -30°C as 100µM stocks.

Oligodeoxynucleotides encoding peptides were codon-optimized for bacterial expression and flanked with linker sequences GGTTCT (GS) at the 5' (all except Head modules) and GGTGGC (GG) at the 3' (all except Tail modules). The possible formation of secondary structures (checked with Oligocalc software) was reduced by adjustment of any self-complementary oligonucleotide sequence (18).

Promoter, RBS and terminator sequences were deduced from the Registry of Standard Biological Parts (19). All sequences of the used ssODNs and primers are listed in Table 1.

Table 1 Used oligodeoxynucletides

Module name	Sequences
Head_GG	5'-Biotin-TEG-AGCCCATGGTTATGGAAAACCTGTATTTTCAGGG-TGGC-OH 3'
GS_EB1_GG	'5 P-GGTTCTCTGATTCGTTTATGGAGCCATCTGATTCACATTTGGTT-TCAGAACCGTCGTCTGAAATGGAAAAAAAAGGGTGGC-P 3'
GS_H5WYG_GG	5' P-GGTTCTGGTCTGTTTCACGCGATTGCCCATTTCATCCACGG-TGGTTGGCATGGTTTAATTCACGGTTGGTATGGCGGTGGC-P 3'
GS_H92_GG	5' P-GGTTCTAAGACACCGAAAAAGGCCAAAAAGCCAAAAACCCC-GAAAAAGGCGAAAAAACCAGGTGGC-P 3'
GS_LAH_GG	5' P-GGTTCTAAAAAGGCACTGCTGGCACTGGCACTCCATCACT-TAGCACACCTTGCTCATCATCTTGCGTTAGCGCTGAAAAAGGCTGG-TGGC-P 3'
GS_NLSV402_GG	5' P-GGTTCTCCAAAAAAGAAACGTAAAGTTCCAAAAAAAAAGCG-CAAAGTCGGTGGC-P 3'
GS_ppTG20_GG	5' P-GGTTCTGGCTTATTTCGTGCGCTGTTGCGTCTGTTACTG-TCTCTGTGGAGATTACTTTTACGTGCGGGTGGC-P 3'
GS_Pr18_GG	5' P-GGTTCTTCTCGTAGTCGGTATTACCGTCAGCGCCAACGTTCTCG-CCGTCGCCGGCGTAGAGGTGGC-P 3'
GS_SPKR4_GG	5' P-GGTTCTAGCCCGAAACGTAGCCCTAAGCGCAGCCCAAAAAG-ATCTCCTAAACGTGGTGGC-P 3'
GS_TAT_GG	5' P-GGTTCTTATGGCCGCAAGAAGCGTCGTCAAAGACGTCGTGG-TGGC-P 3'
GGGS linker	5' OH-GGTGGCGGTTCT-OH 3'
GS_Tail	5' OH-GGTTCTTAATAAGGATCCCATAAGTATGAC-OH 3'
Fw_Head	5' OH-AGCCCATGGTTATGGAAAA-OH 3'
Rev_Tail	5' OH-GTCATACTTATGGGATCCTTATTAA-OH 3'
Head_pTET_J72005	5'Biotin-TEG-AGCCAGATCTTCCCTATCAGTGATAGAGATTGACATC-CCTATCAGTGATAGAGATACTGAGCAC-OH 3'
Complement Head_pTET_J72005	5' OH-GTGCTCAGTATCTCTATCACTGATAGGGATGTCAATCTCTAT-CACTGATAGGGAAGATCTGGCT-OH 3'
BBa_B0034	5' P-AAAGAGGAGAAACACAT-P 3'
BBa_B0035	5' P-ATTAAAGAGGAGAACACAT-P 3'
BBa_J61141	5' P-TGGCTAACTGAGGATCACAT-P 3'
BBa_K376001	5' P-TGGCTAACATAGGGTCACAT-P 3'
Complement BBa_B0034	5' OH-ATGTGTTTCTCCTCTTT-OH 3'
Complement BBa_B0035	5' OH-ATGTGTTCTCCTCTTTAAT-OH 3'
Complement BBa_J61141	5' OH-ATGTGATCCTCAGTTAGCCA-OH 3'
Complement BBa_K376001	5' OH-ATGTGCTCTCCTTAAATTTGTGTGAATTT-OH 3'
Fw_EGFP	5' P-ATGGTGAGCAAGGGCGAGGAG-OH 3'
Rev_EGFP	5' OH-CTTGTACAGCTCGTCCATGCCG-OH 3'

Table 1 Continued

Module name	Sequences
Tail_Term_B0011	5' P-TAAAGAGAATATAAAAAGCCAGATTATTAATCCGGCTTTTTTAT-TATTTAAGCTTCACGC-P 3'
Complement Tail_Term_B0011	5' OH-GCGTGAAGCTTAAATAATAAAAAAGCCGGATTAAT-AATCTGGCTTTTTATATTCTCTTTA-OH 3'
Fw_Head_pTET	5' OH-AGCCAGATCTTCCCTATCAGTGA-OH 3'
Rev_Tail_Term	5' OH-GCGTGAAGCTTAAATAATAAAAAAGC-OH 3'

Methods

Solid support

As a solid support M-280 Streptavidin coated magnetic beads were used and biotinylated universal single-stranded oligodeoxynucleotides (ssHead) were bound to the beads according to manufacturer's protocol. To 1mg of M-280 beads, 400pmol ssHead was added and incubated in an Eppendorf incubator (thermomixer comfort; Eppendorf, Nijmegen, The Netherlands) for 30 minutes at 22°C and 800rpm. To remove unbound ssHead, coated magnetic beads were washed 3 times with 1x B&W buffer (5 mM Tris-base (pH 7.5), 0.5mM EDTA, 1M NaCl) using a 96-well magnetic rack (Life Technologies, Bleijswijk, The Netherlands) to capture the beads during buffer replacements.

Solid-phase assembly of single-stranded oligodeoxynucleotides using RNA ligase

Ligation products were assembled in a stepwise protocol. Ligation reactions (20µL) with thermostable RNA ligase (Epicentre) contained 12.5pmol ssHead coated on M-280 magnetic beads, 1x Thermostable RNA Ligase buffer, 20% (v/v) polyethylene glycol (PEG) 4,000, 50µM ATP, 100 pmol ssODN and 10U-100U thermostable RNA ligase. After vortexing briefly, ligation was performed by incubation for 16hr at 60°C. Ligation reactions (20µL) with T4 RNA ligase (New England Biolabs (NEB), Ipswich, MA, USA) contained 12.5pmol ssHead coated on M-280 magnetic beads, 1x T4 RNA Ligase 1 Reaction buffer, 20% (v/v) PEG 4,000, 100pmol ssODN and 20U T4 RNA ligase. After briefly vortexing, ligation reactions were performed at 22°C or 37°C for 16hrs.

After ligation, samples were washed 2 times with 50µL 1x B&W buffer and 3'-dephosphorylated in 50µL reactions containing 1x FastAP™ Buffer and 1 U FastAP™ Alkaline Phosphatase for 10 minutes at 37°C. Three washes with 50µL 1x B&W buffer were performed, after which a new ligation cycle could be initiated. After ligation of the final ssODN, the products were washed 3 times with 1x B&W and stored at 4°C in 20µL 1x B&W. This protocol is schematically shown in Figure 1a.

Solid-phase assembly of single-stranded oligodeoxynucleotides in the presence of linker oligodeoxynucleotides using DNA ligase

Ligation products were assembled in a stepwise protocol. Ligation reactions (20µL) contained 12.5pmol ssHead coated on M-280 magnetic beads, 50pmol linker, 100pmol ssODN, 1x Rapid Ligation Buffer (part of Rapid DNA Ligation Kit) and 5U T4 DNA ligase (Thermo Fisher Scientific, St. Leon-Rot, Germany). After vortexing briefly, annealing and ligation reactions were performed in a PCR thermocycler. The annealing/ligation PCR program used the following settings: 10min at 42°C, stepwise decrease from 42°C to 16°C in -1 °C/min steps, 30min at 16°C. After ligation, samples were washed 2 times with 50µL 1x B&W buffer and 3'-dephosphorylated in 50µL reactions containing 1x FastAP™ Buffer and 1U FastAP™ Alkaline Phosphatase for 10 minutes at 37°C. Three washes with 50µL 1x B&W buffer were performed, after which a new ligation cycle was initiated. After ligation of the final ssODN, the products were washed 3 times with 1x B&W and stored at 4°C in 20µL 1x B&W. This protocol is schematically shown in Figure 1b.

Solid-phase assembly of double-stranded oligodeoxynucleotides using DNA ligase

First, single-stranded oligodeoxynucleotide modules were made double-stranded. For hybridization, equal amounts of ssODN and its complementary oligodeoxynucleotide were mixed and heated to 70°C for 10min following a cooling down to room temperature within 1 hour. Hybridization of ssHead was performed after coupling to the magnetic beads.

Ligation products were assembled in a stepwise protocol. Ligation reactions (20µL) contained 12.5pmol dsHead coated on M-280 magnetic beads, 100pmol dsON, 1x Rapid Ligation Buffer and 5U T4 DNA ligase (Thermo Fisher Scientific, St. Leon-Rot, Germany). After vortexing briefly, ligation reactions were performed at 16°C for 1 hour. After ligation, samples were washed 2 times with 50µL 1x B&W buffer and 3'-dephosphorylated in 50µL reactions containing 1x FastAP™ Buffer and 1U FastAP™ Alkaline Phosphatase for 10 minutes at 37°C. Three washes with 50µL 1x B&W buffer were performed, after which a new ligation cycle could be initiated. After ligation of the final dsODN, the products were washed 3 times with 1x B&W and stored at 4°C in 20µL 1x B&W. This protocol is schematically shown in figure 1c.

Ligation efficiency determination of the 3 different solid-phase assembly methods by SEC

Ligation efficiency of the 3 different solid-phase assembly methods was determined using size exclusion chromatography (SEC) on an Acquity UPLC® with a BEH400SEC column (Waters, Dublin, Ireland). Ligation products were prepared using the described protocols. After ligation of the final ssODN or dsODN, ligation products were washed

with 1x B&W and re-suspended in 16µL SEC sample buffer (0.1M Tris pH 8.0, 1mM EDTA, 0.1% SDS). To break the biotin-streptavidin bond, samples were incubated for 7min at 95°C and magnetic beads were removed by a 96-well magnetic rack. Supernatants containing ligation products were eluted with 100% eluent A (0.1M Tris pH 8.0, 1mM EDTA) in 15min at a flow rate of 1ml/min. Ligation products were detected at 260nm.

Ligations products were identified using controls (the used ss- and dsHead and ss- dsModules). Ligation efficiency was determined by dividing the peak area of the ligation product by the combined peak area of non-ligated ss- or dsHead and the ligation product(s).

Purification and PCR amplification of ligation products

Of the obtained ligation products (20µL), 2-5µL was loaded on 6 and/or 10% denaturing polyacrylamide TBE-urea gels and the desired ligation product bands were excised. The ligation products were eluted from the gel by passive elution in elution solution (0.3M sodium acetate, 2mM Na_2EDTA, pH 7.6) overnight. Eluted ligation products were precipitated by the addition of 2.5 vol 96% ethanol, cooling down to -80°C for 45 minutes and centrifugation for 12 minutes at 17,000xg. Purified ligation products were re-dissolved in 25µL nuclease-free H_2O, of which 5µL was used in high-fidelity polymerase chain reactions (PCRs). PCRs (50µL) contained 1x *Pfu* Buffer with 20mM $MgSO_4$, 0.2mM dNTP mix, 0.5µM Fw-primer, 0.5µM Rev-primer and 2U *Pfu* polymerase. Primer sequences are listed in Table 1. Reactions were performed in a PCR thermocycler with the following cycling parameters: 2.30 min at 95°C, 30x (0.30min at 95°C, 0.30min at 54°C, 3.00min at 72°C), 10.00min at 72°C and stored at 4°C. Reaction products were analyzed on 2% w/v agarose gels stained with ethidium bromide. Bands were visualized with UV illumination at 302nm.

Construction of multipart combinatorial libraries by assembly of ssODN in the presence of linker ODN

Using the described method for the assembly of ssODN in the presence of linker ODN, a random polypeptide library was assembled. Ligation reactions (20µL) contained 12.5pmol ssHead (Head_GG) coated on M-280 magnetic beads, 50pmol GGGS linker, 1x Rapid Ligation Buffer, 5U T4 DNA ligase and 100pmol of a mix of 9 different ssODN (10pmol of GS_Pr18_GG; GS_SPKR4_GG; GS_H9.2_GG; GS_H5WYG_GG; GS_TAT_GG; GS_ppTG20_GG; GS_EB1_GG; GS_NLSV402_GG and 20 pmol of GS_LAH_GG). Up to six ligation steps with this composition were performed. One additional ligation step was performed for the ligation of the reaction products to a closing single-stranded oligonucleotide, an ssTail (GS_Tail). Ligation products were loaded on a 6% denaturing polyacrylamide TBE-urea gel and the area between the smallest theoretical product (353nt) and the largest theoretical product (623nt) was excised, purified and PCR

amplified as described. PCR products were analyzed on a 2% w/v agarose gel and the same area was excised and purified.

The obtained library was BamHI / NcoI digested, purified and 40ng of the digested products was ligated in 10ng of dephosphorylated pIVEX_EJ (BamHI/NcoI digested and purified from an agarose gel). Half of the reaction mixture was directly used to transform *E.coli* strain DH5α. Bacteria were plated out on LB agar plates containing ampicillin (100µg/mL) and grown overnight at 37°C.

From the obtained LB-agar plates, twenty colonies were inoculated overnight in 5mL LB culture medium containing ampicillin (100µg/mL) and plasmid DNA (pDNA) was harvested. Constructs were BamHI/NcoI digested and analyzed on a 2% w/v agarose gel. In addition, the obtained library constructs were partially sequenced (BaseClear, Leiden, The Netherlands).

Assembly of a protein expression construct containing a small RBS library

Using the described dsODN assembly method, a protein expression construct containing a small RBS library was assembled. To obtain the 5'-phosphorylated EGFP module, PCR was performed, using a 5'-phosphorylated Fw-primer and an unmodified Rev-primer (Table 1) with 1ng pCMV-EGFP vector as a template. PCRs (50µL) contained 1x *Pfu* Buffer with 20mM $MgSO_4$, 0.2mM dNTP mix, 0.5µM Fw-primer, 0.5µM Rev-primer and 2U *Pfu* polymerase. Products were purified from a 1% agarose gel.

First, the assembly of a double-stranded Head and four double-stranded RBS was realized. Ligation reactions (20µL) contained 12.5pmol dsHead (Head_pTET_J72005) coated on M-280 magnetic beads, 1x Rapid Ligation Buffer, 5U T4 DNA ligase and 100pmol of a mix of the four different RBS dsODN (25pmol of BBa_B0034, BBa_B0035, BBa_J61141 and BBa_K376001). For the assembly of the obtained double stranded EGFP module and the formed Head_RBS products, ligation reactions (20µL) contained 6.25pmol dsHead_RBS, 1x Rapid Ligation Buffer, 5U T4 DNA Ligase (Fermentas) and 6.25pmol 5'-phosphorylated ends of EGFP modules. Constructs were finished by the ligation with the double-stranded terminator. Ligation reactions (20µL) contained 6.25pmol dsHead_RBS_EGFP ligation product, 100pmol dsTail (Tail_Term_B0011), 1x Rapid Ligation Buffer and 5U T4 DNA Ligase. Constructs were analyzed and purified from 6% denaturing polyacrylamide TBE-urea gel at the final product height of 800-1000nt, and PCR amplified as described previously.

PCR products were purified from a 2% w/v agarose gel, BglII/HindIII digested and purified. Subsequently, 40ng of the purified product was ligated in 10 ng dephosphorylated BglII/HindIII digested pIVEX-LacZ vector. Half of the reaction mixture was used to directly transform *E.coli* strain XL-1 blue, which were grown overnight at 37°C on LB-agar plates, containing ampicillin (100µg/mL). From the obtained LB-agar plates, 40 colonies were inoculated overnight in 5mL LB culture medium containing ampicillin (100 µg/mL). pDNA was harvested from 2mL of these cultures and analyzed

by HindIII/BglII restriction analysis. The remainder of the bacterial cultures was used for screening of fluorescence. From each overnight culture, 1.15x10^8 bacteria were transferred to a black 96-well plate and bacteria were lysed using 100µL Cellytic B. Fluorescence was determined using a fluorescence well-plate reader (FLUOstar OPTIMA; BMG-Labtech, Offenburg, Germany) with excitation 488nm, emission 520nm filters.

Also, fluorescence was determined by fluorescence microscopy with excitation at 460-500 nm and an EGFP filter (510-560nm). In addition, the obtained library constructs were partially sequenced.

Results

Solid-phase assembly

In our attempt to construct combinatorial gene libraries in a sequence-independent manner without using restriction endonucleases, 3 different ligation methods were tested and compared: (A) ligation of ssODN using RNA ligase, (B) ligation of ssODN using DNA ligase in the presence of linker ODN and (C) blunt-end ligation of dsODN using DNA ligase. For all three strategies ligation was performed on DNA immobilized on magnetic beads to facilitate and accelerate iterative ligations rounds.

For the ligation of ssODN using RNA ligase (Fig. 1a), a universal single-stranded oligodeoxynucleotide (ssHead), modified with 5'-biotin was immobilized on streptavidin-coated magnetic beads. Using RNA ligase, 5'-phophorylated single-stranded oligodeoxynucleotides (ssModules) were ligated stepwise to the ssHead. To ensure proper orientation of the ligated ssModules, and to prevent multiple additions of ssModules, the ssModules were 3'-phosphorylated. In the final ligation step, a universal single-stranded oligodeoxynucleotide (ssTail) was added. One ligation/dephosphorylation cycle took 16 hr.

The ligation of ssODN in the presence of linker ODN with DNA ligase makes use of linkers to stabilize the association of 3'-hydroxyl group and 5'-phosphate group of two ssModules in order to enhance ligation efficiency (Fig. 1b). The used linkers were 12 nucleotides long and were complementary to the last 6 nucleotides of the immobilized ssModule and the first 6 nucleotides of ssModule to be ligated. The linkers supported the assembly of a pre-ligation complex, which can be recognized and joined by DNA ligase. One ligation/dephosphorylation cycle took 1 hr.

Blunt-end ligation of dsDNA made use of double-stranded oligodeoxynucleotides (dsModules) and DNA ligase (Fig. 1c). One ligation/dephosphorylation cycle took 1 hr.

Comparison of the three assembly methods

In order to compare the three different assembly methods, three different sized ss- or dsModules (30, 60 and 90 nucleotides) were ligated onto a ss- or dsHead (36 nucleotides) in one assembly cycle. After ligation, equal amounts of magnetic beads

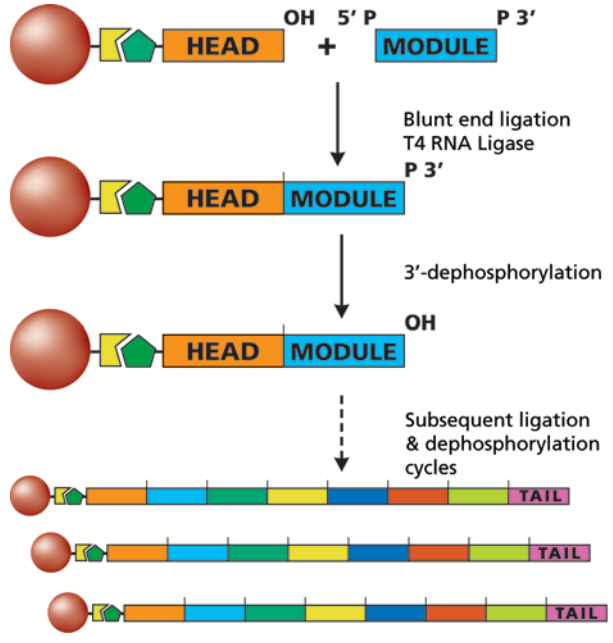


Figure 1a Solid-phase assembly of single-stranded oligodeoxynucleotides. 5'-biotinylated ssHead was immobilized on streptavid coated magnetic beads. Using RNA ligase, 5'-phophorylated ssODN were ligated to the ssHead. To ensure proper orientation and to prevent multiple additions of ssModules, the ssModules were 3'-phosphorylated. By using repetitive ligation/dephosphorylation cycles, it is possible to assemble products of any length and composition. In the final ligation step, a ssTail, containing a double stop codon was added.

were loaded onto a 10% TBE-Urea gel (Fig. 2).

Only faint bands were visible at the height of the expect ligation products (boxes) when using the ssODN assembly method, where the use of thermostable RNA ligase (Epicentre) resulted in higher ligation efficiencies compared to T4 RNA ligase. Clear bands were visible for both the ssODN linker and dsODN assembly methods.

In order to improve ssODN assembly, several different parameters were screened, which included the use of different ligation temperatures and the addition of hexamminecobalt(III) chloride (HCC), betaine and spermidine. Ligation reactions performed with thermostable RNA ligase (Epicentre) at 60°C or T4 RNA ligase (NEB) at 22°C showed the highest ligation efficiency, compared to other RNA ligases (data not shown). Also, the addition of DMSO,HCC, betaine or spermidine did not enhance ligation efficiency (data not shown).

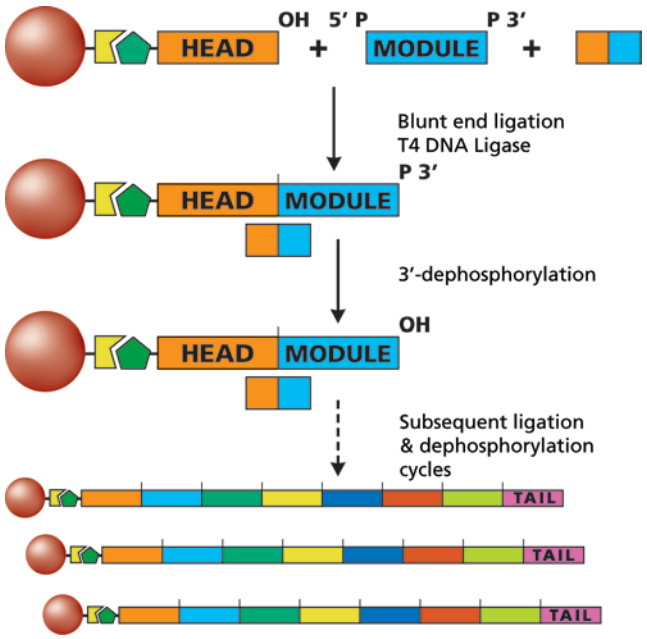


Figure 1b Solid-phase assembly of single-stranded oligodeoxynucleotides using linkers. 5′-biotinylated ssHead was immobilized on streptavid coated magnetic beads. Using linker ODN, the 3′-hydroxyl group and 5′-phosphate group of two ssModules were brought in close proximity. The linkers are complementary to the last 6 nucleotides of the immobilized ssModule and the first 6 nucleotides of ssModule to be ligated. The linkers support the assembly of a pre-ligation complex, which can be recognized and joined by DNA ligase. By using repetitive ligation/dephosphorylation cycles, it is possible to assemble products of any length and composition. In the final ligation step, a ssTail, containing a double stop codon was added.

Ligation efficiency of the three assembly methods was determined by size-exclusion chromatography (SEC). Three different sized ss- or dsModules (45, 60 and 90 nucleotides) were ligated in one assembly cycle to a ss- or dsHead module (36 nucleotides) to create Head-module constructs. The obtained ligation efficiencies are shown in Table 2 and were in line with the observed ligation efficiencies on gel (Fig 2). The ssODN linker and dsODN assembly methods displayed the highest ligation efficiencies with 93% and 55%, respectively. The ssODN assembly method displayed lower ligation efficiencies. The use of thermostable RNA ligase resulted in higher ligation efficiencies compared to normal RNA ligase. The difference in ligation efficiency between the three sized modules is less for the ssODN linker and dsODN assembly methods, compared to the ssODN assembly methods.

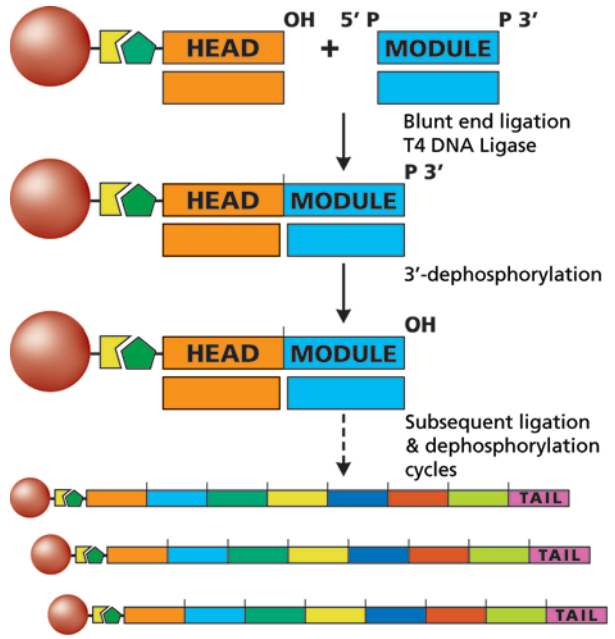


Figure 1c Solid-phase assembly of double-stranded oligodeoxynucleotides using blunt-end ligation. 5′-biotinylated ssHead was immobilized on streptavid coated magnetic beads and ssModules were made double-stranded by hybridization with complementary ODN. Using DNA ligase, 5′-phophorylated dsModules can be liagted to the ssHeads. To ensure proper orientation and to prevent multiple additions of ssModules, the ssModules were 3′-phosphorylated. By using repetitive ligation/dephosphorylation cycles, it is possible to assemble products of any length and composition. In the final ligation step, a dsTail, containing a double stop codon was added.

Table 2 Ligation efficiencies of the three assembly methods determined by SEC. ND: Not determined

Assembly method	Ligation efficiency 45 nt module	Ligation efficiency 60 nt module	Ligation efficiency 90 nt module	Average ligation efficiency
ssODN assembly	ND	5%	17%	11%
ssODN assembly using thermostable RNA ligase	ND	30%	75%	53%
ssODN linker assembly	ND	93%	93%	93%
dsODN assembly	59%	57%	50%	55%

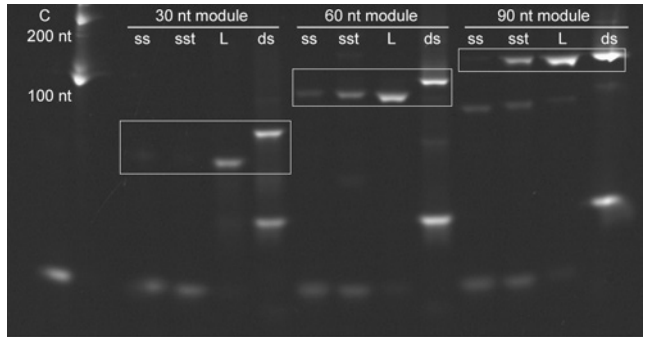


Figure 2 Comparison of the three assembly methods: Assembly of ssODN using RNA ligase (ss); Assembly of ssODN using thermostable RNA ligase (sst); Assembly of ssODN using linker ODN (L); Assembly of dsODN (ds). Ligation reactions were performed with ss- or dsModules of 30, 60 and 90 nucleotides long. Equal amounts of magnetic beads were loaded on a 10% TBE urea gel with: Ladder: Low range RNA ladder. C: 3.1pmol Head.

Construction of multipart combinatorial libraries by ssODN ligation in the presence or absence of linker ODN

For some applications, such as the generation of gene libraries encoding TALEs, elastin-like polymers or multimodular DNA carrier peptides/proteins, a technique to ligate different combinations of DNA modules in a combinatorial fashion is highly desired (20-22). To demonstrate our technology we have tested two conditions to ligate two different ssODN modules in a sequence-independent and/or scarless fashion. Ligation of ssODN using RNA ligase and ligation of dsODN using DNA ligase was used for the seamless fusion of modules. For the ligation of ssODN in the presence of linker ODN using DNA ligase, ssModules were flanked with linker sequences GGTTCT at the 5' and GGTGGC at the 3'. Both assembly methods used the same Head module and Tail module. Both Head and Tail module contain restriction sites for cloning purposes.

First, specific constructs were constructed using both assembly methods for a proof of principle. Two different sized ssModules, module T (45 nucleotides) and module L (90 nucleotides), were ligated onto the ssHead module (36 nucleotides). Multiple assembly cycles were used to ligate multiple T and L modules to each other, creating Head-T_n and Head-L_n were n represents the number of assembly cycles. All constructs were terminated with a Tail module (45 nucleotides). In addition, an alternating construct of both T and L modules, Head-$(T\text{-}L)_3$ was created using ODN linkers. The constructs formed were analyzed on a 6% TBE urea gel, after loading equal amounts of magnetic beads.

For the determination of the ligation efficiency of the subsequent assembly cycles, specific constructs were created using all three assembly methods. In six

subsequent assembly cycles, three different sized ss- or dsModules, (module T, module S (60 nucleotides) and module L) were ligated onto a ss- or dsHead module. The constructs created by ssODN assembly and ssODN linker assembly were terminated with a Tail module in the seventh subsequent assembly cycle. The constructs formed were analyzed by SEC and loaded on a 6% TBE urea gel (Supplementary Fig. 1-7).

The Head-T_2-Tail and Head-L_2-Tail constructs generated using the solid-phase assembly of ssODN with RNA ligase showed expected bands, however the bands were only faintly visible (Fig. 3a, lanes 2 and 4). In contrast Head-T_2-Tail and Head-L_2-Tail constructs , generated using the solid-phase assembly with linker ODN showed the similar expected bands (Fig. 3a, lanes 1 and 3). SEC analysis confirmed the low yield of ligation products with more than 2 modules.

In order to verify product length and attachment of the Tail module, constructs were extracted from gel, purified and amplified by PCR. Only faint products bands could be detected for constructs generated in the absences of linker ODN (Fig. 3b).

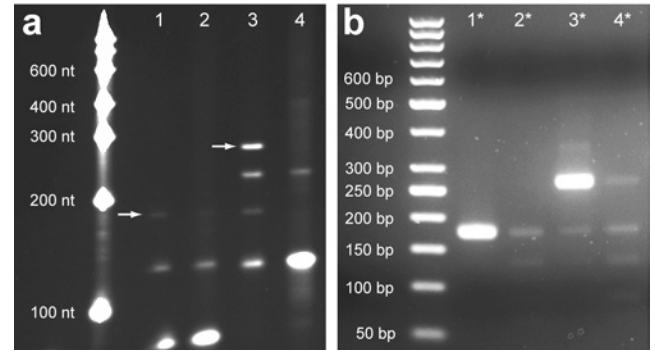


Figure 3 Ligation of ssODN using RNA ligase (a) and subsequent amplification of purified ligation products by PCR (b).

(**a**) 6% TBE urea gel showing the ligation products after ligating 3 modules with either DNA ligase in the presence of linker ODN (lanes 1 and 3) or with RNA ligase (lanes 2 and 4). Formed products are Head-T_2-Tail (lanes 1 and 2) and Head-L_2-Tail (lanes 3 and 4). For size estimation, a low range RNA ladder was loaded.

(**b**) PCR products loaded on 2% w/v agarose gel with a 50bp DNA ladder. * lanes correspond with lanes in panel A; Lane 1* and 3*: Head-T_2-Tail and Head-L_2-Tail using the assembly of ssODN in the presence of linker ODN using DNA ligase, respectively. Lanes 2* and 4*: Head-T_2-Tail and Head-L_2-Tail using the assembly of ssODN using RNA ligase, respectively.

Table 3 Theoretical length of ligation products (in # of nucleotides)

Product	Length	Product	Length
Head-T-Tail	128nt	**Head-L-Tail**	173nt
Head-T2-Tail	173nt	**Head-L2-Tail**	263nt
Head-Tail	83nt		

Figure 4a shows the Head-Tail, Head-T_n-Tail, Head-L_n-Tail (n = 1, 2, 3 and 6) and Head-$(T\text{-}L)_3$-Tail constructs generated using the solid-phase assembly with linker ODN. All constructs showed the expected band pattern. Only the band of Head-T_6-Tail is not clearly visible. SEC analysis confirmed the distribution of ligation products and corresponds to the observed products on gel. Were subsequent ligation cycles with module T yielded ligation products with mostly 2, 3 and 4 modules, subsequent ligation cycles with module s and L yielded ligation products with mostly 5, 6 and 7 modules.

In order to verify product length and attachment of the Tail module, constructs were extracted from gel (boxes Fig. 4a), purified and amplified by PCR (Fig. 4b). Constructs with up to 3 repeating peptide modules and a Tail module were amplified successfully and showed the expected lengths. Product bands could be clearly identified after amplification of constructs generated by assembly using linker ODN. However, amplification of Head-T_6-Tail, Head-L_6-Tail and Head-$(T\text{-}L)_3$-Tail yielded multiple bands, all shorter in size than the expected products. Product bands corresponding to Head-T_1-Tail through Head-T_5-Tail and Head-Tail could be detected after amplification of Head-T_6-Tail construct. Only products corresponding to Head-L_1-Tail and Head-Tail could be detected after amplification of Head-L_6-Tail. After amplification of Head-$(T\text{-}L)_3$-Tail, multiple product bands could be detected corresponding to constructs with multiple T and/or L modules.

Gel electrophoresis and SEC analysis showed the expected band pattern of the constructs generated by solid-phase assembly of dsODN with DNA ligase, in six subsequent assembly cycles with three different sized dsModules (Supplementary Fig. 2, 4, 6 and 7).

After the creation of specific constructs, two random libraries were assembled using the solid-phase assembly of ssODN with linkers. A mix of nine different ssModules was ligated onto a ssHead. Three or six assembly cycles were performed with this mix, generating two libraries Head-MIX_3 and Head-MIX_6. This would theoretically yield libraries with 9^3 = 729 and 9^6 = 531441 different constructs. Both libraries were terminated with a Tail module. The formed constructs were analyzed on a 6% TBE urea gel, after loading equal amounts of magnetic beads.

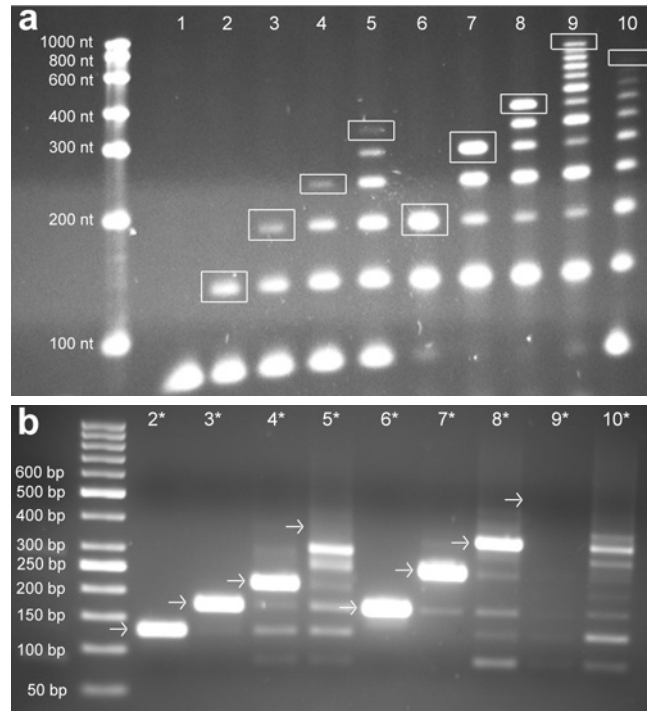


Figure 4 Ligation of ssODN in the presence of linker ODN (a) and subsequent amplification of purified ligation products by PCR (b).
(**a**) 6% TBE urea gel showing the ligation products after ligating 1, 2, 3, 4 and 7 modules with DNA ligase in the presence of linker ODN. Formed products are Head-Tail (lane 1), Head-T_n-Tail, with n=1, 2, 3, and 6 (lanes 2-5), Head-L_n-Tail, with n=1, 2, 3 and 6 (lanes 6-9) and 10 Head-$(T\text{-}L)_3$-Tail (lane 10). For size estimation, a low range RNA ladder was loaded. Boxed constructs were extracted from gel, purified and amplified.
(**b**) PCR products loaded on 2% w/v agarose gel with a 50bp DNA ladder. * lanes correspond with lanes of A; lanes 2*-5*: Head-T_n-Tail with n=1, 2, 3 and 6 respectively; lanes 6*-9*: Head-L_n-Tail with n=1, 2, 3 and 6 respectively; lane 10*: Head-$(T\text{-}L)_3$-Tail. Arrows indicate expected product heights.

Both generated libraries were clearly visible and appeared as smears (Fig. 5a). Separate constructs could still be detected in the lower regions but with increasing size the separate construct bands became gradually obscured in a diffuse smear. By increasing the number of assembly cycles from three to six, the absolute size of the constructs and the average size of the library increased.

After extraction from gel and purification, both libraries were amplified by PCR (Fig. 5b). Also after PCR amplification, smears were detected indicating that all formed

Table 4 Theoretical length of ligation products (in # of nucleotides)

Product	Length	Product	Length
Head-T-Tail	128nt	**Head-L-Tail**	173nt
Head-T_2-Tail	173nt	**Head-L_2-Tail**	263nt
Head-T_3-Tail	218nt	**Head-L_3-Tail**	353nt
Head-T_6-Tail	353nt	**Head-L_6-Tail**	623nt
Head-Tail	83nt	**Head-$(T\text{-}L)_3$-Tail**	488nt

constructs were amplifiable. PCR smears were denser at the bottom of the gel (smaller constructs) than at the top of the gel (larger constructs).

The PCR amplified Head-MIX_6-Tail library was excised from the smallest theoretical 6 module construct (±350bp) to largest theoretical 6 module construct (±700bp), purified and cloned into the BamHI/NcoI-sites of vector pIVEX-EJ. *E.coli* strain DH5α was transformed with the cloned library. To check for cloning bias, all colonies of a single plate were pooled and pDNA was harvested. After restriction enzyme analysis using BamHI/NcoI, the same smear pattern was observed as with the initial PCR amplification (results not shown).

To check for insert size, restriction analysis using BamHI/NcoI was performed on pDNA obtained from twenty colonies after overnight culturing. Two colonies contained Head-Tail inserts, all other eighteen colonies contained inserts corresponding to at least one peptide module and no empty vectors could be detected (Fig. 5c). Sequence analysis of 10 clones confirmed the error-free ligation and random insertion of all used modules in the multipart constructs.

Assembly of a protein expression device containing a small RBS library

A goal in synthetic biology is identification of the precise level of activity of specific parts. As an example, in metabolic engineering the expression levels of individual enzymes are modulated through, e.g. manipulating promoter strengths and ribosome binding sites (RBS) (23). Solid-phase assembly of dsODN modules is ideal for modular screenings, as it allows for seamless, fast and modular assembly.

Using solid-phase assembly of dsODN modules, a small library of RBS's was screened on translation efficiency. From the Registry of Standard Biological Parts catalog (19), the Tet promoter sequence (BBa_J72005), four RBS sequences (BBa_B0034, BBa_B0035, BBa_J61141 and BBa_K376001) and the LuxICDABEG terminator sequence (BBa_B0011) were selected to create an expression device. As a reporter, Enhanced Green Fluorescent Protein (EGFP) was chosen (Fig. 6). To all four RBS sequences, a 5 nucleotide spacer was added to create a spacer between the RBS sequence and the start codon of the EGFP (Fig. 6). The Head module contains the

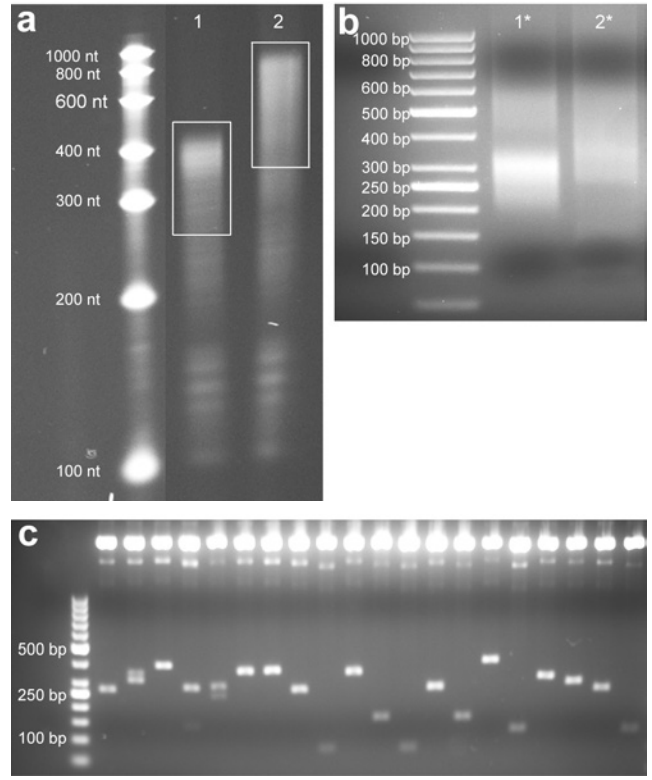


Figure 5 Random multi-domain peptide libraries constructed with the assembly of ssODN using linker ODN.
(**a**) Initial ligation products were loaded on a 6% TBE urea gel and purified at final product height as indicated. Ladder: low range RNA ladder. Lanes 1-2: H-MIX3-T and H-MIX6-T, respectively.
(**b**) Purified ligation products were amplified using PCR and loaded on a 2% agarose gel. Ladder: 50bp DNA ladder. Lanes 1*-2*: PCR amplified ligation products corresponding to lanes 1-2.
(**c**) Plasmid DNA isolated from 20 colonies was digested with BamHI/NcoI and loaded on a 0.7% agarose gel. Eighteen out of twenty colonies contained inserts corresponding to at least one peptide module.

promoter sequence and the Tail module contains a stop codon and the terminator sequences. Both Head and Tail module contain restriction sites for cloning purposes.

First, a mix of the four different ds-modules encoding the four different RBS's was ligated onto the dsHead. Then, the dsEGFP and dsTial modules were sequentially ligated. The formed constructs were analyzed on a 6% denaturing polyacrylamide TBE urea gel, after loading equal amounts of magnetic beads.

Table 5 Modules present in the 10 sequenced constructs

Construct #	Module present	Construct #	Module present
1	EB1-NLSV402	**6**	SPKR4-SPKR4
2	H5WYG-NLSV402	**7**	H9.2-NLSV402
3	EB1-NLSV402-H9.2	**8**	EB1-ppTG20
4	NLSV402-EB1	**9**	NLSV402-Pr18-TAT
5	EB1-ppTG20-SPKR4	**10**	SPKR4-SPKR4

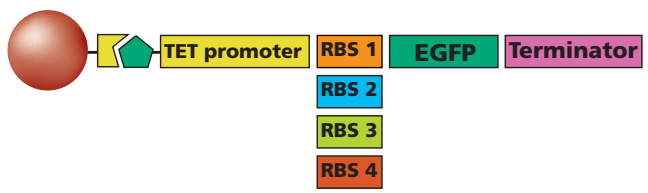


Figure 6 Using the solid-phase assembly of dsODN using DNA ligase, a series of EGFP expression devices was generated. A small library was created with the possibility of 4 different ribosome binding sites coding sequence parts between the promoter and EGFP coding sequence parts.

Figure 7a shows the generated Promoter-RBS_{mix}-EGFP-Terminator constructs using the solid-phase assembly of dsODN. The theoretical size of the library constructs ranges from 858-870 nucleotides and on gel product bands of these sizes were clearly visible. Also bands of intermediate products were visible.

After extraction and purification of the product bands between 800nt and 1000nt, the library constructs were amplified by PCR. The library constructs were amplified successfully and showed bands with expected lengths (Fig. 7b).

The PCR amplified Promoter-RBS_{mix}-EGFP-Terminator library was excised, purified and cloned into the BglII/HindIII-sites of the vector pIVEX-LacZ. *E.coli* strain XL-1 blue was transformed with the cloned library and the number of colonies obtained per transformation was approximately 200.

To check for insert size, restriction analysis using BglII/HindIII was performed on pDNA obtained from 39 colonies after overnight culturing, of which 20 are shown in Figure 7c. Two colonies contained empty vectors. In total, 25 colonies contained pDNA with inserts corresponding to the size of the Promoter-RBS_{mix}-EGFP-Terminator constructs.

To check for construct integrity, fluorescence of the same overnight cultured colonies was determined. A total of 30 overnight cultured colonies showed fluorescence levels above background. Within this group, large differences in fluorescence levels were observed.

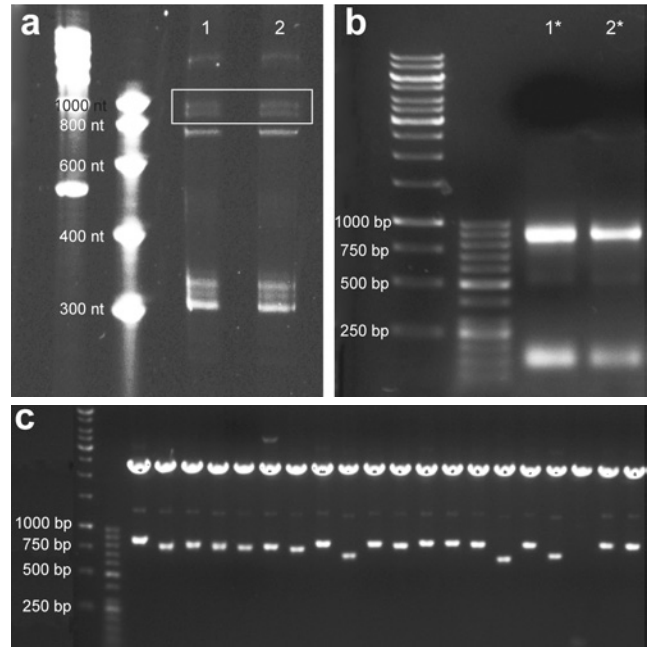


Figure 7 EGFP expression devices containing a small RBS library were created using solid-phase assembly of dsODN using DNA ligase.

(**a**) Initial ligation products were loaded on a 6% TBE urea gel and purified at final product height as indicated. First ladder: High range RNA ladder. Second ladder: low range RNA ladder. Lanes 1-2: Promoter-RBS_{mix}-EGFP-Terminator in duplo.

(**b**) Purified ligation products were amplified using PCR and loaded on a 2% agarose gel. Ladder: 50bp DNA ladder. Lanes 1*-2*: PCR amplified ligation products corresponding to lanes 1-2.

(**c**) Plasmid DNA isolated from 20 colonies was digested with BglII/HindIII and loaded on 0.7% agarose gel to confirm the presence of promoter-RBS-EGFP-Terminator coding sequence.

Table 6 Name, function and sequence of used modules

Module name	Function	Sequence (nt)
Head_pTET_J72005	TET promoter	AGCCAGATCTTCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCAC
BBa_B0034	RBS	AAAGAGGAGAAACACAT
BBa_B0035	RBS	ATTAAAGAGGAGAACACAT
BBa_J61141	RBS	TGGCTAACTGAGGATCACAT
BBa_K376001	RBS	TGGCTAACATAGGGTCACAT
Tail_Term_B0011	Terminator	AAAGAGAATATAAAAAGCCAGATTATTAATCCGGCTTTTTTATTATTTAAGCTTCACGC

To confirm seamless ligation and identification of the different RBS's, 21 out of the 39 obtained constructs were sequenced. Two constructs did not contain an RBS module and the corresponding colonies did not display fluorescence. One construct contained a correctly insert, however did not display fluorescence. The other constructs contained seamless and correct ligated modules which displayed fluorescence (Fig 8). All four different RBS's were found in the constructs and had similar occurrences. The differences between activity of the four RBS's is clearly visible in Figure 8; constructs with the K376001 RBS displayed the highest level of fluorescence and constructs with the J61141 RBS displayed the lowest level of fluorescence.

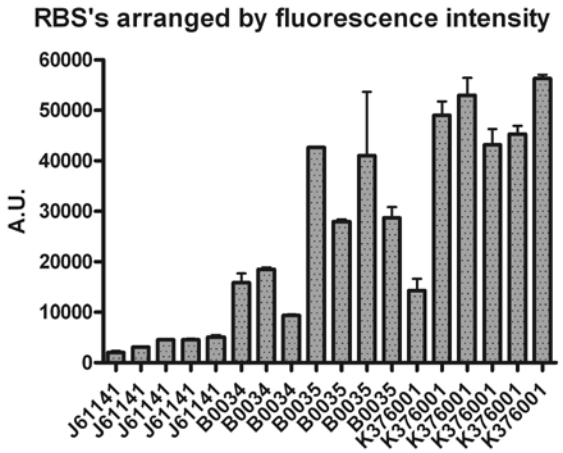


Figure 8 EGFP fluorescence of colonies containing the RBS library constructs, sorted by RBS.

Discussion and conclusion

Although many different methods have been developed for the assembly of multipart gene segments, these methods either introduce scar sequences during ligation or they rely on the use of predefined flanking regions for seamless ligation.

In theory, the solid-phase assembly of ssODN using RNA ligase is the ideal method for the creation of gene constructs. It is simple, it only requires (modified) single-stranded oligodeoxynucleotides and RNA ligase, it allows for seamless and sequence-independent ligation and it uses a modular approach. Also, through the use of a solid support platform, this method is suited for high-throughput usage.

Compared to methods based on restriction endonucleases, the advantages of solid-phase assembly of ssODN using RNA ligase is the ability for scarless ligation,

whereas methods using restriction endonucleases introduce at least a 2bp scar. Moreover, in contrast to methods based on complementary overhangs, solid-phase assembly of ssODN using RNA ligase is truly sequence-independent. Also, ssModules with repetitive or identical sequences can be used whereas methods based on complementary overhangs have difficulties with sequences homology (24). And without the need of restriction sites or flanking regions, module design is straightforward and modules can be re-used. However, methods based on complementary overhangs can assemble multiple parts in one-step and most restriction endonuclease methods are cheaper. Although our proposed ligation method requires 5' and 3' modifications of ODN, which makes this method more expensive, this can be compensated by using ssODN and the general applicability of the ODN.

Direct oligosynthesis methods, such as developed by Blue Heron Biotechnology, also use a solid support-based ligation-mediated assembly process to synthesize DNA (12,25). In each ligation round, the next section of dsDNA anneals to the previously assembled dsDNA, which is attached onto a solid support, through designed sequence overhangs and the new section of dsDNA can be ligated using DNA ligase. Using this technology, DNA fragments of 50bp to 25kb have been synthesized with high accuracy, repetitive sequences, and complex secondary structures (12). However, by using overhangs, re-use of dsDNA modules is not possible and the method becomes sequences depended.

Using solid-phase assembly of ssODN with RNA ligase, we were able to ligate 3 ssModules onto an ssHead in 3 separate assembly cycles. However, the ligation efficiency of the solid-phase assembly of single-stranded modules method is low, with an average of 9%. Due to overlapping retention times of the non ligated ssModule and the ssHead, the ligation efficiency of the 45 nt ssModule could not be determined (Supplementary Fig. 1). The low ligation efficiency may have several causes.

One is the formation of secondary structures in the used modules, which can restrict the accessibility of the 5' and 3' ends of the ssODN for ligation. In fact, we experienced that if the secondary structure involves the 5'- or the 3'-end of the oligonucleotide, ligation is impaired. This can also explain the observed differences in ligation efficiencies between the different modules. The secondary structure formation can be diminished by addition of DMSO, HCC or betaine (26,27). However, the addition of these reagents did not increase ligation efficiency in our hands (data not shown). Another strategy to diminish secondary structure formation is increasing the reaction temperature to >50°C. At present there is only one thermophilic RNA ligase commercially available (Epicentre, TRL8101K). Using this thermostable RNA ligase at 60°C, ligation efficiency increased (with an average of 53%), as do the costs. Secondary structure can also be diminished by altering the nucleotide sequence. Modules encoding proteins can be sequence optimized using different codons. However, the sequences of RBS's, promoters or other regulatory elements cannot be altered without changing their activity.

Besides the formation of secondary structures as a cause for the low ligation efficiency, it is known that RNA ligases have a lower intrinsic activity compared to DNA ligases.

Due to this, one assembly cycle took 16 hours. In order to assemble a construct containing 5 different modules, it takes 4 assembly cycles and thus 4 days. By increasing the assembly cycle time, ligation efficiency could be improved, but also the total assembly time is further increased.

To increase the ligation efficiency of solid-phase assembly of ssODN modules, other parameters may be investigated. Such parameters should be generally aimed at decreasing the random molecular motion in solution of the 5' phosphate and 3' hydroxyl groups and increasing the association of both groups. One could think of the use of nucleotide binding proteins or polymers and enhancement of molecular crowding by addition of PEG, dextrans or BSA. The addition of spermidine, a protein which binds nucleic acids, did not result in enhanced ligation efficiency (results not shown).

To improve the ligation efficiency of the solid-phase assembly of ssODN, a 12 nucleotide linker was introduced. The linker is complementary to the last 6 nucleotides of the immobilized ssModule and the first 6 nucleotides of ssModule to be ligated. The linkers support the assembly of a pre-ligation complex, which stabilizes the 3'-hydroxyl group and 5'-phosphate group of two ssModules and can be ligated by DNA ligase.

The use of linkers greatly improved ligation efficiency, up to an average of 93%, and therefore the assembly cycle time could be decreased to one hour. Compared to Dietrich et al., the ligation efficiency almost doubled, from 35-70% to 80-90% (17). As mentioned above, the ligation efficiency of the 45 nt ssModule could not be determined (Supplementary Fig. 1). However, by using linkers, the assembly method either becomes sequences depended or introduces scars. By using linkers containing the last 6 nucleotides of module A and the first 6 of nucleotides of module B, the construct AB can be joined seamlessly. For the seamless joining of construct BA, a different linker is needed. If one specific construct of 5 modules is needed, 4 different linkers are required. For generating a random library containing 5 different modules, 25 different linkers are required. However, the 12nt linker is shorter than the minimal 20bp overhang/known sequence required in methods for seamless fusion (including SLIC, CPEC and Gibson)(13-15) .

To circumvent this problem universal linkers can be used and random libraries can be constructed using only one linker. However, this will create a 12 nucleotide scar, which becomes a problem for libraries containing RBS's or other regulatory elements, which can be hampered by the introduction of scars, or if the ligated modules are part of a coding gene. Although the 12nt scar is longer than scars introduced by the BioBrick (8bp), BglBrick (6bp) and Golden Gate(4bp) assembly methods, the 12nt scar is sequence independent and can be re-used (7,9,10).

Also with the linker assembly method, secondary structures can hinder ligation. If the secondary structure involves the 5'- or the 3'-end of the oligodeoxynucleotide, the linker cannot bind to its complementary sequence and ligation is impaired.

In this paper we assembled multipart constructs using all three solid-phase assembly methods. Using solid-phase assembly of ssODN, we were able to ligate 3 ssModules onto an ssHead. Seven ssModules could be ligated onto a ssHead module using solid-phase assembly of ssODN with linkers and 6 dsModules could be ligated onto a dsHead module using solid-phase assembly of dsODN. Dietrich et al. only demonstrated the ligation of 4 dsModules onto a Head module (17).

For ssODN assembly, the low ligation efficiency hinders the formation of constructs with more than 3 ssModules (Fig. 3a; Supplementary Fig. 1, 3, 5, 7; Supplementary Table 1, 3, 5).

For ssODN assembly using linkers, after seven subsequent ligation cycles, a distribution is expected of 12%, 37% and 48% for the Head-M5, Head-M6 and Head-M7, respectively (based on 90% ligation efficiency). If the ligation efficiency decreases during the subsequent cycles, the distribution will shift towards lower ligation products. For the S and L module, 52% and 34% of the products contain at least 6 modules after seven ligation cycles, respectively (Fig. 4a; Supplementary Fig. 3, 5, 7; Supplementary Table 3, 5). This demonstrates that the decrease in ligation efficiency for subsequent steps is limited. However, for the T module, only 10% of the products contain at least 5 modules after seven ligation cycles, which shows a great loss in ligation efficiency during subsequent ligation steps (Fig. 4a; Supplementary Fig. 1; Supplementary Table 1). The loss of ligation efficiency seems module specific, as the 'intrinsic' ligation efficiency of the T module is already lower compared to the S and L module. The resolving power of the used SEC system was too low to visualize all expected products (7 assembly cycles could result in 13 different ligation products) and made it impossible to determine ligation efficiency for each individual cycle.

For dsODN assembly, the obtained distribution after six subsequent ligation cycles corresponded to the 50-60% ligation efficiency (Supplementary Fig. 2, 4, 6, 7; Supplementary Table 2, 4, 6). Again, the resolving power of the used SEC system was too low to visualize all expected products.

After gel extraction, the created constructs could be amplified using PCR. A bias towards the amplification of smaller constructs is observed. Separation of all ligation products is required in order to obtain only full length constructs. Here we used gel purification and extraction to obtain full length products, which is laborious and limits automation/high throughput. However, DNA separation techniques using chip-based nanostructures/nanomaterials or microfluidics allow automation and high throughput analysis (28,29). In order to avoid gel extraction, which costs about 16hr, excised bands can directly be used in PCR. Alternatively, desired constructs can be 'picked' from TBE-Urea gel using a needle and inserted into the PCR mix. Both

techniques resulted in the amplification of the desired products (Supplementary Fig. 8). Also, by using a needle, the bias towards the amplification smaller constructs can be decreased.

After demonstrating the creation of specific constructs, a library containing nine different ssModules was created using solid-phase assembly of ssODN modules with linkers (Fig. 5). After assembly and amplification, a smear pattern was visible indication the creation of many constructs with different sizes. Sequence analysis confirmed the ligation via the linker site and the presence of different modules. With the solid-phase assembly of ssDNA modules using linkers method, in only three days the library was created, amplified, ligated and transformed into bacteria.

To create a truly seamless, sequence-independent and fast solid-phase assembly method, we explored the use of double-stranded modules. The advantage of using dsODN is that secondary structures cannot hinder or impede the ligation reaction, which is a problem with methods based on complementary overhangs. The assembly of dsModules using a solid-phase platform was previously described by Dietrich *et al.* and we optimized the ligation efficiency and assembly cycle time of this method (17). Using our method, it took only one hour to ligate two dsModules, whereas the previous method required five hours. The ligation efficiency of the solid-phase assembly of dsModules (50%-60%) is comparable to the method of Dietrich *et al.* (35%-70%). We also demonstrate the use of PCR generated modules, which makes it possible to assemble large (>500bp) modules. We were successful in ligating a PCR generated module of 3100bp onto a dsHead module (results not shown).

To show the possibilities of the solid-phase assembly of dsModules, a small library of RBS's was screened on translation efficiency (Fig. 6). Sequence analysis confirmed the seamless ligation of the used modules and the presence of different RBS's in the constructs. Also, the translation efficiency of the four different RBS could be reviewed, where K376001>B0035>B0034>J61141.

With the solid-phase assembly of dsModules, in only four days the RBS library was created in a truly seamless and sequence-independent manner, amplified, ligated, transformed into bacteria and screened on fluorescence.

In conclusion, we have introduced a sequence-independent ligation method for combinatorial and scarless assembly of multipart gene constructs, which allows fast ligations of multiple fragments in less than a day and will be particularly useful for generation of combinatorial libraries of proteins with repetitive nature, such as TALEs or DARPins. This versatile ligation method will be a useful addition to the toolbox of molecular and synthetic biologists.

Abbreviations

bp: base pairs
ds: double-stranded
EGFP: enhanced green fluorescent protein
LB: Lennox broth
nt: nucleotides
ODN: oligodeoxynucleotides
PAGE: polyacrylamide gel electrophoresis
ss: single-stranded

References

(1) Gordeeva TL, Borschevskaya LN, Sineoky SP. Improved PCR-based gene synthesis method and its application to the Citrobacter freundii phytase gene codon modification. J Microbiol Methods 2010 May;81(2):147-152.

(2) Itaya M, Fujita K, Kuroki A, Tsuge K. Bottom-up genome assembly using the Bacillus subtilis genome vector. Nat Methods 2008 Jan;5(1):41-43.

(3) Blake WJ, Chapman BA, Zindal A, Lee ME, Lippow SM, Baynes BM. Pairwise selection assembly for sequence-independent construction of long-length DNA. Nucleic Acids Res 2010 May;38(8):2594-2602.

(4) Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A modular cloning system for standardized assembly of multigene constructs. PLoS One 2011 Feb 18;6(2):e16765.

(5) Sleight SC, Bartley BA, Lieviant JA, Sauro HM. In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res 2010 May;38(8):2624-2636.

(6) Ellis T, Adie T, Baldwin GS. DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol (Camb) 2011 Feb;3(2):109-118.

(7) Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, et al. BglBricks: A flexible standard for biological part assembly. J Biol Eng 2010 Jan 20;4(1):1.

(8) Zhu B, Cai G, Hall EO, Freeman GJ. In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. BioTechniques 2007 Sep;43(3):354-359.

(9) Engler C, Gruetzner R, Kandzia R, Marillonnet S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 2009;4(5):e5553.

(10) Shetty RP, Endy D, Knight TF,Jr. Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2008 Apr 14;2:5.

(11) Lu Q. Seamless cloning and gene fusion. Trends Biotechnol 2005 Apr;23(4):199-207.

(12) Xiong AS, Peng RH, Zhuang J, Liu JG, Gao F, Chen JM, et al. Non-polymerase-cycling-assembly-based chemical gene synthesis: strategies, methods, and progress. Biotechnol Adv 2008 Mar-Apr;26(2):121-134.

(13) Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA,3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009 May;6(5):343-345.

(14) Li MZ, Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 2007 Mar;4(3):251-256.

(15) Quan J, Tian J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 2011 Feb;6(2):242-251.

(16) Walker A, Taylor J, Rowe D, Summers D. A method for generating sticky-end PCR products which facilitates unidirectional cloning and the one-step assembly of complex DNA constructs. Plasmid 2008 May;59(3):155-162.

(17) Dietrich R, Wirsching F, Opitz T, Schwienhorst A. Gene assembly based on blunt-ended double-stranded DNA-modules. Biotechnol Tech 1998;12(1):49-54.

(18) Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 2007 Jul;35(Web Server issue):W43-6.

(19) Registry of Standard Biological Parts. Available at: www.partsregistry.org.

(20) Filipovska A, Rackham O. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. Mol Biosyst 2012 Mar;8(3):699-708.

(21) MacEwan SR, Chilkoti A. Elastin-like polypeptides: biomedical applications of tunable biopolymers. Biopolymers 2010;94(1):60-77.

(22) Mastrobattista B,van der Aa, Crommelin. Nonviral gene delivery systems: From simple transfection agents to artificial viruses. Drug Discovery Today: Technologies 2005;2(1):103-109.

(23) Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 2006 Aug;24(8):1027-1032.

(24) Hillson N. DNA Assembly Method Standardization for Synthetic Biomolecular Circuits and systems. In: Koeppl H, Setti G, di Bernardo M, Densmore D, editors. Design and Analysis of Biomolecular Circuits New York: Springer; 2011. p. 295.

(25) Hughes RA, Miklos AE, Ellington AD. Gene synthesis: methods and applications. Methods Enzymol 2011;498:277-309.

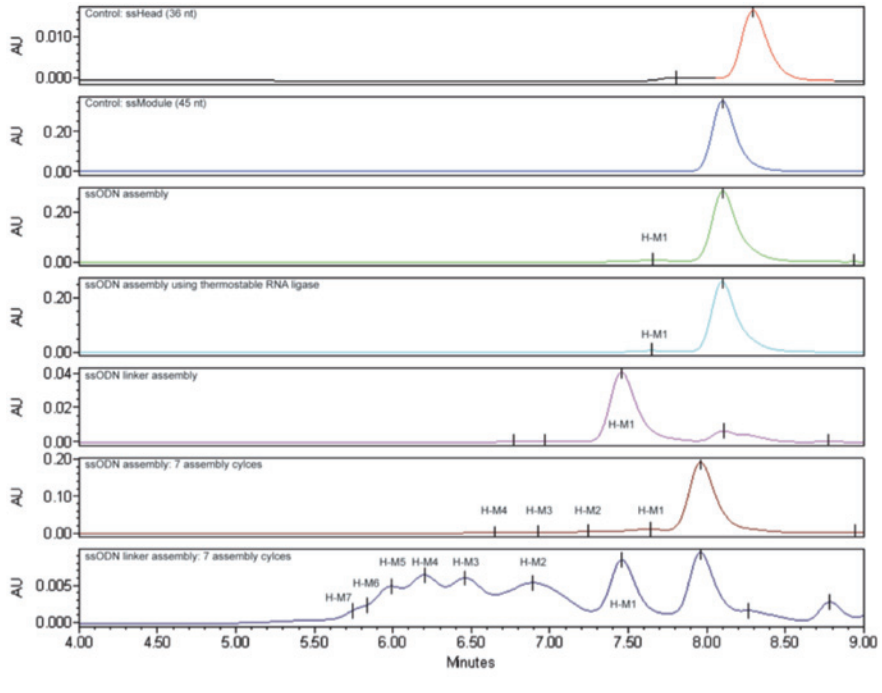
(26) Tessier DC, Brousseau R, Vernet T. Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. Anal Biochem 1986 Oct;158(1):171-178.

(27) Henke W, Herdel K, Jung K, Schnorr D, Loening SA. Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res 1997 Oct 1;25(19):3957-3958.

(28) Sinville R, Soper SA. High resolution DNA separations using microchip electrophoresis. J Sep Sci 2007 Jul;30(11):1714-1728.

(29) Lin YW, Huang MF, Chang HT. Nanomaterials and chip-based nanostructures for capillary electrophoretic separations of DNA. Electrophoresis 2005 Jan;26(2):320-330.

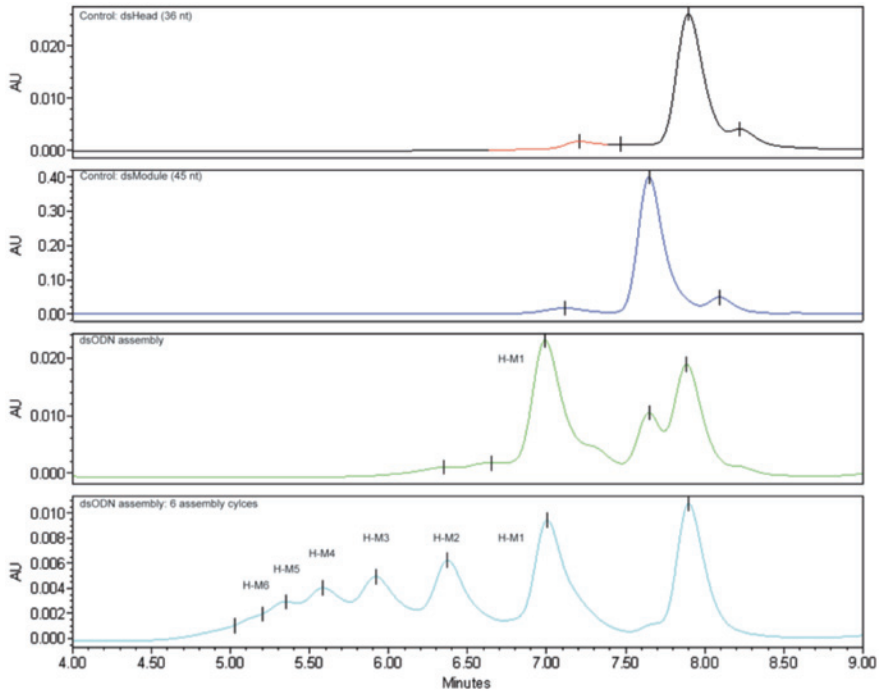
Supplementary material



Supplementary Figure 1 Chromatograms of ligation of the 45nt ssModule (T module) onto the 36nt ssHead in 1 or 6 subsequent ligation cycles using the ssODN assembly method and the ssODN linker assembly method. In the 7th ligation cycle, constructs were terminated with a 45nt Tail module. (H) corresponds to Head module and (M1-M7) corresponds to the number of ligated ssModules.

Supplementary Table 1 Distribution of ligation products. Contribution of ligation products were determined by dividing the peak area of the specific ligation product by the combined peak area of the non-ligated Head and ligated products.

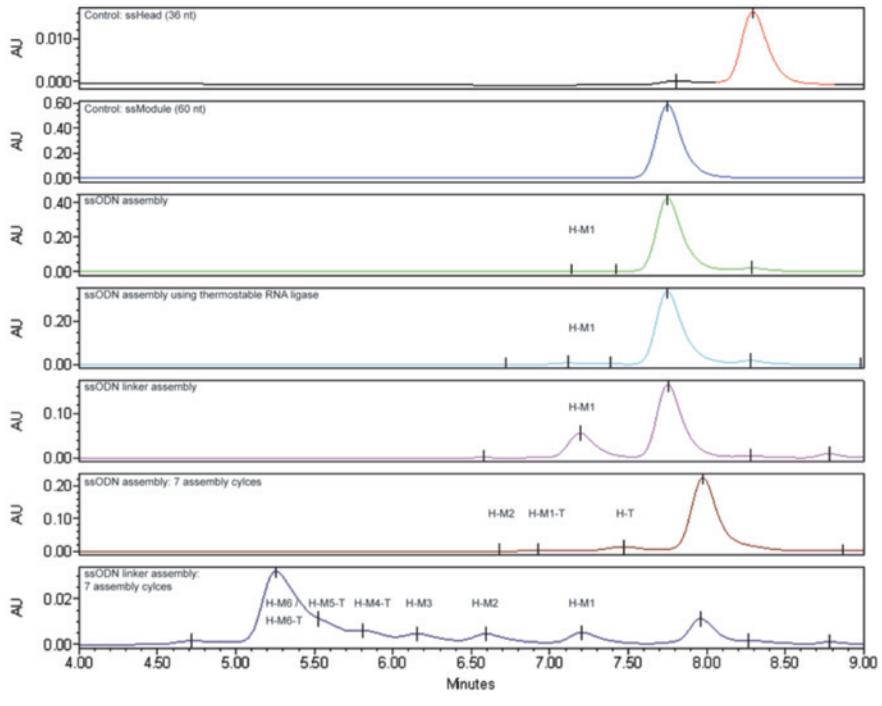
Assembly method	H	H-M1	H-M2	H-M3	H-M4	H-M5	H-M6	H-M7
ssODN assembly: 7 cycles	-	71%	23%	5%	-	-	-	-
ssODN linker assembly: 7 cycles	6%	22%	27%	17%	16%	9%	1%	2%



Supplementary Figure 2 Chromatograms of ligation of the 45nt dsModule (T module) onto the 36nt dsHead in 1 or 6 subsequent ligation cycles using the dsODN assembly method. (H) corresponds to Head module and (M1-M5) corresponds to the number of ligated dsModules.

Supplementary Table 2 Distribution of ligation products. Contribution of ligation products were determined by dividing the peak area of the specific ligation product by the combined peak area of the non-ligated Head and ligated products.

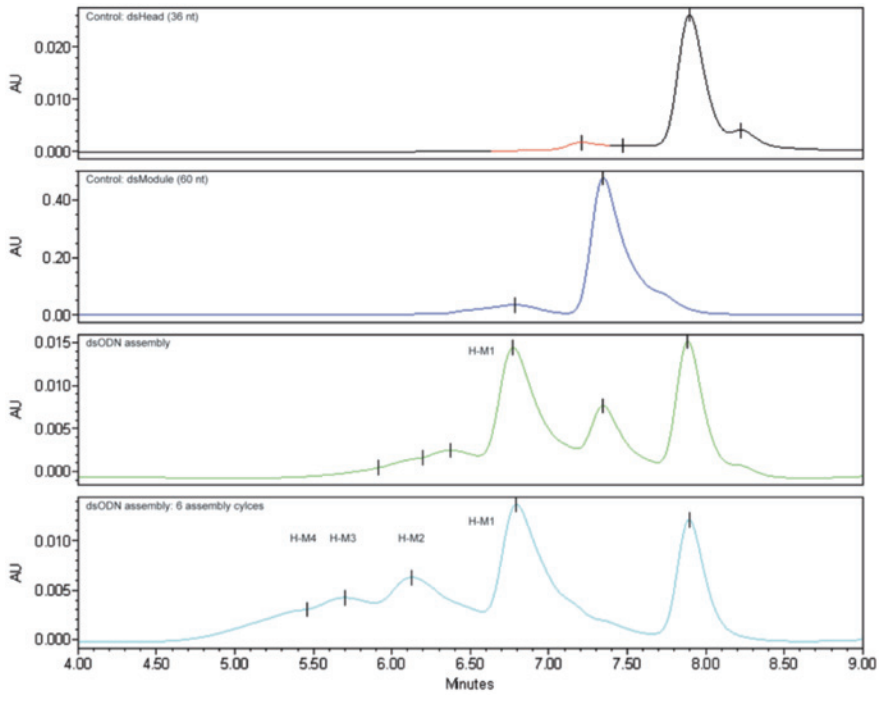
Assembly method	H	H-M1	H-M2	H-M3	H-M4	H-M5
dsODN assembly: 6 cycles	23%	25%	17%	14%	10%	10%



Supplementary Figure 3 Chromatograms of ligation of the 60nt ssModule (S module) onto the 36nt ssHead in 1 or 6 subsequent ligation cycles using the ssODN assembly method and the ssODN linker assembly method. In the 7th ligation cycle, constructs were terminated with a 45nt Tail module. (H) corresponds to Head module, (M1-M6) corresponds to the number of ligated ssModules and (T) corresponds to Tail module.

Supplementary Table 3 Distribution of ligation products. Contribution of ligation products were determined by dividing the peak area of the specific ligation product by the combined peak area of the non-ligated Head and ligated products.

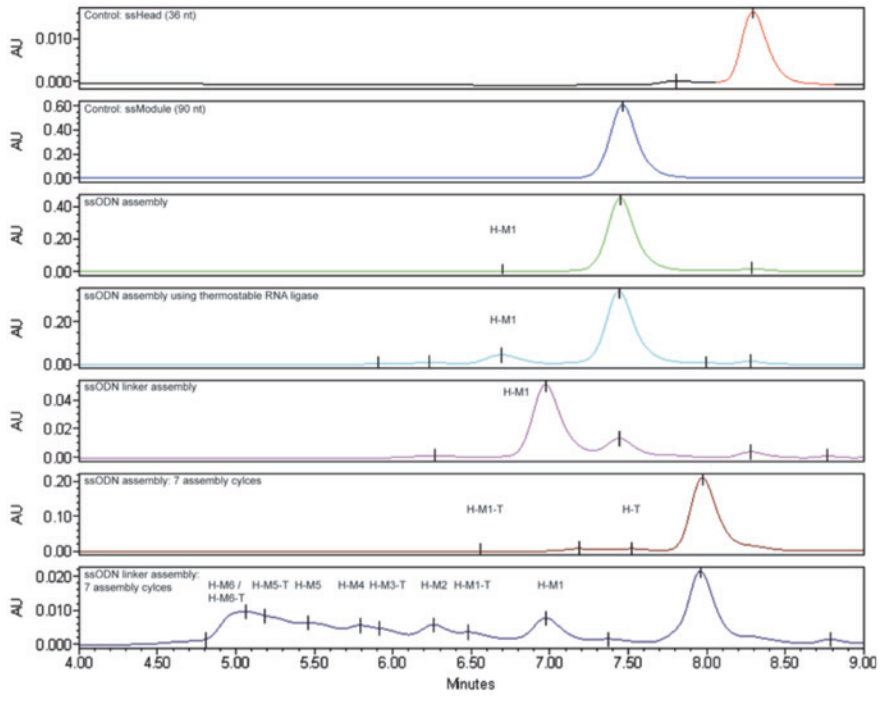
Assembly method	H	H-T	H-M1	H-M1-T	H-M2	H-M3	H-M4-T	H-M5-T	H-M6 / H-M6-T
ssODN assembly: 7 cycles	-	87%	-	13%	1%	-	-	-	-
ssODN linker assembly: 7 cycles	4%	-	9%	-	9%	8%	8%	10%	52%



Supplementary Figure 4 Chromatograms of ligation of the 60 nt dsModule (S module) onto the 36 nt dsHead in 1 or 6 subsequent ligation cycles using the dsODN assembly method. (H) corresponds to Head module and (M1-M4) corresponds to the number of ligated dsModules.

Supplementary Table 4 Distribution of ligation products. Contribution of ligation products were determined by dividing the peak area of the specific ligation product by the combined peak area of the non-ligated Head and ligated products.

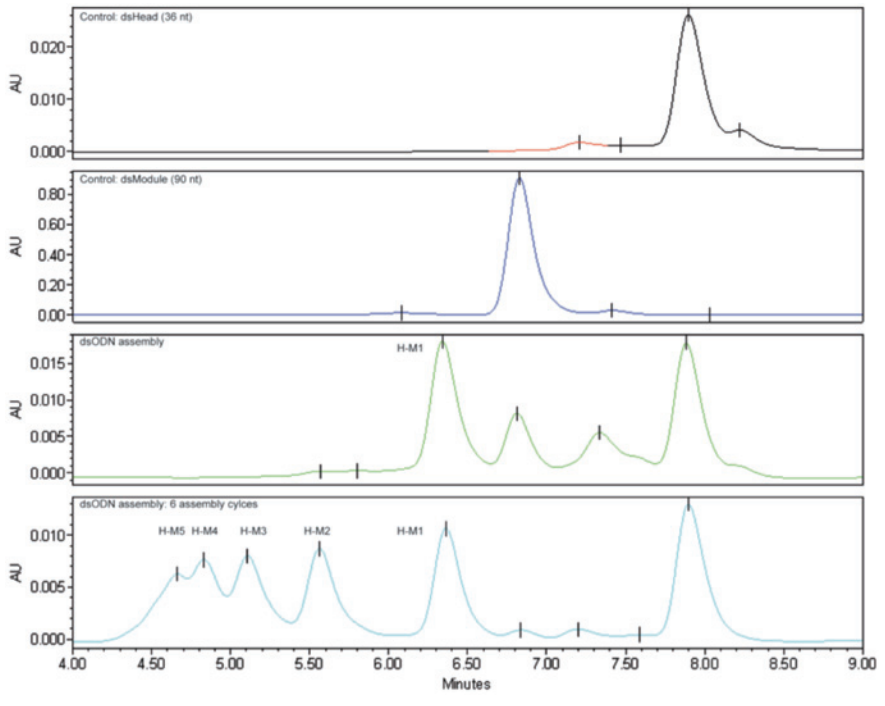
Assembly method	H	H-M1	H-M2	H-M3	H-M4	H-M5
dsODN assembly: 6 cycles	19%	38%	22%	12%	9%	-



Supplementary Figure 5 Chromatograms of ligation of the 90nt ssModule (L module) onto the 36nt ssHead in 1 or 6 subsequent ligation cycles using the ssODN assembly method and the ssODN linker assembly method. In the 7th ligation cycle, constructs were terminated with a 45nt Tail module. (H) corresponds to Head module, (M1-M6) corresponds to the number of ligated ssModules and (T) corresponds to Tail module.

Supplementary Table 5 Distribution of ligation products. Contribution of ligation products were determined by dividing the peak area of the specific ligation product by the combined peak area of the non-ligated Head and ligated products.

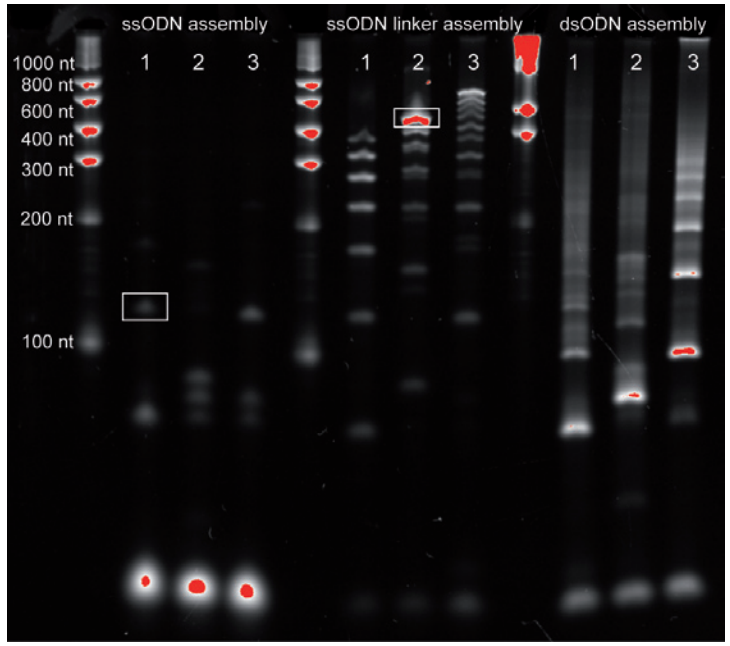
Assembly method	H	H-T	H-M1	H-M1-T	H-M2	H-M3-T	H-M4	H-M5	H-M5-T	H-M6 / H-M6-T
ssODN assembly: 7 cycles	-	49%	48%	4%	-	-	-	-	-	-
ssODN linker assembly: 7 cycles	-	4%	16%	7%	11%	6%	10%	12%	12%	22%



Supplementary Figure 6 Chromatograms of ligation of the 90nt dsModule (S module) onto the 36nt dsHead in 1 or 6 subsequent ligation cycles using the dsODN assembly method. (H) corresponds to Head module and (M1-M5) corresponds to the number of ligated dsModules.

Supplementary Table 6 Distribution of ligation products. Contribution of ligation products were determined by dividing the peak area of the specific ligation product by the combined peak area of the non-ligated Head and ligated products.

Assembly method	H	H-M1	H-M2	H-M3	H-M4	H-M5
dsODN assembly: 6 cycles	23%	19%	17%	16%	13%	12%

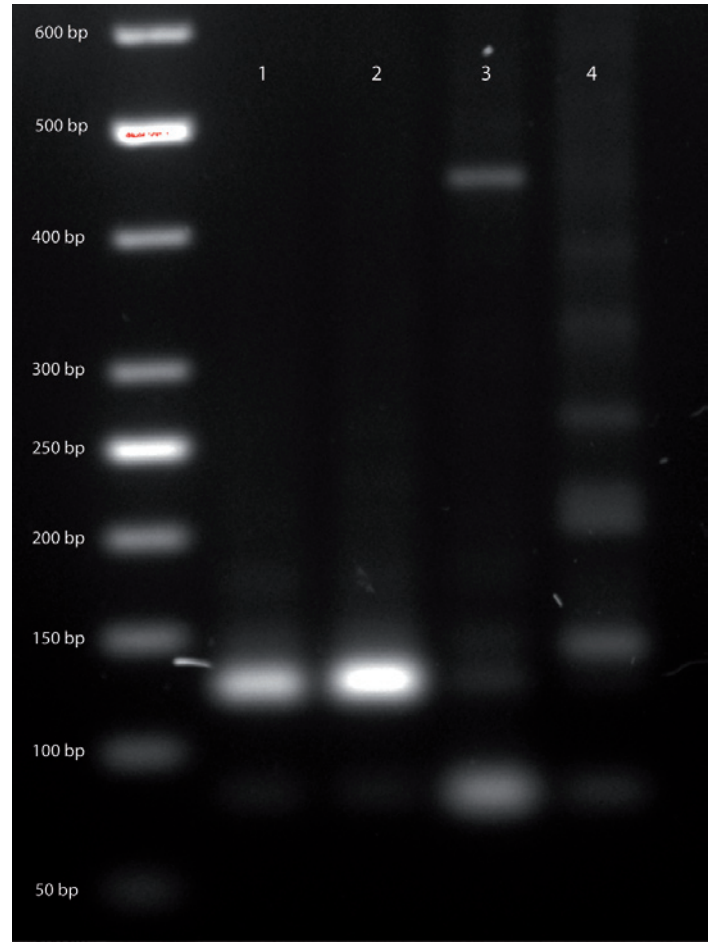


Supplementary Figure 7 Ligation efficiencies of subsequent ligation cycles of the three assembly methods. Ligations of 45nt (1), 60nt (2) and 90nt (3) ss- or dsModules onto 36 nt ss- or dsHead in 6 subsequent ligation cycles. For ssODN assembly and ssODN assembly using linkers, constructs were terminated with a 45nt Tail module in a 7th ligation cycle. Equal amounts of magnetic beads were loaded on a 6% TBE urea gel with: Ladder: Low range RNA ladder.

PCR amplification of ligation products directly from TBE-Urea gels

Of the obtained ligation products (20µL), 2-5µL was loaded on 6 and 10% TBE-urea gels. For 'product picking', a needle was used to pick the desired ligation product from gel and inserted into 50µL PCR reaction mix on ice. After 1min, reactions were preformed in PCR thermocycler.

For 'slice amplification', a scalpel was used to excise the desired ligation product from gel and the gel slice was inserted into 50µL PCR reaction mix on ice. Directly after, reactions were preformed in PCR thermocycler. Reaction products were analyzed on 2% w/v agarose gels stained with ethidium bromide. Bands were visualized with UV illumination at 302nm.



Supplementary Figure 8 PCR products loaded on 2% w/v agarose gel with a 50bp DNA ladder with a 50bp DNA ladder. Lanes 1 and 2: Head-T1-Tail amplified directly from 6% TBE-Urea gel using a needle (lane 1) or gel slice (lane2). Excepted product size: 126 bp. Lanes 3 and 4: Head-S6-Tail amplified directly from 6% TBE-Urea gel using a needle (lane 3) or gel slice (lane 4). Excepted product size: 441bp.

4

Chapter

High-Content Screening of Peptide-Based Non-Viral Gene Delivery Systems

Markus de Raad[a], Erik A. Teunissen[a], Daphne Lelieveld[b], David A. Egan[b] and Enrico Mastrobattista[a],

[a] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands
[b] Cell Screening Center, Department of Cell Biology, University Medical Center Utrecht, Heidelberglaan 100, 3584CX Utrecht, The Netherlands

Journal of Controlled Release, 158 (3), 433-442, (2012)

Abstract

High-content screening (HCS) uses high-capacity automated fluorescence imaging for the quantitative analysis of single cells and cell populations. Here, we developed an HCS assay for rapid screening of non-viral gene delivery systems as exemplified by the screening of a small library of peptide-based transfectants. These peptides were simultaneously screened for transfection efficiency, cytotoxicity, induction of cell permeability and the capacity to transfect non-dividing cells. We demonstrated that HCS is a valuable extension to the already existing screening methods for the *in vitro* evaluation of non-viral gene delivery systems. The added value of screening multiple parameters in parallel, thereby obtaining more information from a single screening event, will accelerate the development of novel gene delivery systems.

Introduction

Plasmid DNAs condensed into nano-sized particles using cationic lipids, polymers or peptides are by far the most extensively studied non-viral gene delivery vectors. Initial screening of such synthetic gene delivery systems is mainly performed *in vitro* on a set of cell lines using single-parameter (reporter gene expression) or dual-parameter (reporter gene expression and cell viability) assays. However, it has become clear that the outcome of transfection experiments is influenced by many different parameters which should therefore be carefully monitored, preferably simultaneously (1).

Currently used assays, which can be microtitre- or flow cytometry based, fall short on screening multiple parameters needed for proper screening of non-viral gene delivery systems *in vitro*. Microtitre-based assays yield averages of transfection efficiencies or cytotoxicities of a given cell population without the possibility to discriminate between differences per cell. An alternative screening method is flow cytometry, which allows high throughput, fluorescent monitoring of multiple events at the single-cell level, but can only be applied to cells in suspension and does not allow monitoring of morphological changes in cell populations or localization/tracking of gene delivery system at the subcellular level (2).

Besides the technical limitations of the used screening methods, many variables (e.g. ratio between DNA and transfectant, incubation time and incubation with/without serum) influence the transfection outcome. In order to minimize the influence of these variables, large standardized parallel screening protocols are needed to properly evaluate new gene delivery systems according to the current standards. Therefore, a high-throughput, multi-parameter screening method that can be standardized is highly desired. High-content screening (HCS) might be a good candidate for this. HCS uses high capacity automated fluorescence imaging for the quantitative analysis of single cells and cell-(sub)populations and it allows the *in vitro* monitoring of living cells in real time. Its main application is in the field of drug discovery. Compared to single or double-parameter based assays, HCS is capable of observing multiple cellular components simultaneously and can collect information from single cells (3). Also, HCS can image live cells non-invasively and allows for subcellular localization (if a confocal platform is used), which cannot be achieved by flow cytometry analysis. Furthermore, HCS can be fully automated which increases analysis speed, standardization and the ability to perform large library screenings (4). These features make the use of HCS in the screening of gene delivery systems very attractive.

Here we report the development of an HCS assay for the evaluation of gene delivery vectors. In order to evaluate and validate the HCS assay, a small library of peptide-based transfectants was used. The HCS assay was set up to screen multiple parameters in parallel, including cell number/relative viability, transfection efficiency,

cell permeability and cell division. Head-to-head comparison of these peptide transfectants revealed important differences between these peptides that would otherwise not have been noted.

By combining the combinatorial multipart gene assembly method described in Chapter 3 and the established HCS assay, we created the tools required for our proposed random, integrative design strategy that selects for multimodular peptides containing combinations of functional traits that are optimal for efficient gene transfer.

Materials and Methods

Materials

pCMV-LacZ plasmid DNA was purchased from the Plasmid Factory, Bielefeld, Germany. pCMV-EGFP plasmid DNA was constructed in house as previously described (5). pCMV-LacZ and pCMV-EGFP are eukaryotic expression plasmids encoding for β-galactosidase and enhanced green fluorescent protein (EGFP), respectively, under the control of the human cytomegalovirus promoter (CMV).

ExGen 500 (22kDA linear polyethylenimine (PEI)) was purchased from Thermo Fisher Scientific, St. Leon-Rot, Germany. Lipofectamine™ 2000, LIVE/DEAD® Fixable Red Dead Cell Stain Kit, anti-green fluorescent protein (GFP) antibody-Alexa Fluor® 594 conjugate (rabbit IgG fraction), DAPI and Click-iT® 5-ethynyl-2'-deoxyuridine (EdU) Alexa Fluor® 647 HCS assay were purchased from Invitrogen, Breda, The Netherlands. DMEM (Dulbecco's modified Eagle´s Medium, with 1g/l glucose, 584mg/ml L-Glutamine), Fetal Bovine Serum (FBS), and Phosphate Buffered Saline (PBS) were purchased from PAA Laboratories GmbH, Pasching, Austria. Chloroquine, paraformaldehyde, amphidicolin and Triton® X-100 were purchased from Sigma-Aldrich, Zwijndrecht, the Netherlands.

Peptides

All peptides were ordered with an N-terminal biotin at crude purity from Genscript, Piscataway, NJ, USA.

Lyophilized peptides were reconstituted in reversed osmosis water with/without the addition of 10-20% v/v acetic acid) to 2mM stock concentrations. Peptides were stored at -30°C.

Cell culture

COS-7 African Green Monkey kidney cells were grown in DMEM supplement with 5% heat inactivated FBS and antibiotics/antimycotics. Cells were maintained at 37°C in a 5% CO_2 humidified air atmosphere and split twice weekly.

Methods

Preparation of peptiplexes, polyplexes and lipoplexes

Complexes were prepared by adding 4 volumes of peptide/polymer (PEI)/lipid (Lipofectamine) solution to 1 volume of plasmid solution (50µg/ml) and mixing immediately by pipetting up and down 10 times. The complexes were incubated for 30 minutes at room temperature. Peptiplexes and polyplexes were prepared in HEPES-buffered saline (HBS, 20mM HEPES, 130mM NaCl, pH 7.4). Lipoplexes were prepared in DMEM. For each experiment, complexes were freshly prepared in triplicate. The ratios of peptide to plasmid DNA are expressed as weight to weight (w/w) ratios, whereas the ratio of PEI to plasmid DNA is expressed as the molar ratio of nitrogens within PEI to phosphates in plasmid DNA (N/P). The ratio of Lipofectamine to plasmid DNA is expressed as the volume of Lipofectamine (µl) to plasmid DNA (µg).

Transfections

Ten thousand COS-7 cells were seeded into black, clear bottom, 96-well tissue culture plates (Greiner Bio-One BV, Alphen a/d Rijn, The Netherlands) 24h prior to transfection. For transfection studies in non-dividing cells, 10,000 COS-7 cells were seeded 48h prior to transfection and medium was replaced by complete medium supplemented with 15µM aphidicolin 24h prior to transfection. Cells were continuously exposed to aphidicolin from at least 16h prior to transfection until time of analysis. Immediately prior to transfection the culture medium was refreshed with 100µl complete medium. Twenty five microlitres of the peptiplex/polyplex/lipoplex samples (corresponding to 0.25µg plasmid DNA/well) was added per well and after 4h incubation, medium was replaced with 100µl complete medium. Cells were incubated for indicated times at 37°C in a 5% CO_2 humidified air atmosphere. To study transfection of nondividing cells, aphidicolin was added to the culture medium prior to transfection at a final concentration of 15µM.

Fixation and nuclear staining

Forty eight hours post transfection, cells were washed once with 100µl PBS and fixed for 30min with 100µl 4% paraformaldehyde in PBS. After fixation, cells were washed once with 100µl PBS and nuclei were stained with 21.8µM DAPI in PBS. After DAPI staining, cells were washed twice with PBS. To each well, 100µl PBS was added and cells were analyzed using a Cellomics Arrayscan V HCS Reader.

Permeability Staining

To determine cell membrane permeability, cells were stained using the LIVE/DEAD© fixable dead cell stain with the red fluorescent reactive dye. After 4h incubation of the complexes with the cells, complexes were removed from cells. After 0h and 4h, cells

were washed with 100µl PBS, 100µl staining solution (1µl dye/1ml PBS) was added and cells were incubated at 37°C, 5% CO_2. After 30min staining solution was removed and cells were washed with 100µl PBS. Then, cells were fixed with 100µl 4% paraformaldehyde in PBS for 30min. After fixation, cells were washed once with 100µl PBS and nuclei were stained with 21.8µM DAPI in PBS. After DAPI staining, cells were washed twice with PBS. To each well, 100µl PBS was added and cells were analyzed by a Cellomics Arrayscan V HCS Reader. Statistical analysis was preformed using the 2-way ANOVA test.

Staining of EGFP transfected dividing and non-dividing cells

To confirm cell cycle arrest in the aphidicolin-treated cells, DNA synthesis was detected using the Click-iT® EdU Alexa Fluor® 647 HCS assay according to the manufacturer's protocol. After 4h incubation of the complexes with the cells, medium was replaced with 100µl complete medium supplemented with 5µM 5-ethynyl-2'-deoxyuridine (EdU) (and 15µM aphidicolin where applicable) and cells were incubated at 37°C in a 5% CO_2 humidified air atmosphere. Forty eight hours post transfection, cells were fixed for 30min with 100µl 4% paraformaldehyde and washed once with 100µl PBS. Cells were then permeabilized for 15min with 100µl 0.25% Triton X-100 in PBS and washed twice with 100µl PBS. Cells were incubated for 1h with 100µl 5% bovine serum albumin (BSA) in PBS at room temperature and washed twice with 100µl PBS. Fifty microliters of anti-GFP- Alexa Fluor® 594 in PBS was added to the cells and incubated for 1h at room temperature. Cells were washed once with 100µl PBS. After antibody staining, 100µl Click-iT® reaction cocktail was added and incubated for 30 min in the dark. Cells were washed once with 100µl Click-iT® reaction rinse buffer and nuclei were stained with 21.8µM DAPI in PBS. After DAPI staining, cells were washed twice with PBS. To each well, 100µl PBS was added and cells were analyzed by a Cellomics Arrayscan V HCS Reader. Statistical analysis was preformed using the 2-way ANOVA test.

High capacity automated fluorescence imaging and image analysis

A Cellomics Arrayscan V HCS Reader (Thermo Fisher Scientific, Waltham, MA, USA) was used to analyze the cells. For the analysis of the HCS data, the Cellomics algorithm Target Activation BioApplication (Thermo Fisher Scientific, Waltham, MA, USA) was used. The Target Activation BioApplication allows for measurements of up to 4 different fluorescent channels. One fluorescent channel is reserved for the identification of individual cells using major cellular organelles/compartments (e.g. nuclei or cytoplasm) and provides the basis for the intracellular region where intensity measurements are made.

In this study, single cells were identified according to the morphology of their nuclei using DAPI staining. After identification of single cells, the fluorescence intensity of the used fluorescent probes was measured in a predefined region. If the fluorescence intensity

is higher than the defined threshold, the single cell is 'positive' or a 'responder' for this fluorescent signal. From these values, the percentage of responders can be calculated.

Results

Setting up an HCS assay for measuring cell number and transfection efficiency

An HCS assay was configured that allowed simultaneous measurement of cell number (by way of fluorescent staining of the nuclei) and reporter gene expression for each individual cell in a 96-well format. For this, COS-7 cells were transfected with pDNA encoding either enhanced green fluorescent protein (EGFP) or β-galactosidase. For plasmid delivery, 22kDa linear PEI and Lipofectamine, two well-known transfectants, were used at ratio's which we previously found to be optimal (PEI: N/P 6; Lipofectamine: 2,5 µl/µg DNA). The cells were imaged on a Cellomics Arrayscan V HCS Reader with the criteria that 4 image fields were collected for each well and these were analyzed by the Target Activation BioApplication.

To measure cell number, cells were identified according to their nuclear morphology (Fig. 1a and b). Viable cells have a nuclear morphology that differs from that of dead or dying cells. In addition, dead cells often detach from tissue culture plates during washing steps. As a consequence, only the viable cells will be counted, which makes cell number a good measure for potential toxic effects due to expose to transfectants. The number of cells counted are shown in Figure 2a. When the total number of cells after transfection is expressed relative to the number of cells of non-treated cells, the relative cell viability is obtained (Fig. 2b). To determine the transfection efficiency, EGFP expressing cells were identified among all the detected cells (Fig. 1c and d). Transfection with PEI-EGFP and Lipofectamine-EGFP resulted in 10% and 15% EGFP positive cells, respectively, whereas cells which were transfected with PEI-LacZ and non-treated cells were not positive for EGFP fluorescence (Fig. 2c).

Two-parameter screening of a peptide library: Relative viability and transfection efficiency

To evaluate the HCS assay for the screening of a small library of peptide-based transfectants for cytotoxicity and transfection efficiency, COS-7 cells were transfected using 70 different DNA condensing peptides with proven transfection capacities (Table 1).

Peptides were diluted and mixed with plasmid DNA encoding EGFP in HBS at different w/w ratio's and COS-7 cells were transfected and analyzed with ArrayScan HCS Reader using the above described Target Activation BioApplication. In Figure 3 the transfection and cytotoxicity results of 14 peptides are displayed, but trends are similar for all 70 peptides (see Supplementary Figures 1 and 2).

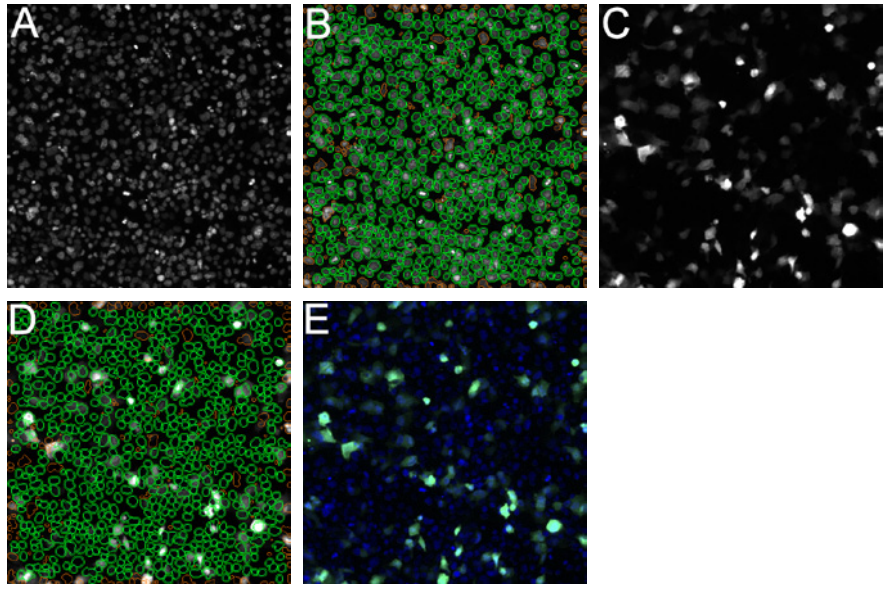


Figure 1 Cell number and transfection efficiency measurements using the Target Activation BioApplication. COS-7 cells were incubated for 4h with 22kDa linear PEI and pCMV-EGFP complexes at N/P ratio of 6. Fourty eight hours after transfection, cells were imaged on a Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. Nuclei were stained with DAPI (a) and cells were identified using their nuclear morphology (b). Included cells are shown outlined in green, rejected cells in red. The % EGFP expressing cells (c) was determined using the identified cells (d). (e) Overlay of nuclei (depicted in blue) and EGFP expression (depicted in green).

Overall, peptides showed little to no cytotoxicity with relative viability between 75% and 100%, with the exception of peptides Hel 11-17Δ1 (LLKLLLKLWKKLLKLLK) and Hel 11-17Δ2 (LKLLLKLWKKLLKLLK), which showed high toxicity at w/w ratios >16. With increasing w/w ratios, toxicity of the peptide complexes increased. In total, 65 peptides showed a transfection efficiency < 0.5% of total cells, 3 peptides showed between 0.5 and 1% , 2 peptides between 5-10% and the positive controls > 10%. The 5 peptides which showed transfection efficiency were: LAH3 (KKALALGLHLAHLALHLALALKKA) (0.5%), ppTG20 (GLFRALLRLLRSLWRLLLRA) (0.6%), LAH4-L3 (KKALLALALHHLALLAHHLALALKKA) (0.8%), LAH4 (KKALLALALHHLAHLALHLALALKKA) (5%) and LAH5 (KKALLALALHHLAHLAHHLALALKKA) (8%). LAH4 and LAH5 showed significant transfection at their optimal w/w, 16 and 32 respectively (1-way ANOVA, P <0.001).

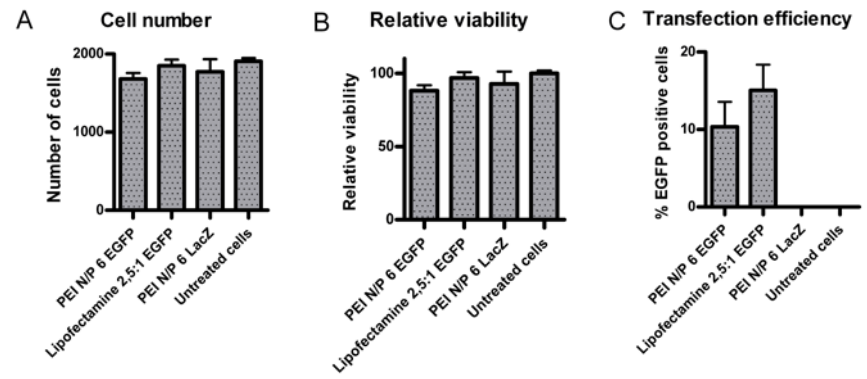


Figure 2 Cell number (a), relative viability (b) and transfection efficiency (c) measured using HCS. COS-7 cells were incubated for 4h with PEI or Lipofectamine complexes. Fourty eight hours after transfection, cells were imaged on the Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. Data are presented as mean+SD. N=3.

Table 1 DNA condensing peptides with proven transfection capacities according to literature

Peptide	Ref	Peptide	Ref	Peptide	Ref
K_8	(6)	CWKKKKKKKKKCWK-KKKKKKKK (CWK2)	(7)	SRSRYYRQRQRSRRRRRR (Pr18)	(8)
K_9	-	CYK-KKKKKKKKKKKKKKKK	(9)	RRRLHRI-HRRQHRSCRRRKRR (Pr21)	(8)
K_{10}	(8)	CWK-KKKKKKKKKKKKKKKC (CWK3)	(10)	PRRRRSSSRPIRRRRPR-RASRR (LMWP1)	(11)
K_{12}	(8)	CWKKKKKKKKCK-KKKKKKKC (CWK4)	(10)	RRRRPRRVSRRRRRRG-GRRRR (LMWP2)	(11)
K_{14}	-	CWKKKKKCKKKKKCK-KKKKC (CWK5)	(10)	KKLALALALALHALALA-LALKKA (LAH)	(12)
K_{16}	(13)	CWKKKKCKKKCKKKCK-KKKC (CWK6)	(10)	KKLAHLALALAL-GLALAHLAKKA (LAH2)	(12)
K_{18}	(8)	R_6	(14)	KKALALGLHLAHLALHLA-LALKKA (LAH3)	(12)
K_{20}	(9)	R_8	(14)	KKALLALALHHLAHLAL-HLALALKKA (LAH4)	(12)
KKKKKKKKKGGC	(15)	R_9	(16)	KKALLALALHHLALLAHHLA-LALKKA (LAH4-L3)	(12)
KWKKKWKKKGGC	(15)	R_{10}	(14)	KKALLALALHHLALLAHL-LALHLKKA (LAH4-L4)	(12)

Table 1 Continued

Peptide	Ref	Peptide	Ref	Peptide	Ref
RRRRKGGC	(15)	R_{12}	(14)	KKKKALAHLHALAAHLHA-LAAAALKKK (LAH4-A6)	(12)
RRRRRRKGGC	(15)	R_{15}	(16)	KKALLALAL-HHLAHLAHHLALALKKA (LAH5)	(12)
RRRRRRRRKGGC	(15)	R_{16}	(14)	KKSAKKTPKKAKKP (H14)	(8)
RWRRRRWRRRKGGC	(15)	R_{18}		KTPKKAKKPKTPKKAKKP (H9-2)	(8)
CKKKKC	(17)	LKALLKALLKAL ($LKAL_3$)	(18)	KTAKKAKKAKTAKKAKKA (H9-2A)	(8)
CKKKKKKC	(17)	KLLKLLLKLWKKLLKLLK (Hel 11-17)	(18)	SPKRSPKRSPKRSPKR ($SPKR_4$)	(19)
CKKKKKKKKC	(17)	LLKLLLKLWKKLLKLLK (Hel 11-17Δ1)	(18)	YSPTSPSY (YY1)	(20)
CKKKKKKKKKKC	(17)	LKLLLKLWKKLLKLLK (Hel 11-17Δ2)	(18)	YSPTSPSYSPTSPSY (YYY2)	(20)
CKKKKHKKKC	(17)	GLFKALLKLLKSLWKLLL-KA (ppTG1)	(21)	MAPKRKSGVSKCETKCTPP (VP1)	(22)
CHKKKKKKHC	(17)	GLFRALLRLLRSLWRLLL-RA (ppTG20)	(21)	ILPWKWPWWPWRR (Indolicidin)	(23)
CHKKKHKKHC	(17)	AAVVLLLWEE (SA2)	(24)	GLFKAIAKFIKGGWKGLIKG (K5)	(23)
CHKHKHKHKC	(17)	AAVVLLLWRR (SA2RR)	-	VRLPPPVRLPPPVRLPPP ($VRLPPP_3$)	(25)
CWK-KKKKKKKKKKKKKKKK (CWK)	(7,9,10)	AAVVLLLWKK (SA2KK)	-	RQIKIWFQNRRMKWKK (Penetratin 43–58)	(26, 26-29)

Screening for a third parameter: Cell permeability

Cell permeability was measured simultaneous with cell number and transfection efficiency. To measure this, cells were stained with the fluorescent LIVE/DEAD® stain (Fig. 4). This fluorescent stain can cross compromised membranes and react with free amines in the interior of the cell, which results in intense fluorescent staining, whereas cells with intact membranes are only dimly stained (30)(30). The emission spectrum of the LIVE/DEAD fluorescent stain is in the red spectrum (Em 615nm), without overlap with DAPI and EGFP spectra, thus allowing simultaneous detection of cell number, transgene expression and permeability.

The percentages of cells stained with LIVE/DEAD© stain directly or 4h after exposure to PEI, Lipofectamine and peptides ppTG20 and LAH5, are shown in Figure 5.

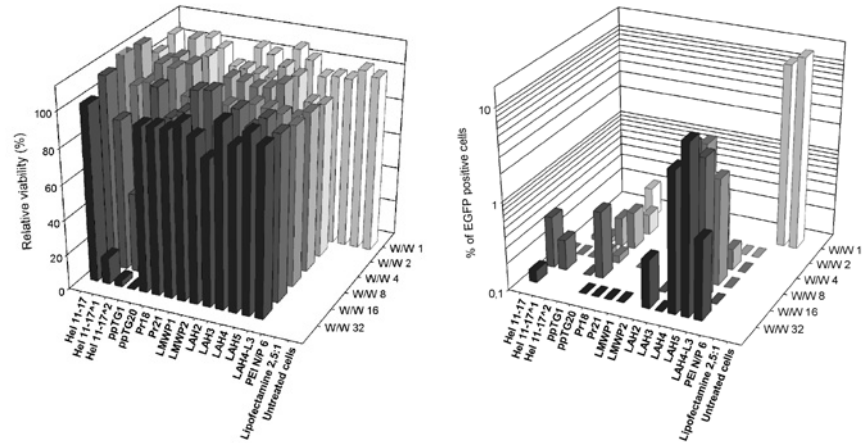


Figure 3 Relative viability (a) and transfection efficiency (b) of 14 DNA condensing peptides. COS-7 cells were incubated for 4h with peptide complexes and 48h after transfection, cells were imaged on the a Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. Positive controls (PEI and Lipofectamine) were tested at their optimal DNA/transfectant ratios (see M&M). N=3.

Similar trends in permeability were observed between directly removed complexes and removal after 4h, with no significant differences. PEI, lipofectamine and both peptides showed different permeability patterns. PEI induced plasma membrane permeability in about 50% of cells, irrespective of the DNA/PEI ratio, whereas with Lipofectamine, the percentage permeable cells increased with increasing ratios. Both peptides displayed the highest percentage of permeable cells (up to 70%).

Screening for a fourth parameter: Transfection of dividing and non-dividing cells

To further expand the HCS assay, an application was developed to detect cells which had divided during the transfection process in order to determine transfection efficiencies of divided vs nondivided cells separately. For this, cells were stained with EdU (Fig. 6). EdU is incorporated into the DNA during active DNA synthesis in the S-phase (31). In non-dividing cells EdU incorporation occurs but at a much lower level, which allows discrimination between divided and non-divided cells based on fluorescent detection of EdU incorporation. As a control for non-dividing cells, cells were treated with aphidicolin to arrest the cell cycle in the early S-phase.

The emission spectrum of EdU Alexa Fluor® 647 fluorescent stain is in the near IR spectrum (Em 670nm), without overlap with DAPI, EGFP and anti-GFP spectra, thus allowing simultaneous detection of cell number, transgene expression and cell division.

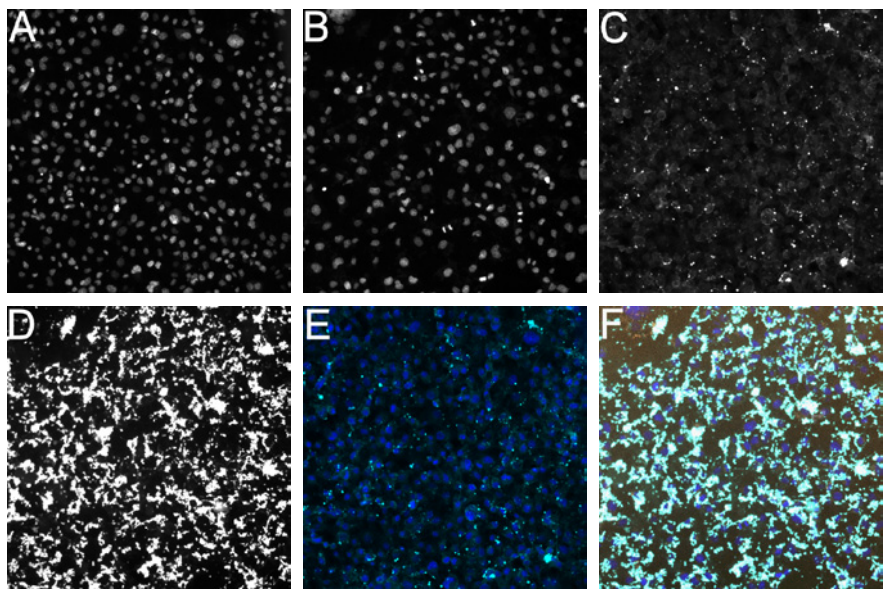


Figure 4 Cell permeability measurements. Images of COS-7 cells untreated (A,C,E) or transfected with ppTG20 at w/w 16, 4h post transfection. COS-7 cells were incubated for 4h with peptide complexes and 4h after transfection, cells were stained with LIVE/DEAD© stain and nuclei were stained with DAPI. Cells were imaged on the Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. a&b: DAPI staining. c&d: LIVE/DEAD© staining. e&f: Composite with nuclei in blue and LIVE/DEAD© staining in green.

PEI, Lipofectamine and the peptides ppTG20 and LAH5 were used for transfection and the amount of cell division was determined 48h after transfection. Figure 7 shows that almost all cells were EdU positive, whereas the aphidicolin-treated cells only showed 20-25% EdU positive cells.

Next, the percentage of transfected cells within the non-divided (EdU negative) and divided (EdU positive) cells was determined (Fig. 8a). The positive controls PEI and Lipofectamine showed transfection efficiencies of 4% and 10%, respectively, whereas the peptides ppTG20 and LAH5 showed transfection efficiencies of about 1%.

The percentage of non-divided cells that had been transfected was low (Fig. 8b). Only peptide LAH5 was able to transfect non-dividing cells.

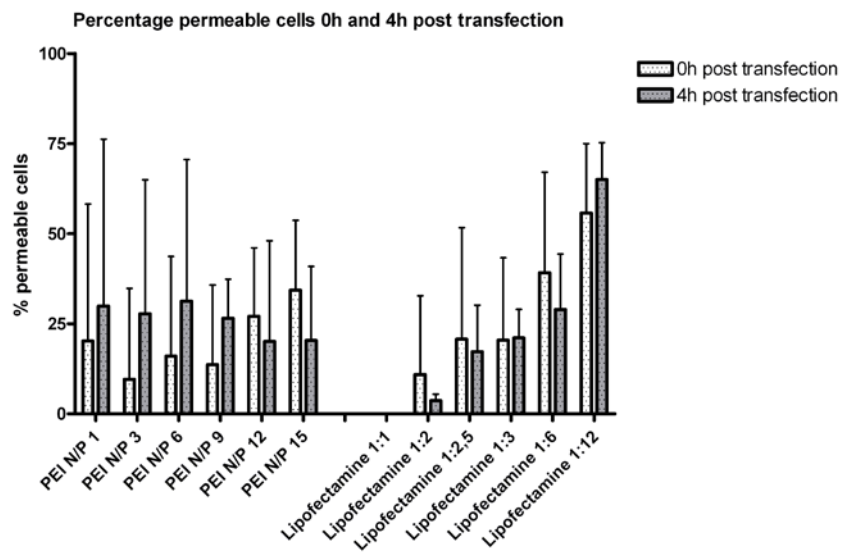


Figure 5 Percentage permeable COS-7 cells 0h and 4h post transfection (with background subtraction) using PEI, Lipofectamine and peptides ppTG20 and LAH5 complexed to DNA at the indicated w/w ratios. COS-7 cells were incubated for 4h with poly-, lipo- and peptide complexes and 0h and 4h after transfection, cells were imaged on the Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. No significant difference between 0h and 4h for PEI, Lipofectamine, ppTG20 and LAH5 (2-way ANOVA). Data are presented as mean+SD. N=3.

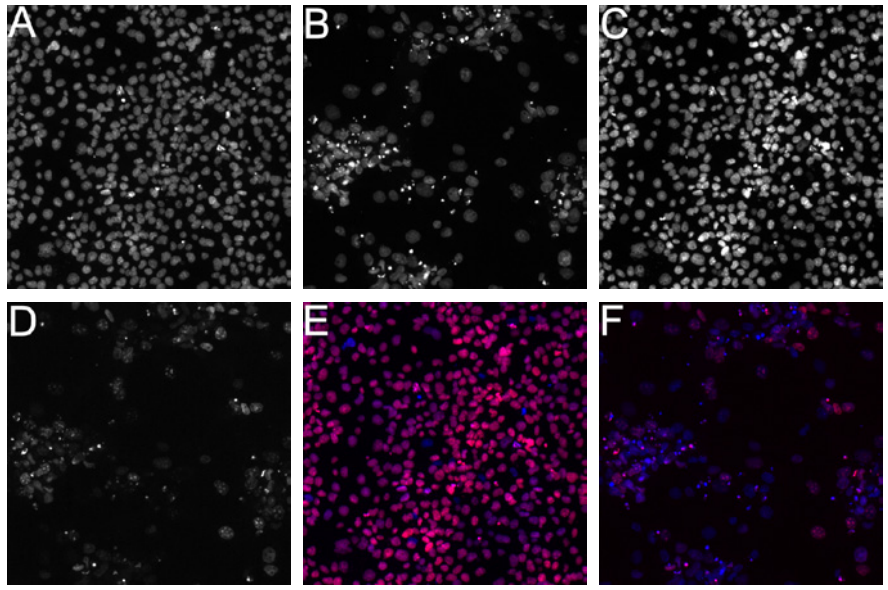


Figure 6 Dividing and non-dividing cells measurements using the Cellomics Arrayscan V HCS Reader. Images of COS-7 cells in the absence (a,c,e) or presence (b,d,f) of aphidicolin 48h post transfection using PEI, Lipofectamine and peptides ppTG20 and LAH5. After COS-7 cells were incubated for 4h with poly-,lipo- and peptide complexes, medium was replaced with complete medium supplemented with 5µM 5-ethynyl-2'-deoxyuridine (EdU) (and 15µM aphidicolin where applicable). Fourty-eight hours after transfection, nuclei were stained with DAPI and cells were imaged on the Cellomics Arrayscan V HCS Reader using the Target Activation BioApplication. a&b: DAPI staining. c&d: EdU staining. e&f: Composite with nuclei in blue and EdU positive cells in purple.

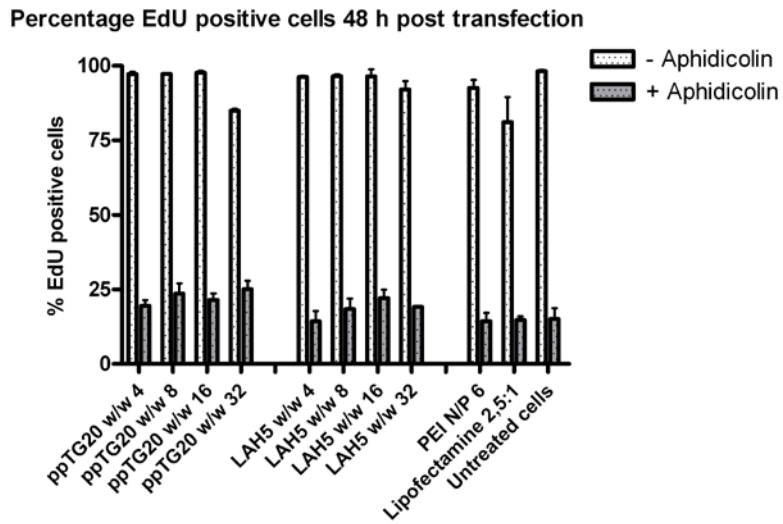


Figure 7 Percentage EdU positive COS-7 cells in the presence (white bars) or absence (gray bars) of aphidicolin 48h post transfection using PEI, Lipofectamine and peptides ppTG20 and LAH5. COS-7 cells were incubated for 4h with poly-, lipo- and peptide complexes and 48h after transfection, cells were imaged on the Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. Data are presented as mean+SD.N=3.

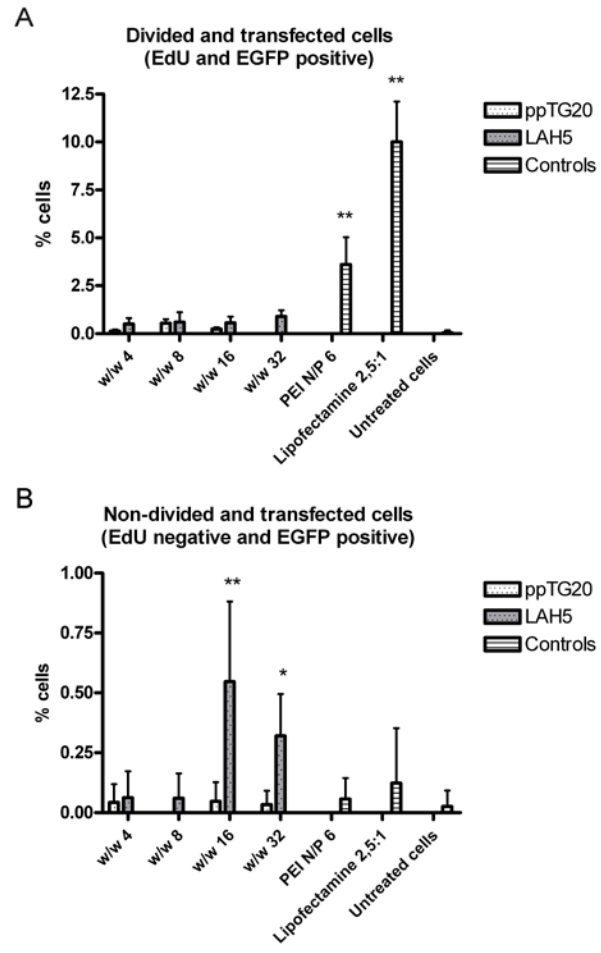


Figure 8 Relative number of transfected COS-7 cells categorized by divided and non-divided. COS-7 cells were transfected in the absence (a) or presence (b) of aphidicolin using PEI, Lipofectamine and peptides ppTG20 and LAH5. Forty-eight hours after transfection, cells were imaged on the Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. The effect of PEI and lipofectamine on dividied & transfected cells and the effect of LAH5 on non-divided & transfected cells is significant at w/w ratios 16 and 32 (2-way ANOVA, *P< 0.05; **P <0.001). Data are presented as mean+SD. N=3.

Discussion and conclusion

In this paper we show that automated fluorescence microscopy can be used for rapid screening of peptide-based transfectants *in vitro*. The availability of high-throughput methods is paramount to the initial screening of novel gene delivery systems as it allows rapid identification of potential candidates for *in vivo* gene delivery applications out of large libraries of transfectants. Current screening methods are often optimized for speed thereby limiting the number of parameters that can be screened to 1 (reporter gene expression) or 2 (reporter gene expression and cell viability). With the HCS method proposed here it is possible to screen for multiple parameters in parallel, which are relative cell viability, transfection efficiency, cell permeability and cell division.

The relative cell viability was obtained by automatic cell counting after fluorescent labeling of the nucleus. Validation of this system with two common transfection systems, PEI and Lipofectamine, showed that the cell viability results obtained with HCS are in good accordance with standard metabolic activity assays such as MTT or XTT assays (results not shown).

The transfection efficiency was determined by direct measurement of EGFP expression in single cells. The advantage of single cell-based analysis over batch analysis is that it yields additional information on the distribution of transgene expression across the cell population, and enables discrimination between high transgene expression in a few transfected cells and low transgene expression over a larger population of cells. The transfection efficiencies obtained with the HCS assay are comparable to results obtained using standard reporter assays (results not shown).

Non-viral gene delivery systems often consist of polycationic molecules required for DNA condensation. Such molecules can adhere to cell membranes which may lead to increased permeability and cytotoxicity. Information about transfectant-induced cell permeability is therefore desired to predict which transfection systems may cause toxicity. By using LIVE/DEAD© stain, which can cross compromised membranes, the induction of cell permeability can be determined. Both tested peptides ppTG20 and LAH5 induced more permeable cells than PEI and Lipofectamine controls. It is known that both peptides can interact and penetrate membranes (12,21).

An often mentioned drawback of non-viral gene delivery systems is their limited capacity to transfect non-dividing cells (32). Dividing cells are much easier to transfect as the temporary breakdown of the nuclear envelope allows access of DNA complexes into the nucleus whereas active transport is needed in the absence of mitosis. In contrast to cell lines in culture, which often divide rapidly, the majority of cells in our body are terminally differentiated and quiescent. An *in vitro* assay to determine the capacity of a gene delivery reagent to transfect non-dividing cells is therefore desired.

By adding EdU to the cells directly after transfection, we were able to identify cells that have passed the S-phase of the cell cycle and therefore are likely to have gone into mitosis. In this way, the transfected cells could be categorized in divided vs. non-divided cells, which clearly demonstrated that transfection levels are highest in cells that have divided. By focusing on the non-divided cells we could demonstrate that the LAH5 peptide showed significantly enhanced transfection of COS-7 cells compared to Lipofectamine or PEI controls, although the absolute transfection level was rather low. This can be caused by the quality of the plasmid DNA used or by differences in the quality of the cell batches. Clearly, transfection of non-dividing cells represents a major bottleneck and should be part of screening procedures for gene delivery agents with final applications *in vivo*.

The expansion of the HCS assay with cell permeability or cell division demonstrates that multi-parameter screenings are possible. Although we only tested the system screening 3 parameters simultaneously, it is possible to further expand the assay with an extra parameter, because only 4 of the 5 channels of the Cellomics Arrayscan V HCS Reader were used. Besides the addition of extra parameter(s), also other algorithms can be used to measure a wide variety of targets and cellular processes. Interesting applications for the field of gene delivery are cellular receptor binding, internalization, endosomal escape, nuclear-cytoplasmic translocation and particle localization. For example, Rawlinson et al. used HCS for analysis of the cytotoxicity of pDMAEMA in a seven-parameter assay (33).

HCS allows for the imaging of live cells, since temperature and CO_2 levels can be monitored and controlled. This allows for long-term analysis and individual cell tracking over time (34)(34). However, some cellular imaging assays require fixation and/or permeabilization of cells. In this paper, fixation was required for detection of permeable cells and fixation and permeabilization was required in order to distinguish dividing and non-dividing cells. For comparison, all experiments were preformed with fixed cells.

Flow cytometry can also detect multiple parameters at the single cell level using fluorescent techniques, but as this requires the use of cells suspension spatio-temporal and morphological information of single cells will be lost. With HCS, this information is available and thus enables the measurement of changes in morphology or location of single cells in time (35). Also, HCS systems can be equipped with a confocal fluorescence microscope which enables the localization of gene delivery systems at the subcellular level (3). However, HCS requires the use of adherent cells and cells which have contact inhibition growth characteristics, otherwise overlapping cells will complicate the analysis.

Another advantage of HCS compared to standard 2-parameter read-out assays and flow cytometry is automation. HCS uses both automated image acquisition and image analysis, which decreases analysis time. Furthermore, with the use of robotics

the entire analysis process can be automated, which makes it a high throughput system. For example, the Cellomics Arrayscan V HCS Reader used in this paper was equipped with a robotic arm that could automatically load, measure and re-load well-plates. Most plates were analyzed in the evening/at night and about 20 plates (20 x 96 = 1920 wells) could be analyzed in a single run (4,36).

Standardization of *in vitro* screenings of non-viral gene delivery systems is very important as there are many experimental parameters that can greatly affect the outcome of such screenings (1). To demonstrate the importance of standardized screenings we have randomly selected 70 different peptide-based transfectants from the literature that were reported to have good transfection efficiencies on cells in culture. However, since these experiments were performed on different cell types, and under different experimental conditions, it is difficult, if not impossible to make a direct comparison of these peptide-based transfectants and rank them according to their transfection efficiency. Therefore, we have used our HCS method to screen these peptides under standardized conditions for comparability. The screening conditions were chosen in compliance with the proposed standard protocol for transfection screening which consists of a 4h incubation of transfectants with cells in culture in the presence of serum followed by reporter gene read out 48h after transfection. The results are striking: although all peptides were reported to have good transfection efficiencies that are comparable to the included positive controls, we only observed transfection of COS-7 cells in 5 out of the 70 peptides screened under the tested conditions. This difference with previously reported results can be explained by the various and often artificial conditions under which such systems are being tested. For example, 40 out of 70 peptides were tested for transfection in the absence or in the presence of low concentrations of serum. This is often done since serum can have a detrimental effect on the DNA/peptide complex stability and may cause aggregation or dissociation of the transfection complexes or may inhibit the interaction with cells, not taking into account that such artificial conditions are poorly predictive for the *in vivo* situation. The results presented here clearly show the need for standardized conditions to screen gene delivery systems in order to improve inter-experimental comparability.

In conclusion, we have demonstrated that automated fluorescence microscopy can be used for high throughput screening of peptide-based transfectants for multiple read-outs, including transfection efficiency, cell viability and the capacity to transfect non-dividing cells. This method will be a valuable extension to the already existing screening methods because the simultaneous measurement of multiple parameters provides more information than can be obtained from a single screening event. Eventually, this should lead to a better *in vitro*/*in vivo* correlation of transfection results obtained with non-viral gene delivery systems.

In combination with the solid-phase platform for combinatorial gene assembly (Chapter 3), the developed HCS assay will be used for the evaluation of multimodular peptide libraries on cell viability and transfection efficiency.

Abbreviations

EGFP: enhanced green fluorescent protein
HCS: high-content screening
kDa: kilo dalton
pDNA: plasmid DNA

References

(1) van Gaal EV, van Eijk R, Oosting RS, Kok RJ, Hennink WE, Crommelin DJ, et al. How to screen non-viral gene delivery systems in vitro? J Control Release 2011 Sep 25;154(3):218-232.

(2) Gasparri F, Cappella P, Galvani A. Multiparametric cell cycle analysis by automated microscopy. J Biomol Screen 2006 Sep;11(6):586-598.

(3) Carpenter AE. Image-based chemical screening. Nat Chem Biol 2007 Aug;3(8):461-465.

(4) Dragunow M. High-content analysis in neuroscience. Nat Rev Neurosci 2008 Oct;9(10):779-788.

(5) van Gaal EV, Oosting RS, Hennink WE, Crommelin DJ, Mastrobattista E. Junk DNA enhances pEI-based non-viral gene delivery. Int J Pharm 2010 May 5;390(1):76-83.

(6) Vaysse L, Arveiler B. Transfection using synthetic peptides: comparison of three DNA-compacting peptides and effect of centrifugation. Biochim Biophys Acta 2000 Apr 6;1474(2):244-50.

(7) Wadhwa MS, Collard WT, Adami RC, McKenzie DL, Rice KG. Peptide-mediated gene delivery: influence of peptide structure on gene expression. Bioconjug Chem 1997 Jan-Feb;8(1):81-8.

(8) Schwartz B, Ivanov MA, Pitard B, Escriou V, Rangara R, Byk G, et al. Synthetic DNA-compacting peptides derived from human sequence enhance cationic lipid-mediated gene transfer in vitro and in vivo. Gene Ther 1999 Feb;6(2):282-92.

(9) McKenzie DL, Collard WT, Rice KG. Comparative gene transfer efficiency of low molecular weight polylysine DNA-condensing peptides. J Pept Res 1999 Oct;54(4):311-8.

(10) McKenzie DL, Kwok KY, Rice KG. A potent new class of reductively activated peptide gene delivery agents. J Biol Chem 2000 Apr 7;275(14):9970-7.

(11) Kharidia R, Friedman KA, Liang JF. Improved gene expression using low molecular weight peptides produced from protamine sulfate. Biochemistry (Mosc) 2008 Oct;73(10):1162-8.

(12) Kichler A, Leborgne C, Marz J, Danos O, Bechinger B. Histidine-rich amphipathic peptide antibiotics promote efficient delivery of DNA into mammalian cells. Proc Natl Acad Sci U S A 2003 Feb 18;100(4):1564-8.

(13) Harbottle RP, Cooper RG, Hart SL, Ladhoff A, McKay T, Knight AM, et al. An RGD-oligolysine peptide: a prototype construct for integrin-mediated gene delivery. Hum Gene Ther 1998 May 1;9(7):1037-47.

(14) Futaki S, Suzuki T, Ohashi W, Yagami T, Tanaka S, Ueda K, et al. Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. J Biol Chem 2001 Feb 23;276(8):5836-40.

(15) Plank C, Tang MX, Wolfe AR, C. SF,Jr. Branched cationic peptides for gene delivery: role of type and number of cationic residues in formation and in vitro activity of DNA polyplexes. Hum Gene Ther 1999 Jan 20;10(2):319-32.

(16) Kim HH, Lee WS, Yang JM, Shin S. Basic peptide system for efficient delivery of foreign genes. Biochim Biophys Acta 2003 May 12;1640(2-3):129-36.

(17) McKenzie DL, Smiley E, Kwok KY, Rice KG. Low molecular weight disulfide cross-linking peptides as nonviral gene delivery carriers. Bioconjug Chem 2000 Nov-Dec;11(6):901-9.

(18) Niidome T, Takaji K, Urakawa M, Ohmori N, Wada A, Hirayama T, et al. Chain length of cationic alpha-helical peptide sufficient for gene delivery into cells. Bioconjug Chem 1999 Sep-Oct;10(5):773-80.

(19) Fortunati E, Ehlert E, van Loo ND, Wyman C, Eble JA, Grosveld F, et al. A multi-domain protein for beta1 integrin-targeted DNA delivery. Gene Ther 2000 Sep;7(17):1505-15.

(20) Suzuki M. The heptad repeat in the largest subunit of RNA polymerase II binds by intercalating into DNA. Nature 1990 Apr 5;344(6266):562-5.

(21) Rittner K, Benavente A, Bompard-Sorlet A, Heitz F, Divita G, Brasseur R, et al. New basic membrane-destabilizing peptides for plasmid-based gene delivery in vitro and in vivo. Mol Ther 2002 Feb;5(2):104-14.

(22) Murray KD, Etheridge CJ, Shah SI, Matthews DA, Russell W, Gurling HM, et al. Enhanced cationic liposome-mediated transfection using the DNA-binding peptide mu (mu) from the adenovirus core. Gene Ther 2001 Mar;8(6):453-60.

(23) Hsu CH, Chen C, Jou ML, Lee AY, Lin YC, Yu YP, et al. Structural and DNA-binding studies on the bovine antimicrobial peptide, indolicidin: evidence for multiple conformations involved in binding to membranes and DNA. Nucleic Acids Res 2005;33(13):4053-64.

(24) van Hell AJ, Fretz MM, Crommelin DJ, Hennink WE, Mastrobattista E. Peptide nanocarriers for intracellular delivery of photosensitizers. J Control Release 2010 Feb 15;141(3):347-353.

(25) del Pozo-Rodriguez A, Pujals S, Delgado D, Solinis MA, Gascon AR, Giralt E, et al. A proline-rich peptide improves cell transfection of solid lipid nanoparticle-based non-viral vectors. J Control Release 2009 Jan 5;133(1):52-9.

(26) Elmquist A, Lindgren M, Bartfai T, Langel U. VE-cadherin-derived cell-penetrating peptide, pVEC, with carrier functions. Exp Cell Res 2001 Oct 1;269(2):237-44.

(27) Fischer PM, Zhelev NZ, Wang S, Melville JE, Fahraeus R, Lane DP. Structure-activity relationship of truncated and substituted analogues of the intracellular delivery vector Penetratin. J Pept Res 2000 Feb;55(2):163-72.

(28) Kilk K, Magzoub M, Pooga M, Eriksson LE, Langel U, Graslund A. Cellular internalization of a cargo complex with a novel peptide derived from the third helix of the islet-1 homeodomain. Comparison with the penetratin peptide. Bioconjug Chem 2001 Nov-Dec;12(6):911-6.

(29) Scheller A, Oehlke J, Wiesner B, Dathe M, Krause E, Beyermann M, et al. Structural requirements for cellular uptake of alpha-helical amphipathic peptides. J Pept Sci 1999 Apr;5(4):185-94.

(30) Perfetto SP, Chattopadhyay PK, Lamoreaux L, Nguyen R, Ambrozak D, Koup RA, et al. Amine reactive dyes: an effective tool to discriminate live and dead cells in polychromatic flow cytometry. J Immunol Methods 2006 Jun 30;313(1-2):199-208.

(31) Salic A, Mitchison TJ. A chemical method for fast and sensitive detection of DNA synthesis in vivo. Proc Natl Acad Sci U S A 2008 Feb 19;105(7):2415-2420.

(32) Brunner S, Sauer T, Carotta S, Cotten M, Saltik M, Wagner E. Cell cycle dependence of gene transfer by lipoplex, polyplex and recombinant adenovirus. Gene Ther 2000 Mar;7(5):401-407.

(33) Rawlinson LA, O'Brien PJ, Brayden DJ. High content analysis of cytotoxic effects of pDMAEMA on human intestinal epithelial and monocyte cultures. J Control Release 2010 Aug 17;146(1):84-92.

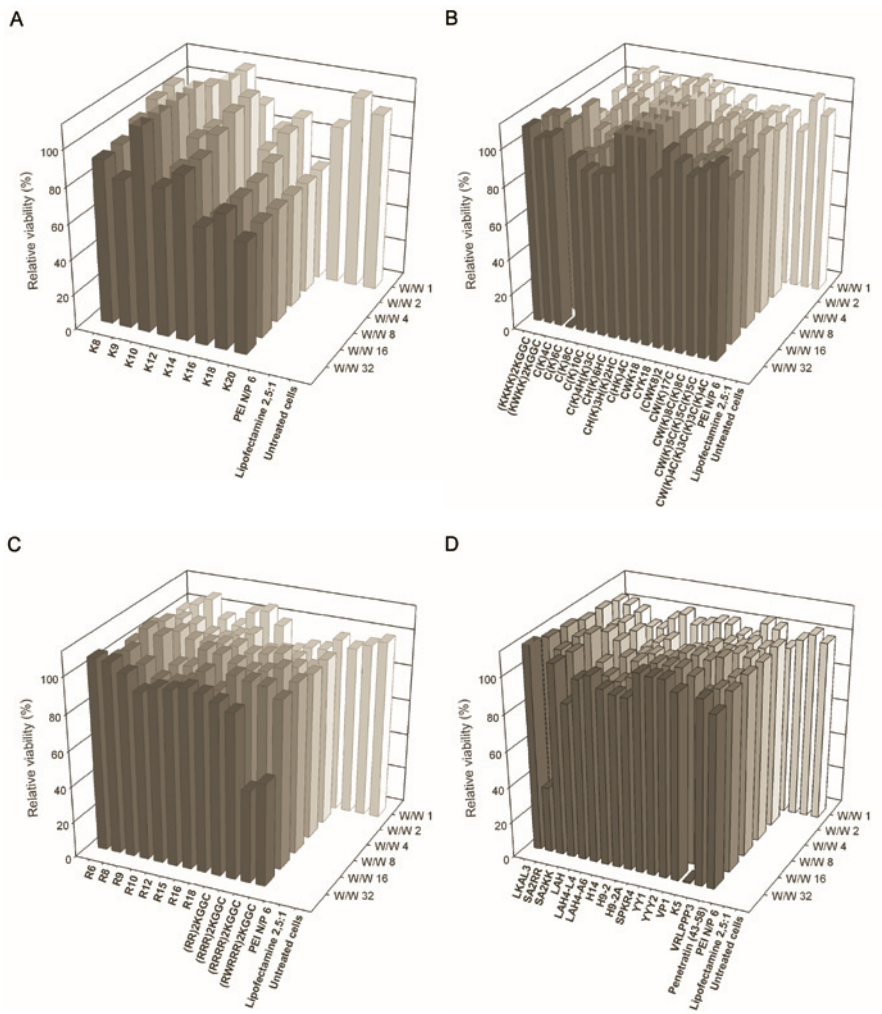
(34) Neumann B, Held M, Liebel U, Erfle H, Rogers P, Pepperkok R, et al. High-throughput RNAi screening by time-lapse imaging of live human cells. Nat Methods 2006 May;3(5):385-390.

(35) Zanella F, Lorens JB, Link W. High content screening: seeing is believing. Trends Biotechnol 2010 May;28(5):237-245.

(36) Shariff A, Kangas J, Coelho LP, Quinn S, Murphy RF. Automated image analysis for high-content screening and analysis. J Biomol Screen 2010 Aug;15(7):726-734.

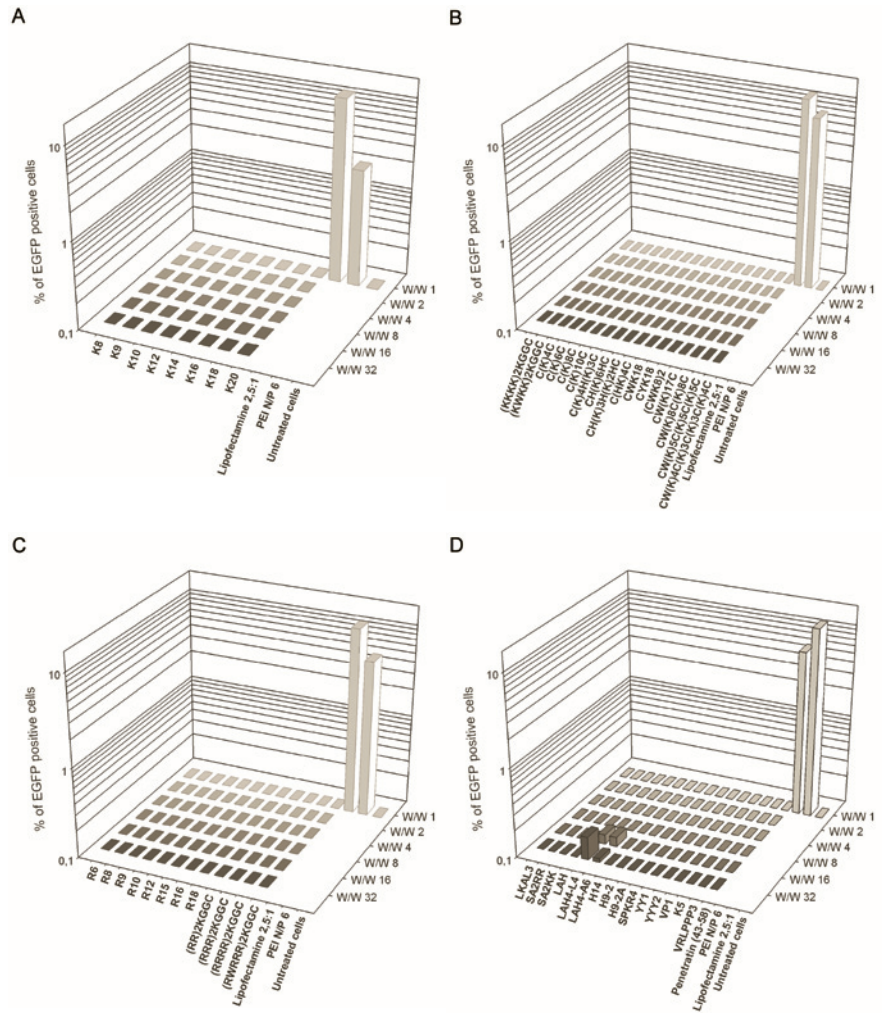
Supplementary material

Relative viability of DNA binding peptides



Supplementary Figure 1 Relative viability of 56 DNA condensing peptides (A-D). COS-7 cells were incubated for 4h with peptide complexes and 48h after transfection, cells were imaged on the a Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. Positive controls (PEI and Lipofectamine) were tested at their optimal DNA/transfectant ratios (see M&M). N=3

Transfection efficiency of DNA binding peptides



Supplementary Figure 2 Transfection efficiency of 56 DNA condensing peptides. COS-7 cells were incubated for 4h with peptide complexes and 48h after transfection, cells were imaged on the a Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. Positive controls (PEI and Lipofectamine) were tested at their optimal DNA/transfectant ratios (see M&M). N=3

5

Chapter

Multimodular Peptide Libraries for Gene Delivery

Markus de Raad[a], Erik A. Teunissen[a] and Enrico Mastrobattista[a]

[a] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

Abstract

Compared to viral vectors, non-viral gene delivery systems are inefficient in delivering their therapeutic nucleic acids intact inside the target cell and in facilitating transgene expression. In order to improve non-viral vectors, we propose to randomly combine functional domains needed for gene delivery from existing or *de novo* designed peptides into single, multimodular constructs. Here, we demonstrated the generation of a gene library encoding multimodular peptides and the subsequent recombinant expression and purification of individual peptide constructs. Unfortunately, expression of the recombinant peptides in *E. coli* did not result in sufficient peptide yields to screen the multimodular peptides for their transfection efficiency.

Introduction

Gene therapy is defined as the introduction and expression of genetic material in cells of an individual for therapeutic benefit. Successful gene therapy requires efficient and specific delivery of therapeutic nucleic acids, without adversely affecting cellular viability or function (1).

Over the past twenty years, most promising gene therapy strategies have relied on viral delivery systems. Compared to viral vectors, non-viral gene delivery systems are inefficient in delivering their therapeutic nucleic acids intact inside the target cell and in facilitating transgene expression. Viruses are highly complex systems, adepts at infecting host cells and have evolved numerous strategies to deliver their genetic material into the appropriate cellular compartment(s) (2). By mimicking viruses, and incorporating multiple functionalities essential for gene delivery into non-viral vectors, the transfection efficiency of non-viral gene delivery systems can potentially be enhanced. Examples of this strategy are the incorporation of functional peptide/protein domains in existing synthetic vectors and designing multifunctional modular polymers or chimeric proteins (3-5). However, the integration of different functionalities into a single system is quite complex and compatibility can be problematic. For example, high vector stability is beneficial in the extracellular environment, but could compromise dissociation of the nucleic acid payload once it has been taken up by target cells (6). Moreover, in case of the multifunctional polymers and chimeric proteins, not much is known about the optimal sequential alignment of functional domains within the macromolecular structure for optimal gene transfer. Issues with complexity and compatibility can be circumvented if a more random, integrative approach is followed that selects for non-viral vectors consisting of combinations of functional traits that together show efficient gene delivery (7).

We propose to combine functional domains needed for gene delivery from existing or *de novo* designed peptides randomly into single, multimodular constructs (7). For instance, peptides derived from DNA condensing proteins can be combined with *de novo* designed membrane disrupting peptides and viral derived nuclear localization signals (NLS). By combining multiple peptides into a single backbone, we create an exactly defined construct, thereby eliminating compositional variations, and achieve reproducibility at the molecular level, which is highly favorable from a pharmaceutical point of view (8). Because it is difficult to rationalize beforehand which combinations and/or positions of the functional domains will be optimal for gene delivery, all possible combinations of domains should be generated (7). By using a modular assembly strategy for combinatorial gene construction, a gene library can be generated that codes for multimodular peptides. This gene library can then be transformed into suitable host cells for recombinant expression and after subsequent purification, the multimodular peptide constructs can be screened for transfection efficiency.

The main objectives of this study are: (i) To demonstrate that the previously developed sequence-independent ligation method can be used to generate diverse combinatorial gene libraries encoding single-chain, multimodular polypeptides with DNA condensing, cell penetrating and nuclear localization properties (Chapter 3) (9). (ii) To investigate if such multimodular peptide constructs can be recombinantly produced in bacteria and purified in good yields for further use as gene carriers.

Materials and Methods

Materials

All chemicals and reagents were purchased from Sigma-Aldrich Chemie B.V. (Zwijndrecht, The Netherlands), unless stated otherwise. Dynabeads M-280 Streptavidin were purchased from Dynal (Oslo, Norway). Restriction enzymes, FastAP™ (Thermosensitive Alkaline Phosphatase), Generuler™ DNA and Riboruler™ RNA ladders, Phusion polymerase, Rapid DNA Ligation Kit, GeneJET™ Plasmid Miniprep Kit, isopropyl β-D-1-thiogalactopyranoside (IPTG), RNase A, DNase I, Spectra™ Multicolor Low Range Protein Ladder and HisPur™ Ni^{2+}-NTA Resin were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Precast denaturing polyacrylamide TBE-urea gels (6%) and gradient (4–12%) NuPAGE™ Novex Bis-Tris mini-gels were purchased from Invitrogen (Breda, The Netherlands). Precast Criterion™ Tris-Tricine peptide gels (16.5%) were purchased from Bio-Rad (Hercules, CA, USA).

The used culture media contained: Lennox Broth (LB) (tryptone (10g/L), yeast extract (5g/L) and NaCl (5g/L)), LB-agar (tryptone (10g/L), yeast extract (5g/L), NaCl (5g/L) and agar (15g/L)) and Terrific Broth (TB) (tryptone (12g/L), yeast extract (24g/L), K_2HPO_4 (9.4g/L), KH_2PO_4 (2.2g/L) and glycerol (8mL/L)). All used solid and liquid media contained glucose (1% w/v) and ampicillin (100µg/mL).

Synthetic oligodeoxynucleotides

All oligodeoxynucleotides were synthesized by Eurogentec S.A. (Seraing, Belgium). Single-stranded oligodeoxynucleotides (ssODN) modules were synthesized with 5'- and 3'- phosphates and PAGE purified at a 40nmol scale or with 5'biotin-TEG and PAGE purified at a 200nmol scale. ssODN linkers and primers were synthesized unmodified and SePOP desalted at a 40nmol scale. Oligodeoxynucleotides were dissolved in nuclease-free water and stored at -30°C as 100µM stocks.

Oligodeoxynucleotides encoding peptides were codon-optimized for bacterial expression and flanked with linker sequences GGTTCT (GS) at the 5' (all except Head modules) and GGTGGC (GG) at the 3' (all except Tail modules). The possible formation of secondary structures (checked with Oligocalc software) was reduced by adjustment of any self-complementary oligonucleotide sequence (10). All sequences of the used ssODNs and primers are listed in Table 1.

Methods

Construction of multimodular peptide libraries by solid phase combinatorial multipart gene assembly

Using the method for the assembly of ssODN in the presence of linker ODN, a random polypeptide library was assembled (Chapter 3) (9). Ligation reactions (20µL) contained 12.5pmol ssHead (Head_GG) coated on M-280 magnetic beads, 50pmol GGGS linker, 1x Rapid Ligation Buffer, 5U T4 DNA ligase and 100pmol of a mix of 9 different ssODNs (11,1pmol of GS_Pr18_GG; GS_SPKR4_GG; GS_H9.2_GG; GS_H5WYG_GG; GS_TAT_GG; GS_ppTG20_GG; GS_Mu_GG; GS_NLSV402_GG and GS_LAH5_GG). Five ligation steps with this composition were performed. One additional ligation step was performed for the ligation of the reaction products to a terminating ssTail (GS_Tail). Ligation products were loaded on a 6% denaturing polyacrylamide TBE-urea gel and the area between the smallest theoretical product (±150 nucleotides) and the largest theoretical product (±600 nucleotides) was excised, purified and PCR amplified (as described in Chapter 3) (9). PCR products were analyzed on a 2% w/v agarose gel and the area between ±300-600bp was excised and purified.

The library obtained was XbaI/ HindIII digested, purified and 48-140ng of the digested products were ligated in 100ng of dephosphorylated pSCherry2 (XbaI/ HindIII digested and purified from an agarose gel). Half of the reaction mixture was used directly to transform *E.coli* strain DH5α (Invitrogen). Bacteria were plated out on LB-agar plates and grown overnight at 37°C. From the obtained LB plates, all colonies were harvested using 2ml LB. Harvested colonies were grown in 10ml LB for 4hr at 37°C and plasmid DNA (pDNA) was harvested.

Recombinant protein production in *E. coli*

Fifty nanogram library pDNA was used to transform *E. coli* strain BL21-Gold(DE_3) (Stratagene, Santa Clara, CA, USA). Bacteria were plated out on LB-agar and grown overnight at 37°C. From the plates obtained, colonies were inoculated overnight in 5mL LB. Glycerol stocks were prepared and pDNA was harvested from all overnight cultures. pDNA of the library constructs obtained were NaeI digested and analyzed on a 1% w/v agarose gel. In addition, pDNA of the library constructs obtained were sequenced for insert identification with full coverage of the insert sequences (BaseClear, Leiden, The Netherlands).

For protein expression, starter cultures were grown overnight in 5-30ml LB from the obtained glycerol stocks. Next, cultures were diluted to an optical density at 600nm (OD_{600}) of 0.1 in 30-550mL Terrific Broth. The diluted cultures were grown at 37°C to OD_{600} of 0.4-0.8, when protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.1mM final concentration. After 5h, bacteria were harvested by centrifugation (5,000xg, 15 min) and stored at -20°C.

Table 1 Used oligodeoxynucletides

Name	Sequences
Head_GG	5'-Biotin-TEG-GTTAGCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACA-TAATGGGTGGC-OH 3'
Mu	5' P-GGTTCTCGTCGTGCTCATCATCGGCGTAGACGTGCATCTCATCGCCGTATGCGT-CGCGGAGGTGGTGGC-P 3'
H5WYG	5' P-GGTTCTGGTCTGTTTCACGCGATTGCCCATTTCATCCACGGTGGTTGGCATGG-TTTAATTCACGGTTGGTATGGCGGTGGC-P 3'
H9.2	5' P-GGTTCTAAGACACCGAAAAAGGCCAAAAAGCCAAAAACCCCGAAAAAGGC-GAAAAAACCAGGTGGC-P 3'
LAH5	5' P-GGTTCTAAAAAGGCACTGCTGGCACTGGCACTCCATCACTTAGCACACCTTGCT-CATCATCTTGCGTTAGCGCTGAAAAAGGCTGGTGGC-P 3'
NLSV402	5' P-GGTTCTCCAAAAAAGAAACGTAAAGTTCCAAAAAAAAAGCGCAAAGTCGG-TGGC-P 3'
ppTG20	5' P-GGTTCTGGCTTATTTCGTGCGCTGTTGCGTCTGTTACTGTCTCTGTGGAGAT-TACTTTTACGTGCGGGTGGC-P 3'
Pr18	5' P-GGTTCTTCTCGTAGTCGGTATTACCGTCAGCGCCAACGTTCTCGCCGTCGC-CGGCGTAGAGGTGGC-P 3'
SPKR4	5' P-GGTTCTAGCCCGAAACGTAGCCCTAAGCGCAGCCCAAAAAGATCTCCTAAACG-TGGTGGC-P 3'
TAT (47-57)	5' P-GGTTCTTATGGCCGCAAGAAGCGTCGTCAAAGACGTCGTGGTGGC-P 3'
GGGS linker	5' OH-GGTGGCGGTTCT-OH 3'
GS_Tail	5' OH-GGTTCTCATCACCATCACCATCACCATCACTAATAAAAGCTTAAAGAG-OH 3'
Fw_Head	5' OH-AGCCCATGGTTATGGAAAA-OH 3'
Rev_Tail	5' OH-GTCATACTTATGGGATCCTTATTAA-OH 3'

Bacterial lysis

The bacterial pellets obtained were thawed at 4°C, suspended in 5ml/g wet pellet native lysis buffer (50mM NaH_2PO_4, 300mM NaCl at pH 8.0) containing 0.2mg/ml lysozyme, 0.05% (w/v) sodium deoxycholate and one tablet Complete™ Protease Inhibitor Cocktail Tablets (Roche, Almere, The Netherlands) per 50ml native lysis buffer, and incubated for 30min on an orbital shaker (200rpm) at 20°C. Next, bacteria were snap-frozen using liquid nitrogen. After defrosting, DNase I (5ug/ml) and RNase A (10ug/ml) were added and incubated at 20°C for 30min on an orbital shaker (200rpm). Bacteria were centrifuged at 12,000xg for 45min at 4°C to pellet the inclusion bodies and supernatant (=soluble fraction) was saved for purification/analysis. Then, inclusion bodies were snap-frozen using liquid nitrogen and denaturing lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole and 8M urea at pH 8.0) was added. After 60min incubation on an orbital shaker (200rpm) at room temperature, inclusion bodies were homogenized by a probe sonicator (Labsonic P,

equipped with a 3-mm diameter probe; Sartorius AG, Goettingen, Germany) set at 320W and 24kHz for 2 cycles of 30 seconds with a 0.5-second interval. The lysates were centrifuged at 10,000xg for 30min at room temperature to pellet the cellular debris. The supernatant was saved for purification/analysis.

Protein purification

The multimodular peptides produced were purified from the supernatants obtained after lysis using Ni^{2+}-NTA affinity chromatography. The supernatants from the soluble fraction and the lysed inclusion bodies were either processed separately (library screening) or combined (large scale expression). To the supernatants, HisPur™ Ni^{2+}-NTA slurry was added (25µl slurry for 30ml cultures and 200µl slurry for 550ml cultures) and the suspension was mixed on an end-over-end rotator for 1h at room temperature (RT). Then, the lysate/resin mixture was loaded on polypropylene columns and washed with 8 resin bed volumes wash buffer (50mM NaH_2PO_4, 300mM NaCl, 8M urea and 40mM imidazole at pH 8.0). Subsequently, HIS-tagged modular peptides were eluted with 3 x 1 resin bed volumes of elution buffer (50mM NaH_2PO_4, 300mM NaCl, 8M urea and 300mM imidazole at pH 8.0).

Quantity and purity of the modular peptides was analyzed by SDS-PAGE and Western Blotting. For SDS-PAGE, samples were run on Criterion™ 16.5% Tris-Tricine peptide gels and visualized by Krypton™ Protein Stain and/or PageBlue™ Protein Staining Solution (based on Coomassie G-250, Thermo Fisher Scientific). For Western Blotting, samples were run on 4–12% gradient NuPAGE™ Novex Bis-Tris mini-gels and proteins were electro-transferred to a nitrocellulose membrane using the I-Blot Dry Blotting system (Invitrogen). Membranes were preblocked overnight at 4°C with 5% BSA in Tris-buffered saline containing 0.1% Tween-20 (TBS-T) and subsequently incubated for 2h with mouse monoclonal anti-polyhistidine antibody diluted 1:1000 in 5% BSA in TBS-T. After washing with TBS-T, membranes were incubated for 1h at RT with goat anti-mouse peroxidase-conjugated secondary antibody (Thermo Fischer Scientific) diluted 1:1000 in 5% BSA in TBS-T. Proteins were detected using SuperSignal West Femto Chemiluminescent Substrate (Thermo Fischer Scientific) and a Gel Doc Imaging system equipped with a XRS camera and Quantity One analysis software (Bio-Rad).

For buffer exchange and concentration, modular peptide solutions were washed with HEPES-buffered saline (HBS) (20mM HEPES, 130mM NaCl, pH 7.4), using Vivaspin centrifugal concentrators (Sartorius Stedim Biotech S.A., Aubagne, France) with a molecular weight membrane cut-off of 3kDa.

Protein concentration was determined with the MicroBCA™ protein assay (Pierce Biotechnology, Rockford, IL, USA) against lysozyme as a standard protein.

Results

Construction of a gene library encoding multimodular peptides by solid phase combinatorial multipart gene assembly

The gene library encoding the multimodular peptides was constructed in a combinatorial manner by ligation of single-stranded oligodeoxynucleotides (ssODN) using DNA ligase in the presence of linker ssODN (Chapter 3) (9). Briefly, a universal ssODN (Head) modified with 5'-biotin was immobilized on streptavidin-coated magnetic beads. Using linker ODN, which were complementary to the last 6 nucleotides of the immobilized ssODN and the first 6 nucleotides of ssODN to be ligated, DNA ligase was used to ligate ssODN in a stepwise manner. In the final ligation step, a universal single-stranded oligodeoxynucleotide (Tail) was added.

For the generation of the multimodular peptide library, 9 different functional peptides were selected: 4 DNA condensing peptides ($SPKR_4$, Pr18, Mu and H9.2), 4 membrane disrupting and/or endosomolytic peptides (LAH5, ppTG20, H5WYG and TAT(47-57)) and 1 NLS peptide (NLSV402) (Table 2). All library constructs started with a universal Head, which contained the RBS and start codon ATG. All library constructs were terminated with a universal Tail, which contained a 8xHIS-tag followed by double stop codon TAATAA. Both Head and Tail contained restriction sites for cloning purposes. The Head, Tail and all peptides were reverse translated into ssODN sequences (using the *E.coli* codon usage table, and flanked with linker sequences GGTTCT (encoding Gly-Ser) at the 5' (except Head) and GGTGGC (encoding Gly-Gly) at the 3' (except Tail) (Table 2) (11).

Five ligation cycles were performed with a mix containing equal amounts of ssODN encoding all 9 functional peptides, generating a library of Head-MIX_5-Tail. This would theoretically yield a library with $(9^1)+(9^2)+(9^3)+(9^4)+(9^5)=66{,}429$ different constructs with $9^5=59{,}049$ full length constructs. The generated constructs were analyzed in a 6% TBE urea gel, after loading equal amounts of magnetic beads.

Both libraries obtained were clearly visible and appeared as smears, with highest intensity between 400 and 600 nucleotides (Fig. 1a). No individual constructs were visible, only the non-ligated Tail was visible.

After extraction from gel and purification, both libraries were amplified by PCR to generate dsDNA products and enough material for further steps (Fig. 1b). Also after PCR amplification, smears were detected indicating that all constructs formed were amplifiable.

PCR showed an amplification bias towards the smaller templates (Fig. 1b). The PCR amplified Head-MIX_5-Tail library was excised from the smallest theoretical 5 module construct (±300bp) to the largest theoretical 5 module construct (±600bp), purified and cloned into the XbaI/HindIII-sites of vector pSCherry2. *E.coli* strain DH5α

Table 2 Function, name, and amino acid sequence of functional peptides used. Nucleotide sequence of the corresponding ssODN constructs to produce modular gene constructs are also indicated

Function	Peptide	AA sequence	ssDNA sequence	Ref
DNA condensation	Pr18	GSSRSRYYRQRQRSRRRRRRGG	ggttcttctcgtagtcggtattaccgtcagcgccaacgttctcgccgtcgccggcgtagaggtggc	(12)
	$SPKR_4$	GSSPKRSPKRSPKRSPKRGG	ggttctagcccgaaacgtagccctaagcgcagcccaaaaagatctcctaaacgtggtggc	(13)
	H9.2	GSKTPKKAKKPKTPKKAKKPGG	ggttctaagacaccgaaaaaggccaaaaagccaaaaaccccgaaaaaggcgaaaaaaccaggtggc	(12)
	Mu	GSRRAHHRRRRASHRRMRRGG	ggttctcgtcgtgctcatcatcggcgtagacgtgcatctcatcgccgtatgcgtcgcggaggtggtggc	(14)
Membrane disruption/ translocation	TAT (47-57)	GSYGRKKRRQRRRGG	ggttcttatggccgcaagaagcgtcgtcaaagacgtcgtggtggc	(15)
	ppTG20	GSGLFRALLRLLRSLWRLLLRAGG	ggttctggcttatttcgtgcgctgtgcgtctgttactgtctctgtggagattactttttacgtgcgggtggc	(16)
	LAH5	GSKKALLALALHHLAHLAHHLALALKKAGG	ggttctaaaaaggcactgctggcactggcactccatcacttagcacaccttgctcatcatcttgcgttagcgctgaaaaaggctggtggc	(17)
Endosomolytic	H5WYG	GSGLFHAIAHFIHGGWHGLIHGWYGGG	ggttctggtctgtttcatgcgattgccccacttcatccatggtggttggcatggttttaattcacggttggtatggcggtggc	(18)
NLS	NLSV402	GSPKKKRKVPKKKRKVGG	ggttctccaaaaaagaaacgtaaagttccaaaaaaaaagcgcaaagtcggtggc	(19)
Head	Head_GG	MGG	gttagc**tctaga**aataattttgtttaacttaagaaggagatatacataATGggtggc	-
Tail	GS_Tail	GSHHHHHHHH	ggttctcatcaccatcaccatcaccatcacTAATAA**aagctt**aaagag	-

Underlined sequences correspond to GGGS linker sites. ssODN sequences in capitals correspond to the start codon in the Head and the double stop codons in the Tail. ssODN sequences in bold correspond to XbaI site in the Head and HindIII site in the Tail.

was transformed with the cloned library. To obtain the pDNA library, all colonies of a single plate were pooled and pDNA was isolated.

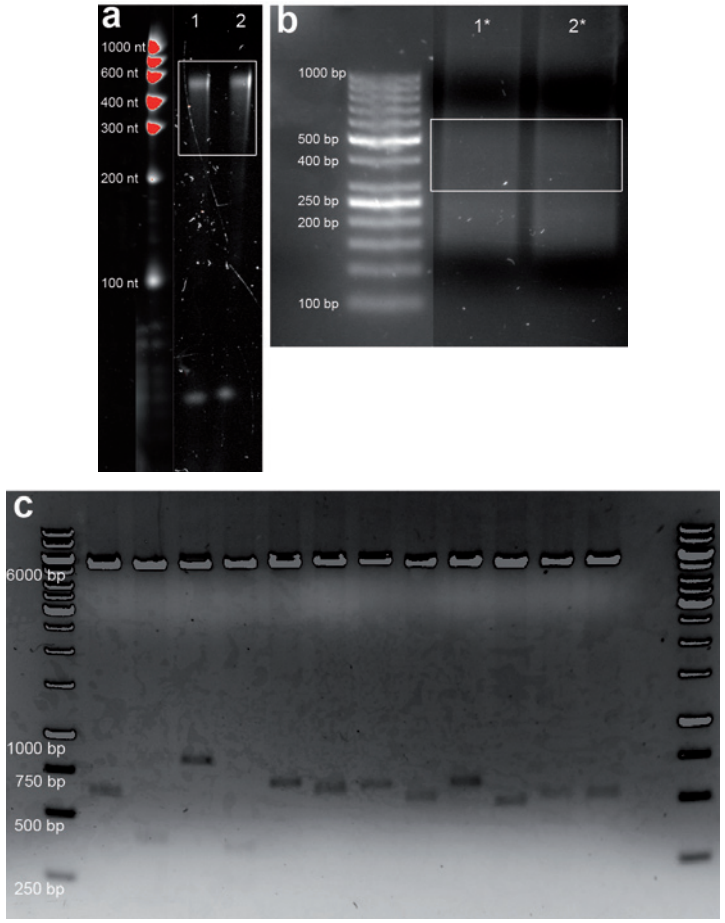


Figure 1 Gene library construction.
(**a**) Initial ligation products were loaded on a 6% TBE-urea gel and purified at final product height as indicated. Ladder: low range RNA ladder. Lanes 1 and 2: Head-MIX5-Tail in duplo.
(**b**) Purified ligation products were amplified using PCR, loaded on a 2% agarose gel and purified at final product height as indicated. Ladder: 50bp DNA ladder. Lanes 1* and 2*: PCR-amplified ligation products corresponding to lanes 1 and 2.
(**c**) Plasmid DNA isolated from 12 colonies was digested with NaeI and loaded on 0.7% agarose gel. Vector backbone appears as a single band of ±5300bp whereas library inserts appear between 500-950bp, except for inserts containing a sequence coding for Pr18, which is also digested by NaeI and will appear below 500bp.

Library screening

The library pDNA obtained was used for the transformation of *E.coli* strain BL21-Gold(DE_3). To confirm library presence and determine construct size, restriction analysis was performed on pDNA obtained from 180 colonies after overnight culturing (Fig. 1c, not all data shown). Only 103 out of 180 (57%) of screened colonies contained correct pDNA (data not shown). All 103 positive colonies contained inserts corresponding to constructs with at least one functional peptide. Sequence analysis of 37 colonies confirmed random insertion of all used functional peptides (Fig. 2 and Table 3). The sequences of 4 constructs could not reliably be identified.

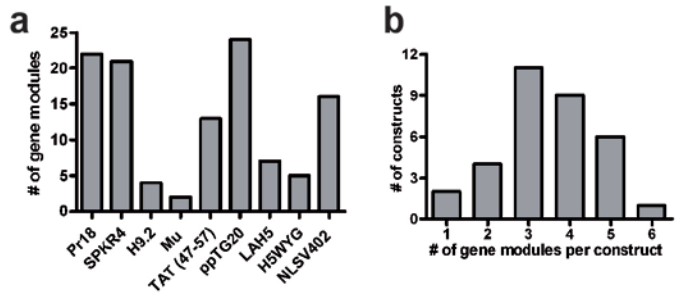


Figure 2 Library composition. (a) The frequency of gene modules identified by sequencing of all 37 constructs. (b) The number of ligated gene modules per construct.

After sequence analysis, the 37 sequenced colonies were screened for protein expression. To confirm the production of the multimodular peptides, Western Blotting was performed on crude bacterial lysates (Fig. 3). For the negative control, non-induced bacteria were used and for the positive control, bacteria expressing Enhanced Green Fluorescence Protein (EGFP) were used.

Twenty-two out of thirty-seven colonies (57%) produced detectable amounts of peptide constructs. All peptides produced were in the range of the expected theoretical sizes (3 -20kDa).

Table 3 Number, name and peptide composition of all 37 sequenced colonies

#	Name	Composition	#	Name	Composition
1	CmL 1.2	TAT-LAH5-TAT-ppTG20	20	CmL 1.84	NLS404-Pr18-SPKR-Pr18
2	CmL 1.5	SPKR-Pr18-ppTG20	21	CmL 1.86	Pr18-ppTG20
3	CmL 1.11	NLSV402-SPKR-Pr18-SPKR-Pr18	22	CmL 2.42	Pr18-H5WYG
4	CmL1.13	LAH5-ppTG20-Pr18-Mu	23	CmL 2.51	SPKR-NLSV402-ppTG20-H9.2-NLSV402-H5WYG
5	CmL1.15	TAT-LAH5-TAT-ppTG20	24	CmL 2.52	NI
6	CmL 1.22	TAT-ppTG20-SPKR-TAT	25	CmL 2.53	Pr18-TAT-NLS402-TAT
7	CmL 1.24	Pr18-ppTG20-Pr18	26	CmL 2.71	Pr18-ppTG20-ppTG20-NLSV402
8	CmL 1.25	H5WYG-SPKR-ppTG20	27	CmL 2.78	Pr18-Pr18-SPKR
9	CmL 1.29	NLS402	28	CmL 2.86	SPKR-TAT
10	CmL 1.33	Pr18-NLS402	29	CmL 1.91	Pr18-ppTG20-ppTG20
11	CmL 1.34	NLSV402-ppTG20-LAH5-ppTG20-SPKR	30	CmL 1.103	Pr18-NLSV402-NLSV402-ppTG20
12	CmL 1.43	TAT-SPKR-SPKR-NLSV402	31	CmL 1.104	TAT-H9.2-SPKR-NLSV402-ppTG20
13	CmL 1.47	NI	32	CmL 1.106	SPKR-ppTG20-ppTG20
14	CmL 2.3	H9.2-Pr18-NLSV402	33	CmL 1.109	Pr18-H5WYG-ppTG20-Pr18-NLSV402
15	CmL 1.57	ppTG20-LAH5-LAH5	34	CmL 1.116	TAT-ppTG20-H5WYG
16	CmL1.62	SPKR-ppTG20-Mu	35	CmL 1.127	Pr18-ppTG20-LAH5
17	CmL1.66	NI	36	CmL 1.132	TAT-SPKR-SPKR-SPKR-NLSV402
18	CmL 1.74	NI	37	CmL 1.136	Pr18
19	CmL 1.77	SPKR-SPKR-ppTG20-H9.2-SPKR			

(NI) Not identified.

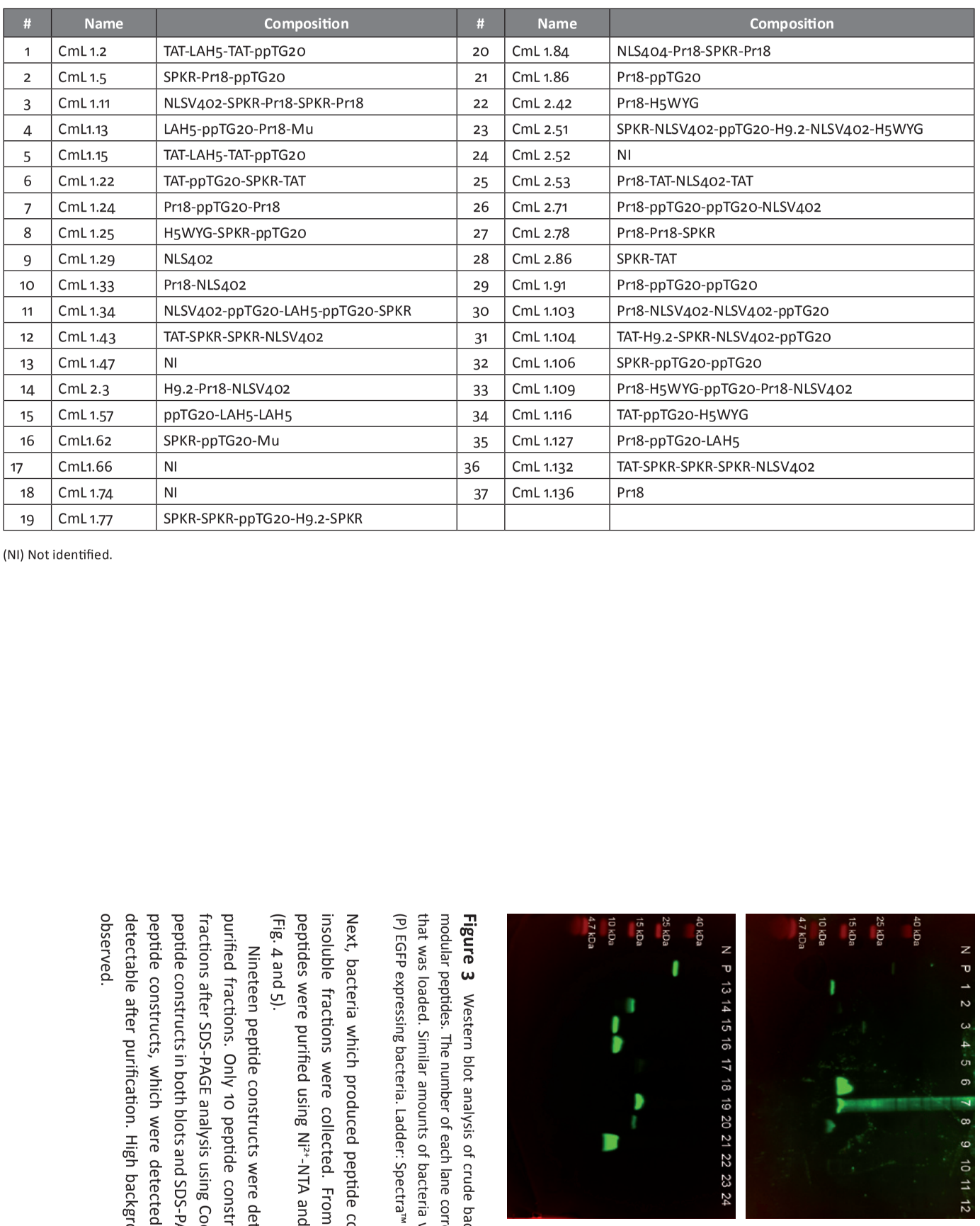


Figure 3 Western blot analysis of crude bacterial lysates to confirm production of multimodular peptides. The number of each lane corresponds to the sequenced colonies (see Table 2) that was loaded. Similar amounts of bacteria were loaded on gel. (N) non-induced bacteria, (P) EGFP expressing bacteria. Ladder: Spectra™ Multicolor Low Range Protein Ladder.

Next, bacteria which produced peptide constructs were lysed and the soluble and insoluble fractions were collected. From these fractions, the produced modular peptides were purified using Ni^{2+}-NTA and analyzed by SDS-PAGE and Western Blot (Fig. 4 and 5).

Nineteen peptide constructs were detected by Western blotting in one or both purified fractions. Only 10 peptide constructs were visible in one or both purified fractions after SDS-PAGE analysis using Coomassie G-250 staining. All corresponding peptide constructs in both blots and SDS-PAGE gels appeared at the same height. Five peptide constructs, which were detected in the crude bacterial lysates, were not detectable after purification. High background levels of non-specific proteins were observed.

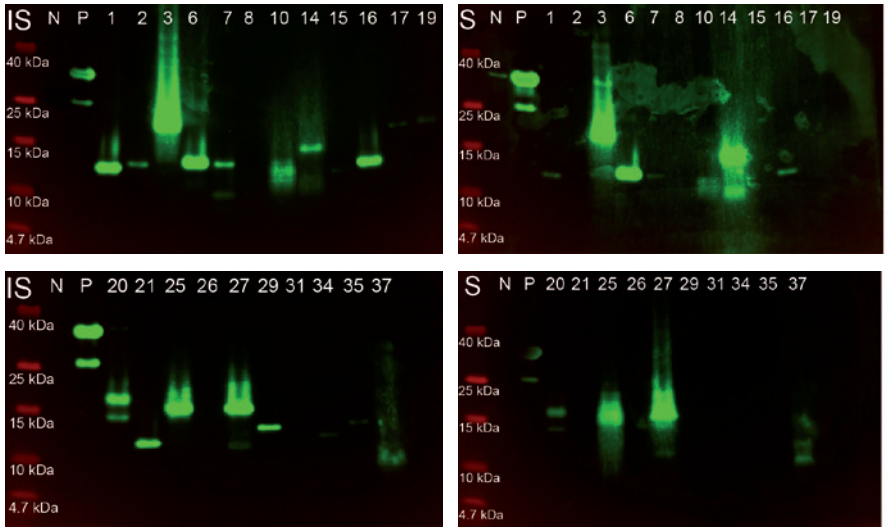


Figure 4 Western blot analysis of purified multimodular peptides obtained from the soluble (S) and insoluble (IS) fraction. The number of each lane corresponds to the sequenced colonies (see Table 2) that was loaded. Similar volumes of elution fraction were loaded on gel. (N) non-induced bacteria, (P) EGFP expressing bacteria. Ladder: Spectra™ Multicolor Low Range Protein Ladder.

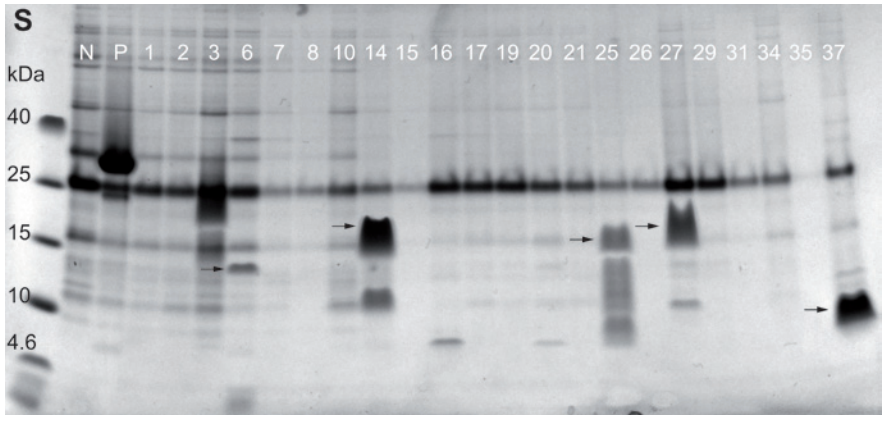


Figure 5 SDS-PAGE analysis of purified multimodular peptides obtained from the soluble (S) and insoluble (IS) fraction visualized using Coomassie G-250 (PageBlue™ Protein Staining Solution). The number of each lane corresponds to the sequenced colonies (see Table 2) that was loaded. Similar volumes of elution fraction were loaded on gel. (N) non-induced bacteria, (P) EGFP expressing bacteria. Ladder: Spectra™ Multicolor Low Range Protein Ladder.

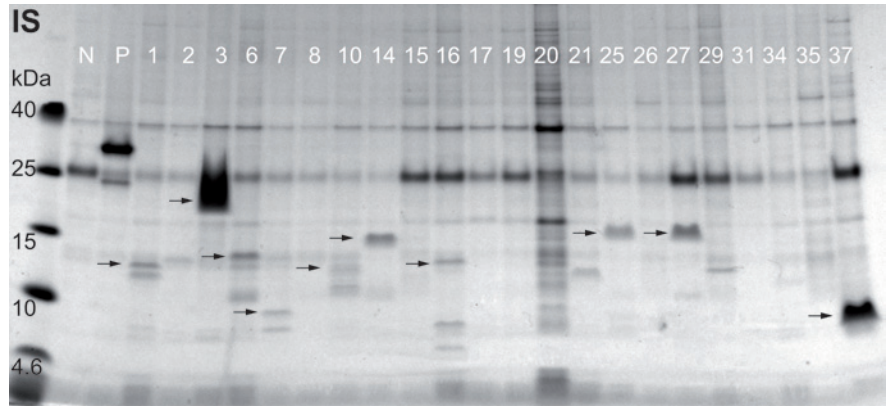


Figure 5 Continued.

Large scale expression

In order to obtain the higher quantities of the peptide constructs needed to perform cellular assays (e.g. transfections assays), 7 peptide constructs were selected and expressed at 0.55L scale: CmL 1.2, CmL 1.11, CmL 1.22, CmL 1.86, CmL 1.136, CmL 2.53 and CmL 2.78. After expression, purification and concentration, the selected peptide constructs were analyzed using SDS-PAGE and the yield was quantified (Fig. 6 and Table 4).

All purified peptide constructs were detectable after SDS-PAGE analysis using Coomassie G-250 staining, except for CmL 1.86 which was not visible on SDS-PAGE. After an additional concentration/dialysis step using Vivaspin centrifugal concentrators, only 3 out of 7 peptides could be recovered, namely CmL 1.22, 1.136 and 2.78. High background levels of non-specific proteins were observed, which made quantification of the yield of the peptide constructs difficult. Due to the low yields and purity, no cellular assays were performed.

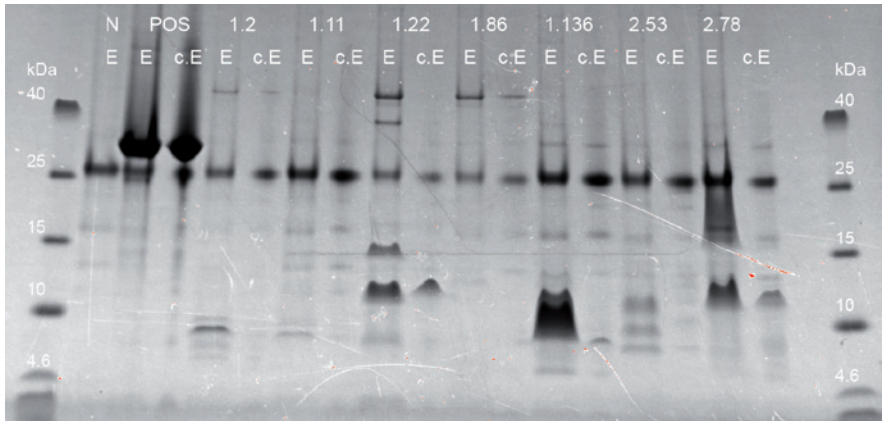


Figure 6 SDS-PAGE analysis of purified multimodular peptides after large scale expression visualized using Coomassie G-250 (PageBlue™ Protein Staining Solution). Similar volumes of elution fraction were loaded on gel. (E) elution fraction, (c.E) elution fraction after concentration, (N) non-induced bacteria, (POS) EGFP expressing bacteria. Ladder: Spectra™ Multicolor Low Range Protein Ladder.

Table 4 Total protein yield before and after concentration from a 0.55L bacterial culture, harvested 5h after protein induction

	Yield before concentration (µg)	Yield before concentration (nmol)	Yield after concentration (µg)	Yield after concentration (nmol)
CmL 1.2	20	2.0	21,1	2.1
CmL 1.11	50	5.0	40,4	4.0
CmL 1.22	53	5.3	22,4	2.2
CmL 1.86	70	7.0	27,6	2.8
CmL 1.136	111	11.1	41,8	4.2
CmL 2.53	89	8.9	38,4	3.8
CmL 2.78	123	12.3	25,0	2.5
POS	315	10.4	247,1	8.1
NEG	24	-	ND	-

Discussion and conclusion

In this study, we showed the generation of a gene library encoding multimodular peptides and the subsequent recombinant expression and purification of individual peptide constructs from this library as part of an integrative strategy to screen and select functional peptide vectors for gene delivery.

First, gene libraries were assembled. The solid-phase platform described in Chapter 3 for combinatorial multipart gene assembly allows for sequence independent and combinatorial ligation of ssODN, which is required to generate the multimodular peptide libraries (9). Nine different functional peptides were selected to construct the library. Four DNA condensing peptides ($SPKR_4$, Pr18, Mu and H9.2), 2 membrane disrupting and/or endosomalytic peptides (H5WYG and TAT(47-57)) and 1 NLS peptide (NLSV402) were selected from literature with proven functional capacities, whereas the 2 peptides LAH5 and ppTG20 were selected based on performance in transfection studies (20).

After cloning, the pDNA library obtained was transformed into BL21-Gold(DE3) bacteria and resulted in colonies containing unique inserts (Table 3). However, about 45% of the obtained colonies contained mutations, mostly (partial) removal of the entire promoter/RBS stretch (data not shown). This could be caused by plasmid toxicity or toxicity of the encoded multimodular peptide, despite the presence of transcriptional repressors. Recombinant proteins can be toxic to the host, either by interference with a cellular function or by their physical properties (such as CPP, which can interact with the membrane), and are thus strongly selected against. The "stress" caused by toxic protein may lead to increased mutational alterations of its respective gene. Kawe et al. showed that *E.coli* finds efficient ways of promoter deletions in the expression plasmid, leading to the exclusive growth of non-producing bacteria, when expression is not well repressed in the newly transformed bacterial cell (21). By using stronger repressors, promoter deletions can be prevented (21,22). For this reason, a first screen on bacterial colonies for the presences of correct pDNA was performed before recombinant protein expression. Also, in the generated pDNA library a bias towards certain peptides was observed. Seventy five percent of all peptide domains indentified in the 37 constructs were either Pr18, $SPKR_4$, ppTG20 or NLSV402. This apparent bias could be caused by intrinsic differences in the ligation efficiencies of the used ssODN encoding the 9 different peptides (9).

Western blot analysis of crude bacteria after recombinant expression revealed that only 21 cultures produced detectable amounts of multimodular peptide constructs. The high cationic charge density and the linear structure of the multimodular peptides make them highly susceptible to proteolytic degradation (23,24). This can result in the removal of the C-terminal HIS-tag from the peptide construct, which will then be undetectable on a Western blot. Truncation at the

C-terminal end was also observed with similar other modular peptide constructs (Fig. 7). Two different constructs (C1 and C2) were expressed and detected in crude bacterial lysates using Western Blot analysis. Both constructs had a similar size and contained a N-terminal HIS-tag and a C-terminal HA-tag. C2 was visible on both blots, indicating the presence of both N- and C-terminal tags, whereas C1 was only visible on the blot using the anti-HIS antibody. Also, the size of C1 was smaller compared to C2. Besides proteolytic degradation, mutations in the pDNA could result in truncations or loss of the HIS-tag.

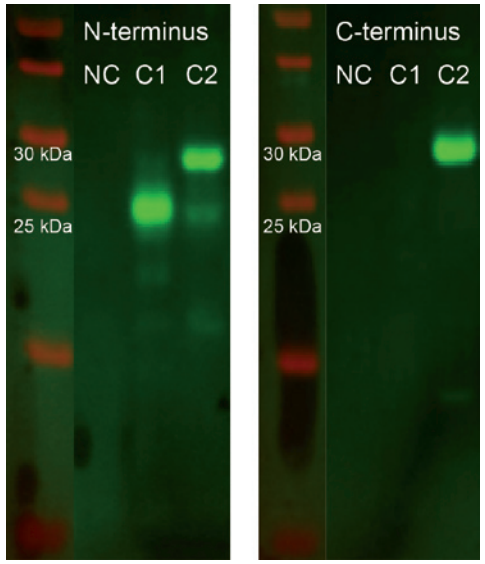


Figure 7 Western blot analysis of crude bacterial lysates to determine in vivo degradation. Similar amounts of bacteria were loaded on the gel. (NC) non-induced bacteria, (C1) construct 1, (C2) construct 2. Ladder: Spectra™ Multicolor Low Range Protein Ladder.

After identifying which bacteria produced detectable amounts of multimodular peptides, the produced multimodular peptides were purified from both the soluble and insoluble fraction. Out of the 21 cultures, only 8 peptide constructs were detectable using SDS-PAGE. By using the more sensitive Western blot analysis with antibodies against the HIS-tag of the recombinant protein, the number of detectable peptide constructs increased to 19, which indicates that the overall yield of the produced multimodular peptides was low. The majority (11 out of 19) of multimodular peptides produced inside the bacteria were present in both the soluble and insoluble fractions. Eight of the multimodular peptides produced inside the bacteria were present only in the insoluble fraction.

The low peptide yields can have several causes. Most probably they result from host-related toxicity and proteolytic degradation during recombinant production.

The low expression levels could be caused by problems at the transcriptional/translation level or could be caused by toxicity of the produced multimodular peptides. At the transcriptional/translation level, secondary structures in the mRNA can lead to slowed or even fully blocked translation initiation (25). Also, the use of rare codons can lead to premature translation termination. We tried to decrease secondary structure formation and rare codon usage by altering the nucleotide sequence and by codon optimizing the sequences encoding the peptides where possible. Also, in the pSCherry vector backbone used, the tRNA genes of the six rare codons are incorporated and supply rare tRNA's. At the protein level, several of the membrane disrupting peptides used are known to have an antimicrobial effect. This makes them potentially toxic to the producing bacteria (16,17,23). There was no correlation between construct size and peptide expression levels. Also, no specific peptide or peptide combination caused excessive toxicity.

In order to reduce proteolytic degradation and toxicity of the produced multimodular peptides, we tried to optimize our protein expression protocol by adjusting several parameters, including temperature, inducer concentration, different repressors and media type (data not shown). The optimal conditions derived from these experiments were used in this study. Although we applied the optimized expression protocol, it could be that some peptide constructs require different optimal conditions for expression. Small differences in the amino acid sequence or the length of the construct can already transform a protein that is easily expressed into a protein that fails to be expressed (26). Normally, expression conditions are optimized for each protein individually, which is not applicable for library screening. This can result in suboptimal expression conditions for a subpopulation of constructs (26,27).

The goal of this study was to screen the multimodular peptide library for their transfection efficiency. For a standard *in vitro* transfection experiment, about 50µg of peptide is needed to transfect 10^5 cells in culture with molar ratios between 1 and 200. In order to obtain a sufficient amount of multimodular peptide, 7 constructs were selected for 'large' scale expression. The 16 fold increase in culture size resulted in multimodular peptides yields between 20-120µg. However, single step IMAC purification resulted in co-purification of significant amounts of bacterial proteins, which contributed to 20-90% of the total yield. Different purification methods could be used to decrease the level of contaminations, for example mAb- or streptavidin based purification, but this would not result in higher yields (28,29). Although direct recombinant expression of small, positively charged peptides (e.g. antimicrobial peptides) is possible, yields differ greatly and depend on peptide size and amino acid composition (30). Compared to the peptides used in our study, most directly expressed peptides contain low charge densities.

A frequently used strategy to reduce toxicity to the expression host and to decrease proteolytic degradation of the expressed peptide is fusion to a carrier protein (23,29,31,32). As in nature, where host defense peptides are usually expressed as inactive precursors, fusion to a carrier protein protects both the peptide and their host during protein synthesis, which can increase peptide yield (24,30). However, most fusion proteins used need to be removed after expression, either by chemical or enzymatic cleavage, which can result in lower yields due to inefficient cleavage or by degradation during cleavage (29). Furthermore, the yields differ greatly between peptides and no fusion protein is universal (23,32).

In contrast to our results, other groups were able to recombinantly produce modular peptide constructs for gene delivery purposes in *E.coli* with yields ≥2mg/L. For example, the modular nanotransporters (MNT) by Slastnikova et al., multifunctional DNA carriers by Glover et al. and the biopolymer FP–(DCE)n–NLS–CS–TM by Canine et al. (33-35). However, almost all of these multimodular peptide constructs contain 1 or more functional proteins or protein domains in their native, folded conformation, which makes them less susceptible to proteolytic degradation. Furthermore, these functional proteins can act as fusion partners and reduce the linear structure of the constructs which decreases proteolytic degradation and improves expression. Only the designer biomimetic vector (DBV) reported by Mangipudi et al. is comparable to our multimodular peptides, with a size of 11.4 kDa. It contains a similar DNA condensing peptide, NLS peptide and endosomolytic peptide (36). The only difference at the construct level is the presence of a targeting peptide in the DBV. This allows the formation of dimers *in vivo*, which could reduce the proteolytic degradation of modular construct. Also, all groups used different repressors (pREP4 or pLysS), different *E.coli* strains (BL21(DE_3), M15 or BL21-Rosetta2) and different culture conditions, including expression temperature, culture medium and inducer concentrations. All groups did use the same T7 promoter. The many variables between the expression parameters make comparison very difficult and, as mentioned earlier, optimal culture conditions differ from peptide to peptide.

Comparable to the functional peptides used in gene delivery, antimicrobial peptides (AMPs) are small (<100 AA), positively charged peptides with membrane permeabilization capacities (37). Most AMPS are recombinantly expressed using either *E.coli* (82%) or Yeasts (16%) (38). Due to their susceptibility to proteolytic degradation and host toxicity, almost all (94%) AMPs are expressed as fusions with other proteins in *E.coli* (38). Fusion proteins can enhance solubility (e.g. SUMO, GST and TrxA), can drive expression to inclusion bodies protecting them from intracellular degradation (e.g. KSI and TrpΔLE) and can be used as affinity tags (c-myc, FLAG and HA) (30,39). However, no single affinity tag is ideal and the effect on the fusion protein varies according to which peptide is fused to it (37).

An alternative to recombinant production of the multimodular peptide libraries is chemical synthesis. However, synthesis of peptides is severely limited by decreased reaction yields with increasing peptide lengths, especially for the range tested here. Longer peptides can be generated by linkage of shorter fragments; however, this is time consuming, inefficient and an expensive process (40).

Overall, we demonstrated the generation of a genetic library encoding multimodular peptides and the subsequent recombinant expression and purification of individual peptide constructs. Unfortunately, direct expression of the recombinant peptides in *E. coli* did not result in sufficient peptide yields to screen the multimodular peptides for transfection efficiency.

Abbreviations

bp:	base pairs
EGFP:	enhanced green fluorescent protein
HIS-tag:	polyhistidine tag
IMAC:	immobilized metal ion affinity chromatography
IPTG:	isopropyl β-D-1-thiogalactopyranoside
LB:	Lennox broth
kDa:	kilo dalton
NLS:	nuclear localization signal
nt:	nucleotides
ODN:	oligodeoxynucleotides
PAGE:	polyacrylamide gel electrophoresis
SDS-PAGE:	sodium dodecyl sulfate polyacrylamide gel electrophoresis
ss:	single-stranded
TB:	terrific broth

Literature

(1) Glover DJ. Artificial viruses: exploiting viral trafficking for therapeutics. Infect Disord Drug Targets 2012 Feb;12(1):68-80.

(2) Jang JH, Lim KI, Schaffer DV. Library selection and directed evolution approaches to engineering targeted viral vectors. Biotechnol Bioeng 2007 Oct 15;98(3):515-24.

(3) McCarthy HO, Wang Y, Mangipudi SS, Hatefi A. Advances with the use of bio-inspired vectors towards creation of artificial viruses. Expert Opin Drug Deliv 2010 Apr;7(4):497-512.

(4) Macewan SR, Chilkoti A. Harnessing the power of cell-penetrating peptides: Activatable carriers for targeting systemic delivery of cancer therapeutics and imaging agents. Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology 2013;5(1):31-48.

(5) Torchilin VP. Cell penetrating peptide-modified pharmaceutical nanocarriers for intracellular drug and gene delivery. Biopolymers - Peptide Science Section 2008;90(5):604-610.

(6) Miyata K, Nishiyama N, Kataoka K. Rational design of smart supramolecular assemblies for gene delivery: chemical challenges in the creation of artificial viruses. Chem Soc Rev 2012 Apr 7;41(7):2562-2574.

(7) Mastrobattista B,van der Aa, Crommelin. Nonviral gene delivery systems: From simple transfection agents to artificial viruses. Drug Discovery Today: Technologies 2005;2(1):103-109.

(8) Dimarco RL, Heilshorn SC. Multifunctional materials through modular protein engineering. Adv Mater 2012;24(29):3923-3940.

(9) de Raad M, Kooijmans SA, Teunissen EA, Mastrobattista E. A Solid-Phase Platform for Combinatorial and Scarless Multipart Gene Assembly. ACS Synth Biol 2013 Mar 15.

(10) Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 2007 Jul;35(Web Server issue):W43-6.

(11) Henaut A, Danchin A. Analysis and predictions from Escherichia coli sequences, or E. coli in silico. Escherichia Coli and Salmonella: Cellular and Molecular Biology 1996:2047-2066.

(12) Schwartz B, Ivanov MA, Pitard B, Escriou V, Rangara R, Byk G, et al. Synthetic DNA-compacting peptides derived from human sequence enhance cationic lipid-mediated gene transfer in vitro and in vivo. Gene Ther 1999 Feb;6(2):282-92.

(13) Fortunati E, Ehlert E, van Loo ND, Wyman C, Eble JA, Grosveld F, et al. A multi-domain protein for beta1 integrin-targeted DNA delivery. Gene Ther 2000 Sep;7(17):1505-15.

(14) Murray KD, Etheridge CJ, Shah SI, Matthews DA, Russell W, Gurling HM, et al. Enhanced cationic liposome-mediated transfection using the DNA-binding peptide mu (mu) from the adenovirus core. Gene Ther 2001 Mar;8(6):453-60.

(15) Ignatovich IA, Dizhe EB, Pavlotskaya AV, Akifiev BN, Burov SV, Orlov SV, et al. Complexes of plasmid DNA with basic domain 47-57 of the HIV-1 Tat protein are transferred to mammalian cells by endocytosis-mediated pathways. J Biol Chem 2003 Oct 24;278(43):42625-36.

(16) Rittner K, Benavente A, Bompard-Sorlet A, Heitz F, Divita G, Brasseur R, et al. New basic membrane-destabilizing peptides for plasmid-based gene delivery in vitro and in vivo. Mol Ther 2002 Feb;5(2):104-14.

(17) Kichler A, Leborgne C, Marz J, Danos O, Bechinger B. Histidine-rich amphipathic peptide antibiotics promote efficient delivery of DNA into mammalian cells. Proc Natl Acad Sci U S A 2003 Feb 18;100(4):1564-8.

(18) Midoux P, Kichler A, Boutin V, Maurizot JC, Monsigny M. Membrane permeabilization and efficient gene transfer by a peptide containing several histidines. Bioconjug Chem 1998 Mar-Apr;9(2):260-7.

(19) Ritter W, Plank C, Lausier J, Rudolph C, Zink D, Reinhardt D, et al. A novel transfecting peptide comprising a tetrameric nuclear localization sequence. J Mol Med 2003 Nov;81(11):708-17.

(20) de Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E. High-content screening of peptide-based non-viral gene delivery systems. J Control Release 2012 Mar 28;158(3):433-442.

(21) Kawe M, Horn U, Plückthun A. Facile promoter deletion in Escherichia coli in response to leaky expression of very robust and benign proteins from common expression vectors. Microbial Cell Factories 2009;8.

(22) Saïda F, Uzan M, Odaert B, Bontems F. Expression of highly toxic genes in E. coli: Special strategies and genetic tools. Current Protein and Peptide Science 2006;7(1):47-56.

(23) Li Y. Carrier proteins for fusion expression of antimicrobial peptides in Escherichia coli. Biotechnol Appl Biochem 2009;54(1):1-9.

(24) Vassilevski AA, Kozlov SA, Grishin EV. Antimicrobial peptide precursor structures suggest effective production strategies. Recent Patents on Inflammation and Allergy Drug Discovery 2008;2(1):58-63.

(25) Lui BH, Cochran JR, Swartz JR. Discovery of improved EGF agonists using a novel in vitro screening platform. J Mol Biol 2011;413(2):406-415.

(26) Peti W, Page R. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. Protein Expr Purif 2007;51(1):1-10.

(27) Noguère C, Larsson AM, Guyot J-, Bignon C. Fractional factorial approach combining 4 Escherichia coli strains, 3 culture media, 3 expression temperatures and 5 N-terminal fusion tags for screening the soluble expression of recombinant proteins. Protein Expr Purif 2012;84(2):204-213.

(28) Terpe K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 2003;60(5):523-533.

(29) Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 2006;48(1):1-13.

(30) Li Y. Recombinant production of antimicrobial peptides in Escherichia coli: A review. Protein Expr Purif 2011;80(2):260-267.

(31) Francis DM, Page R. Strategies to optimize protein expression in E. coli. Current Protocols in Protein Science 2010(SUPPL. 61):5.24.1-5.24.29.

(32) Esposito D, Chatterjee DK. Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol 2006;17(4):353-358.

(33) Canine BF, Wang Y, Hatefi A. Biosynthesis and characterization of a novel genetically engineered polymer for targeted gene transfer to cancer cells. J Controlled Release 2009;138(3):188-196.

(34) Glover DJ, Su MN, Mechler A, Martin LL, Jans DA. Multifunctional protein nanocarriers for targeted nuclear gene delivery in nondividing cells. FASEB Journal 2009;23(9):2996-3006.

(35) Slastnikova TA, Rosenkranz AA, Gulak PV, Schiffelers RM, Lupanova TN, Khramtsov YV, et al. Modular nanotransporters: A multipurpose in vivo working platform for targeted drug delivery. International Journal of Nanomedicine 2012;7:467-482.

(36) Mangipudi SS, Canine BF, Wang Y, Hatefi A. Development of a genetically engineered biomimetic vector for targeted gene transfer to breast cancer cells. Molecular Pharmaceutics 2009;6(4):1100-1109.

(37) Parachin NS, Mulder KC, Viana AAB, Dias SC, Franco OL. Expression systems for heterologous production of antimicrobial peptides. Peptides 2012;38(2):446-456.

(38) Li Y, Chen Z. RAPD: A database of recombinantly-produced antimicrobial peptides. FEMS Microbiol Lett 2008;289(2):126-129.

(39) Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnology Journal 2012;7(5):620-634.

(40) Mersich C, Jungbauer A. Generation of bioactive peptides by biological libraries. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences 2008;861(2):160-170.

6

Chapter

EGFR Targeted Multimodular Peptide Libraries for Gene Delivery

Markus de Raad[a], Raimond Heukers[b], Erik A. Teunissen[a], Paul M.P. van bergen en Henegouwen[b] and Enrico Mastrobattista[a]

[a] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands
[b] Division Cell Biology, Department of Biology, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Abstract

Here, we demonstrated the generation of a gene library encoding EGFR targeted multimodular peptides by fusing peptides to the C-terminus of the anti-EGFR biparatopic nanobody 7D12-9G8. After the expression and subsequent purification of 4 individual fusion constructs, we showed that the modified nanobody was still able to specifically bind to EGFR with similar affinity as the unmodified nanobody. The constructs produced were able to induce splice correction via the delivery of splice-correcting ON in HeLa pLuc/705 cells.

Introduction

In Chapter 5, we demonstrated the generation of a gene library encoding multimodular peptides and the subsequent recombinant expression and purification of individual peptide constructs. However, the yields of the peptides were not sufficient to screen these constructs for transfection efficiency.

Fusion proteins are proteins that are genetically fused to the target protein or peptide and are frequently used to increase the expression yields of peptides (1-3).

Fusion proteins can increase the expression yield by protecting the peptides from intracellular proteolysis or by reducing toxicity of the peptides towards the host (1,4). Fusion proteins are utilized for affinity purification, enhancing solubility, enabling detection of the recombinant fusion construct or for promoting the formation of inclusion bodies. However, after recombinant expression and purification, the fusion protein has no function and removal is desired to avoid potential loss of function and/or structure of the peptides. Fusion partners can be removed via enzymatic or chemical cleavage, but will result in lower yields of the final peptide due to inefficient cleavage, loss during handling or by degradation during cleavage (5,6). For the multimodular peptide libraries, this problem can be circumvented by using fusion proteins with functionalities suited for gene delivery, such as specific or non-specific DNA binding proteins and protein ligands for specific cell targeting (7-10).

Nanobodies are camelid-derived single-domain antibody fragments consisting of the variable domain of heavy chain antibodies (11). As a fusion protein, nanobodies have many favorable characteristics, including high solubility, intrinsic stability, the capacity to refold after denaturation while retaining their binding capacity, in addition to their modular nature and high expression yields in bacteria (11,12). Also, nanobodies have previously been used as fusion partner (11). All these characteristics make nanobodies attractive fusion proteins. From a gene delivery point of view, nanobodies display high target specificity, high binding affinity and have proven to be suitable as targeting ligands in nanoparticulate drug carrier systems (12-14).

In order to increase the expression yield of the multimodular peptide libraries, the biparatopic nanobody 7D12-9G8, which has specificity for the human epidermal growth factor receptor (hEGFR), was selected as an N-terminal fusion protein. The biparatopic nanobody contains two linked nanobodies, 7D12 and 9G8, directed against two non-overlapping epitopes on domain III of the hEGFR (15). hEGFR is implicated in several human epithelial malignancies and is nowadays recognized as target for therapy. The biparatopic nanobody has a higher binding avidity than the monovalent nanobodies (15). Also, upon binding of 7D12-9G8 to EGFR, both the nanobody and the receptor are rapidly internalized, followed by intracellular transport to and entrapment in lysosomes (16,17). By N-terminally adding this nanobody to the multimodular peptide libraries containing endosomolytic and membrane destabilizing

peptides, endosomal escape of the nanobody/nucleic acids complexes can be facilitated, which is required to deliver the nucleic acid cargo at their specific intracellular target site. Also, the endosomal escape of the nanobody itself can be promoted, which is beneficial for immunotoxins or other immuno-fusion constructs with intracellular targets (11).

The multimodular peptide libraries will be fused to the C-terminus of the 7D12-9G8 anti-EGFR nanobody (Fig. 1c), generating EGFR targeted multimodular peptide libraries. Two different libraries will be generated, one with and one without a cathepsin D substrate between the nanobody and the library. The cathepsin D substrate is recognized and cleaved by the endoprotease cathepsin D, which is abundant in late endosomes, and facilitates dissociation of the nanobody from the peptides after internalization (18).

In this study, the aim is to increase the yield of the multimodular peptide libraries recombinantly produced in *E. coli* by C-terminal fusion of the biparatopic anti-EGFR nanobody 7D12-9G8 to the peptide library. After recombinant expression and purification, the constructed targeted libraries were tested for their capacity to deliver an exon-skipping ON into the cytosol of EGFR-expressing target cells.

Materials and Methods

Materials

All chemicals and reagents were purchased from Sigma-Aldrich Chemie B.V. (Zwijndrecht, The Netherlands), unless stated otherwise. Dynabeads M-280 Streptavidin were purchased from Dynal (Oslo, Norway). Restriction enzymes, FastAP™ (Thermosensitive Alkaline Phosphatase), Generuler™ DNA and Riboruler™ RNA ladders, Phusion polymerase, Rapid DNA Ligation Kit, GeneJET™ Plasmid Miniprep Kit, isopropyl β-D-1-thiogalactopyranoside (IPTG), RNase A, DNase I, PageRuler™ Prestained Protein Ladder and HisPur™ NI^{2+}-NTA Resin were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Precast denaturing polyacrylamide TBE-urea gels (6%) and Lipofectamine™ 2000 were purchased from Invitrogen (Breda, The Netherlands). Precast Criterion™ XT Bis-Tris gels (12%) and Criterion™ TBE gels (4-20%) were purchased from Bio-Rad (Hercules, CA, USA). DMEM (Dulbecco's modified Eagle's Medium, with 1g/l glucose, 584mg/ml L-Glutamine), Fetal Bovine Serum (FBS), and Phosphate Buffered Saline (PBS) were purchased from PAA Laboratories GmbH (Pasching, Austria). The Luciferase Assay System was from Promega (Madison, WI, USA). The used bacterial culture media: Lennox Broth (LB) (tryptone (10g/L), yeast extract (5g/L) and NaCl (5g/L)), LB-agar (tryptone (10g/L), yeast extract (5g/L), NaCl (5g/L) and agar (15g/L)) and Terrific Broth (TB) (tryptone (12g/L), yeast extract (24g/L), K_2HPO_4 (9.4g/L), KH_2PO_4 (2.2g/L) and glycerol (8mL/L)) were purchased from Sigma-Aldrich. Al used solid and liquid media contained glucose (1% w/v) and ampicillin (100µg/mL).

The pET28a-(7D12-9G8) encoding the biparatopic nanobody 7D12-9G8 (30kDa) was kindly provided by Dr. van Bergen en Henegouwen (Cell Biology, Department of Biology, Utrecht University).

Synthetic oligodeoxynucleotides

All oligodeoxynucleotides were synthesized by Eurogentec S.A. (Seraing, Belgium). Single-stranded oligodeoxynucleotides (ssODN) modules were synthesized with 5'- and 3'- phosphates and PAGE purified at a 40nmol scale or with 5'biotin-TEG and PAGE purified at a 200nmol scale. ssODN linkers and primers were synthesized unmodified and SePOP desalted at a 40nmol scale. Oligodeoxynucleotides were dissolved in nuclease-free water and stored at -30°C as 100 µM stocks.

Oligodeoxynucleotides encoding peptides were codon-optimized for bacterial expression and flanked with linker sequences GGTTCT (GS) at the 5' (all except Head modules) and GGTGGC (GG) at the 3' (all except Tail modules). The possible formation of secondary structures (checked with Oligocalc software) was reduced by adjustment of any self-complementary oligonucleotide sequence (19).

The sequence of the splice-correcting oligonucleotide ON-705 was 5'-CCU-CUUACCUCAGUUACA-3'and was fully 2'-O-methylated. ON-705 was purified using RP-HPLC, dissolved in nuclease-free water and stored at -30°C as 100 µM stock. All sequences of the used ssODN and primers are listed in Table 1.

Synthetic peptides

All synthetic peptides were ordered with an N-terminal biotin at crude purity from Genscript (Piscataway, NJ, USA). Lyophilized peptides were reconstituted in reversed osmosis water with/without the addition of 10-20% (v/v acetic acid) to 2 mM stock concentrations. Peptides were stored at -30 °C

Methods

Construction of pCmB destination vector for multimodular peptide libraries

As basis for the destination vector pCmB, the plasmid pSCherry2 (Eurogentec) was used. A DNA fragment containing the biparatopic 7D12-9G8 nanobody coding sequence was obtained by digestion of the pET28a-(7D12-9G8) with XbaI/HindIII. The fragment was ligated into the XbaI/HindIII digested pSCherry2 vector. The resulting plasmid was the library destination vector pCmB.

Construction of multimodular peptide libraries by solid phase combinatorial multipart gene assembly

Using the method for the assembly of ssODN in the presence of linker ODN, a random polypeptide library was assembled (Chapter 3) (20). Ligation reactions (20µL) contained 12.5pmol ssHead (Head_+C_GG or Head_-C_GG) coated on M-280 magnetic beads, 50 pmol GGGS linker, 1x Rapid Ligation Buffer, 5U T4 DNA ligase and 100 pmol of a mix of 9 different ssODN (11,1 pmol of GS_Pr18_GG; GS_SPKR4_GG; GS_H9.2_GG; GS_H5WYG_GG; GS_TAT_GG; GS_ppTG20_GG; GS_Mu_GG; GS_NLSV402_GG and GS_LAH5_GG). Five ligation steps with this composition were performed. One additional ligation step was performed for the ligation of the reaction products to a terminating ssTail (GS_Tail). Ligation products were loaded on a 6% denaturing polyacrylamide TBE-urea gel and the area between the smallest theoretical product (±150 nucleotides) and the largest theoretical product (±600 nucleotides) was excised, purified and PCR amplified (as described in Chapter 3) (20). PCR products were analyzed on a 2% w/v agarose gel and the area between ±200-600bp was excised and purified.

The library obtained was BamHI/ HindIII digested, purified and 27-140ng of the digested products were ligated in 100ng of dephosphorylated pCmB (BamHI/ HindIII digested and purified from an agarose gel), to create a C-terminal fusion between the biparatopic nanobody 7D12-9G8 and the library inserts. Half of the reaction mixture was used directly to transform *E.coli* strain DH5α (Invitrogen). Bacteria were plated out on LB-agar plates and grown overnight at 37°C. From the obtained LB plates, all colonies were harvested using 2ml LB. Harvested colonies were grown in 10ml LB for 4hr at 37°C and plasmid DNA (pDNA) was harvested.

Recombinant protein production in *E. coli*

Fifty nanogram library pDNA was used to transform *E. coli* strain BL21-Gold(DE_3) (Stratagene, Santa Clara, CA, USA). Bacteria were plated out on LB-agar and grown overnight at 37°C. From the plates obtained colonies were inoculated overnight in 5mL LB. Glycerol stocks were prepared and pDNA was harvested from all overnight cultures. pDNA of the library constructs obtained were XbaI/HindIII digested and analyzed on a 1% w/v agarose gel. In addition, pDNA of the library constructs obtained were sequenced for insert identification with full coverage of the insert sequences (inserts were fully sequenced; BaseClear, Leiden, The Netherlands).

For protein expression, starter cultures were grown overnight in 5-50ml LB from the obtained glycerol stocks. Next, cultures were diluted to an optical density at 600nm (OD_{600}) of 0.1 in 30-500mL Terrific Broth. The diluted cultures were grown at 37°C to OD_{600} of 0.4-0.8, when protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.1mM final concentration. After 5 h, bacteria were harvested by centrifugation (5,000xg, 15 min) and stored at –20°C.

Table 1 Used oligodeoxynucletides

Name	Sequences
Head_+C_GG	5'-Biotin-TEG-GTTATTGGATCCGGCGGTGGCGGTAGCGGTGGCTTCTTAGGGTT-TGGTGGC-OH 3'
Head_-C_GG	5'-Biotin-TEG-GTTATTGGATCCGGTGGGTCTGGTTCTGGTGGC-OH 3'
Mu	5' P-GGTTCTCGTCGTGCTCATCATCGGCGTAGACGTGCATCTCATCGCCGTATGCGT-CGCGGAGGTGGTGGC-P 3'
H5WYG	5' P-GGTTCTGGTCTGTTTCACGCGATTGCCCATTTCATCCACGGTGGTTGGCATGG-TTTAATTCACGGTTGGTATGGCGGTGGC-P 3'
H9.2	5' P-GGTTCTAAGACACCGAAAAAGGCCAAAAAGCCAAAAACCCCGAAAAAGGC-GAAAAAACCAGGTGGC-P 3'
LAH5	5' P-GGTTCTAAAAAGGCACTGCTGGCACTGGCACTCCATCACTTAGCACACCTT-GCTCATCATCTTGCGTTAGCGCTGAAAAAGGCTGGTGGC-P 3'
NLSV402	5' P-GGTTCTCCAAAAAAGAAACGTAAAGTTCCAAAAAAAAAGCGCAAAGTCGG-TGGC-P 3'
ppTG20	5' P-GGTTCTGGCTTATTTCGTGCGCTGTTGCGTCTGTTACTGTCTCTGTGGAGAT-TACTTTTACGTGCGGGTGGC-P 3'
Pr18	5' P-GGTTCTTCTCGTAGTCGGTATTACCGTCAGCGCCAACGTTCTCGCCGTCGC-CGGCGTAGAGGTGGC-P 3'
SPKR4	5' P-GGTTCTAGCCCGAAACGTAGCCCTAAGCGCAGCCCAAAAAGATCTCC-TAAACGTGGTGGC-P 3'
TAT (47-57)	5' P-GGTTCTTATGGCCGCAAGAAGCGTCGTCAAAGACGTCGTGGTGGC-P 3'
GGGS linker	5' OH-GGTGGCGGTTCT-OH 3'
GS_Tail	5' OH-GGTTCTCATCACCATCACCATCACCATCACTAATAAAAGCTTAAAGAG-OH 3'
Fw_Head_+C	5' OH-GTTATTGGATCCGGCGGT-OH 3'
Fw_Head_-C	5' OH-GTTATTGGATCCGGTGGGT-OH 3'
Rev_Tail	5' OH-GTCATACTTATGGGATCCTTATTAA-OH 3'

Bacterial lysis

The bacterial pellets obtained were thawed at 4°C, suspended in 5ml/g wet pellet denaturing lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole and 8M urea at pH 8.0), and incubated for ±8h on an orbital shaker (200rpm) at room temperature (RT). Next, bacteria were snap-frozen using liquid nitrogen and incubated for ±16h on an orbital shaker (200rpm) at RT. Bacteria were centrifuged at 10,000xg for 30min at 4°C to pellet the cellular debris and supernatant was used for purification/analysis.

Protein purification

The targeted multimodular peptides produced were purified from the supernatants obtained after lysis using Ni^{2+}-NTA affinity chromatography. To the supernatants, HisPur™ Ni^{2+}-NTA slurry was added (25µl slurry for 30ml cultures and 250µl slurry for

550ml cultures) and the suspension was mixed on an end-over-end rotator for ±2.5h at RT. Then, the lysate/resin mixture was loaded on polypropylene columns and washed with 5 resin bed volumes wash buffer (50mM NaH_2PO_4, 300mM NaCl, 8M urea and 40mM imidazole at pH 8.0). Subsequently, His-tagged modular peptides were eluted with 3x 1 resin bed volumes of elution buffer (50mM NaH_2PO_4, 300mM NaCl, 8M urea and 300mM imidazole at pH 8.0).

Quantity and purity of the targeted modular peptides was analyzed by SDS-PAGE. For SDS-PAGE, samples were run on Criterion™ XT Bis-Tris 12% gels or 4–12% gradient NuPAGE™ Novex Bis-Tris mini-gels and visualized by PageBlue™ Protein Staining Solution (based on Coomassie G-250, Thermo Fisher Scientific).

For buffer exchange and refolding, targeted modular peptide solutions were dialysed successively against PBs containing 4M, 2M, 1M and 0M urea, using Slide-A-Lyzer MINI Dialysis Device (Thermo Fisher Scientific) with a molecular weight membrane cut-off of 10kDa.

Protein concentration was determined with the MicroBCA™ protein assay (Pierce Biotechnology, Rockford, IL, USA) against DNaseI as a standard protein.

Cell culture

HeLa pLuc/705 cells were grown in DMEM supplemented with 10% heat inactivated FBS, hygromycin (200µg/mL) and antibiotics/antimycotics. The Her14 cells (expressing high levels of EGFR) and the NIH 3T3 clone 2.2 (abbreviated as 3T3 2.2, expressing low levels of EGFR) cells were grown in DMEM supplemented with 7.5% (v/v) FBS without antibiotics (21). All cells were maintained at 37 °C in a 5% CO2 humidified air atmosphere.

Binding assay

Eight thousand NIH3T3 2.2 or HER14 cells were seeded into 96-well tissue culture plates 48h prior to binding assay. The cells were pre-incubated with binding medium (2% BSA in completed medium) for 10 min on ice and subsequently incubated with a concentration range (2.5-2000nM) of targeted multimodular peptide constructs in binding medium for 2h on ice. The cells were washed with ice cold PBS, fixed in 4% PFA and subsequently, the autofluorescence was quenched with 100mM glycine in PBS for 15 minutes. Bound nanobody was detected with rabbit-anti-VHH (1:1000, developed in house) in PBS containing 2% BSA (PBA) followed by goat-anti-rabbit IRDye800CW (Li-COR Biosciences, Cambridge, United Kingdom, 1:500 in PBA), which were incubation for 1h at RT each. Finally, cells were washed twice with PBA and fluorescence was quantified using the Odyssey Infrared Imager (LI-COR, Lincoln, NE, USA). Apparent binding affinity (kD) and for one site specific binding was determined by curve fitting.

Preparation of peptiplexes and lipoplexes

Complexes were prepared by adding 4 volumes of fusion-peptide/lipid (Lipofectamine) solution to 1 volume of ON-705 solution (5-200nM) and mixing immediately by pipetting up and down 10 times. The complexes were incubated for 30min at room temperature. Targeted multimodular peptide complexes were prepared in PBS. Lipoplexes were prepared in DMEM according to manufactures protocol. For each experiment, complexes were freshly prepared in triplicate. The ratios of fusion constructs/peptide to ON-705 are expressed as molar ratios.

Quantification of free 2'-OMe RNA by gel electrophoresis

Complexes were prepared as described above. Twenty five microlitres of complex samples, containing 5pmol ON-705 and targeted multimodular peptide constructs at different molar ratios (0-40), were subjected to electrophoresis on non-denaturing PAGE on Criterion™ TBE 4-20% gels and ON-705 was visualized by post staining using GelRed™ Nucleic Acid Gel Stain (Biotium, Hayward, CA, USA). The amount of free ON-705 was quantified using Image Lab™ software (BioRad).

Splice correction assay

Ten thousand HeLa pLuc/705 cells were seeded into 96-well tissue culture plates 24h prior to transfection. Immediately prior to transfection the culture medium was refreshed with 100µl serum-free medium. Twenty five microliters of the peptiplex/lipoplex samples (corresponding to 0,05- 5 pmol ON-705/well) was added per well and after 4h incubation, medium was replaced with 100µl complete medium. Cells were incubated for indicated times at 37°C in a 5% CO_2 humidified air atmosphere. Twenty-four hours after transfection, cells were washed with 100µL PBS and lysed with 100µL lysis buffer (Luciferase Cell Culture Lysis Reagent (Promega)). A freeze-thaw cycle was performed by incubating cells for 1h at -80°C. Luciferase activity was measured on FLUOstar OPTIMA microplate reader, equipped with a luminescence light guide (BMG LabTech, Offenburg, Germany), by mixing 50µL of cell lysate with 50µL of Luciferase Assay Reagent (Promega). Luciferase expression is expressed as relative light units (RLU).

Results

Construction of a gene library encoding targeted multimodular peptide constructs by solid phase combinatorial multipart gene assembly

A gene library encoding multimodular peptides was constructed in a combinatorial manner by ligation of single-stranded oligodeoxynucleotides (ssODN) using DNA ligase in the presence of linker ssODN (Chapter 3) (20). Briefly, a ssODN (Head)

modified with 5'-biotin was immobilized on streptavidin-coated magnetic beads. Using linker ODN, which were complementary to the last 6 nucleotides of the immobilized ssODN and the first 6 nucleotides of ssODN to be ligated, DNA ligase was used to ligate ssODN in a stepwise manner. In the final ligation step, a single-stranded oligodeoxynucleotide (Tail) was added.

For the generation of multimodular peptide libraries, 9 different functional peptides were selected: 4 DNA condensing peptides ($SPKR_4$, Pr18, Mu and H9.2), 4 membrane disrupting and/or endosomolytic peptides (LAH5, ppTG20, H5WYG and TAT(47-57)) and 1 NLS peptide (NLSV402) (Table 2).

By using 2 different Heads (Head+C and Head-C), 2 different libraries were created. Libraries created with the Head+C (referred to as +C libraries) contain a linker sequence GSGGGGS and the sequence GGFLGF designed for cleavage by the abundant endosomal endoprotease cathepsin D (Table 2) (18). Libraries created with the Head-C (referred to as -C library) contain only a linker sequence GSGGSGS (Table 2). All library constructs were terminated with a universal Tail, which contained a 8xHIS-tag followed by a double stop codon TAATAA. Both Heads and Tail contained restriction sites for cloning purposes. The Heads, Tail and all peptides were reverse translated into ssODN sequences (using the *E.coli* codon usage table, and flanked with linker sequences GGTTCT (encoding Gly-Ser) at the 5' (except Head) and GGTGGC (encoding Gly-Gly) at the 3' (except Tail) (Table 2)(22).

Five ligation cycles were performed with a mix containing equal amounts of ssODN encoding all 9 functional peptides, generating a library of Head-MIX_5-Tail. This would theoretically yield a library with $(9^1)+(9^2)+(9^3)+(9^4)+(9^5)=66,429$ different constructs with $9^5=59,049$ full length constructs. The generated constructs were analyzed on a 6% TBE urea gel, after loading equal amounts of magnetic beads.

Both libraries obtained were clearly visible and appeared as smears, with highest intensity between 400 and 600 bases (Fig. 1a). Individual non-ligated Heads and Head-MIX_5 constructs were still visible in both libraries.

After extraction from gel and purification, both libraries were amplified by PCR to generate dsDNA products and enough material for further steps (Fig. 1b). After PCR amplification, smears were detected indicating that all constructs formed were amplifiable. PCR showed an amplification bias towards the smaller templates, with clear bands corresponding to constructs encoding only 1 or 2 peptides (Fig. 1b).

The PCR amplified Head-MIX_5-Tail libraries were excised from ±200bp to ±600bp, purified and cloned into the BamHI/HindIII-sites of vector pCmB, thereby fusing the gene library in frame to the C terminus of the biparatopic nanobody 7D12-9G8. Two different targeted libraries were created, one library with a cathepsin D site between the nanobody and peptide(s) and one without (Fig. 1c). *E.coli* strain DH5α was transformed with both cloned libraries. To obtain the pDNA libraries, all colonies of a single plate were pooled and pDNA was harvested.

Table 2 Function, name, and amino acid sequence of functional peptides used. Nucleotide sequence of the corresponding ssODN constructs to produce modular gene constructs are also indicated

Function	Peptide	AA sequence	ssDNA sequence	Ref.
DNA condensation	Pr18	GSSRSRYYRQRQRSRRRRRRGG	ggttcttctcgtagtcggtattaccgtcagcgccaacgttctcgccgtcgc-cggcgtagaggtggc	(23)
	$SPKR_4$	GSSPKRSPKRSPKRSPKRGG	ggttctagcccgaaacgtagccctaagcgcagcccaaaaagatctcc-taaacgtggtggc	(7)
	H9.2	GSKTPKKAKKPKTPKKAKKPGG	ggttctaagacaccgaaaaaggccaaaaagccaaaaaccccgaaaaaggc-gaaaaaaccaggtggc	(23)
	Mu	GSRRAHHRRRRASHRRMRRGG	ggttctcgtcgtgctcatcatcggcgtagacgtgcatctcatcgccgtatgcgt-cgcggaggtggtggc	(24)
Membrane disruption/ translocation	TAT (47-57)	GSYGRKKRRQRRRGG	ggttcttatggccgcaagaagcgtcgtcaaagacgtcgtggtggc	(25)
	ppTG20	GSGLFRALLRLLRSLWRLLLRAGG	ggttctggcttatttcgtgcgctgttgcgtctgttactgtctctgtggagat-tacttttacgtgcgggtggc	(26)
	LAH5	GSKKALLALALHHLAHLAHHLALALKKAGG	ggttctaaaaaggcactgctggcactggcactccatcacttagcacaccttgct-catcatcttgcgttagcgctgaaaaaggctggtggc	(27)
Endosomolytic	H5WYG	GSGLFHAIAHFIHGGWHGLIHGWYGGG	ggttctggtctgtttcatgcgattgcccacttcatccatggtggttggcatggtt-taattcacggttggtatggcggtggc	(28)
NLS	NLSV402	GSPKKKRKVPKKKRKVGG	ggttctccaaaaaagaaacgtaaagttccaaaaaaaaagcgcaaagtcgg-tggc	(29)
Head+C	Head_+C_GG	GSGGGGS*GGFLGF*GG	gttatt**ggatcc**ggcggtggcggtagcggtggcttcttagggtttggtggc	-
Head-C	Head_-C_GG	GSGGSGSGG	gttatt**ggatcc**ggtgggtctggttctggtggc	-
Tail	GS_Tail	GSHHHHHHHH	ggttctcatcaccatcaccatcaccatcacTAATAA**aagctt**aaagag	-

Underlined sequences correspond to GGGS linker sites. The amino acids sequence in italics corresponds to a cathepsin D cleavage site. ssODN sequence in capitals corresponds to the double stop codons in the Tail. ssODN sequence in bold corresponds to BamHI site in the Head and HindIII site in the Tail.

Library screening

The obtained +C and -C library pDNA was used for the transformation of *E.coli* strain BL21-Gold(DE_3). To confirm library presence and determine construct size, restriction analysis was performed on pDNA obtained from 24 colonies from the +C library and 24 colonies from the -C library, after overnight culturing (Fig. 1d). Library inserts appear between 900-1400bp (nanobody+library insert). After restriction enzyme analysis, 42 out of 48 (88%) of screened colonies contained DNA fragments corresponding to the expected sizes. All 42 positive colonies contained inserts corresponding to constructs with at least one functional peptide. Sequence analysis confirmed the error-free ligation and random insertion of the used functional peptides (Table 3, not all data shown).

After pDNA analysis, 12 colonies from +C library and 12 colonies from -C library were screened for protein expression. After expression, bacterial lysis and Ni^{2+}-NTA

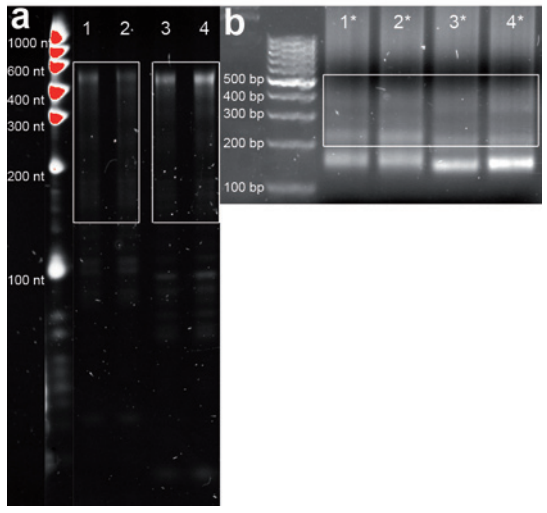


Figure 1 Gene library construction. (a) Initial ligation products were loaded on a 6% TBE-urea gel and purified at final product height as indicated. Ladder: low range RNA ladder. Lanes 1 and 2: Head+C-MIX_5-Tail in duplo. Lanes 3 and 4: Head-C-MIX_5-Tail in duplo (b) Purified ligation products were amplified using PCR and loaded on a 2% agarose gel. Ladder: 100bp DNA ladder. Lanes 1*-4*: PCR-amplified ligation products corresponding to lanes 1-4. (c) Schematic representation of the generated targeted multimodular peptide constructs. The constructs consist of a N-terminal biparatopic 7D12-9G8 nanobody, a cathepsin D cleavage site in case of +C libraries and 1-5 functional peptides.(d) Plasmid DNA isolated from 24 colonies from the -C library were digested with XbaI/HindIII and loaded on 1% agarose gel. Ladder: 1kb and 50bp DNA ladders. Vector backbone appears as a single band of ±5800bp whereas library inserts appear between 900-1400bp.

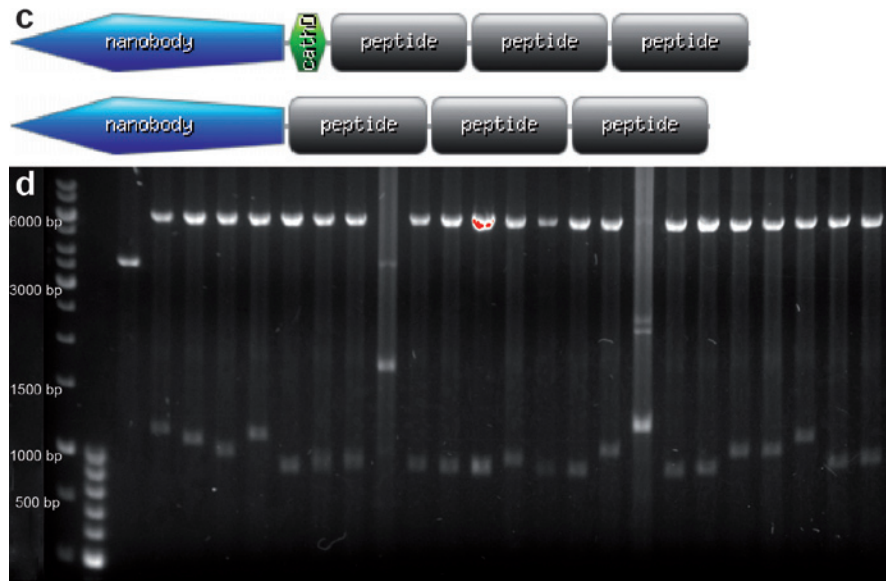


Figure 1 Continued.

purification, the production of the targeted multimodular peptides was analyzed by SDS-PAGE (Fig. 2). For the negative control, non-induced bacteria were used.

Fifteen peptide constructs were detected after SDS-PAGE analysis and staining with Coomassie G-250. All produced peptides were in range of the expected theoretical sizes (33-50kDa).

Large scale expression and characterization of the targeted multimodular peptides

In order to obtain higher quantities of the targeted peptide constructs needed to perform cellular assays, 4 peptide constructs were selected and expressed at 0.5L scale: +C6, +C8, +C9 and -C7. After expression and purification, targeted peptide constructs were refolded in a step-wise manner by dialysis against PBS containing 4M, 2M, 1M and 0M urea, consecutively using Slide-A-Lyzer MINI Dialysis Devices (Thermo Fisher Scientific). The cleared lysate, the eluted targeted peptide constructs and refolded targeted peptide constructs were analyzed using SDS-PAGE and the yield was quantified (Fig. 3). Also, pDNA corresponding to the 4 constructs was sequenced to identify peptide composition (Table 3).

All purified and refolded targeted peptide constructs were detected after SDS-PAGE using Coomassie staining. All refolded constructs had a purity of 65%-95% at a yield of 9-25 nmol/L *E.coli*.

To determine the ability of the targeted multimodular peptides to bind specifically to EGFR over-expressing cells, cells expressing high levels of EGFR (Her14) and cells expressing low levels of EGFR (3T3 2.2) were incubated with the fusion constructs. The biparatopic nanobody 7D12-9G8 without peptides was used as control.

All 4 fusion constructs were able to specifically bind to EGFR over-expressing cells and had binding affinities similar to that of the nanobody without peptides, except for -C7. (Fig. 4a and 4b). However, the fusion constructs had also a higher binding affinity to the low EGFR expressing 3T3 2.2 cells compared to the nanobody control.

Splice correction mediated by targeted multimodular peptides

After determining the ability to bind specifically to EGFR over-expressing cells, the targeted multimodular peptides were evaluated for the delivery of 2'-OMe ON-705 in

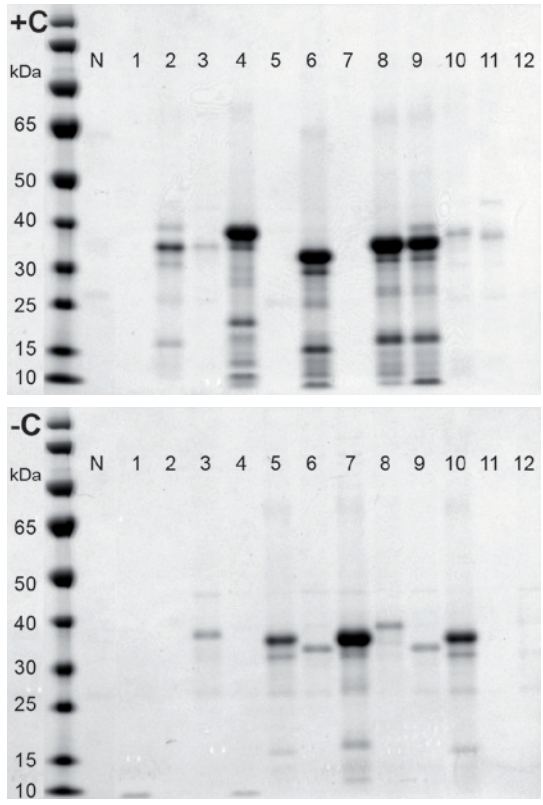


Figure 2 SDS-PAGE analysis of purified targeted multimodular peptides visualized using Coomassie G-250 (PageBlue™ Protein Staining Solution). Similar volumes of elution fraction were loaded on gel. (N) non-induced bacteria. Ladder: PageRuler™ Prestained Protein Ladder.

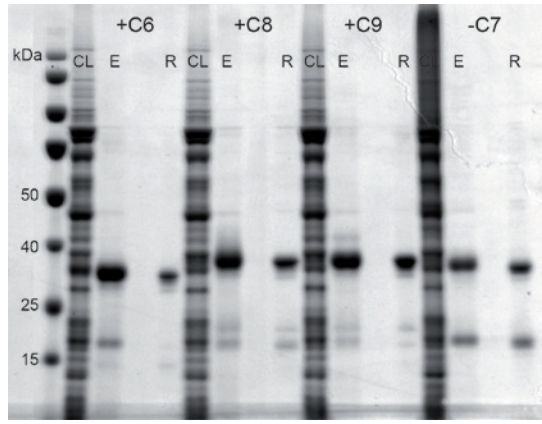


Figure 3 SDS-PAGE analysis of cleared lysate, purified and refolded +C6, +C8, +C9 and -C7 constructs visualized using Coomassie G-250 (PageBlue™ Protein Staining Solution). Similar volumes were loaded on gel. (CL) cleared lyate, (E) elution fraction, (R) refolded fraction. Ladder: PageRuler™ Prestained Protein Ladder.

Table 3 Peptide composition and total protein yield after refolding

Name	Composition	Yield after refolding (µg)	Yield after refolding (nmol)
+C6	TAT	155	4.5
+C8	H9.2	390	11.2
+C9	LAH5-NLSV402	380	10.9
-C7	H5WYG	440	12.6

HeLa pLuc/705 cells, using the splice correction assay. This assay is based on the HeLa pLuc/705 cell line, which is stably transfected with a luciferase-encoding gene interrupted by a mutated β-globin intron 2 (30,31). The mutation causes aberrant splicing of luciferase pre-mRNA resulting in the synthesis of non-functional luciferase (30). The aberrant splice site can be masked with the antisense 2'-OMe oligonucleotide 705 (ON-705), which redirects splicing towards the correct mRNA and consequently restores luciferase activity.

First, to validate the splice correction assay and to optimize assay parameters (e.g. ON-705 concentration), Lipofectamine was used to transfect the cells in the presences or absences of serum (Fig. 5). The maximum splice correction was obtained with 20-200nM ON-705 and transfection efficiency was similar in the presence or absences of serum.

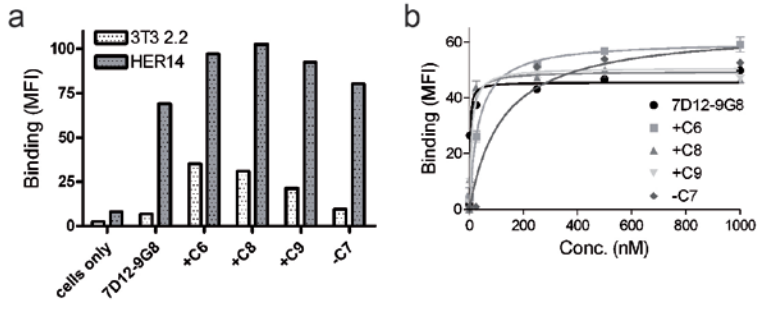


Figure 4 Cell binding of +C6, +C8, +C9 and -C7 fusion constructs. (a) HER14 and 3T3 2.2 cells were incubated with 2 µM of either +C6, +C8, +C9, -C7 or 7D12-9G8. Bound constructs or nanobody were detected with rabbit anti-VHH, followed by goat-anti-rabbit IRDye800CW and fluorescence was quantified with an Odyssey scanner. (b) HER14 cells were incubated with 0-1000nM of either +C6, +C8, +C9, -C7 (dashed lines) or Nanobody 7D12-9G8 without fusion peptides (solid line). Bound constructs or nanobody were detected with rabbit anti-VHH, followed by goat-anti-rabbit IRDye800CW and fluorescence was quantified with an Odyssey scanner. Data are presented as mean + SD, N=3.

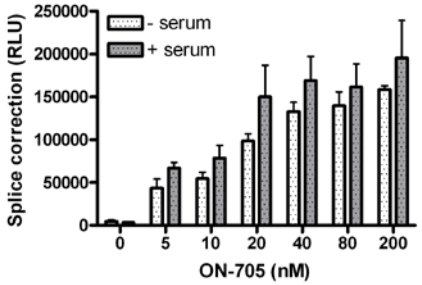


Figure 5 Lipofectamine mediated ON-705 splice correction. HeLa pLuc/705 cells were incubated for 4h with Lipofectamine complexes in the precensce or absence of serum. Twenty-four hours after transfection luciferase activity was measured. Data are presented as mean+SD, N=3.

To validate that HeLa pLuc/705 express EGFR, the expression level of EGFR was determined by Western Blotting (Fig. 6). EGFR levels of HeLa pLuc/705 were comparable with other EGFR over-expressing cells HER14, MDA and 14C (32,33).

After validation of EGFR expression by HeLa pLuc/705, complexation of ON-705 by the fusion constructs was studied using gel electrophoresis. Therefore, 3 fusion constructs +C6, +C8 and -C7 were mixed with ON-705 in PBS at different molar ratio's and the complexes formed were subjected to electrophoresis on a non-denaturing

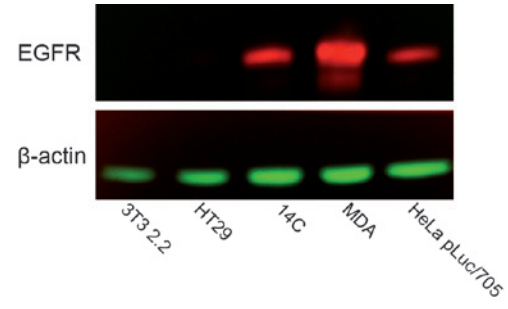


Figure 6 EGFR expression levels in cells. Cell lysates of 3T3 2.2 (low EGFR expressing cells), HT29 (moderate EGFR expressing cells), 14C, MDA (both high EGFR expressing cells) and HeLa pLuc/705 were collected and subjected to SDS-PAGE. Levels of total EGFR and β-actin (loading control) were detect by Western Blotting.

PAGE TBE 4-20% gel (Fig. 7). Percentage of free ON-705 was calculated by dividing the amount of free ON-705 after complexation by the amount of free ON-705 of the control.

Both +C6 and +C8 were able to complex a percentage of ON-705 at all molar ratios, whereas -C7 was unable to complex ON-705 at all.

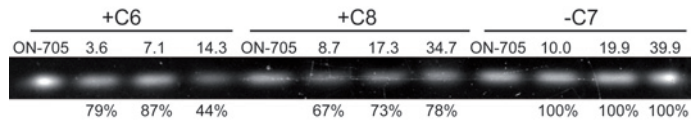


Figure 7 Quantification of free 2'-OMe RNA after complexation of ON-705 by +C6, +C8 and -C7 fusion constructs using gel electrophoresis. Different amounts of targeted multimodular peptide constructs were mixed with 5pmol ON-705, yielding 3.6:1-39.9:1 molar ratios, and subjected to electrophoresis on a non-denaturing PAGE TBE 4-20% gel. Displayed percentages are the percentage of free ON-705. (ON-705) 5pmol ON-705 only.

Next, the 4 fusion constructs +C6, +C8, +C9 and -C7 were mixed with ON-705 in PBS at different molar ratio's and HeLa pLuc/705 cells were transfected in the absence of serum (Fig. 8). The constructs +C8, +C9 and -C7 were able to induce splice correction at their highest molar ratios tested. Only +C6 was unable to induce splice correction.

To compare the splice correction efficiencies of the targeted multimodular peptide constructs with those of other peptide-based transfectants, HeLa pLuc/705 cells were transfected using 3 synthetic peptides with proven transfection capacities

using pDNA, namely ppTG20, LAH4 and LAH5 (26,27,34). The 3 peptides were mixed with of ON-705 in HBS at different molar ratios and HeLa pLuc/705 cells were transfected (Fig. 9). Only LAH4 was able to induce splice correction at molar ratio's of 100:1 and 200:1.

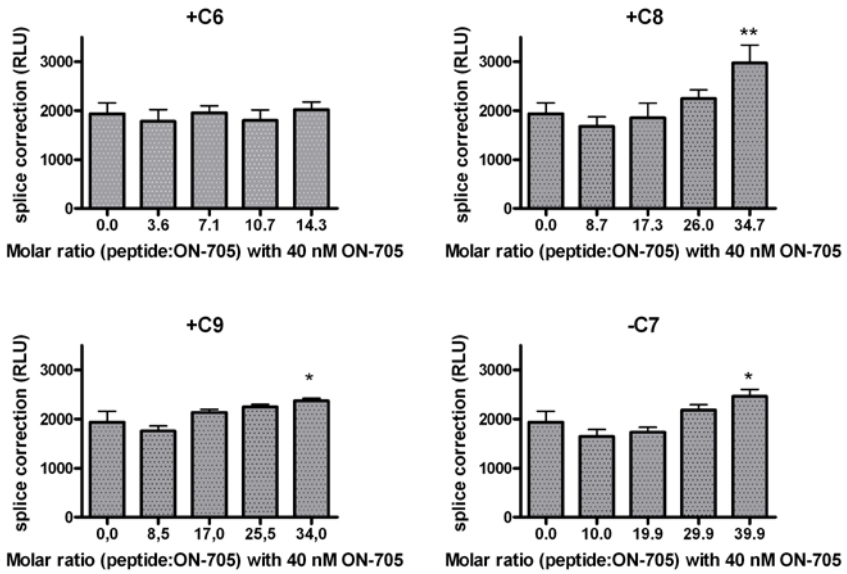


Figure 8 +C6, +C8, +C9 and -C7 fusion construct mediated ON-705 splice correction. HeLa pLuc/705 cells were incubated for 4h with the targeted multimodular peptide construct complexes in the absence of serum. Twenty-four hours after transfection luciferase activity was measured. The differences between +C8, +C9 and -C7 fusion constructs in complex with ON-705 treated and untreated cells were considered significant at their highest molar ratios (2-way ANOVA, *P<0.05; **P<0.01). Data are presented as mean+SD, N=3.

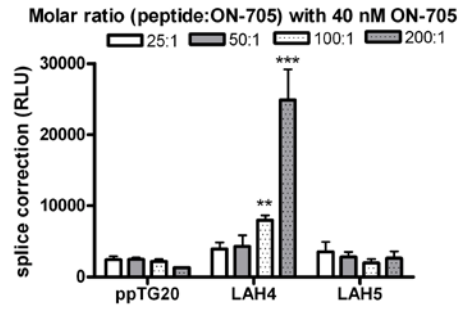


Figure 9 ppTG20, LAH4 and LAH5 mediated ON-705 splice correction. HeLa pLuc/705 cells were incubated for 4h with peptide complexes in the absence of serum. Twenty-four hours after transfection luciferase activity was measured. The differences between LAH4 in complex with ON-705 treated and untreated cells were considered significant at molar ratios 100: and 200:1 (2-way ANOVA, **P<0.01; ***P<0.001). Data are presented as mean+SD, N=3.

Discussion and conclusion

In this study, we showed the generation of a gene library encoding EGFR targeted multimodular peptides and the subsequent recombinant expression and purification of individual peptide constructs from this library. The generated libraries were evaluated for their ability to induce splice-correction by delivering the ON-705 into HeLa pLuc/705 cells as part of an integrative strategy to simultaneously select functional peptides for gene delivery and facilitating endosomal escape of nanobodies. Two different combinatorial gene libraries, encoding 9 different functional peptides ($SPKR_4$, Pr18, Mu, H9.2, H5WYG, TAT(47-57), LAH5, ppTG20 and NLSV402) were generated using the solid-phase platform for combinatorial multipart gene assembly (Chapter 3 and 5)(20). Libraries created with the Head+C contained a cathepsin D substrate sequence between the nanobody and the peptides, whereas in libraries created with Head-C the cathepsin D substrate sequence was omitted. After internalization, the nanobody has no function and dissociation of the peptides/ON-705 complexes from the nanobody could be beneficial (18).

After cloning, the pDNA library obtained was transformed into BL21-Gold(DE3) bacteria and resulted in colonies containing unique inserts (Fig. 1d). Also, about 88% of screened colonies contained correct pDNA, which is almost 2 times higher compared to pDNA obtained from colonies transformed with multimodular peptide libraries without a fusion protein (Chapter 5). This indicates that the use of the nanobody as fusion partner reduces the toxicity of the encoded multimodular peptides. Sequence analysis confirmed the error-free ligation and random insertion of the functional peptides used (Table 3, not all data shown).

In total, 24 colonies (12 colonies from both the +C and -C libraries) were screened for protein expression. After purification, only 15 fusion constructs were detectable using SDS-PAGE (Fig. 2). Furthermore, expression levels varied greatly between the constructs. The use of the nanobody as fusion protein resulted in higher expression levels compared to expression levels without fusion protein in Chapter 5, (9-25nmol/L *E.coli* vs 2-9nmol/L *E.coli*, respectively). Although the use of the nanobody as fusion protein resulted in a 1-12.5 fold increase of the recombinant protein yield, it still required an increase of bacterial culture volume to 0.5L in order to produce sufficient amounts of fusion constructs to perform cellular assays. By using a fusion protein, we tried to reduce the toxicity of the multimodular peptides and tried to reduce their susceptibility for proteolytic degradation (2,35). Although host toxicity was reduced, the increase in recombinant protein yield was modest. Screening of other anti-EGFR nanobodies (e.g. EgA1 or 9G8 alone) as fusion partners could be beneficial for increasing the expression yield (12,15).

To obtain a sufficient amount of EFGR targeted multimodular peptides to perform cellular assays, 4 constructs were selected for large scale expression (0.5L), namely +C6, +C8, +C9 and -C7, and resulted in yields between 150-440µg (4.5-12.6nmol) after dialysis/refolding. However, increasing the culture volume limits high-throughput production and purification.

To determine whether the addition of C-terminal multimodular peptides influences the selectivity and binding affinity of the 7D12-9G8 biparatopic nanobody for EGFR, cell association studies were performed with all 4 constructs. All fusion constructs were able to specifically bind to EGFR over-expressing cells, although all constructs displayed a relatively high non-specific binding compared to the free nanobody (Fig. 4a). This is probably caused by the positive charge of the peptides, which are able to adsorb onto negatively charged membranes. Also, the binding affinities of all constructs were in the same range as the free nanobody, except for -C7 (Fig. 4b). A deviation in the fit caused by an outlier is the cause of the lower binding curve of -C7 (Fig. 4b). These observations demonstrate that nanobodies can be modified by fusing cationic functional peptides to their C-terminus without affecting binding specificity and affinity.

After binding EGFR and receptor mediated internalization, the multimodular peptides in the fusion construct should facilitate endosomal escape in order to avoid lysosomal entrapment and to deliver their nucleic acid cargo. The delivery of a nucleic acid cargo into the cytosol can be determined with a splice correction assay. This assay is based on the HeLa pLuc/705 cell line, which is stably transfected with a mutated luciferase which causes aberrant splicing of luciferase pre-mRNA resulting in the synthesis of non-functional luciferase (30). By cytosolic delivery of ON-705, the splice site can be masked and splicing is redirected, which restores luciferase activity. All 4 fusion constructs +C6, +C8, +C9 and -C7 were mixed with ON-705 at different

molar ratios and screened for splice correction. Only +C8, +C9 and -C7 were able to induce splice correction at their highest molar ratios tested (Fig.6). Moreover, splice correction induced by +C8, +C9 and -C7 was approximately 100-fold lower compared to Lipofectamine (Fig. 5). The ability of the fusion constructs to induce splice correction was also compared to that of 3 synthetic peptides with proven pDNA transfection capacities (34). Only LAH4 was able to induce splice correction at high molar ratios (100: and 200:1)and the induced splice correction was 3-10 fold higher compared to +C8, +C9 and -C7.

The observed low splice correction of the fusion constructs might be caused by low EGFR expression on the HeLa pLuc/705 cells. However, Western Blot analysis confirmed that the HeLa cells express high levels of EGFR, comparable to other EGFR positive cells lines such as HER14, MDA and 14C (Fig. 5)(32).

Another cause of the low splice correction could be the lack of complexation of ON-705 by the fusion constructs. In previous studies, complexation of ssON was a prerequisite for efficient delivery and splice correction (31,36). We tried to assess complexation of ON-705 by 3 out of the 4 fusion constructs by quantifying the amount of free 2'-OMe RNA using gel electrophoresis, but it did not give conclusive results. The strongest interaction with ON-705 was displayed by +C6, which was unable to induce splice correction, whereas -C7 was unable to complex ON-705 but did induce splice correction. Additional analysis techniques (e.g. dye replacement assay, DLS) should be used to determine if the fusion constructs are able to form complexes with ON-705.

Besides the ability to form complexes with ON-705, another obstacle is the amount of fusion constructs needed for complex formation with the nucleic acid cargo. Other studies showed that chemically modified peptides can form complexes with ssONs when at least a 3 fold molar excess of peptide over nucleic acid cargo is used (37-39). However, unmodified peptides display more modest complexation and a 10-20 fold molar excess is required to form peptide/ONs complexes (31,40). In case of the fusion constructs, due to the high amount of cationic peptides required, the amount of nanobody in the formed complexes is also high. Since internalization via EGFR is limited by the number of receptors present on the cell surface, high amounts of nanobody can cause saturation of EGFR and uptake of complexes will be limited. We estimated that at the highest molar ratios used in the splice correction assay, the concentration of nanobody is about 10-30 fold higher than the concentration needed for saturation of EGFR. The optimal molar ratios for condensation of ONs and the optimal molar ratios for internalization are not within the same order of magnitude, which probably contributes to the low splice correction displayed by the fusion constructs.

Another cause of the low splice correction could be the selected constructs. Out of a library containing (theoretically) over 66,000 different fusion constructs, only 4

were screened for their ability to induce splice correction. More candidates will have to be screened before firm conclusions can be drawn about the usefulness of targeted multimodular peptides for gene delivery.

Other research groups created similar fusion constructs with a N-terminal protein for targeting purposes and C-terminal peptide(s) for nucleic acid complexation and/or endosomal escape (Chapter 2)(41-44). Almost all demonstrated complexation/binding of nucleic acids (pDNA or siRNA) by the fusion constructs and were able to transfect cells *in vitro*. However, the cationic peptides used for nucleic acid complexation, the molar ratios and the amounts of nucleic acids differ. Compared to the peptides in our nanobody fusion constructs, the charge density of the peptides used by the other groups is higher. Also, we used higher molar ratios and lower amounts of nucleic acid cargo. Furthermore, transfection assay parameters were also different (e.g. incubation time, serum concentration and cells lines).

In conclusion, we demonstrated the generation of a gene library encoding EGFR targeted multimodular peptides by fusing peptides to the C-terminus of the anti-EGFR biparatopic nanobody 7D12-9G8. The fusion of the nanobody to the peptide library increased the recombinant expression yield of the fusion constructs. After the expression and subsequent purification of 4 individual fusion constructs, we showed that the modified nanobody was still able to specifically bind to EGFR with similar affinity as the unmodified nanobody. The constructs produced were able to mediate splice correction via the delivery of splice-correcting ON-705, although not very efficiently. However, more constructs need to been screened to identify other functional peptide combinations for gene delivery and for facilitating endosomal escape of nanobodies.

Abbreviations

bp:	base pairs
hEGFR:	human epidermal growth factor receptor
HIS-tag:	polyhistidine tag
IMAC:	immobilized metal ion affinity chromatography
IPTG:	isopropyl β-D-1-thiogalactopyranoside
kDa:	kilo dalton
LB:	Lennox broth
NLS:	nuclear localization signal
ODN:	oligodeoxynucleotide
ON-705:	2'-O-methylated splice-correcting oligonucleotide 705
PAGE:	polyacrylamide gel electrophoresis
pDNA:	plasmid DNA
RT:	room temperature
SDS-PAGE:	sodium dodecyl sulfate polyacrylamide gel electrophoresis
ss:	single-stranded
TB:	terrific broth

Literature

(1) Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnology Journal 2012;7(5):620-634.

(2) Li Y. Carrier proteins for fusion expression of antimicrobial peptides in Escherichia coli. Biotechnol Appl Biochem 2009;54(1):1-9.

(3) Francis DM, Page R. Strategies to optimize protein expression in E. coli. Current Protocols in Protein Science 2010(SUPPL. 61):5.24.1-5.24.29.

(4) Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. Mol Cell Biochem 2008;307(1-2):249-264.

(5) Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 2006;48(1):1-13.

(6) Kenig M, Peternel S, Gaberc-Porekar V, Menart V. Influence of the protein oligomericity on final yield after affinity tag removal in purification of recombinant proteins. Journal of Chromatography A 2006;1101(1-2):293-306.

(7) Fortunati E, Ehlert E, van Loo ND, Wyman C, Eble JA, Grosveld F, et al. A multi-domain protein for beta1 integrin-targeted DNA delivery. Gene Ther 2000 Sep;7(17):1505-15.

(8) Gao S, Simon MJ, Morrison III B, Banta S. Bifunctional chimeric fusion proteins engineered for DNA delivery: Optimization of the protein to DNA ratio. Biochimica et Biophysica Acta - General Subjects 2009;1790(3):198-207.

(9) Govindarajan S, Sivakumar J, Garimidi P, Rangaraj N, Kumar JM, Rao NM, et al. Targeting human epidermal growth factor receptor 2 by a cell-penetrating peptide-affibody bioconjugate. Biomaterials 2012;33(8):2570-2582.

(10) Napoli A, Zivanovic Y, Bocs C, Buhler C, Rossi M, Forterre P, et al. DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of Sulfolobus solfataricus. Nucleic Acids Res 2002;30(12):2656-2662.

(11) Muyldermans S. Single domain camel antibodies: current status. J Biotechnol 2001;74(4):277-302.

(12) Roovers RC, Laeremans T, Huang L, De Taeye S, Verkleij AJ, Revets H, et al. Efficient inhibition of EGFR signalling and of tumour growth by antagonistic anti-EGFR Nanobodies. Cancer Immunology, Immunotherapy 2007;56(3):303-317.

(13) Altintas I, Heukers R, Van Der Meel R, Lacombe M, Amidi M, Van Bergen En Henegouwen PMP, et al. Nanobody-albumin nanoparticles (NANAPs) for the delivery of a multikinase inhibitor 17864 to EGFR overexpressing tumor cells. J Controlled Release 2013;165(2):110-118.

(14) Van Der Meel R, Oliveira S, Altintas I, Haselberg R, Van Der Veeken J, Roovers RC, et al. Tumor-targeted Nanobullets: Anti-EGFR nanobody-liposomes loaded with anti-IGF-1R kinase inhibitor for cancer treatment. J Controlled Release 2012;159(2):281-289.

(15) Roovers RC, Vosjan MJWD, Laeremans T, El Khoulati R, De Bruin RCG, Ferguson KM, et al. A biparatopic anti-EGFR nanobody efficiently inhibits solid tumour growth. International Journal of Cancer 2011; 129(8):2013-2024.

(16) Hofman EG, Bader AN, Voortman J, Van Den Heuvel DJ, Sigismund S, Verkleij AJ, et al. Ligand-induced EGF receptor oligomerization is kinase-dependent and enhances internalization. J Biol Chem 2010; 285(50):39481-39489.

(17) Heukers R, Vermeulen JF, Fereidouni F, Bader AN, Voortman J, Roovers RC, et al. EGFR endocytosis requires its kinase activity and N-terminal transmembrane dimerization motif. J Cell Sci 2013 Aug 13.

(18) Haines AM, Irvine AS, Mountain A, Charlesworth J, Farrow NA, Husain RD, et al. CL22 - a novel cationic peptide for efficient transfection of mammalian cells. Gene Ther 2001 Jan;8(2):99-110.

(19) Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 2007 Jul;35 (Web Server issue):W43-6.

(20) de Raad M, Kooijmans SA, Teunissen EA, Mastrobattista E. A Solid-Phase Platform for Combinatorial and Scarless Multipart Gene Assembly. ACS Synth Biol 2013 Mar 15.

(21) Honegger AM, Dull TJ, Felder S, Van Obberghen E, Bellot F, Szapary D, et al. Point mutation at the ATP binding site of EGF receptor abolishes protein-tyrosine kinase activity and alters cellular routing. Cell 1987;51(2):199-209.

(22) Henaut A, Danchin A. Analysis and predictions from Escherichia coli sequences, or E. coli in silico. Escherichia Coli and Salmonella: Cellular and Molecular Biology 1996:2047-2066.

(23) Schwartz B, Ivanov MA, Pitard B, Escriou V, Rangara R, Byk G, et al. Synthetic DNA-compacting peptides derived from human sequence enhance cationic lipid-mediated gene transfer in vitro and in vivo. Gene Ther 1999 Feb;6(2):282-92.

(24) Murray KD, Etheridge CJ, Shah SI, Matthews DA, Russell W, Gurling HM, et al. Enhanced cationic liposome-mediated transfection using the DNA-binding peptide mu (mu) from the adenovirus core. Gene Ther 2001 Mar;8(6):453-60.

(25) Ignatovich IA, Dizhe EB, Pavlotskaya AV, Akifiev BN, Burov SV, Orlov SV, et al. Complexes of plasmid DNA with basic domain 47-57 of the HIV-1 Tat protein are transferred to mammalian cells by endocytosis-mediated pathways. J Biol Chem 2003 Oct 24;278(43):42625-36.

(26) Rittner K, Benavente A, Bompard-Sorlet A, Heitz F, Divita G, Brasseur R, et al. New basic membrane-destabilizing peptides for plasmid-based gene delivery in vitro and in vivo. Mol Ther 2002 Feb;5(2):104-14.

(27) Kichler A, Leborgne C, Marz J, Danos O, Bechinger B. Histidine-rich amphipathic peptide antibiotics promote efficient delivery of DNA into mammalian cells. Proc Natl Acad Sci U S A 2003 Feb 18;100(4):1564-8.

(28) Midoux P, Kichler A, Boutin V, Maurizot JC, Monsigny M. Membrane permeabilization and efficient gene transfer by a peptide containing several histidines. Bioconjug Chem 1998 Mar-Apr;9(2):260-7.

(29) Ritter W, Plank C, Lausier J, Rudolph C, Zink D, Reinhardt D, et al. A novel transfecting peptide comprising a tetrameric nuclear localization sequence. J Mol Med 2003 Nov;81(11):708-17.

(30) Kang S-, Cho M-, Kole R. Up-regulation of luciferase gene expression with antisense oligonucleotides: Implications and applications in functional assay development. Biochemistry (N Y) 1998;37(18):6235-6239.

(31) Mäe M, EL Andaloussi S, Lundin P, Oskolkov N, Johansson HJ, Guterstam P, et al. A stearylated CPP for delivery of splice correcting oligonucleotides using a non-covalent co-incubation strategy. J Controlled Release 2009;134(3):221-227.

(32) Brenner JC, Graham MP, Kumar B, Saunders LM, Kupfer R, Lyons RH, et al. Genotyping of 73 UM-SCC head and neck squamous cell carcinoma cell lines. Head and Neck 2010;32(4):417-426.

(33) van der Meel R, Oliveira S, Altintas I, Heukers R, Pieters EH, van Bergen En Henegouwen PM, et al. Inhibition of Tumor Growth by Targeted Anti-EGFR/IGF-1R Nanobullets Depends on Efficient Blocking of Cell Survival Pathways. Mol Pharm 2013 Oct 7;10(10):3717-3727.

(34) de Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E. High-content screening of peptide-based non-viral gene delivery systems. J Control Release 2012 Mar 28;158(3):433-442.

(35) Vassilevski AA, Kozlov SA, Grishin EV. Antimicrobial peptide precursor structures suggest effective production strategies. Recent Patents on Inflammation and Allergy Drug Discovery 2008;2(1):58-63.

(36) Morris MC, Gros E, Aldrian-Herrada G, Choob M, Archdeacon J, Heitz F, et al. A non-covalent peptide-based carrier for in vivo delivery of DNA mimics. Nucleic Acids Res 2007;35(7).

(37) Ezzat K, El Andaloussi S, Zaghloul EM, Lehto T, Lindberg S, Moreno PMD, et al. PepFect 14, a novel cell-penetrating peptide for oligonucleotide delivery in solution and as solid formulation. Nucleic Acids Res 2011;39(12):5284-5298.

(38) Lehto T, Abes R, Oskolkov N, Suhorutšenko J, Copolovici D-, Mäger I, et al. Delivery of nucleic acids with a stearylated (RxR)4 peptide using a non-covalent co-incubation strategy. J Controlled Release 2010;141(1):42-51.

(39) Oskolkov N, Arukuusk P, Copolovici D-, Lindberg S, Margus H, Padari K, et al. NickFects, phosphorylated derivatives of transportan 10 for cellular delivery of oligonucleotides. International Journal of Peptide Research and Therapeutics 2011;17(2):147-157.

(40) Said Hassane F, Abes R, El Andaloussi S, Lehto T, Sillard R, Langel U, et al. Insights into the cellular trafficking of splice redirecting oligonucleotides complexed with chemically modified cell-penetrating peptides. J Controlled Release 2011;153(2):163-172.
(41) Winkler J, Martin-Killias P, Plückthun A, Zangemeister-Wittke U. EpCAM-targeted delivery of nanocomplexed siRNA to tumor cells with designed ankyrin repeat proteins. Molecular Cancer Therapeutics 2009;8(9):2674-2683.
(42) Canine BF, Wang Y, Hatefi A. Evaluation of the effect of vector architecture on DNA condensation and gene transfer efficiency. J Controlled Release 2008;129(2):117-123.
(43) Zhang T, Wang C-, Zhang W, Gao Y-, Yang S-, Wang T-, et al. Generation and characterization of a fusion protein of single-chain fragment variable antibody against hemagglutinin antigen of avian influenza virus and truncated protamine. Vaccine 2010;28(23):3949-3955.
(44) Song E, Zhu P, Lee S-, Chowdhury D, Kussman S, Dykxhoorn DM, et al. Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. Nat Biotechnol 2005;23(6):709-717.

7

Chapter

EGFP fused multimodular peptide libraries for gene delivery

Markus de Raad[a], Erik A. Teunissen[a] and Enrico Mastrobattista[a]

[a] Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

Abstract

Currently used non-viral gene delivery systems are inefficient at delivering therapeutic nucleic acids intact inside the target cell and in facilitating transgene expression. To improve non-viral vectors, we propose to randomly combine functional domains needed for gene delivery from existing or *de novo* designed peptides into single, multimodular constructs. In order to improve recombinant expression yields of the multimodular peptides, enhanced green fluorescent protein (EGFP) was fused to the multimodular peptides. In addition, the presence of EGFP would allow us to trace these peptide-based gene delivery systems inside cells by fluorescence microscopy. Here, we demonstrated the creation of a gene library encoding EGFP fused multimodular peptides by fusing peptides to the C-terminus of EGFP. Fusion of EGFP to the multimodular peptide libraries allowed direct monitoring of recombinant protein expression. After the expression and subsequent purification of 13 individual fusion constructs, we showed that the EGFP fused peptide constructs were able to complex pDNA. Unfortunately, the produced EGFP fused peptide constructs were unable to mediate transfection of COS-7 cells with a reporter gene.

Introduction

Recombinant production of *de novo* peptides is often difficult due to often unpredictable toxicity issues to the host. This can be related to the particular structure of a peptide as well as to its biological properties. Fusion of a *de novo* peptide sequence to an existing protein may alleviate toxicity and is therefore a strategy often applied for the recombinant production of *de novo* peptides that are too large to be produced synthetically.

In Chapter 6, we demonstrated that by fusing multimodular peptides to the C-terminus of an anti-EGFR biparatopic nanobody functional fusion constructs could be generated. Also, the recombinant yield of the targeted multimodular peptide libraries was increased 2-12.5 fold compared to multimodular peptides constructs without a fusion partner. Despite the increased yield, still relatively large culture volumes (0.5L) are required to obtain sufficient amounts of the multimodular peptides for further analysis and for functional testing, e.g. for *in vitro* transfections. Large culture volumes hinder high-throughput production and thus high-throughput screening of the multimodular peptide libraries.

A change to a more robust fusion protein might result in higher expression yields of the multimodular peptides. Ideally, the fusion protein of choice possesses functionalities suitable for gene delivery, thereby forming an integral part of the multimodular peptide construct. Otherwise, removal of the fusion protein after recombinant expression and purification is needed, which will result in lower yields of the final multimodular peptides due to inefficient cleavage, losses during processing or by degradation during cleavage (2,3).

Enhanced green fluorescent protein (EGFP) might be a suitable candidate for this purpose. EGFP and other fluorescent proteins have been widely used to study e.g. protein localization, translocation and interactions in living systems (4). For example, fluorescent proteins have been used in microscopy to visualize endocytic pathways utilized by various viruses (5). Fusion of EGFP to the multimodular peptide libraries would allow monitoring of localization and trafficking of the formed nucleic acid/peptide complexes *in vitro*. Besides the functionality of EGFP in visualizing the gene delivery process at the (sub)cellular level, the use of EGFP may also be advantageous for the recombinant production process and subsequent purification steps. EGFP allows direct monitoring of recombinant protein expression and purification (6). Also, the high expression levels in *E. coli* and resistance to intracellular proteases makes EGFP a suitable fusion partner (7). Furthermore, EGFP has previously been used as a fusion partner for potentially toxic peptides, such as the cell penetrating peptide TAT, the antimicrobial peptide sarcotoxin IA and other small peptides with yields of 2.5-10mg/L bacterial culture (7-9).

The multimodular peptide libraries will be fused to the C-terminus of the EGFP, generating EGFP fused multimodular peptide libraries (Fig. 1c). Two different libraries

will be generated, one with and one without a cathepsin D cleavage site between the nanobody and the library. The cathepsin D cleavage site is recognized and cleaved by the endoprotease cathepsin D, which is abundant in late endosomes, and enables dissociation of the nanobody from the peptides after internalization (10).

In this study, the aim is to further increase the recombinant yield of the multimodular peptide libraries by C-terminal fusion of the fluorescent protein EGFP to the peptide library. After recombinant expression and purification, the created EGFP libraries were evaluated for cell toxicity, cell association and transfection efficiency using the HCS assay developed as described in Chapter 4.

Materials and Methods

Materials

All chemicals and reagents were purchased from Sigma-Aldrich Chemie B.V. (Zwijndrecht, The Netherlands), unless stated otherwise. Dynabeads M-280 Streptavidin were purchased from Dynal (Oslo, Norway). Restriction enzymes, FastAP™ (Thermosensitive Alkaline Phosphatase), Generuler™ DNA and Riboruler™ RNA ladders, Phusion polymerase, Rapid DNA Ligation Kit, GeneJET™ Plasmid Miniprep Kit, isopropyl β-D-1-thiogalactopyranoside (IPTG), RNase A, DNase I, PageRuler™ Prestained Protein Ladder and HisPur™ NI^{2+}-NTA Resin were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Precast denaturing polyacrylamide TBE-urea gels (6%), Lipofectamine™ 2000 and DAPI were purchased from Invitrogen (Breda, The Netherlands). Precast Criterion™ XT Bis-Tris gels (12%) and Criterion™ TBE gels (4-20%) were purchased from Bio-Rad (Hercules, CA, USA). DMEM (Dulbecco's modified Eagle's Medium, with 1g/l glucose, 584mg/ml L-Glutamine), Fetal Bovine Serum (FBS), and Phosphate Buffered Saline (PBS) were purchased from PAA Laboratories GmbH (Pasching, Austria). The used bacterial culture media: Lennox Broth (LB) (tryptone (10g/L), yeast extract (5g/L) and NaCl (5g/L)), LB-agar (tryptone (10g/L), yeast extract (5g/L), NaCl (5g/L) and agar (15g/L)) and Terrific Broth (TB) (tryptone (12g/L), yeast extract (24g/L), K_2HPO_4 (9.4g/L), KH_2PO_4 (2.2g/L) and glycerol (8mL/L)) were purchased from Sigma-Aldrich. Al used solid and liquid media contained glucose (1% w/v) and ampicillin (100µg/mL).

pCMV-mRFP-actin plasmid DNA was constructed as previously described and was kindly provided by Dr. Schiffelers (Department of Clinical Chemistry and Hematology, University Medical Center Utrecht) (11). pCMV-EGFP plasmid DNA was constructed in house as previously described (12). pCMV-mRFP-actin and pCMV-EGFP are eukaryotic expression plasmids encoding for the fusion of monomeric red fluorescent protein (mRFP) to actin and enhanced green fluorescent protein (EGFP), respectively, under the control of the human cytomegalovirus promoter (CMV).

Synthetic oligodeoxynucleotides

All oligodeoxynucleotides were synthesized by Eurogentec S.A. (Seraing, Belgium). Single-stranded oligodeoxynucleotides (ssODN) modules were synthesized with 5'- and 3'- phosphates and PAGE purified at a 40nmol scale or with 5'biotin-TEG and PAGE purified at a 200nmol scale. Linker ODN and primers were synthesized unmodified and SePOP desalted at a 40nmol scale. ssODNwere dissolved in nuclease-free water and stored at -30°C as 100µM stocks.

ssODN encoding peptides were codon-optimized for bacterial expression and flanked with linker sequences GGTTCT (GS) at the 5' (all except Head modules) and GGTGGC (GG) at the 3' (all except Tail modules). The possible formation of secondary structures (checked with Oligocalc software) was reduced by adjustment of any self-complementary oligonucleotide sequence (13). All sequences of the used ssODNs and primers are listed in Table 1.

Methods

Construction of pCmE destination vector for multimodular peptide libraries

As basis for the destination vector pCmE, the plasmid pSCherry2 (Eurogentec) was used. A DNA fragment containing the EFGP coding sequence was obtained by digestion of the pIVEXem-EGFP with XbaI/BamHI. The fragment was ligated into the XbaI/BamHII digested pSCherry2 vector. The resulting plasmid was the library destination vector pCmE and was used as control for the recombinant expression of EGFP without fused peptides.

Construction of multimodular peptide libraries by solid phase combinatorial multipart gene assembly

Using the method for the assembly of ssODNs in the presence of linker ODNs, a random polypeptide library was assembled (Chapter 3) (14). Ligation reactions (20µL) contained 12.5pmol ssHead (Head_+C_GG or Head_-C_GG) coated on M-280 magnetic beads, 50pmol GGGS linker, 1x Rapid Ligation Buffer, 5U T4 DNA ligase and 100 pmol of a mix of 9 different ssODNs (11,1 pmol of GS_Pr18_GG; GS_SPKR4_GG; GS_H9.2_GG; GS_H5WYG_GG; GS_TAT_GG; GS_ppTG20_GG; GS_Mu_GG; GS_NLSV402_GG and GS_LAH5_GG). Five ligation steps with this composition were performed. One additional ligation step was performed for the ligation of the reaction products to a terminating ssTail (GS_Tail). Ligation products were loaded on a 6% denaturing polyacrylamide TBE-urea gel and the area between the smallest theoretical product (±150 nucleotides) and the largest theoretical product (±600 nucleotides) was excised, purified and PCR amplified (as described in Chapter 3) (14). PCR products were analyzed on a 2% w/v agarose gel and the area between ±200-600bp was excised and purified.

The library obtained was BamHI/ HindIII digested, purified and 42-180ng of the digested products were ligated in 100ng of dephosphorylated pCmE (BamHI/ HindIII digested and purified from an agarose gel), to create a C-terminal fusion between EGFP and the library inserts. Half of the reaction mixture was used directly to transform *E.coli* strain DH5α (Invitrogen). Bacteria were plated out on LB-agar plates and grown overnight at 37°C. From the obtained LB plates, all colonies were harvested using 2ml LB. Harvested colonies were grown in 10ml LB for 4hr at 37°C and plasmid DNA (pDNA) was harvested.

Table 1 Used oligodeoxynucletides

Name	Sequences
Head_+C_GG	5'-Biotin-TEG-GTTATTGGATCCGGGCGGTGGCGGTAGCGGTGGCTTCT-TAGGGTTTGGTGGC
Head_-C_GG	5'-Biotin-TEG-GTTATTGGATCCGGGCGGTGGCGGTTCTGGTGGC
Mu	5' P-GGTTCTCGTCGTGCTCATCATCGGCGTAGACGTGCATCTCATCGCCGTAT-GCGTCGCGGAGGTGGTGGC-P 3'
H5WYG	5' P-GGTTCTGGTCTGTTTCACGCGATTGCCCATTTCATCCACGGTGGT-TGGCATGGTTTAATTCACGGTTGGTATGGCGGTGGC-P 3'
H9.2	5' P-GGTTCTAAGACACCGAAAAAGGCCAAAAAGCCAAAAACCCC-GAAAAAGGCGAAAAAACCAGGTGGC-P 3'
LAH5	5' P-GGTTCTAAAAAGGCACTGCTGGCACTGGCACTCCATCACTTAGCA-CACCTTGCTCATCATCTTGCGTTAGCGCTGAAAAAGGCTGGTGGC-P 3'
NLSV402	5' P-GGTTCTCCAAAAAAGAAACGTAAAGTTCCAAAAAAAAAGCGCAAAGT-CGGTGGC-P 3'
ppTG20	5' P-GGTTCTGGCTTATTTCGTGCGCTGTTGCGTCTGTTACTGTCTCTGTGGA-GATTACTTTTACGTGCGGGTGGC-P 3'
Pr18	5' P-GGTTCTTCTCGTAGTCGGTATTACCGTCAGCGCCAACGTTCTCGCCGTCG-CCGGCGTAGAGGTGGC-P 3'
SPKR4	5' P-GGTTCTAGCCCGAAACGTAGCCCTAAGCGCAGCCCAAAAAGATCTCC-TAAACGTGGTGGC-P 3'
TAT (47-57)	5' P-GGTTCTTATGGCCGCAAGAAGCGTCGTCAAAGACGTCGTGGTGGC-P 3'
GGGS linker	GGTGGCGGTTCT
GS_Tail	GGTTCTCATCACCATCACCATCACCATCACTAATAAAAGCTTAAAGAG
Fw_Head_+C	GTTATTGGATCCGGGCGG
Fw_Head_-C	GTTATTGGATCCGGGCGG
Rev_Tail	GTCATACTTATGGGATCCTTATTAA

Flow cytometric analysis of EGFP-peptide libraries

The EGFP fluorescences of library transformed bacteria was determined using flow cytometry. Fifty nanogram library pDNA was used to transform *E. coli* strain BL21-Gold(DE_3) (Stratagene, Santa Clara, CA, USA). Bacteria were plated out on LB-agar and grown overnight at 37°C. From the obtained plates, all colonies were harvested using 1ml LB. Harvested colonies were grown in 5ml LB for 1 hr at 37°C. Then, protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.1mM final concentration. After 2.5h, bacteria were harvested by centrifugation (5,000xg, 10 min at 4°C) and resuspended in PBS at a final concentration of $2x10^7$ bacteria/ml. Bacteria were analyzed using a BD Influx flow cytometer (Becton Dickinson, Brussels, Belgium) and non-induced bacteria were used to gate the fluorescence threshold.

Recombinant protein production in *E. coli*

Fifty nanogram library pDNA was used to transform *E. coli* strain BL21-Gold(DE_3). Bacteria were plated out on LB-agar and grown overnight at 37°C. From the plates obtained, colonies were inoculated overnight in 5mL LB. Glycerol stocks were prepared and pDNA was harvested from all overnight cultures. pDNA of the library constructs obtained was XbaI/HindIII digested and analyzed on a 1% w/v agarose gel. In addition, pDNA of the library constructs obtained was sequenced for insert identification with full coverage of the insert sequences (inserts were fully sequenced; BaseClear, Leiden, The Netherlands).

For protein expression, starter cultures were grown overnight in 5ml LB from the obtained glycerol stocks. Next, cultures were diluted to an optical density at 600nm (OD_{600})of 0.1 in 50mL Terrific Broth. The diluted cultures were grown at 37°C to OD_{600} of 0.4-0.8, when protein expression was induced by addition of IPTG at 0.1mM final concentration. After 5h, bacterial cultures were screened on fluorescence and harvested by centrifugation (5,000xg, 15min) and stored at -20°C.

For fluorescence screening, $19x10^7$ bacteria form each culture were transferred to a black 96-well plate and bacteria were lysed using 100µL Cellytic B. Fluorescence was determined using a fluorescence well-plate reader (FLUOstar OPTIMA; BMG-Labtech, Offenburg, Germany) with excitation 488nm, emission 520nm filters.

Bacterial lysis

The bacterial pellets obtained were thawed at 4°C, suspended in 20ml/g wet pellet CelLytic B cell lysis reagent containing 0.2mg/ml lysozyme, 0.05% (w/v) sodium deoxycholate, 300mM NaCl, 10mM $MgCl_2$ and one tablet Complete™ Protease Inhibitor Cocktail Tablets (Roche, Almere, The Netherlands) per 50ml CelLytic B, and incubated for 30min on an orbital shaker (200rpm) at 4°C. Next, bacteria were snap-frozen using liquid nitrogen. After defrosting, DNase I (5ug/ml) and RNase A

(10ug/ml) were added and incubated at 4°C for 60min on an orbital shaker (200rpm). Bacteria were centrifuged at 10,000xg for 30min at 4°C to pellet the cellular debris and supernatant was used for further processing.

Protein purification

The EGFP fused multimodular peptides produced were purified from the supernatants obtained after lysis using Ni^{2+}-NTA affinity chromatography. To an equal volume of wash buffer (50mM NaH_2PO_4, 300mM NaCl and 40mM imidazole at pH 7.4) HisPur™ Ni^{2+}-NTA slurry was added (125µl slurry for 50ml cultures) and the suspension was mixed on an end-over-end rotator for 1hr at 4°C. Then, the lysate/resin mixture was loaded on polypropylene columns and washed with 8 resin bed volumes wash buffer. Subsequently, HIS-tagged modular peptides were eluted with 3 x 1 resin bed volumes of elution buffer (50mM NaH_2PO_4, 300mM NaCl, 300mM imidazole and 20% (v/v) glycerol at pH 7.4).

Quantity and purity of the EGFP fused modular peptides was analyzed by SDS-PAGE. For SDS-PAGE, samples were run on Criterion™ XT Bis-Tris 12% gels and visualized by PageBlue™ Protein Staining Solution (based on Coomassie G-250, Thermo Fisher Scientific).

For buffer exchange and concentration, modular peptide solutions were washed with HEPES-buffered saline (HBS, 20mM HEPES, 130mM NaCl, pH 7.4) supplemented with 20% (v/v) glycerol, using Vivaspin centrifugal concentrators (Sartorius Stedim Biotech S.A., Aubagne, France) with a molecular weight membrane cut-off of 10kDa. Protein concentration was determined with the MicroBCA™ protein assay (Pierce Biotechnology, Rockford, IL, USA) against DNase I as a standard protein.

Cell culture

COS-7 African Green Monkey kidney cells were grown in DMEM supplement with 5% heat inactivated FBS and antibiotics/antimycotics. Cells were maintained at 37°C in a 5% CO_2 humidified air atmosphere and split twice weekly.

Preparation of peptiplexes and lipoplexes

Complexes were prepared by adding 4 volumes of fusion-peptide or polymer (PEI) solution to 1 volume of pDNA solution (50µg/ml) and mixing immediately by pipetting up and down 10 times. The complexes were incubated for 30min at room temperature. Peptiplexes were prepared with increasing amounts (0-2nmol) of EGFP fused multimodular peptide constructs. The ratio of PEI to plasmid DNA is expressed as the molar ratio of nitrogens within PEI to phosphates in plasmid DNA (N/P). Peptiplexes and polyplexes were prepared in HBS. For each experiment, complexes were freshly prepared.

DNA gel retardation assay

Complexes were prepared as described above. Twenty five microlitres of the peptiplex samples, containing 0.25µg pDNA and different amounts (0-2nmol) of EGFP fused multimodular peptide constructs, were subjected to electrophoresis on a 0.8% agarose gel (containing 1% ethidium bromide for visualization).

Transfections

Ten thousand COS-7 cells were seeded into black, clear bottom, 96-well tissue culture plates (Greiner Bio-One BV, Alphen a/d Rijn, The Netherlands) 24h prior to transfection. Immediately prior to transfection, the culture medium was refreshed with 100µl complete medium. Twenty five microlitres of the peptiplex and polyplex samples (corresponding to 0.25µg pDNA/well) was added per well and after 24h incubation, medium was replaced with 100µl complete medium. Cells were incubated for indicated times at 37°C in a 5% CO_2 humidified air atmosphere.

Fixation and nuclear staining

Forty eight hours post transfection cells were washed once with 100µl PBS and fixed for 30min with 100µl 4% paraformaldehyde in PBS. After fixation, cells were washed once with 100µl PBS and nuclei were stained with 21.8µM DAPI in PBS. After DAPI staining, cells were washed once with PBS. To each well 100µl PBS was added and cells were analyzed using a Cellomics Arrayscan V HCS Reader.

High capacity automated fluorescence imaging and image analysis

A Cellomics Arrayscan V HCS Reader (Thermo Fisher Scientific, Waltham, MA, USA) was used to analyze the cells. For the analysis of the HCS data, the Cellomics algorithm Target Activation BioApplication (Thermo Fisher Scientific, Waltham, MA, USA) was used. The Target Activation BioApplication allows for measurements of up to 4 different fluorescent channels. One fluorescent channel is reserved for the identification of individual cells using major cellular organelles/compartments (e.g. nuclei or cytoplasm) and provides the basis for the intracellular region where intensity measurements are made.

In this study, single cells were identified according to the morphology of their nuclei using DAPI staining. After identification of single cells, the fluorescence intensity of the used fluorescent probes was measured in a predefined region. If the fluorescence intensity is higher than the defined threshold, the single cell is 'positive' or a 'responder' for this fluorescent signal. From these values, the percentage of responders can be calculated.

Results

Construction of a gene library encoding EGFP fused multimodular peptides by solid phase combinatorial multipart gene assembly

A gene library encoding multimodular peptides was constructed in a combinatorial manner by ligation of single-stranded oligodeoxynucleotides (ssODN) using DNA ligase in the presence of linker ssODN (Chapter 3) (14). Briefly, a ssODN (Head) modified with 5'-biotin was immobilized on streptavidin-coated magnetic beads. Using linker ODN, which were complementary to the last 6 nucleotides of the immobilized ssODN and the first 6 nucleotides of ssODN to be ligated, DNA ligase was used to ligate ssODN in a stepwise manner. In the final ligation step, a single-stranded oligodeoxynucleotide (Tail) was added.

For the generation of multimodular peptide libraries, 9 different functional peptides were selected: 4 DNA condensing peptides ($SPKR_4$, Pr18, Mu and H9.2), 4 membrane disrupting and/or endosomolytic peptides (LAH5, ppTG20, H5WYG and TAT(47-57)) and 1 NLS peptide (NLSV402) (Table 2).

By using 2 different Heads (Head+C and Head-C), 2 different libraries were created. Libraries created with the Head+C (referred to as +C libraries) contain a linker sequence DPGGGGS and the sequence GGFLGF designed for cleavage by the abundant endosomal endoprotease cathepsin D (Table 2) (10). Libraries created with the Head-C (referred to as -C library) contain only a linker sequence DPGGGGS (Table 2). All library constructs were terminated with a universal Tail, which encodes a 8xHIS-tag followed by a double stop codon TAATAA. Both Heads and Tail contained restriction sites for cloning purposes. The Heads, Tail and all peptides were reverse translated into ssODN sequences (using the *E.coli* codon usage table, (15)) and flanked with linker sequences GGTTCT (encoding Gly-Ser) at the 5' (except Head) and GGTGGC (encoding Gly-Gly) at the 3' (except Tail) (Table 2).

Five ligation cycles were performed with a mix containing equal amounts of ssODN encoding all 9 functional peptides, generating a library of Head-MIX_5-Tail. This would theoretically yield a library with $(9^1)+(9^2)+(9^3)+(9^4)+(9^5)=66,429$ different constructs with $9^5=59,049$ full length constructs. The generated constructs were analyzed on a 6% TBE urea gel, after loading equal amounts of magnetic beads. Both libraries obtained were clearly visible and appeared as smears, with highest intensity between 400 and 600 nucleotides (Fig. 1a). Individual non-ligated Heads and Head-MIX, constructs were still visible in both libraries.

After extraction from gel and purification, both libraries were amplified by PCR using primers Fw_Head_+C or Fw_Head_-C and Rev_Tail to generate dsDNA products and enough material for further steps (Fig. 1b). After PCR amplification, smears were detected indicating that all formed constructs were amplifiable. PCR showed an amplification bias towards the smaller templates, with clear bands corresponding to constructs encoding only 1 or 2 peptides (Fig. 1b).

Table 2 Function, name, and amino acid sequence of the functional peptides used. Nucleotide sequence of the corresponding ssODN constructs to produce modular gene constructs are also indicated

Function	Peptide	AA sequence	ssDNA sequence	Ref.
DNA condensation	Pr18	GSSRSRYYRQRQRSRRRRRRGG	ggttcttctcgtagtcggtattaccgtcagcgccaacgttctcgccgtcgc-cggcgtagaggtggc	(16)
	$SPKR_4$	GSSPKRSPKRSPKRSPKRGG	ggttctagcccgaaacgtagccctaagcgcagcccaaaaagatctcc-taaacgtggtggc	(17)
	H9.2	GSKTPKKAKKPKTPKKAKKPGG	ggttctaagacaccgaaaaaggccaaaaagccaaaaaccccgaaaaaggc-gaaaaaaccaggtggc	(16)
	Mu	GSRRAHHRRRRASHRRMRRGG	ggttctcgtcgtgctcatcatcggcgtagacgtgcatctcatcgccgtatgcgt-cgcggaggtggtggc	(18)
Membrane disruption/ translocation	TAT (47-57)	GSYGRKKRRQRRRGG	ggttcttatggccgcaagaagcgtcgtcaaagacgtcgtggtggc	(19)
	ppTG20	GSGLFRALLRLLRSLWRLLLRAGG	ggttctggcttatttcgtgcgctgttgcgtctgttactgtctctgtggagat-tacttttacgtgcgggtggc	(20)
	LAH5	GSKKALLALALHHLAHLAHHLALALKKAGG	ggttctaaaaaggcactgctggcactggcactccatcacttagcacacct-tgctcatcatcttgcgttagcgctgaaaaaggctggtggc	(21)
Endosomolytic	H5WYG	GSGLFHAIAHFIHGGWHGLIHGWYGGG	ggttctggtctgtttcatgcgattgcccacttcatccatggtggttggcatggtt-taattcacggttggtatggcggtggc	(22)
NLS	NLSV402	GSPKKKRKVPKKKRKVGG	ggttctccaaaaaagaaacgtaaagttccaaaaaaaaaagcgcaaagtcgg-tggc	(23)
Head+C	Head_+C_GG	DPGGGGS*GGFLGF*GG	gttatt**ggatcc**gggcggtggcggtagcggtggcttcttagggtttggtggc	-
Head-C	Head_-C_GG	DPGGGGSGG	gttatt**ggatcc**gggcggtggcggttcggtggc	-
Tail	GS_Tail	GSHHHHHHHH	ggttctcatcaccatcaccatcaccatcacTAATAA**aagctt**aaagag	-

Underlined sequences correspond to GGGS linker sites. The amino acids sequences in italics correspond to a cathepsin D cleavage site. ssODN sequences in capitals correspond to the double stop codons in the Tail. ODN sequences in bold correspond to BamHI site in the Head and HindIII site in the Tail.

The PCR amplified Head-MIX$_5$-Tail libraries were excised from ±200bp to ±800bp, purified and cloned into the BamHI/HindIII-sites of vector pCmE, thereby fusing the gene library in frame to the C terminus of EGFP. Two different targeted libraries were created, one library with a cathepsin D site between EGFP and peptide(s) and one without (Fig. 1c). *E.coli* strain DH5α was transformed with both cloned libraries. To obtain the pDNA libraries, all colonies of a single plate were pooled and pDNA was harvested.

Library screening

To determine EGFP fluorescence of the generated libraries, *E.coli* strain BL21-Gold(DE$_3$) was transformed with the obtained +C library or pCmE pDNA (encoding for EGFP). All colonies of a single plate were pooled and induced with 0.1 mM IPTG (except for non-induced bacteria). After 2.5 h, EGFP fluorescence was detected using flow cytometry (Fig. 2).

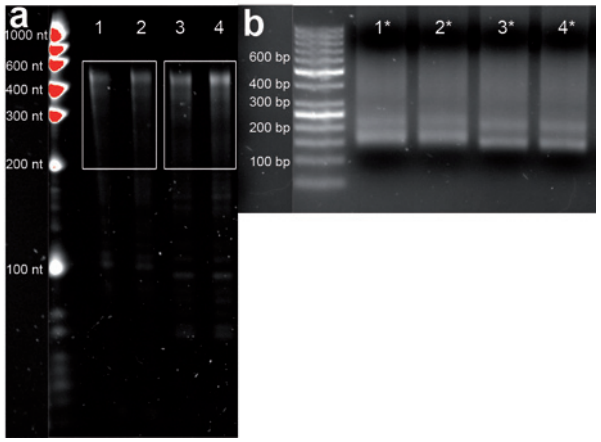


Figure 1 Gene library construction. (a) Initial ligation products were loaded on a 6% TBE-urea gel and purified at final product height as indicated. Ladder: low range RNA ladder. Lanes 1 and 2: Head+C-MIX5-Tail in duplo. Lanes 3 and 4: Head-C-MIX5-Tail in duplo (b) Purified ligation products were amplified using PCR and loaded on a 2% agarose gel. Ladder: 50bp DNA ladder. Lanes 1*-4*: PCR-amplified ligation products corresponding to lanes 1-4. (c) Schematic representation of the generated EGFP fused multimodular peptide constructs. The constructs consist of an N-terminal EGFP, a cathepsin D cleavage site in case of +C libraries and 1-5 functional peptides. (d) Plasmid DNA isolated from 24 colonies from the -C library were digested with XbaI/HindIII and loaded on 1% agarose gel. Ladder: 1kb and 50bp DNA ladder. Vector backbone appears as a single band of ±5700bp whereas library inserts appear between 850-1400bp.

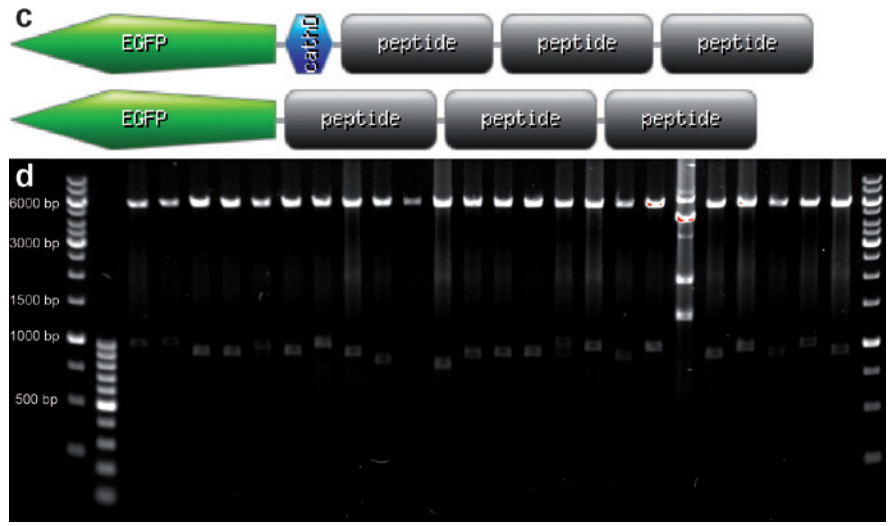


Figure 1 Continued.

Both bacteria expressing EGFP (Fig. 2b) and the +C library (Fig. 2c) contained a sub population of bacteria expressing EGFP, whereas non-induced bacteria (transformed with pCmE) did not contain an EGFP expressing population (Fig. 2a). The observed fluorescence intensity displayed by the EGFP expressing bacteria was higher compared to +C library expressing bacteria.

To continue with the recombinant expression of the EGFP fused multimodular peptide libraries, the obtained +C and -C library pDNA was again used for the transformation of *E.coli* strain BL21-Gold(DE$_3$). To confirm library presence and determine construct size, restriction analysis was performed on pDNA obtained from 24 colonies from the +C library and 24 colonies from the -C library, after overnight culturing (Fig. 1d, not all data shown).

Library inserts appear between 850-1400bp (EGFP+library insert). After restriction enzyme analysis, 68 out of 72 (94%) of screened colonies contained DNA fragments corresponding to the expected sizes (data not shown). All 68 positive colonies contained inserts corresponding to constructs with at least one functional peptide. Sequence analysis confirmed the error-free ligation and random insertion of the used gene modules encoding functional peptides.

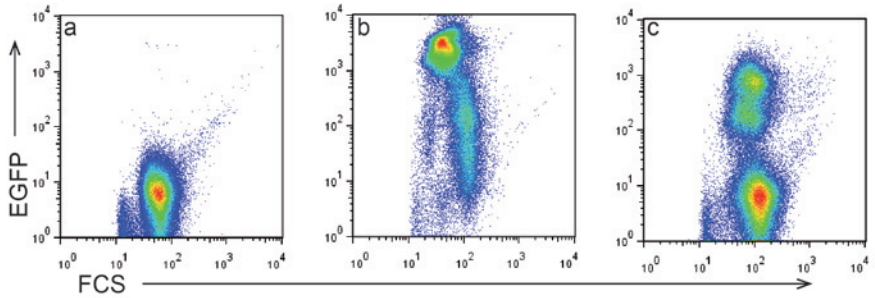


Figure 2 Flow cytometric analysis of EGFP-peptide library. Dot plots of FSC versus EGFP fluorescence representing (a) non-induced bacteria transformed with pCmE, (b) EGFP expressing bacteria and (c) +C library expressing bacteria, 2.5h after induction with 0.1mM IPTG.

After pDNA analysis, 48 colonies with correct pDNA from both +C and -C libraries were screened for EGFP fluorescence to detect protein expression. After expression and bacterial lysis, EGFP fluorescence was measured (Fig. 3).

Ten colonies displayed no/low levels of EGFP fluorescences, whereas twenty-one colonies displayed equal or higher EGFP fluorescence levels compared to EGFP expressing bacteria.

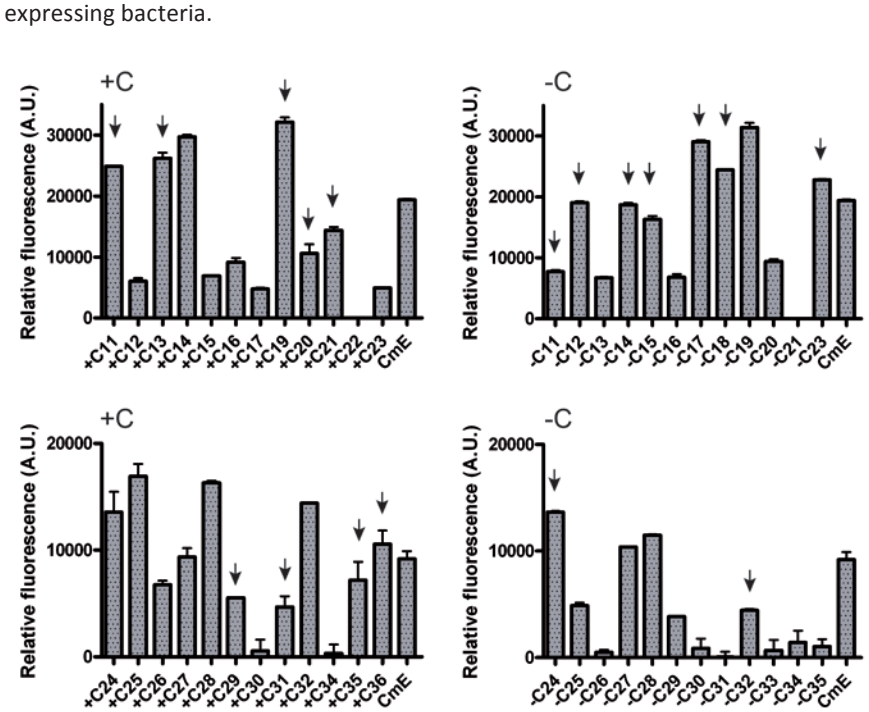


Figure 3 EGFP fluorescence of crude bacterial lysates from both +C and -C libraries to confirm production of EGFP fused multimodular peptides (with background subtraction). The number of each bar corresponds to individual colonies. Similar amounts of bacteria were analyzed. (CmE) EGFP without fusion peptides. Arrows indicate selected colonies for protein expression. Data are presented as mean + SD, N=2.

Large scale expression of the EGFP fused multimodular peptides

After pDNA analysis and EGFP fluorescence screening, 9 colonies from +C library and 9 colonies from -C library were selected for protein expression at 0.05L scale. Colonies were selected based on displayed fluorescence and library insert size. All selected colonies displayed EGFP fluorescence and contained the largest inserts within their libraries. After expression, bacterial lysis, Ni^{2+}-NTA purification, dialysis and concentration,

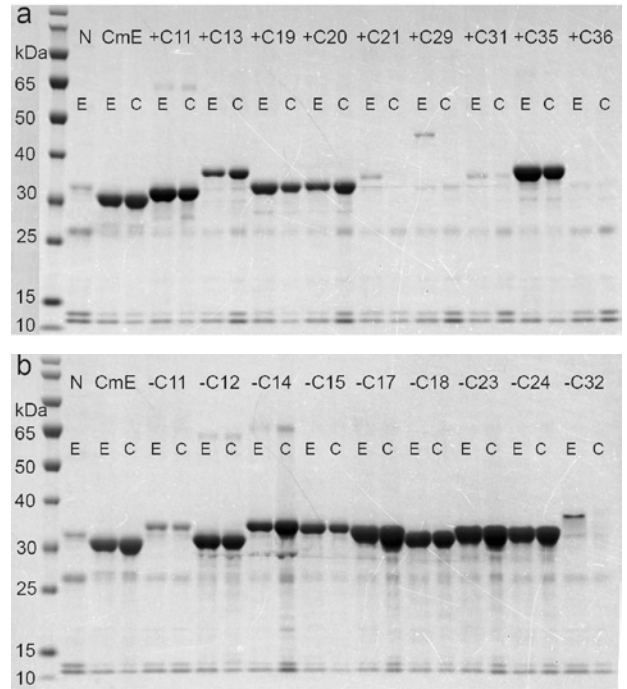


Figure 4 SDS-PAGE analysis of purified, dialyzed and concentrated EGFP fused multimodular peptide constructs from +C (a) and -C (b) libraries visualized using Coomassie G-250 (PageBlue™ Protein Staining Solution). Similar volumes were loaded on gel. (E) purified constructs before dialysis and concentration, (C) purified constructs after dialysis and concentration, (N) non-induced bacteria, (CmE) EGFP without fusion peptides. Ladder: PageRuler™ Prestained Protein Ladder.

the production of the EGFP fused multimodular peptides was detected by SDS-PAGE and staining with Coomassie G-250 (Fig. 4). For the negative control, non-induced bacteria were used and for the positive control, EGFP expressing bacteria were used. Also, all constructs were sequenced to identify peptide composition and the yield was quantified (Table 3).

After Ni^{2+}-NTA purification, all peptide constructs were detectable on SDS-PAGE, except for +C36, which was not visible. All peptides produced were in the range of the expected theoretical sizes (30-50kDa). Three peptide constructs were lost during dialysis and concentration using Vivaspin centrifugal concentrators, namely +C21, +C29 and -C32. The yields of the concentrated constructs ranged from 22 to 320nmol peptide/L bacterial culture.

Table 3 Name, peptide composition, and total protein yield after dialysis of recombinantly expressed constructs. (NI) Not identified

Name	Composition	Yield after dialysis (µg)	Yield after dialysis (nmol)
+C11	Pr18	316	9.1
+C13	H9.2-NLSV402	92	2.7
+C19	SPKR	223	6.4
+C20	NI	102	2.9
+C21	TAT-H5WYG	41	1.2
+C29	SPKR-SPKR-H5WYG-SPKR-H9.2	37	1.1
+C31	TAT-H5WYG	46	1.3
+C35	SPKR-SPKR	391	11.3
+C36	Pr18-ppTG20	53	1.5
-C11	LAH5	107	3.1
-C12	Pr18	476	13.8
-C14	NLSV402-Pr18	234	6.8
-C15	TAT-NLSV402	319	9.2
-C17	H9.2	379	10.9
-C18	Mu	466	13.5
-C23	H9.2	553	16.0
-C24	SPKR	433	12.5
-C32	H9.2-H5WYG	143	4.1
CmE	-	228	7.5

DNA complexation and transfection efficiency of EGFP fused multimodular peptides

Next, 13 EGFP fused multimodular peptide constructs were screened for their capacity to bind to pDNA and to mediate gene delivery by measuring reporter gene expression *in vitro*. The 13 EGFP fused peptide constructs were selected based on their yields after dialysis and concentration. The EGFP peptides were tested at the highest peptide/pDNA ratios possible within the experimental setup used.

Increasing amounts of EGFP peptides were mixed with pDNA encoding mRFP in HBS. Formation of complexes was determined using the gel retardation assay. Thus, the complexes formed were subjected to electrophoresis in a 0.8% agarose gel (Fig. 5).

All EGFP constructs, except for -C14, were able to retard pDNA. Only the constructs +C13, +C35, -C12, -C17, -C18, -C23, and -C24 were able to reach complete retardation of pDNA (i.e. pDNA remains in loading slot of gel) at the highest peptide amount tested.

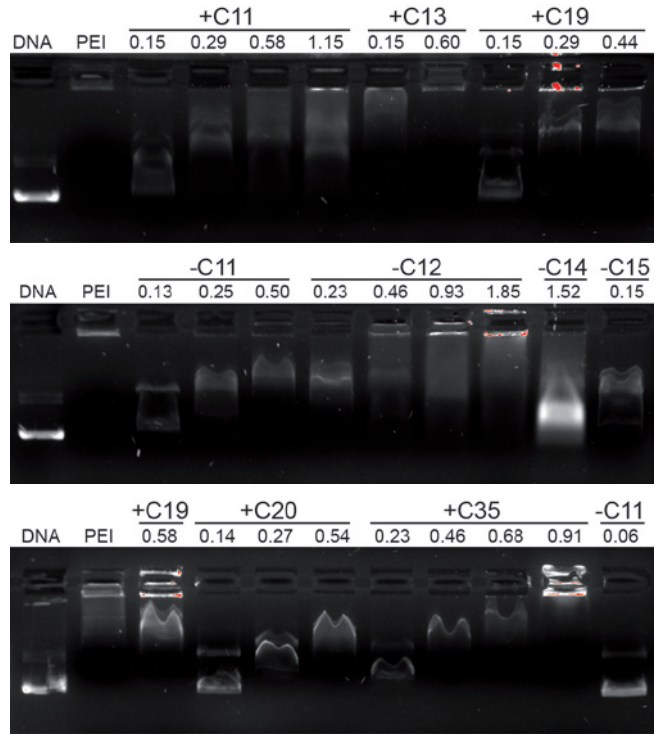


Figure 5 Gel retardation analysis of complexation of pDNA by EGFP fused multimodular peptide constructs. Different amounts (0-2nmol) of EGFP constructs were mixed with 0.25µg pDNA and subjected to electrophoresis in a 0.8% agarose gel. (DNA) 0.25µg pDNA only, (PEI) PEI polyplexes at N/P 6.

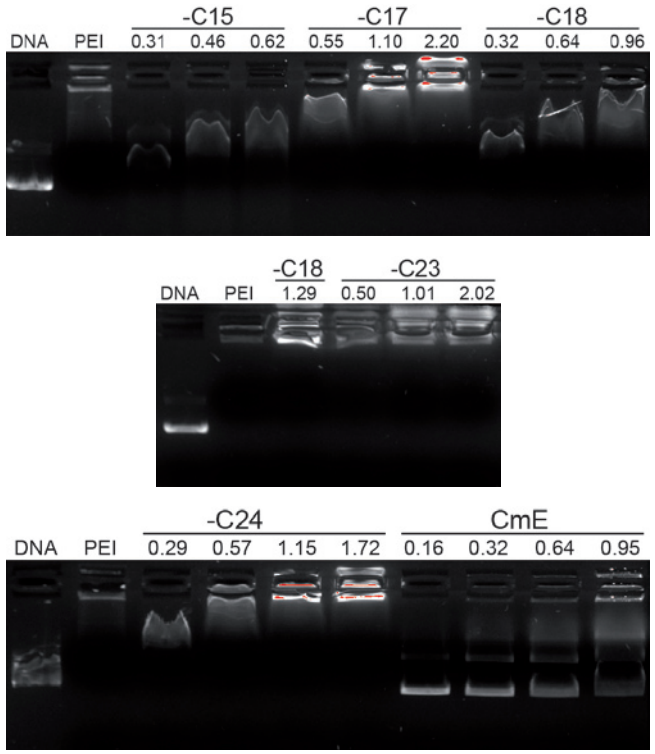


Figure 5 Continued.

For determining the transfection efficiency of the EGFP fusion peptides, increasing amounts of EGFP fusion peptides were mixed with pDNA encoding mRFP in HBS, after which COS-7 cells were transfected and analyzed with ArrayScan HCS Reader using the previously described Target Activation BioApplication with minor modifications (Chapter 4) (24). To determine the transfection efficiency, mRFP expressing cells instead of EGFP expressing cells were identified among all the detected cells. To determine association of the complexes with cells, EGFP positive cells were identified among all the detected cells. In Figure 6 the cytotoxicity, complex association and transfection efficiency results of all EGFP fused multimodular peptides are displayed. Overall, constructs showed little to no cytotoxicity with a relative viability between 70% and 100%. With increasing amounts of peptides, toxicity of the EGFP peptide complexes increased.

Ten out of thirteen constructs displayed association with cells. With increasing amounts of peptides, the percentage EGFP positive cells increased, except for +C35. None of the selected EGFP fusion constructs was able to transfect COS-7 cells, whereas the positive control PEI showed a transfection efficiency of 17% of total cells.

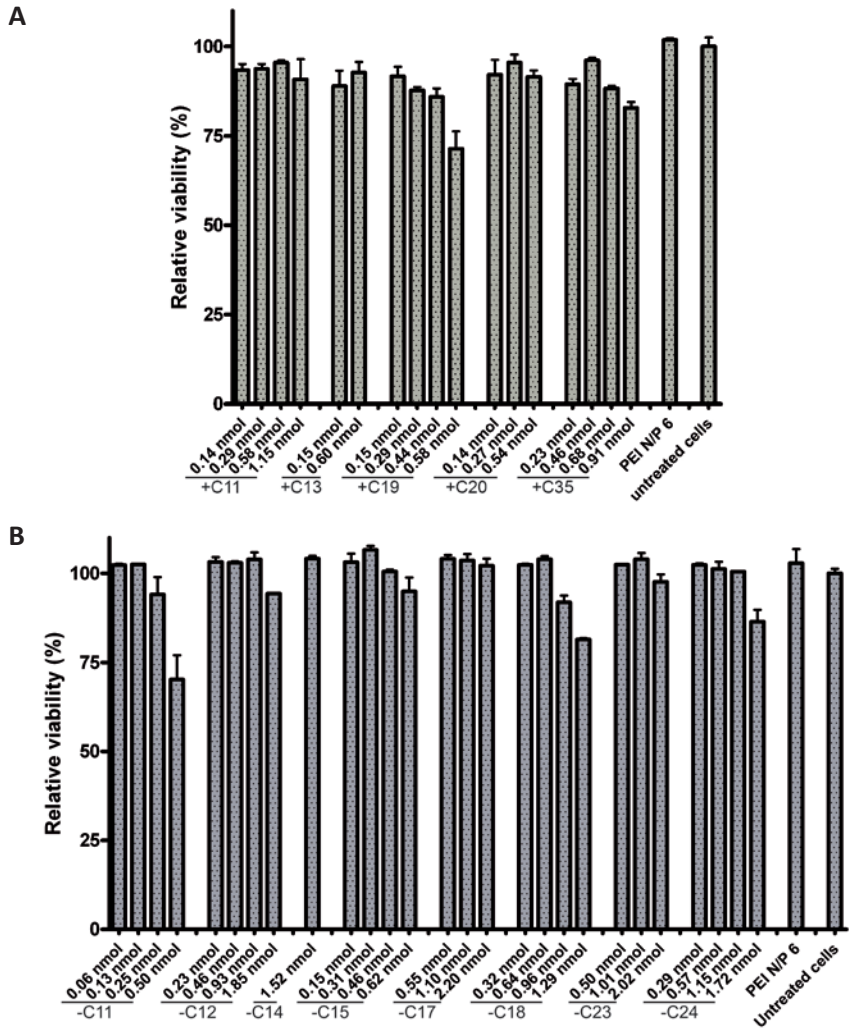
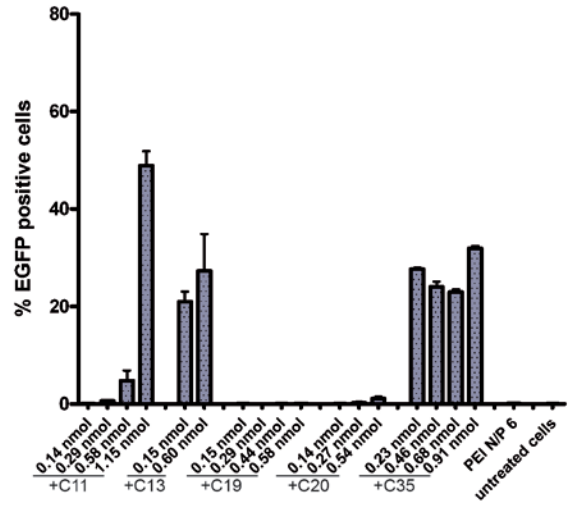


Figure 6 Relative viability (A,B), complex association (C,D) and transfection efficiency (E,F) of +C (A,C,E) and -C (B,D,F) EGFP fusion constructs. COS-7 cells were incubated for 24h with peptide complexes and 48 h after transfection cells were imaged on the a Cellomics Arrayscan V HCS Reader and analyzed by the Target Activation BioApplication. Positive control (PEI) were tested at their optimal DNA/transfectant ratios (see M&M). Data are presented as mean + SD, N=2.

C



D

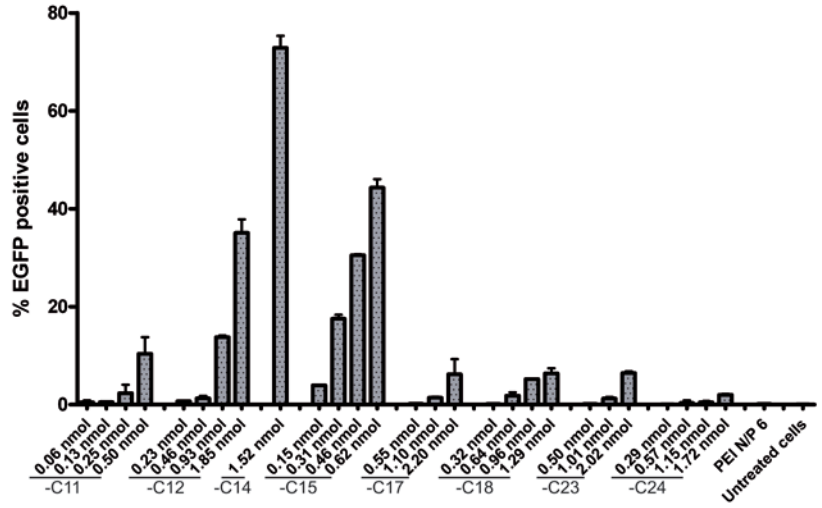
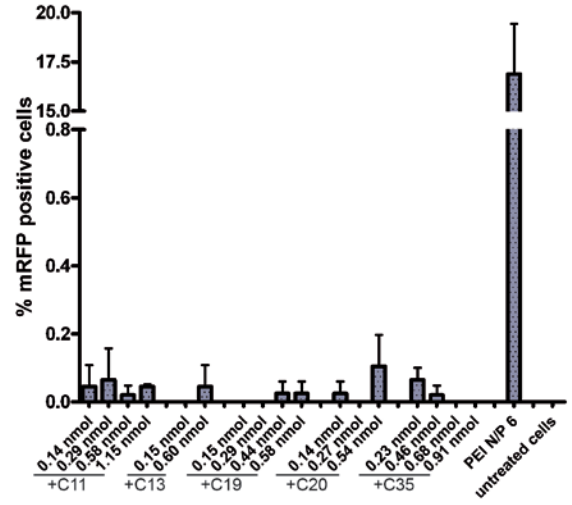


Figure 6 Continued

E



F

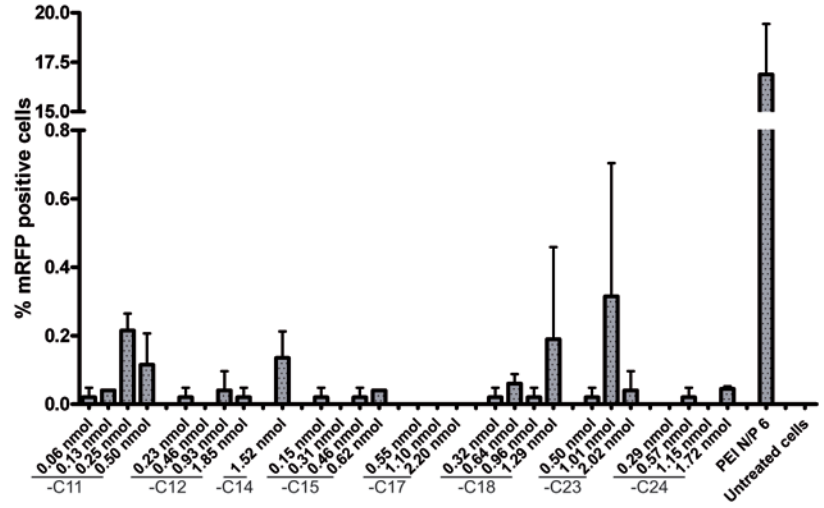


Figure 6 Continued

Discussion and conclusion

In this study, we report on the generation of a gene library encoding EGFP fused multimodular peptides and the subsequent recombinant expression and purification of individual peptide constructs from this library. The generated libraries were evaluated for their ability to deliver pDNA into COS-7 cells as part of an integrative strategy to screen and select functional peptide vectors for gene delivery.

Two different combinatorial gene libraries, encoding 9 different functional peptides ($SPKR_4$, Pr18, Mu, H9.2, H5WYG, TAT(47-57), LAH5, ppTG20 and NLSV402) were generated using the solid-phase platform for combinatorial multipart gene assembly (Chapter 3, 5 and 6) (14). Libraries created with the Head+C contained a cathepsin D cleavage site between EGFP and the peptides, whereas in libraries created with Head-C the cathepsin D cleavage site was omitted. After internalization, dissociation of the peptides/pDNA complexes from EGFP could be beneficial (10).

After cloning, the pDNA library obtained was transformed into BL21-Gold(DE3) bacteria. Flow cytometric analysis of induced +C library showed a subpopulation of fluorescent bacteria, confirming the expression of EGFP (Fig. 2c). The higher fluorescence intensity of bacteria expressing EGFP compared to bacteria expressing +C library could be caused by differences in expression levels or folding of the EGFP fusion constructs. Lower expression levels of library constructs could be caused by toxicity/intolerance of the construct towards the host or degradation of the constructs. Due to the presence of the C-terminal peptide(s), correct folding of EGFP could partially be impaired.

Transformation of the pDNA libraries into BL21-Gold(DE3) bacteria resulted in colonies containing unique inserts (Fig. 1d). About 94% of screened colonies contained correct pDNA, which is 2 times higher compared to pDNA obtained from colonies transformed with multimodular peptide libraries without a fusion protein (Chapter 5). This indicates that the use of EGFP as fusion partner reduces the toxicity of the encoded multimodular peptides and has a similar effect as the anti-EGFR nanobody as fusion partner (Chapter 6). In total, 48 colonies (24 colonies from both the +C and -C libraries) were screened for EGFP fluorescence as an indication of protein expression. About 20% of the screened colonies did not display EGFP fluorescence, despite containing the correct pDNA. This could be caused by mutations in the pDNA that may occur during production or by impaired folding of the EGFP domain of the fusion construct.

After pDNA analysis and EGFP fluorescence screening, 9 colonies from +C library and 9 colonies from -C library were selected for protein expression at 0.05L scale. After purification, almost all fusion constructs were detectable using SDS-PAGE (Fig. 4). Only +C36 could not be detected, despite the fluorescence observed in the bacterial host (Fig. 3). C-terminal truncation of the EGFP fusion construct, either by proteolytic degradation or by pDNA mutation may have caused the loss of the HIS-tag

located at the C-terminus, with subsequent failure of purification. Expression levels between the constructs varied greatly. In general, constructs which displayed low EGFP fluorescence also displayed low recombinant protein yields (Fig. 3 and 4). Furthermore, after dialysis and concentration 3 constructs could not be recovered, namely +C21, +C29 and -C32. Adsorption of the positively charged fusion constructs onto the negatively charged dialysis membrane could be a cause of the low recovery. By using EGFP as a fusion protein, we tried to increase the yield of multimodular peptides. The use of EGFP resulted in higher expression levels compared to expression levels of nanobody fusion constructs and constructs without fusion protein in chapter 5 and 6 (22-320nmol/L *E.coli* vs 9-25nmol/L *E.coli* vs 2-9nmol/L *E.coli*, respectively). Because of the increased recombinant protein yield, an expression scale of 0.05L resulted in sufficient amounts of fusion constructs to perform cellular assays.

Besides improving recombinant expression yields, EGFP allows direct monitoring of recombinant protein expression. Fluorescence levels after recombinant protein expression correlated to recombinant yields, where high fluorescence resulted in high recombinant protein expression and can therefore be used as a predictive tool. However, truncations caused by proteolytic degradation or mutations will result in false positive bacteria. Using EGFP as a C-terminal fusion partner could circumvent this.

The use of EGFP as fusion protein allows for pre-screening and selection of bacteria containing high levels of fusion constructs in a high-throughput fashion (6,25,26). Fluorescence activated cell sorting (FACS) can be applied to isolate single bacteria within large populations. Using FACS, multiple selection rounds can be performed to isolate specific populations or different subpopulations can be identified and isolated (25). Other, more straightforward screening methods are in-gel GFP fluorescence detection or the use of a fluorescent plate reader to identify fluorescent bacteria in 96 or 384 well plates (27).

Thirteen EGFP fused multimodular peptide constructs were screened for their capacity to bind to pDNA. Increasing amounts of EGFP peptides were mixed with pDNA encoding mRFP in HBS and the complexes formed were subjected to electrophoresis. Almost all EGFP constructs were able to bind to and retain pDNA, except for -C14. However, only 7 constructs were able to retain pDNA completely. According to literature, the peptides used in the library are all able to complex and completely retain 0.25µg pDNA at concentrations starting from 0.1-0.5nmol (16-23). This is in line with our observations where pDNA was completely retained at concentrations starting at 0.5nmol. In general, EGFP fusion constructs are less efficient at retaining pDNA compared to peptides without fusion constructs in literature.

After determining the pDNA binding capacity, the 13 EGFP fusion constructs were tested for their ability to mediate DNA transfection into COS-7 cells. Simultaneously, the cytotoxicity and complex association of the fusion constructs was determined. Overall, the EGFP fusion constructs were non-toxic.

Cell-associated EGFP fluorescence was detected 48h after transfection, indicating the presence of EGFP peptides or EGFP peptide/DNA complexes either attached or inside COS-7 cells. This at least indicates that some constructs are able to bind cells, even without the presence of a receptor-binding domain in the construct. The overall cationic charge of the EGFP fusion peptides may facilitate cell adsorption and subsequent internalization as is the case for many polycationic transfectants. However, results obtained are difficult to interpret. Not all constructs displayed fluorescence and large differences were observed between constructs. Also, the different amounts of fusion constructs used hinders direct comparisons.

None of the fusion constructs was able to transfect COS-7 cells. The observed lack of transfection can be caused by insufficient amounts of EGFP fusion constructs present in the formed complexes. In Chapter 4, the two peptides ppTG20 and LAH5, present in the currently created libraries, were tested for their transfection efficiency. ppTG20 showed a transfection efficiency of 0.8% using 1.5 nmol peptide/0.25 µg pDNA and LAH5 showed a transfection efficiency of 8% using 2.7 nmol peptide/0.25 µg pDNA (24). The amounts of fusion constructs used in this study ranged from 0.06-2.2 nmol/0.25 µg pDNA, which may be in some cases too low to detect transfection.

Another cause of the lack of transfection could be the particle stability. Although EGFP constructs were able to bind to and retain pDNA, the EGFP fusion constructs may be less efficient at retaining pDNA compared to peptides without fusion constructs. The formed complexes could be unstable and dissociate before they are taken up by cells and would thus be unable to transfect.

Also, the co-purification and inadequate removal of contaminants during protein purification can hinder transfection. DNases and RNases are used in the purification process. However, if these nucleases are not completely removed after purification, the reporter pDNA can be (partially) degraded during particle formation, thereby lowering the amount of pDNA available to transfect cells. Besides DNases and RNAses, lipopolysaccharide (LPS) contamination can influence the transfection efficiency. LPS are found in the outer membrane of Gram-negative bacteria (including *E.coli*) and could be co-purified with the EGFP-fusion constructs. If LPS levels are too high (≥50,000 EU), transfection can be inhibited.

Another possible reason for the lack of transfection could be the particular constructs used. Out of a library containing (theoretically) over 66,000 different fusion constructs, only 13 were screened for their ability to deliver pDNA into cells (28).

More candidates need to be screened using higher amounts of fusion constructs before definite conclusions can be drawn about whether EGFP fused multimodular peptides can be used for gene delivery.

In conclusion, we demonstrated the generation of a gene library encoding EGFP fused multimodular peptides by fusing peptides to the C-terminus of EGFP. The use of EGFP fused to the peptide library increased the recombinant expression yield of the fusion constructs and allowed direct monitoring of recombinant protein expression. After the expression and subsequent purification of 13 individual fusion constructs, these constructs were able to bind to and complex pDNA. Unfortunately, none of the selected fusion constructs was able to transfect COS-7 cells. More constructs from the library have to been screened to identify other functional peptide combinations for efficient gene delivery.

Abbreviations

bp:	base pairs
EGFP:	enhanced green fluorescent protein
EU:	endotoxin unit
FACS:	fluorescence activated cell sorting
HIS-tag:	polyhistidine tag
IMAC:	immobilized metal ion affinity chromatography
IPTG:	isopropyl β-D-1-thiogalactopyranoside
kDa:	kilo dalton
LB:	Lennox broth
mRFP:	monomeric red fluorescent protein
NLS:	nuclear localization signal
ODN:	oligodeoxynucleotide
PAGE:	polyacrylamide gel electrophoresis
pDNA:	plasmid DNA
SDS-PAGE:	sodium dodecyl sulfate polyacrylamide gel electrophoresis
ss:	single-stranded
TB:	terrific broth

References

(1) Mastrobattista B,van der Aa, Crommelin. Nonviral gene delivery systems: From simple transfection agents to artificial viruses. Drug Discovery Today: Technologies 2005;2(1):103-109.

(2) Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 2006;48(1):1-13.

(3) Kenig M, Peternel S, Gaberc-Porekar V, Menart V. Influence of the protein oligomericity on final yield after affinity tag removal in purification of recombinant proteins. Journal of Chromatography A 2006;1101(1-2):293-306.

(4) Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA. Fluorescent proteins and their applications in imaging living cells and tissues. Physiol Rev 2010;90(3):1103-1163.

(5) Sieczkarski SB, Whittaker GR. Dissecting virus entry via endocytosis. J Gen Virol 2002;83(7):1535-1545.

(6) Newstead S, Kim H, Von Heijne G, Iwata S, Drew D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 2007;104(35):13936-13941.

(7) Skosyrev VS, Rudenko NV, Yakhnin AV, Zagranichny VE, Popova LI, Zakharov MV, et al. EGFP as a fusion partner for the expression and organic extraction of small polypeptides. Protein Expr Purif 2003;27(1): 55-62.

(8) Caron NJ, Torrente Y, Camirand G, Bujold M, Chapdelaine P, Leriche K, et al. Intracellular delivery of a Tat-eGFP fusion protein into muscle cells. Molecular Therapy 2001;3(3):310-318.

(9) Skosyrev VS, Kulesskiy EA, Yakhnin AV, Temirov YV, Vinokurov LM. Expression of the recombinant antibacterial peptide sarcotoxin IA in Escherichia coli cells. Protein Expr Purif 2003;28(2):350-356.

(10) Haines AM, Irvine AS, Mountain A, Charlesworth J, Farrow NA, Husain RD, et al. CL22 - a novel cationic peptide for efficient transfection of mammalian cells. Gene Ther 2001 Jan;8(2):99-110.

(11) Pacholsky D, Vakeel P, Himmel M, Löwe T, Stradal T, Rottner K, et al. Xin repeats define a novel actin-binding motif. J Cell Sci 2004;117(22):5257-5268.

(12) van Gaal EV, Oosting RS, Hennink WE, Crommelin DJ, Mastrobattista E. Junk DNA enhances pEI-based non-viral gene delivery. Int J Pharm 2010 May 5;390(1):76-83.

(13) Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 2007 Jul;35(Web Server issue):W43-6.

(14) de Raad M, Kooijmans SA, Teunissen EA, Mastrobattista E. A Solid-Phase Platform for Combinatorial and Scarless Multipart Gene Assembly. ACS Synth Biol 2013 Mar 15.

(15) Henaut A, Danchin A. Analysis and predictions from Escherichia coli sequences, or E. coli in silico. Escherichia Coli and Salmonella: Cellular and Molecular Biology 1996:2047-2066.

(16) Schwartz B, Ivanov MA, Pitard B, Escriou V, Rangara R, Byk G, et al. Synthetic DNA-compacting peptides derived from human sequence enhance cationic lipid-mediated gene transfer in vitro and in vivo. Gene Ther 1999 Feb;6(2):282-92.

(17) Fortunati E, Ehlert E, van Loo ND, Wyman C, Eble JA, Grosveld F, et al. A multi-domain protein for beta1 integrin-targeted DNA delivery. Gene Ther 2000 Sep;7(17):1505-15.

(18) Murray KD, Etheridge CJ, Shah SI, Matthews DA, Russell W, Gurling HM, et al. Enhanced cationic liposome-mediated transfection using the DNA-binding peptide mu (mu) from the adenovirus core. Gene Ther 2001 Mar;8(6):453-60.

(19) Ignatovich IA, Dizhe EB, Pavlotskaya AV, Akifiev BN, Burov SV, Orlov SV, et al. Complexes of plasmid DNA with basic domain 47-57 of the HIV-1 Tat protein are transferred to mammalian cells by endocytosis-mediated pathways. J Biol Chem 2003 Oct 24;278(43):42625-36.

(20) Rittner K, Benavente A, Bompard-Sorlet A, Heitz F, Divita G, Brasseur R, et al. New basic membrane-destabilizing peptides for plasmid-based gene delivery in vitro and in vivo. Mol Ther 2002 Feb;5(2):104-14.

(21) Kichler A, Leborgne C, Marz J, Danos O, Bechinger B. Histidine-rich amphipathic peptide antibiotics promote efficient delivery of DNA into mammalian cells. Proc Natl Acad Sci U S A 2003 Feb 18;100(4):1564-8.

(22) Midoux P, Kichler A, Boutin V, Maurizot JC, Monsigny M. Membrane permeabilization and efficient gene transfer by a peptide containing several histidines. Bioconjug Chem 1998 Mar-Apr;9(2):260-7.

(23) Ritter W, Plank C, Lausier J, Rudolph C, Zink D, Reinhardt D, et al. A novel transfecting peptide comprising a tetrameric nuclear localization sequence. J Mol Med 2003 Nov;81(11):708-17.

(24) de Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E. High-content screening of peptide-based non-viral gene delivery systems. J Control Release 2012 Mar 28;158(3):433-442.

(25) Alfasi S, Sevastsyanovich Y, Zaffaroni L, Griffiths L, Hall R, Cole J. Use of GFP fusions for the isolation of Escherichia coli strains for improved production of different target recombinant proteins. J Biotechnol 2011;156(1):11-21.

(26) Schellenberger V, Wang C-, Geething NC, Spink BJ, Campbell A, To W, et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. Nat Biotechnol 2009; 27(12): 1186-1190.

(27) Müller-Lucks A, Bock S, Wu B, Beitz E. Fluorescent in situ folding control for rapid optimization of cell-free membrane protein synthesis. PLoS ONE 2012;7(7).

(28) Butash KA, Natarajan P, Young A, Fox DK. Reexamination of the effect of endotoxin on cell proliferation and transfection efficiency. BioTechniques 2000;29(3):610-619.

8

Chapter

Summarizing Discussion and Perspectives

Summarizing Discussion and Perspectives

Gene therapy forms an attractive approach for the therapeutic intervention of a wide range of diseases through the delivery of therapeutic nuclei acids into target cells. In order to be effective, the therapeutic nucleic acids have to reach their specific site of action inside target cells. Due to their physicochemical characteristics, nucleic acids are unable to perform this task themselves and require the assistance of a delivery system.

Naturally derived or synthetic peptides offer an attractive platform for non-viral gene delivery. They may aid in the protection of NA against nuclease attack, mediate cell-specific binding, facilitate cellular uptake and cytosolic release and enable intracellular routing. In **Chapter 2** we give an overview of functional peptides used for gene delivery, either individually or in multifunctional/multi-component vectors. Despite the progress which has been made towards the generation of more efficient peptide vectors, several remaining challenges were discussed in **Chapter 2**.

Since gene delivery is a multistep process, it is essential that multiple functionalities are united into a single delivery system. Therefore, the development of novel multifunctional peptide vectors will be pivotal for improving peptide based gene delivery. In this thesis, we have tried to obtain novel multifunctional peptides that are able to efficiently deliver therapeutic nucleic acids into target cells. For this, we have proposed a design strategy that follows a random, integrative approach that selects for multimodular peptides containing combinations of functional traits that are optimal for efficient gene transfer. This strategy requires a combinatorial, modular design setup for the construction of multimodular peptide libraries and complementary high-throughput production and subsequent screening for the selection of most effective multimodular peptides for gene delivery (Fig. 1 in **Chapter 1**). For library construction, we developed a novel ligation method for combinatorial and scarless gene assembly. For the high-throughput generation of multimodular peptides, we explored the possibilities for high-throughput recombinant production of multimodular peptides in *E.coli* and their subsequent purification. For high-throughput screening, we developed an high-content screening (HCS) assay for the screening of multimodular peptides for their gene delivery capacity. Here we will discuss the benefits/drawbacks of our proposed random design strategy and the challenges and solutions for the 3 sections required for this strategy, namely (i) library construction, (ii) recombinant production and purification and (iii) screening for gene delivery capacity. Furthermore, other possible alternative routes for the generation of multimodular peptide libraries and their complementary high-throughput production and subsequent screening will be discussed.

Random versus rational

Generating novel multifunctional peptides for gene delivery by fusing multiple peptides into a single vector gives rise to different and new challenges. Some functionalities within the vector can be conflicting and knowledge on the optimal configuration (i.e. the composition and positions of the functional peptides in the vector) is lacking. Also, the selection of functional peptides to be used can be difficult. Selecting the best individual performing peptides from within their functional class would be logical. However, combining the best individual performing peptides into a single vector does not automatically yield the best performing multifunctional peptide vector.

Rational design has been the main strategy to design and optimize multifunctional peptide vectors. However, with the above mentioned challenges in mind, we propose a design strategy that follows an integrative, high-throughput approach. The advantage of this strategy is that in principle large numbers of multimodular peptide candidates can be screened for their gene delivery capacity without *a priori* knowledge of their exact composition. In this high-throughput approach problems with conflicting/opposing functionalities will be easily out-selected and the identification of the structure/activity relationships is accelerated.

Rational design offers a starting point for vector construction and allows for specific optimization. Integration of rational design elements into our design strategy could be beneficial when structure/activity relationships are studied. For example, in the generated libraries specific peptides can be built in at specific positions and libraries can be generated so that all constructs start with a CPP or end with a specific peptide.

A general concern of fusing multiple functional peptides into a single vector, is whether the individual peptides retain their respective function when assembled into a single vector. Although we did not investigate this ourselves, it is known from literature that individual peptides in a multimodular vector retain their respective function when properly designed (1).

Library construction

Genetic modular assembly strategies allow for generation of gene libraries that encode multifunctional peptide vectors in a fast, simple, and combinatorial way. Although several methods and standards have been developed, these all suffer from drawbacks such as the introduction of scar sequences during ligation or the need for specific flanking sequences or *a priori* knowledge of the final sequence to be obtained. In **Chapter 3**, we have introduced a sequence-independent ligation method for combinatorial and scarless assembly of multipart gene constructs, which allows fast ligation of multiple fragments in less than a day. This method is based on the ligation of single-stranded (ss) or double-stranded (ds) ODN and PCR products immobilized on a solid support.

To show proof of concept, a library containing nine different modules was created using solid-phase assembly of ssODN with linkers. Sequence analysis confirmed the ligation via the linker site and the presence of different modules. In only three days the library was created, amplified, ligated and transformed into bacteria. These results showed that the ligation method is suitable for the generation of gene libraries encoding multifunctional peptide vectors.

A possible concern for the generation of gene libraries is the ligation bias. Not all ssODN possess similar ligation efficiencies and this could cause an overrepresentation or underrepresentation of certain peptide domains within the library. A bias towards specific peptide domains was observed in the generated library in **Chapter 5**. Seventy-five percent of all peptide domains indentified in 37 constructs, were either Pr18, $SPKR_4$, ppTG20 or NLSV402. Without a bias, the percentage should be around 44%. This can be circumvented by adjusting the concentration of ssODN in the ligation mix. Switching to dsODN assembly should also reduce the ligation bias as it circumvents formation of hairpin loop structures as is the case in ssODN. Throughout this thesis, all gene libraries were assembled using the solid-phase assembly of ssODN with linkers. As a consequence, peptide modules are linked by the amino acids GGGS. Although this linker is flexible and frequently used, its effect on the functionality of the individual peptides and on nucleic acid complexation/particle formation is unknown (2). To generate libraries without using linkers, the dsODN assembly strategy should be used.

Recombinant production and purification

After gene library construction, the library has to be transformed into a suitable host organism and subsequently expressed and purified in a high-throughput manner.

Transformation of *E.coli*

The generated gene libraries from **Chapter 5, 6 and 7** were transformed in *E.coli* BL21-Gold(DE3) for recombinant expression. After transformation of the library encoding multimodular peptides without a fusion protein, about 45% of the obtained bacterial colonies contained mutations in the pDNA (**Chapter 5**). This could be caused by plasmid toxicity or toxicity of the encoded multimodular peptide, despite the presence of transcriptional repressors. The induced “stress” by the toxic plasmid/protein may lead to increased mutational alterations of its respective gene. As a consequence, bacterial colonies were first screened for the presences of correct pDNA before recombinant protein expression, which is time consuming and restricts high-throughput expression. Using stronger transcription repressors (pREP4 or pLysS) or promoters with low basal activity (araBAD) the amount of colonies with mutations could be decreased (3,4). On the other hand, the high toxicity/low tolerance displayed by the multimodular peptides can be used as a prescreening step. Toxic or poorly

tolerated multimodular peptides will be difficult to express and will result in low recombinant protein yields. If bacteria containing pDNA encoding these toxic/poorly tolerated multimodular peptides can be eliminated from the population, a library of well tolerated multimodular peptides, with higher recombinant protein yields, can be obtained. Such a selection can be performed by inducing recombinant protein expression directly after transformation. Expression of toxic/poorly tolerated multimodular peptides will cause bacterial growth arrest and the inability to form colonies, whereas bacteria expressing well tolerated multimodular peptides will grow normally and be able to form colonies. However, mutations may result in false positive colonies, similar to the mutations observed in **Chapter 5**.

Another solution for the observed toxicity is fusion of the multimodular peptides to a carrier protein. Fusion of the multimodular peptides to the anti EGFR biparatopic nanobody 7D12-9G8 and EGFP resulted in only 5-10% bacterial colonies containing mutated pDNA, indicating the reduced toxicity of the encoded multimodular peptides (**Chapter 6 and 7**).

Recombinant expression of multimodular peptides without fusion proteins

Library transformation was followed by recombinant protein expression and purification. Recombinant expression of the library encoding multimodular peptides without a fusion partner resulted in low total protein yields (**Chapter 5**). Possible causes are the toxicity/low tolerance of the multimodular peptides towards the host, blockage of mRNA translation and proteolytic degradation of the multimodular peptides during recombinant production.

In order to reduce blockage of translation, we tried to codon optimize the sequences encoding the peptides and supplied the tRNAs of six rare codons in *E.coli*. To reduce toxicity and proteolytic degradation, we tried to optimize our protein expression protocol by adjusting several parameters, including temperature, inducer concentration, different repressors and media type. The optimal conditions we found here were used throughout this thesis. Also, the bacterial culture volumes were increased to 0.5L. Despite these adjustments, recombinant expression of the multimodular peptides did not result in sufficient yields to screen the peptides for their transfection efficiency in a high-throughput fashion.

Within the used setup of recombinant production using *E.coli*, other adjustments are possible to improve the recombinant protein yield. Using different promoters and *E.coli* strains, changing culture setup to increase bacterial culture densities in combination with different culture conditions could result in higher expression yields (3,5). By using (fractional) factorial approaches, screening of multiple variables (e.g. *E. coli* strains, culture media, expression temperatures and promoters) for optimal recombinant protein yield can be performed (6). However, optimization of expression conditions needs to be performed for each protein individually, which is not applicable

for library screening. Small differences in the amino acid sequence or the length of the construct can transform a protein that is easily expressed into a protein that fails to be expressed (5). Optimized conditions for one protein can result in suboptimal expression conditions for a subpopulation of constructs (5,6).

Other groups were able to recombinantly produce modular peptide constructs for gene delivery purposes in *E.coli* with yields ≥2mg/L. However, almost all of these multimodular peptide constructs contain 1 or more functional proteins or protein domains in their native, folded conformation, which makes them less susceptible to proteolytic degradation and can improve expression. To our knowledge, only one other group produced a comparable multimodular peptide construct, with a size of 11.4 kDa containing similar DNA condensing, NLS and endosomolytic peptides (7). However, the group did not report the obtained yield of their multimodular peptide. Also, all groups used different expression parameters, which makes comparisons very difficult.

Although direct recombinant expression of small, positively charged peptides (e.g. antimicrobial peptides) is possible, yields differ widely and depend on peptide size and amino acid composition (8). Almost all (94%) antimicrobial peptides produced in *E.coli* were expressed as fusion proteins (9).

Recombinant expression of multimodular with fusion proteins

We tried to improve the recombinant protein yield within our setup by using fusion constructs. The introduction of the anti EGFR biparatopic nanobody 7D12-9G8 and EGFP as fusion proteins did result in higher protein yields (**Chapter 6 and 7**). Using the nanobody as fusion partner, protein yield increased 1-12.5 fold, whereas using EGFP resulted in a 2.5-35 fold increase of the protein yield. A large difference in yield was observed between the 2 used fusion proteins. This indicates that the effect of the fusion protein varies (10). More commonly used fusion proteins (e.g. SUMO, GST and KSI) could further enhance the protein yield (8,11). However, these fusion proteins have no function after expression and removal is required, which will result in lower yields of the final multimodular peptides due to inefficient cleavage, loss during handling or by degradation during cleavage (12,13). A possible solution for inefficient removal of non-functional fusion proteins are inducible autoproteolytic fusion proteins. During recombinant protein production, the fusion protein enhances the expression of the fusion construct. After purification, autoproteolysis can be induced and the fusion protein removes itself from the protein of interest. Examples of inducible autoproteolytic fusion proteins are EDDIE, cysteine protease domain (CPD) and inteins (8,14,15).

The biparatopic nanobody 7D12-9G8 and EGFP were selected as fusion proteins based on their functionality. In principle, many different fusion proteins can be selected as long as they fulfill two requirements: (i) They should possess a functionality

suited for gene delivery; (ii) They should easily be expressed in *E.coli*. Whereas EGFP fulfills both requirements, the biparatopic nanobody 7D12-9G8 is more difficult to express, which could explain the lower increase in recombinant protein yield. Other anti EGFR nanobodies with higher recombinant expression yields could be used as fusion proteins, such as EgA1 (16). Besides nanobodies, other functional proteins could be used as fusion proteins, such as coiled coils and albumin for particle formation or DARPins and affibodies for specific cell targeting (16-19).

Not all possible causes of low protein expression could be eliminated. In order to reduce possible blockage of translation, we tried to codon optimize the sequences encoding the peptides. However, this was only possible to a certain extent. For the assembly of the gene libraries, the ssModules used needed to be codon optimized to prevent unwanted hairpin formation that would impede ligation, which required the introduction of rare codons and repetitive usage of certain codons. After codon optimization for ligation, codon usage was also optimized for recombinant expression, if possible. Furthermore, due to the library formation, the secondary structure of mRNA cannot be optimized for each construct. As a result, combinations of certain modules could result in the formation of mRNA with secondary structures which cannot be translated.

High-throughput expression and purification

Despite the increase in recombinant protein yield by using fusion proteins, bacterial culture volumes still had to be increased in order to obtain sufficient amounts of fusion constructs to perform cellular assays. For the targeted multimodular peptide constructs, a volume of 0.5L was required whereas the EGFP fused multimodular peptide constructs required a minimal volume of 0.05L. These culture volumes are not compatible with high-throughput expression and purification systems (20-22). For example, in order to screen 5% of the generated targeted multimodular peptide library, about 3300 constructs have to be expressed and purified. For culture volumes ≤ 5ml, 24 or 96 deep well plates can be used which are compatible with liquid handling robots for culturing, automated lysis and purification (20,21). When using 96 well plates, about 35 plates are needed, whereas 138 plates are already needed when using 24 well plates. However, in case of 0.5L cultures volumes, 3300 2L Erlenmeyer flasks are required and culturing bacteria will take about 200 days.

Bias

Another consequence of the low expression yields is the creation of a bias in the screening of library constructs. Fusion constructs can only be screened for transfection efficiency if a sufficient amount of fusion protein is obtained. Due to the large variation in yields obtained, not all constructs can be used for screening for transfection efficiency and the library size decreases. With a smaller library, the

chances of finding promising candidates decrease. Furthermore, fusion constructs containing a high amount of CPP could be toxic for the bacteria through the membrane destabilizing properties of these peptides and thus be troublesome to express. However, a high amount of CPP could be beneficial for gene delivery.

By using the 3 currently explored setups (multimodular peptides without a fusion partner, nanobody and EGFP fused multimodular peptides), a bias towards non-toxic/well tolerated multimodular peptides is created.

Screening for gene delivery capacity

Recombinant expression and subsequent purification yielded sufficient amounts of both the nanobody- and EGFP fused multimodular peptide constructs to screen them for gene delivery capacity.

Delivery of ON-705 by targeted multimodular peptide libraries

Four different targeted multimodular peptide constructs, namely +C6, +C8, +C9 and -C7, were tested for their ability to induce splice correction via the delivery of a splice-correcting oligonucleotide in HeLa pLuc/705 cells (**Chapter 6**). The 3 constructs +C8, +C9 and -C7 were able to induce splice correction, although not very efficiently.

EGFR over-expression

The low levels of splice correction could be caused by a low degree of uptake by the HeLa pLuc/705 cells. However, the modified nanobody 7D12-9G8 in the targeted multimodular peptide constructs was still able to specifically bind to EGFR with similar affinity as the unmodified nanobody (**Chapter 6**). Also, Western Blot analysis confirmed EGFR over-expression by HeLa pLuc/705 cells (**Chapter 6).** Although binding and internalization studies were not performed with the HeLa pLuc/705 cells, we expected that the targeted fusion constructs would be able to bind to the cells and be internalized via the EGF receptor. According to literature, the nanobody 7D12-9G8 is rapidly internalized after binding to the EGF receptor (23).

Complexation of ON-705 by targeted multimodular peptides

Another cause of the low splice correction efficiency displayed by the constructs is complexation of the splice-correcting ON-705 and the inability to form complexes. It is known from the literature that small peptides are unable to efficiently form complexes with ON-705 and other single-stranded oligonucleotides when electrostatic interaction are the only driving force (24,25). Although constructs +C6 and +C8 were able to interact with the ON-705 at molar ratios of, respectively, 4:1 and 35:1, complexation of ON-705 was only modest (**Chapter 6**). Higher molar ratios are needed to complex all ON-705. However, the high molar ratios required give rise to another obstacle. Due to the high molar ratios, the amount of nanobody in the complexes

increases. Since internalization via EGFR is limited by the amount of receptors present on the cell surface, high amounts of nanobody can cause saturation of EGFR and uptake of complexes will thereby not increase proportionally. The optimal molar ratios for complexation of ONs and the optimal molar ratios for internalization are not within the same order of magnitude, which probably contributes to the low splice correction displayed by the fusion constructs.

Improving complexation of ON-705 by targeted multimodular peptides

In order to improve complexation and to decrease the amount of construct needed for complexation, the targeted multimodular peptides can be modified by stearylation or by introduction of other fatty acids groups at the C-terminus (25). However, chemical modification will further decrease fusion construct yield and is not compatible with the high-throughput setup. Instead of complexation of ON-705 through electrostatic interactions, ON-705 can be covalently attached to the fusion constructs, forming conjugates (26). The nucleic acid cargo can be linked 1:1 to the fusion constructs, which decreases the necessary amount of fusion construct. However, conjugation of nucleic acids is laborious, expensive and dissociation of the nucleic acids after internalization can be problematic.

Alternatively, instead of splice correcting ONs other therapeutic nucleic acids could be delivered by the targeted multimodular peptides. Peptides are able to condensate pDNA into small complexes whereas smaller ssONs interact more disorderly with cationic peptides. Since less peptides are needed to retain pDNA complexed compared to ONs, the chance of EGFR saturation by free nanobody will be less of a problem and may therefore lead to better uptake of the pDNA/peptide complexes required for transfection.

High-content screening (HCS)

The high-content screening (HCS) assay developed is able to simultaneously screen for cytotoxicity, transfection efficiency, induction of cell permeability and the capacity to transfect non-dividing cells (**Chapter 4**). For a proof-of-principle, a small library of synthetic peptides- was screened for transfection when complexed with EGFP-encoding pDNA. Two peptides which were able to transfect COS-7 cells were selected for use in the gene libraries, namely LAH5 and ppTG20. These results demonstrate that automated fluorescence microscopy can be used for high-throughput screening of gene delivery systems for multiple read-outs and will be a valuable extension to the already existing screening methods.

EGFP fused multimodular peptide libraries

In **Chapter 7**, 13 EGFP-fused multimodular peptide constructs were screened using the developed HCS assay. All constructs exhibited only low toxicity. Furthermore,

EGFP fluorescence was detected 48h after transfection, indicating the presence of EGFP fused multimodular peptides or EGFP peptide/pDNA complexes. This indicates that EGFP peptide/pDNA complexes can adhere to cells without the presence of a specific receptor binding domain. This is presumably mediated by the net positive surface charge on these complexes (**Chapter 7**). However, results obtained are difficult to interpret, mainly due to the use of different amounts of fusion construct and to the different intrinsic fluorescence levels of the fusion constructs.

Although almost all EGFP fused multimodular peptide constructs were able to complex pDNA, none of the fusion constructs was able to transfect COS-7 cells (**Chapter 7**). The lack of transfection could have been caused by the low ratios of fusion peptide/pDNA. With the screening of the small library of peptide-based transfectants in **Chapter 4**, LAH5 displayed a transfection efficiency of 8% at a ratio of 2.7 nmol peptide/0.25 µg pDNA. The EGFP fused construct -C11 also contained LAH5, but complexes were formed with a maximal ratio of 0.5 nmol EGFP-peptide/0.25 µg pDNA and displayed a transfection efficiency of 0.2%. However, when complexes of LAH5 were formed at a ratio of 0.35 nmol peptide/0.25 µg pDNA, the transfection efficiency dropped to 0.1-0.2% (**Chapter 4**). To screen the EGFP fused multimodular peptide constructs for their transfection efficiency, a range between 0.5-3 nmol EGFP-peptide/0.25 µg pDNA should be used, which is similar to the range used for screening the library of peptide-based transfectants in **Chapter 4**. Also, the proposed ratios are comparable to ratios used for the screening of other multimodula r peptide constructs in the literature (1,7,27). However, yields of the EGFP fusion constructs ranged from 1-16 nmol and should be increased in order to enable the screening of all fusion constructs.

Low numbers

A more general explanation for the low level of splice correction could be the particular fusion constructs used. Only 4 out of 66,000 targeted multimodular peptide constructs and only 13 out of 66,000 EGFP fused multimodular peptide constructs were screened for splice correction or pDNA delivery, respectively. In order to increase the chances of indentifying efficient constructs, more candidates need to be screened.

Alternatives and perspectives

The aim of this thesis was to select multimodular peptides that are able to efficiently deliver therapeutic nucleic acids into target cells from a large library of multimodular peptides. Despite our efforts, the development of novel multimodular peptides for targeted and efficient gene delivery was not successful.

Although our proposed random, integrative design strategy did not result in novel multimodular peptides with promising prospects for successful gene delivery, the rationale behind the strategy is still valid. We were able to construct gene libraries encoding multimodular peptides using a novel sequence-independent ligation method for combinatorial and scarless gene assembly. Also, a HCS was developed for high-throughput screening of transfectants for cytotoxicity, transfection efficiency, induction of cell permeability and the capacity to transfect non-dividing cells. The bottleneck was the high-throughput production and purification of the multimodular peptides. Whereas the yields of multimodular peptides without a fusion protein were insufficient to continue with cellular assays, nanobody- or EGFP fused multimodular peptides required large culture volumes incompatible with high-throughput expression and purification in order to obtain sufficient amounts of fusion constructs. To avoid this bottleneck, alternative routes are proposed without compromising the random and integrative design strategy.

Expression platforms

The most straightforward alternative is switching to a different expression platform. In *E. coli*, the most commonly used production strategies are the intracellular production of recombinant proteins, where the proteins are expressed in the periplasm or in the cytoplasm. However, intracellular recombinant protein synthesis can cause toxicity towards the host. By secretion of the recombinantly expressed protein into the culture medium, the toxicity caused by the protein can be reduced. Also, by secretion of the protein of interest, purification can be simplified. Fusion of secretion signals to the recombinant protein allows for secretory expression in *E.coli* (8,28).

Instead of *E.coli*, other Gram-negative bacteria, such as *Pseudomonas* and *Corynebacterium*, can be used (28). The main advantage of these bacteria over *E.coli* is their rapid growth.

Another attractive bacterium for recombinant protein production is *Lactococcus lactis* (28,29). These Gram-positive bacteria are able to secrete the recombinant protein into the culture medium in high amounts and do not contain an outer membrane consisting of lipopolysaccharides (LPS) (29).

Yeasts such as *Saccharomyces cerevisiae* and *Pichia pastoris* are widely used as expression systems and have been extensively applied as hosts for the production of biopharmaceuticals and industrial enzymes (10,30). The advantages of yeasts over *E.coli* are their ability to secrete the recombinant protein into the culture medium and their non-pathogenic nature (30). Furthermore, yeasts are frequently used for the production of antimicrobial peptides, with yields reaching 750 mg/L without the need for fusion proteins (9,10).

Another expression platform that can be used as an alternative is *in vitro* transcription/translation (IVTT) (31,32). IVTT systems are based on crude extracts or lysates containing the necessary components for transcription and protein translation. The main advantage of IVTT systems is that they do not rely on the use of living cells and therefore do not require maintenance of cell viability (31). However, the obtained recombinant protein yield varies and production of extracts is laborious and expensive.

Prescreening of the multimodular library

Instead of switching to a different expression platform, expanding the current high-throughput production and purification setup with a prescreening step is an attractive alternative to increase recombinant protein yield. The goal of a prescreening step is the identification and selection of promising candidates at an early stage to decrease library size, thereby allowing the use of larger culture volumes.

The earlier proposed prescreening for the removal of toxic/poorly tolerated multimodular peptides using reduced growth as a read-out, can be a possible screening tool. On the other hand, more robust and high-throughput selections can be used. Flow cytometry and fluorescence activated cell sorting (FACS) are able to analyze and isolate individual bacteria within large populations in a relatively short time (33). Both techniques can be applied to bacteria, as demonstrated in **Chapter 7**, where flow cytometry was used for the detection of fluorescent bacteria after transformation of *E.coli* BL21-Gold (DE3) with a pDNA library encoding EGFP fused multimodular peptides. Prescreening based on recombinant protein production levels can be used to select bacteria containing multimodular peptides that are well tolerated and yield high amounts of multimodular peptides. Using a fluorescent protein as fusion partner allows direct monitoring of recombinant protein expression. However, by using a prescreening based on host tolerance, promising candidates with optimal characteristics for gene delivery might be outselected. Efficient transfection is often accompanied by a certain degree of cytotoxicity as is the case for both synthetic and biological transfection systems. Setting up a prescreen for host cell tolerance, may therefore lead to loss of interesting candidates.

Different screening assays

The above mentioned alternative strategies are based on improving recombinant protein yield. Due to the low recombinant protein yield, screening of the peptide libraries for transfection efficiency was either not possible or required large culture volumes in order to produce sufficient material. This can also be turned around, by selecting screening methods which require less multimodular peptides.

For the evaluation of the capacity of delivering therapeutic nucleic acids into cells, transfection studies have to performed. However, the transfection assays used required the delivery of relative large amounts of nucleic acids in order to detect

transfection either via transgene expression or via splice correction. By increasing the amount of nucleic acid, the amount of multimodular peptide required for delivery also has to be increased. The use of more sensitive assays would allow the use of smaller amounts of multimodular peptide to detect transfection. For example, the reverse transfection technology can be used. In a reverse transfection assay, cells are added directly to a glass slide or well plate containing the transfectants and examined on e.g. reporter gene expression (34,35). By using a microarrayer robot, transfectants can be printed onto a glass slide to form a microarray and large numbers of transfectants can be screened in a single study while low amounts of tranfectants are needed (34,35). Alternatively, well- or microtiter plates can be used (36,37).

Instead of using reporter gene expression to evaluate transfectants for gene delivery, other assays can be used for studying intermediate steps in the entire gene delivery process, such as specific cell binding, internalization or inducing endosomal escape/cytosolic delivery. For example, to screen for efficient delivery of cargo into the cytosol, the split-EGFP assay can be used. The split-EGFP assay is based on the spontaneous, auto-assembly capacity of two non fluorescent portions of GFP (GFP 1-10 and GFP 11) to restore a fully fluorescent GFP (38,39). The GFP11 fragment is fused to the peptide/protein of interest and the GFP 1–10 detector fragment is stably expressed in the target cells. If the fusion peptide/protein reaches the cytosol intact, GFP functionality is restored and the cells fluoresce. By fusing the GFP11 fragment to the N or C terminus of the multimodular peptides, uptake and subsequent endosomal escape can be studied.

Alternatively, instead of measuring the biological activity of a reporter cargo, absolute amounts of internalized multimodular peptides with or without cargo can be measured using MALDI-TOF MS (40). Also, mass spectrometry can be used to identify the peptide sequence of the peptide sequences of the multimodular peptide constructs.

Multi-component complexes

Another strategy which requires less multimodular peptides and could enhance gene delivery capacity is the formation of multi-component complexes. Instead of forming complexes with nucleic acids by using a single multimodular peptide, different multimodular peptides can be mixed and multi-component complexes are obtained. Also, polymers or lipids can be mixed with multimodular peptides to form ternary complexes. Several studies showed an increase in transfection efficiency of multi-component complexes containing a peptide and polymer/lipid compared to the polymer or lipid complexes without peptide (41-43). However, by using multiple components to form complexes, the risk of compositional variations and poor reproducibility arises.

Library construction using native chemical ligation (NCL)

An alternative to recombinant production of multifunctional peptide libraries, is peptide synthesis and native chemical ligation (NCL) (44,45). Using NCL, peptides can be ligated in a combinatorial manner directly into multimodular peptide constructs (45). The peptide building blocks can either be synthesized or recombinantly expressed. The main advantage of direct assembly of multimodular peptides is that the production does not require living organisms. Therefore, there is no bias towards the selection of non-toxic, well tolerated multimodular peptides. However, the aggregation of intermediate products or ligation of “difficult peptide sequences” can cause a ligation bias, which is similar to the ligation bias observed in the generation of gene libraries (**Chapter 3)** (44-46). Also, low yields, high costs and large time consumption are other challenges associated with the synthesis of multimodular peptide libraries (44-46).

Alternative applications

In this thesis, we aimed at creating novel multimodular peptides for gene delivery through the use of our proposed random, integrative design strategy. However, such a strategy is not restricted to the creation of multimodular peptides for gene delivery. Using the sequence-independent ligation method for combinatorial and scarless gene assembly, libraries encoding all kind of modular polypeptides or proteins can be designed. For example, the widespread use of antibiotics has generated a strong evolutionary pressure for the emergence of bacteria that either have an inherent resistance to a particular antibiotic or have the capacity to acquire such resistance (47). Therefore, novel classes of antibacterial compounds need to be generated. The modular nature of antimicrobial peptides allows for the generation of multimodular libraries encoding novel antimicrobial peptides. Also, combinations of peptides could potentially increase their antimicrobial activity. By using single-molecule PCR, IVTT and a 96- or 384 well based bacterial growth assay, libraries of novel antimicrobial peptides can be screened in a high-throughput fashion (48,49).

Cystine-knot miniproteins and DARPins have emerged as new classes of molecules for therapeutic and diagnostic applications, including protein binding and receptor binding/blocking (17,50,51). The modular nature of these proteins allows for the creation of combinatorial libraries for the generation of e.g. novel targeting ligands. By using bacterial/yeast display systems or IVTT in combination with genotype-phenotype fusion display, large libraries can be screened in a high-throughput manner (20,52,53).

In conclusion, in this thesis we tried to obtain novel multimodular peptides that are able to efficiently deliver therapeutic nucleic acids into target cells. For this, a strategy was followed which involved: (i) Gene library construction encoding multimodular

peptides for gene delivery, (ii) High-throughput recombinant production and purification of these peptides and (iii). High-throughput screening for gene delivery capacity on cells in culture (Fig. 1 in **Chapter 1**). Two out of the three steps to achieve this goal were successful: (i) we created a novel sequence-independent ligation method for combinatorial and scarless gene assembly; (iii) we developed a HCS assay for the simultaneous screening of gene delivery systems' cytotoxicity, transfection efficiency, induction of cell permeability and the capacity to transfect non-dividing cells. Unfortunately, we did not succeed in optimizing the recombinant expression of the multimodular peptide library to levels required for high-throughput handling. Nevertheless, the insights obtained provide good starting points to further improve high-throughput recombinant protein expression which should ultimately enable the high-content screening of multimodular peptides for gene delivery as initially proposed in this thesis.

Abbreviations

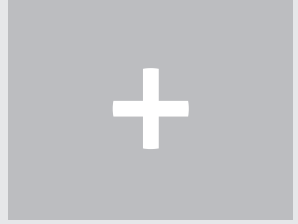
CPP: cell penetrating peptide
ds: double-stranded
hEGFR: human epidermal growth factor receptor
kDa: kilo dalton
HCS: high-content screening
IVTT: *in vitro* transcription/translation
NCL: native chemical ligation
ON-705: 2'-O-methylated splice-correcting oligonucleotide 705
ODN: oligodeoxynucleotides
pDNA: plasmid DNA
ss: single-stranded

References

(1) Glover DJ, Su MN, Mechler A, Martin LL, Jans DA. Multifunctional protein nanocarriers for targeted nuclear gene delivery in nondividing cells. FASEB Journal 2009;23(9):2996-3006.
(2) Chen X, Zaro JL, Shen W. Fusion protein linkers: Property, design and functionality. Adv Drug Deliv Rev (0).
(3) Francis DM, Page R. Strategies to optimize protein expression in E. coli. Current Protocols in Protein Science 2010(SUPPL. 61):5.24.1-5.24.29.
(4) Kawe M, Horn U, Plückthun A. Facile promoter deletion in Escherichia coli in response to leaky expression of very robust and benign proteins from common expression vectors. Microbial Cell Factories 2009;8.
(5) Peti W, Page R. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. Protein Expr Purif 2007;51(1):1-10.
(6) Noguère C, Larsson AM, Guyot J-, Bignon C. Fractional factorial approach combining 4 Escherichia coli strains, 3 culture media, 3 expression temperatures and 5 N-terminal fusion tags for screening the soluble expression of recombinant proteins. Protein Expr Purif 2012;84(2):204-213.
(7) Mangipudi SS, Canine BF, Wang Y, Hatefi A. Development of a genetically engineered biomimetic vector for targeted gene transfer to breast cancer cells. Molecular Pharmaceutics 2009;6(4):1100-1109.
(8) Li Y. Recombinant production of antimicrobial peptides in Escherichia coli: A review. Protein Expr Purif 2011;80(2):260-267.
(9) Li Y, Chen Z. RAPD: A database of recombinantly-produced antimicrobial peptides. FEMS Microbiol Lett 2008;289(2):126-129.
(10) Parachin NS, Mulder KC, Viana AAB, Dias SC, Franco OL. Expression systems for heterologous production of antimicrobial peptides. Peptides 2012;38(2):446-456.
(11) Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnology Journal 2012;7(5):620-634.
(12) Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif 2006;48(1):1-13.
(13) Kenig M, Peternel S, Gaberc-Porekar V, Menart V. Influence of the protein oligomericity on final yield after affinity tag removal in purification of recombinant proteins. Journal of Chromatography A 2006;1101(1-2):293-306.
(14) Ke T, Liang S, Huang J, Mao H, Chen J, Dong C, et al. A novel PCR-based method for high throughput prokaryotic expression of antimicrobial peptide genes. BMC Biotechnology 2012;12.
(15) Wright O, Yoshimi T, Tunnacliffe A. Recombinant production of cathelicidin-derived antimicrobial peptides in Escherichia coli using an inducible autocleaving enzyme tag. New Biotechnology 2012;29(3):352-358.
(16) Altintas I, Heukers R, Van Der Meel R, Lacombe M, Amidi M, Van Bergen En Henegouwen PMP, et al. Nanobody-albumin nanoparticles (NANAPs) for the delivery of a multikinase inhibitor 17864 to EGFR overexpressing tumor cells. J Controlled Release 2013;165(2):110-118.
(17) Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, et al. High-affinity binders selected from designed ankyrin repeat protein libraries. Nat Biotechnol 2004;22(5):575-582.
(18) Holliger P, Hudson PJ. Engineered antibody fragments and the rise of single domains. Nat Biotechnol 2005;23(9):1126-1136.
(19) Tripet B, Yu L, Bautista DL, Wong WY, Irvin RT, Hodges RS. Engineering a de novo-designed coiled-coil heterodimerization domain for the rapid detection, purification and characterization of recombinantly expressed peptides and proteins. Protein Eng 1996;9(11):1029-1042.
(20) Glökler J, Schütze T, Konthur Z. Automation in the high-throughput selection of random combinatorial libraries-different approaches for select applications. Molecules 2010;15(4):2478-2490.
(21) Lin C-, Moore PA, Auberry DL, Landorf EV, Peppler T, Victry KD, et al. Automated purification of recombinant proteins: Combining high-throughput with high yield. Protein Expr Purif 2006;47(1):16-24.
(22) Lesley SA. High-throughput proteomics: Protein expression and purification in the postgenomic world. Protein Expr Purif 2001;22(2):159-164.

(23) Heukers R, Vermeulen JF, Fereidouni F, Bader AN, Voortman J, Roovers RC, et al. EGFR endocytosis requires its kinase activity and N-terminal transmembrane dimerization motif. J Cell Sci 2013 Aug 13.
(24) Mäe M, EL Andaloussi S, Lundin P, Oskolkov N, Johansson HJ, Guterstam P, et al. A stearylated CPP for delivery of splice correcting oligonucleotides using a non-covalent co-incubation strategy. J Controlled Release 2009;134(3):221-227.
(25) Lehto T, Kurrikoff K, Langel Ü. Cell-penetrating peptides for the delivery of nucleic acids. Expert Opinion on Drug Delivery 2012;9(7):823-836.
(26) Endoh T, Ohtsuki T. Cellular siRNA delivery using cell-penetrating peptides modified for endosomal escape. Adv Drug Deliv Rev 2009 Jul 25;61(9):704-9.
(27) Wang Y, Mangipudi SS, Canine BF, Hatefi A. A designer biomimetic vector with a chimeric architecture for targeted gene transfer. J Controlled Release 2009;137(1):46-53.
(28) Chen R. Bacterial expression systems for recombinant protein production: E. coli and beyond. Biotechnol Adv 2012;30(5):1102-1107.
(29) Morello E, Bermúdez-Humarán LG, Llull D, Solé V, Miraglio N, Langella P, et al. Lactococcus lactis, an efficient cell factory for recombinant protein production and secretion. J Mol Microbiol Biotechnol 2008;14(1-3):48-58.
(30) Waegeman H, Soetaert W. Increasing recombinant protein production in Escherichia coli through metabolic and genetic engineering. Journal of Industrial Microbiology and Biotechnology 2011;38(12): 1891-1910.
(31) Mersich C, Jungbauer A. Generation of bioactive peptides by biological libraries. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences 2008;861(2):160-170.
(32) Murray CJ, Baliga R. Cell-free translation of peptides and proteins: From high throughput screening to clinical production. Curr Opin Chem Biol 2013;17(3):420-426.
(33) Alfasi S, Sevastsyanovich Y, Zaffaroni L, Griffiths L, Hall R, Cole J. Use of GFP fusions for the isolation of Escherichia coli strains for improved production of different target recombinant proteins. J Biotechnol 2011;156(1):11-21.
(34) Bailey SN, Wu RZ, Sabatini DM. Applications of transfected cell microarrays in high-throughput drug discovery. Drug Discov Today 2002;7(18):S113-S118.
(35) Wu RZ, Bailey SN, Sabatini DM. Cell-biological applications of transfected-cell microarrays. Trends Cell Biol 2002;12(10):485-488.
(36) Fujita S, Ota E, Sasaki C, Takano K, Miyake M, Miyake J. Highly efficient reverse transfection with siRNA in multiple wells of microtiter plates. Journal of Bioscience and Bioengineering 2007;104(4):329-333.
(37) Hook B, Schagat T. Reverse Transfection Using FuGENE® 6 and FuGENE® HD. Available at: http://nld.promega.com/resources/articles/pubhub/reverse-transfection-using-fugene-6-and-fugene-hd/.
(38) Avilov SV, Moisy D, Munier S, Schraidt O, Naffakh N, Cusack S. Replication-competent influenza a virus that encodes a split-green fluorescent protein-tagged PB2 polymerase subunit allows live-cell imaging of the virus life cycle. J Virol 2012;86(3):1433-1448.
(39) Kaddoum L, Magdeleine E, Waldo GS, Joly E, Cabantous S. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. BioTechniques 2010;49(4):727-736.
(40) Aubry S, Aussedat B, Delaroche D, Jiao C-, Bolbach G, Lavielle S, et al. MALDI-TOF mass spectrometry: A powerful tool to study the internalization of cell-penetrating peptides. Biochimica et Biophysica Acta - Biomembranes 2010;1798(12):2182-2189.
(41) Akita H, Tanimoto M, Masuda T, Kogure K, Hama S, Ninomiya K, et al. Evaluation of the nuclear delivery and intra-nuclear transcription of plasmid DNA condensed with μ (mu) and NLS-μ by cytoplasmic and nuclear microinjection: A comparative study with poly-L-lysine. J Gene Med 2006;8(2):198-206.
(42) Gadi J, Ruthala K, Kong K-, Park HW, Kim MH. The third helix of the Hoxc8 homeodomain peptide enhances the efficiency of gene transfer in combination with lipofectamine. Mol Biotechnol 2009; 42(1):41-48.
(43) Rajagopalan R, Xavier J, Rangaraj N, Rao NM, Gopal V. Recombinant fusion proteins TAT-Mu, Mu and Mu-Mu mediate efficient non-viral gene delivery. J Gene Med 2007;9(4):275-286.
(44) Corradin G, Kajava AV, Verdini A. Long synthetic peptides for the production of vaccines and drugs: A technological platform coming of age. Science Translational Medicine 2010;2(50).

(45) Raibaut L, Ollivier N, Melnyk O. Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 2012;41(21):7001-7015.
(46) Lindgren J, Ekblad C, Abrahmsén L, Eriksson Karlström A. A Native Chemical Ligation Approach for Combinatorial Assembly of Affibody Molecules. ChemBioChem 2012;13(7):1024-1031.
(47) Breukink E, de Kruijff B. Lipid II as a target for antibiotics. Nature Reviews Drug Discovery 2006;5(4):321-323.
(48) Lui BH, Cochran JR, Swartz JR. Discovery of improved EGF agonists using a novel in vitro screening platform. J Mol Biol 2011;413(2):406-415.
(49) Huber R, Palmen TG, Ryk N, Hillmer A-, Luft K, Kensy F, et al. Replication methods and tools in high-throughput cultivation processes - recognizing potential variations of growth and product formation by on-line monitoring. BMC Biotechnology 2010;10.
(50) Moore SJ, Leung CL, Cochran JR. Knottins: Disulfide-bonded therapeutic and diagnostic peptides. Drug Discovery Today: Technologies 2012;9(1):e3-e11.
(51) Simon M, Stefan N, Plückthun A, Zangemeister-Wittke U. Epithelial cell adhesion molecule-targeted drug delivery for cancer therapy. Expert Opinion on Drug Delivery 2013;10(4):451-468.
(52) Daugherty PS. Protein engineering with bacterial display. Curr Opin Struct Biol 2007;17(4):474-480.
(53) Biyani M, Husimi Y, Nemoto N. Solid-phase translation and RNA-protein fusion: A novel approach for folding quality control and direct immobilization of proteins using anchored mRNA. Nucleic Acids Res 2006;34(20).



Appendices

Nederlandse samenvatting

Curriculum vitae

List of publications

Dankwoord | *Acknowledgements*

Nederlandse Samenvatting

Gentherapie

Gentherapie kan worden gedefinieerd als de introductie van lichaamsvreemde nucleïnezuren in een cel met de intentie om een pathologisch proces te voorkomen, te stoppen of terug te draaien. Gentherapie vormt een aantrekkelijke strategie voor de therapeutische interventie van een brede groep ziektes, zoals erfelijke aandoeningen, stofwisselingsziektes, chronische stoornissen en kanker (1,2). Gentherapie kan worden toegepast via drie verschillende routes: gen additie/vervanging, gen modulatie/ ”knockdown” en gencorrectie/reparatie (1,3).

Genadditie/vervanging wordt gebruikt als supplement voor een ontbrekend of therapeutisch eiwit en wordt vooral toegepast om monogenetische defecten te verhelpen. Een andere toepassing is “suicide therapy”, waarbij genexpressie leidt tot geïnduceerde celdood (1). Met genmodulatie/”knockdown” wordt het afzwakken of de totale inhibitie van genexpressie door middel van interventie van transcriptie ofwel translatie bedoeld en kan bijvoorbeeld worden toegepast om het nadelige effect van een oncogen ongedaan te maken. Gencorrectie/reparatie is een methode om wild-type functies in dominant negatieve mutaties te herstellen (3). Om gentherapie via de drie verschillende routes toe te passen, kunnen verschillende soorten therapeutische nucleïnezuren worden gebruikt. Voor gen additie/vervanging kan plasmide DNA (pDNA) of messenger RNA (mRNA) gebruikt worden. Voor genmodulatie/”knockdown” wordt vooral small interfering-RNA (siRNA) gebruikt. Bij gencorrectie/reparatie kunnen enkelstrengs oligonucleotiden of kleine DNA fragmenten worden gebruikt. De verschillende therapeutische nucleïnezuren hebben ook hun eigen specifieke plek van werking binnenin de cel. Waar pDNA en oligonucleotiden of kleine DNA fragmenten voor genmodulatie in de celkern actief zijn, hebben siRNA en mRNA hun plaats van werking in het cytosol.

Om gentherapie te laten werken, moeten de therapeutische nucleïnezuren in de juiste doel cellen (de cellen die behandeling nodig hebben) worden afgeleverd en hun specifieke plaats van werking binnenin de cel, de celkern of het cytosol, bereiken. Maar door hun hoge ladingsdichtheid en hoog moleculair gewicht zijn nucleïnezuren niet in staat om cellulaire membranen te penetreren en moeten worden geholpen om hun plaats van werking binnenin de cel te bereiken (4). Om te zorgen dat de therapeutische nucleïnezuren worden opgenomen door de doel cellen zijn afgiftesystemen nodig. Maar behalve de opname van de therapeutische nucleïnezuren door de doel cellen zijn er nog meer extra- en intracellulaire barrières die moeten worden overwonnen (Fig. 1, **Hoofdstuk 2**). De afgiftesystemen, zogenaamde vectoren, moeten bijvoorbeeld de nucleïnezuren beschermen tegen degradatie, in staat zijn om te kunnen binden aan en worden opgenomen door de doel cellen en het ontsnappen

aan endosomale ophoping/afbraak. Geraffineerde afgiftesystemen die verschillende essentiële eigenschappen bevatten zijn dus nodig om de doel cellen te kunnen bereiken, om opgenomen te worden door de doel cellen en intracellulair de therapeutische nucleïnezuren op de plaats van werking af te geven.

Virale en niet-virale vectoren

Genafgiftesystemen kunnen worden verdeeld in twee categorieën: virale en niet-virale (synthetische) vectoren. Virale vectoren zijn afgeleid van virussen, in tegenstelling tot niet-virale vectoren die gebaseerd zijn op (synthetische) macromoleculen zoals lipiden en polymeren.

Ongeveer 70% van alle klinische onderzoeken met gentherapie worden uitgevoerd met virale vectoren (5). Onlangs heeft de EMA voor het eerst een viraal gentherapeutisch product alipogene tiparvovec (merknaam Glybera®) voor de behandeling van lipoproteine lipase deficiëntie toegelaten tot de Europese markt (6). Virussen zijn zeer complex en gespecialiseerd in het binnendringen van cellen om hun genetische materiaal af te leveren en de geïnfecteerde cel aan te zetten tot de productie van nieuwe viruspartikels (7). Virale vectoren zijn daardoor ook zeer efficiënt in het afleveren van therapeutische nucleïnezuren en overtreffen de niet-virale vectoren op dit gebied. Maar virale vectoren hebben ook nadelen, zoals hun immunotoxiciteit en de kans op het ontstaan van kanker bij een misplaatste integratie van het virale DNA in het genoom van de gastheercel, ook wel insertionele mutagenese genoemd (8). Hierdoor zijn niet-virale vectoren op de voorgrond getreden als mogelijk alternatief voor virale afgiftesystemen.

De meeste niet-virale afgiftesystemen zijn gebaseerd op zelfvormende complexen van nucleïnezuren met positief geladen moleculen, zoals polymeren, lipiden en peptiden. De complexen worden gevormd door middel van ladingsinteractie (9). In het algemeen zijn niet-virale afgiftesystemen beperkt immunotoxisch, zijn niet pathogeen en bezitten dus niet de nadelen van hun virale tegenhangers (10).

Verder kunnen niet-virale afgiftesystemen relatief goedkoop worden geproduceerd en kunnen zeer grootte nucleïnezuren dragen (11). En van uit een farmaceutische oogpunt zijn de verschillende componenten van niet-virale vectoren makkelijker te karakteriseren en te evalueren op veiligheid. Maar vergeleken met virale vectoren is efficiëntie van genafgifte van niet-virale vectoren laag.

Peptiden gebaseerde vectoren

Peptiden gebaseerde vectoren voor genafgifte bezitten voordelen vergeleken met vectoren gebaseerd op polymeren of lipiden. Systemen gebaseerd op polymeren of lipiden kunnen cytotoxisch zijn (12). Verder is het vaak nodig om de gebruikte polymeren of lipiden te functionaliseren met liganden voor doelgerichte afgifte, endosomale ontsnappingsmiddelen of polyethyleen glycol (PEG) om efficiënte

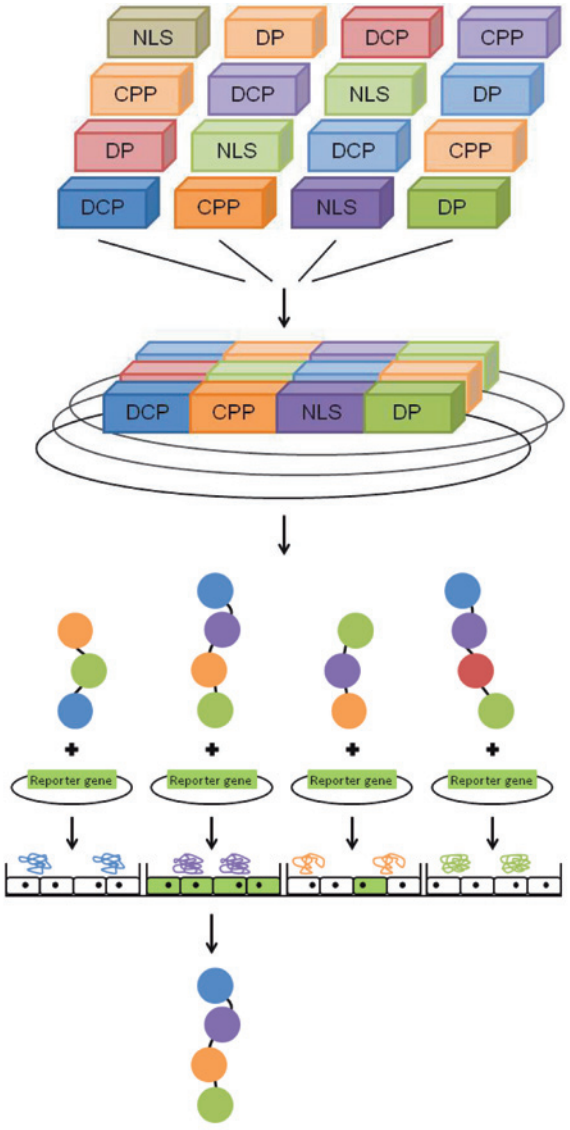
genafgifte mogelijk te maken. Reproduceerbare incorporatie en/of het vastmaken van dit soort functionele componenten is moeilijk en resulteert vaak in compositionele variaties, wat vanuit een farmaceutisch oogpunt nadelig is. Daarentegen bieden afgiftesystemen gebaseerd op peptiden een veelzijdig platform voor efficiënte genafgifte. Peptiden zijn biologisch afbreekbaar, biocompatibel en kunnen vanwege hun enorme diversiteit aan primaire en secundaire structuren vele biologische functionaliteiten vervullen, waaronder ook die functionaliteiten die nodig zijn voor genafgifte, zoals het kunnen complexeren van nucleïnezuren, het kunnen destabiliseren van cel- en endosomale membranen, en het induceren van transport door de kernporieën. Maar omdat de effectieve afgifte van nucleïnezuren voor therapeutische doeleinden een meerstaps proces is waarbij verschillende funcionaliteiten sequentieel geactiveerd dienen te worden is het noodzaak verschillende functionele componenten in een dragersysteem te verenigen. Door het combineren en samenvoegen van peptiden met verschillende functionaliteiten die nodig zijn voor efficiënte genafgifte in één enkele peptide keten kan het ideale vector worden gecreëerd. Hierdoor worden compositionele variaties geëlimineerd, wordt de farmaceutische formulering vereenvoudigd en wordt reproduceerbaarheid op moleculair niveau bereikt.

Doel

Het doel van dit proefschrift is om uit een grote bibliotheek van multimodulaire peptiden die kandidaten te selecteren die in staat zijn om therapeutische nucleïnezuren efficiënt te kunnen afleveren op de juiste plek binnenin de doel cellen.

Om dit doel te bereiken hebben we een integrale aanpak gebruikt die alle mogelijke combinaties van verschillende functionele peptiden test op geschiktheid om therapeutische nucleïnezuren efficiënt af te geven. Dit proces is schematisch weergegeven in Figuur 1. Door op een genniveau willekeurig peptiden met eigenschappen die nodig zijn voor genafgifte (o.a. DNA condenserende peptiden en cel penetrerende peptiden) te combineren, wordt er een combinatoriële bibliotheek gegenereerd die codeert voor multimodulaire peptiden. Deze bibliotheek wordt dan gebruikt voor de recombinante productie van multimodulaire peptiden op grote schaal, welke vervolgens geëvalueerd worden op hun efficiëntie van genafgifte. Meerdere evaluatie-/selectierondes worden dan uitgevoerd om de meeste efficiënte multimodulaire peptiden voor genafgifte te verkrijgen. Door gebruik te maken van deze strategie hopen we peptiden gebaseerde vectoren te verbeteren en te optimaliseren.

We hebben om verscheidende redenen gekozen voor recombinante productie van deze multimodulaire peptiden. Chemische synthese van lange peptiden (langer dan 50 aminozuren) is lastig en leidt vaak tot lage opbrengsten en heterogeen materiaal (13,14). En door gebruik te maken van recombinante eiwit synthese wordt de precisie van de cellulaire machinerie gehanteerd om reproduceerbare en exact gespecificeerde peptiden te maken (15).



Figuur 1 De voorgestelde integrale strategie om optimale combinaties van functionele peptiden voor efficiënte genafgifte te selecteren door middel van combinatoriële eiwit technieken. Functionele peptiden (DNA condenserende peptiden (DCP), cel penetrerende peptiden (CPP), nucleaire lokalisatie signaal peptiden (NLS) en doelgerichte peptiden (DP)) worden willekeurig gecombineerd op een genetisch niveau tot een combinatoriële gen bibliotheek coderend voor multimodulaire peptiden. De genbibliotheek wordt dan in een expressie plasmide ingebracht en tot expressie gebracht in een geschikte bacteriestam. Na recombinante expressie en

zuivering op grote schaal, worden de multimodulaire peptiden direct geëvalueerd op genafgifte en de meest efficiënte kandidaten geselecteerd.

Ontwikkeling van een ligatie methode om genbibliotheken te construeren

Voor de recombinante expressie van bibliotheken van multifunctionele polypeptiden voor genafgifte was het noodzakelijk eerst een combinatoriële genbibliotheken te maken. Maar ondanks dat er verschillende methoden bestaan voor de universele assemblage van genetische segmenten, kleven aan alle methoden één of meerdere nadelen. Sommige methoden introduceren zogenaamde 'littekens', waarmee de introductie van extra sequenties bedoeld wordt, of hebben voorkennis nodig van welke sequenties aan elkaar gezet moeten worden. In **Hoofdstuk 3** wordt een methode geïntroduceerd voor de sequentie onafhankelijke ligatie van meerdere gensegmenten zonder de introductie van littekens. Het grote voordeel is dat deze methode combinatorieël is, waardoor alle combinaties gemaakt kunnen worden tussen twee of meerdere genetische segmenten. De methode is gebaseerd op de ligatie van enkel- of dubbelstrengs oligodeoxynucleotiden geïmmobiliseerd op een drager. Verschillende variabelen zijn getest om de geïmmobiliseerde drager ligatie te optimaliseren. We hebben gedemonstreerd dat door middel van deze methode specifieke constructen en combinatoriële genbibliotheken kunnen worden gegenereerd.

In vitro evaluatie van peptiden gebaseerde vectoren

Om de geconstrueerde multimodulaire peptiden snel en in grote hoeveelheid tegelijkertijd te evalueren op genafgifte in cellen, hebben we een nieuwe assay ontwikkeld (beschreven in **Hoofdstuk 4**). Deze assay is gebaseerd op ' High-Content Screening' (HCS) en maakt gebruik van geautomatiseerde fluorescentie microscopie voor de kwantitatieve analyse van individuele cellen en celpopulaties. De fluorescentie microscoop maakt foto's van de cellen bij specifieke golflengtes, waardoor specifieke kleuren kunnen worden onderscheiden. De celkern van de cellen wordt blauw gekleurd om de cellen te kunnen herkennen en te tellen, zodat de toxiciteit van de gebruikte vectoren kan worden vastgesteld. Als de peptide vectoren een bepaald pDNA afleveren in de cellen, dan kleuren de cellen groen en aan de hand daarvan kan de genafgifte efficiëntie worden bepaald. Andere kleuren kunnen hierna nog worden gebruikt om andere processen in de cellen te bestuderen. De toepassing van de ontwikkelde HCS assay wordt laten zien door middel van de evaluatie van een kleine bibliotheek peptidengebaseerde vectoren. De vectoren werden tegelijkertijd beoordeeld op cytotoxiciteit, gen afgifte, inductie van cel permeabiliteit en genafgifte in niet delende cellen. We hebben gedemonstreerd dat de ontwikkelde HCS assay een waardevolle uitbreiding is op de bestaande evaluatie methodes.

Construeren van multimodulaire peptidenbibliotheken en de evaluatie van deze bibliotheken op genafgifte

Na de ontwikkeling van de methoden om genbibliotheken te construeren en om vectoren te evalueren op genafgifte in cellen, zijn er multimodulaire peptiden bibliotheken gemaakt, zoals vermeld in **Hoofdstuk 5**. Na het construeren van de genbibliotheek zijn individuele multimodulaire peptiden recombinant geproduceerd in *E.coli* en gezuiverd. De opbrengsten van de individuele peptiden waren na recombinante productie echter niet hoog genoeg om deze te evalueren op genafgifte in cellen. De lage opbrengst kan verschillende oorzaken hebben gehad. Door de positieve lading en de celpenetrerende eigenschappen van sommige peptiden zijn de deze toxisch voor de bacteriën, waardoor de bacteriën dood gaan of de peptiden afbreken en dus de opbrengst daalt. Ondanks pogingen om de recombinante productie te verbeteren (door o.a. verschillende groei media of verschillende bacterie stammen te gebruiken) is het niet gelukt om voldoende peptiden te produceren.

De fusie van een peptide aan een eiwit is een strategie die vaak gebruikt wordt om de opbrengst van recombinant geproduceerde peptiden te verhogen. In **Hoofdstuk 6** hebben we doelgerichte multimodulaire peptidenbibliotheken voor genafgifte geconstrueerd door de anti-epidermale groeifactor receptor (EGFR) biparatopische nanobody 7D12-9G8 te koppelen aan de multimodulaire peptidenbibliotheek. De EGFR is een receptoreiwit dat in hoge mate voorkomt op het celoppervlak van verschillende soorten tumoren. De nanobody 7D12-9G8 herkent en bindt aan EGFR en kan dus gebruikt worden om onze multimodulaire peptiden specifiek te richten naar tumorcellen. De multimodulaire peptiden zijn gekoppeld aan de C-terminus van de nanobody (Fig. 1c, **Hoofdstuk 6**). Na recombinante expressie en daaropvolgende zuivering van 4 individuele fusieconstructen, hebben we aangetoond dat het gemodificeerde nanobody 7D12-9G8 nog steeds in staat is om EGFR te binden en is de opbrengst 2-12,5x hoger dan zonder fusie-eiwit. De fusieconstructen zijn in staat om oligonucleotiden te binden en af te geven in HeLa pLuc/705 cellen.

In **Hoofdstuk 7** hebben we het “enhanced green fluorescent protein” (EGFP) gekoppeld aan de multimodulaire peptiden om de recombinante opbrengst te verhogen. Naast het verhogen van de opbrengst kunnen we EGFP ook gebruiken om de multimodulaire peptide vectoren te volgen in cellen met behulp van fluorescentie microscopie. EGFP (zoals de naam al zegt) is een fluorescerend eiwit met een groene kleur. De multimodulaire peptiden worden gekoppeld aan de C-terminus van het EGFP (Fig. 1c, **Hoofdstuk 7**). Na recombinante expressie en daaropvolgende zuivering van 13 individuele fusie constructen, is de opbrengst 10-30x verhoogd dan zonder fusie-eiwit. We hebben ook aangetoond dat de fusieconstructen in staat om pDNA te binden en complexeren. Helaas waren de fusieconstructen niet in staat om pDNA in COS-7 cellen af te geven.

In conclusie, in dit proefschrift hebben we geprobeerd om uit een grote bibliotheek van multimodulaire peptiden kandidaten te selecteren die in staat zijn om op een efficiënte manier therapeutische nucleïnezuren af te leveren in doel cellen. Om dit te bereiken, was een strategie gekozen die de volgende onderdelen bevat: (i) Het construeren van genbibliotheken coderend voor multimodulaire peptiden voor genafgifte, (ii) recombinante productie en zuivering van deze multimodulaire peptiden op grote schaal en (iii) de evaluatie van deze peptide op genafgifte in cellen (Fig. 1). Twee van de drie onderdelen zijn met succes uitgevoerd: (i) we hebben een nieuwe methode gecreëerd voor sequentie onafhankelijke en combinatoriële ligatie van gensegmenten zonder de introductie van littekens; (iii) we hebben een HCS assay ontwikkeld voor de gelijktijdige beoordeling van genafgiftesystemen op cytotoxiciteit, genafgifte, inductie van cel permeabiliteit en genafgifte in niet delende cellen. Het is echter niet gelukt om de recombinante productie van de multimodulaire peptiden te optimaliseren zodat hoeveelheden kunnen worden geproduceerd die noodzakelijk zijn voor verdere analyse. Ondanks dat biedt de opgedane kennis genoeg aanknopingspunten om door te gaan om de recombinante eiwit expressie te verbeteren en de grootschalige analyse van vele duizenden multimodulaire peptiden voor gentherapie doeleinden mogelijk te maken.

References

(1) Kay MA. State-of-the-art gene-based therapies: The road ahead. Nature Reviews Genetics 2011;12(5): 316-328.

(2) Kaiser J. Gene therapists celebrate a decade of progress. Science 2011;334(6052):29-30.

(3) Hsu CYM, Uludag H. Nucleic-acid based gene therapeutics: Delivery challenges and modular design of nonviral gene carriers and expression cassettes to overcome intracellular barriers for sustained targeted expression. J Drug Target 2012;20(4):301-328.

(4) Mann A, Thakur G, Shukla V, Ganguli M. Peptides in DNA delivery: current insights and future directions. Drug Discov Today 2008 Feb;13(3-4):152-60.

(5) Gene Therapy Clinical Trials Worldwide. Available at: http://www.wiley.com//legacy/wileychi/genmed/clinical/.

(6) Wirth T, Ylä-Herttuala S. History of gene therapy. Gene (0).

(7) Jang JH, Lim KI, Schaffer DV. Library selection and directed evolution approaches to engineering targeted viral vectors. Biotechnol Bioeng 2007 Oct 15;98(3):515-24.

(8) Mingozzi F, High KA. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 2011 May;12(5):341-355.

(9) Glover DJ, Lipps HJ, Jans DA. Towards safe, non-viral therapeutic gene expression in humans. Nat Rev Genet 2005 Apr;6(4):299-310.

(10) Douglas KL. Toward development of artificial viruses for gene therapy: a comparative evaluation of viral and non-viral transfection. Biotechnol Prog 2008 Jul-Aug;24(4):871-83.

(11) De Laporte L, Cruz Rea J, Shea LD. Design of modular non-viral gene therapy vectors. Biomaterials 2006 Mar;27(7):947-54.

(12) Canine BF, Hatefi A. Development of recombinant cationic polymers for gene therapy research. Adv Drug Deliv Rev 2010 Dec 30;62(15):1524-1529.

(13) Corradin G, Kajava AV, Verdini A. Long synthetic peptides for the production of vaccines and drugs: A technological platform coming of age. Science Translational Medicine 2010;2(50).

(14) Raibaut L, Ollivier N, Melnyk O. Sequential native peptide ligation strategies for total chemical protein synthesis. Chem Soc Rev 2012;41(21):7001-7015.

(15) Dimarco RL, Heilshorn SC. Multifunctional materials through modular protein engineering. Adv Mater 2012;24(29):3923-3940.

Curriculum Vitae

Markus de Raad was born on March 13th 1984 in Gouda, The Netherlands. In 2002, he graduated from secondary school at Roncalli Scholengemeenschap in Bergen op Zoom, after which he started the Bachelor's program in Chemistry at Hogeschool Utrecht. In order to receive his Bachelor degree in Research & Development he performed two internships, from January 2005 until July 2005 at the department of Veterinary Pharmacy, Pharmacology and Toxicology at Utrecht University and from November 2005 until July 2006 at the department of Pharmaceutics at Utrecht University. In 2006, he started the Master's program Drug Innovation at Utrecht University. During his Master's, he performed a 12-month research project under the supervision of Dr. Maryam Amidi and Assoc. Prof. Dr. Enrico Mastrobattista at the department of Pharmaceutics at Utrecht University, focusing on the development of antigen producing vesicles for vaccination purposes. After this, he performed a 9-month research project under the supervision of Dr. Koen Oosterhuis and Prof. Dr. Ton Schumacher at the department of Immunology at the Netherlands Cancer Institute (NKI), where he focused on DNA vaccination by DNA tattooing. In 2010, Markus obtained his Master's degree and started a Ph.D. project at the department of Pharmaceutics at Utrecht University supervised by Prof. Dr. Daan Crommelin, Prof. Dr. Peter Rottier and Assoc. Prof. Dr. Enrico Mastrobattista. He worked at a project in which peptide libraries were produced and evaluated on their capacity for gene delivery. The results of his Ph.D. project are described in this thesis.

List of Publications

Publications from this thesis

De Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E. High-content screening of peptide-based non-viral gene delivery systems. J Controlled Release 2012;158(3):433-442.

De Raad M, Kooijmans SAA, Teunissen EA, Mastrobattista E. A solid-phase platform for combinatorial and scarless multipart gene assembly. ACS Synthetic Biology 2013;2(6):316-326.

De Raad M, Teunissen EA, Mastrobattista E. Peptide vectors for gene delivery: from single peptides to multifunctional peptide nanocarriers. *Submitted for publication.*

Other publications

De Wolf HK, **De Raad M**, Snel C, Van Steenbergen MJ, Fens MHAM, Storm G, et al. Biodegradable poly(2-dimethylamino ethylamino)phosphazene for in vivo gene delivery to tumor cells. Effect of polymer molecular weight. Pharm Res 2007;24(8):1572-1580.

Amidi M, **De Raad M**, De Graauw H, Van Ditmarsch D, Hennink WE, Crommelin DJA, et al. Optimization and quantification of protein synthesis inside liposomes. J Liposome Res 2010;20(1):73-83.

Amidi M, **De Raad M**, Crommelin DJA, Hennink WE, Mastrobattista E. Antigen-expressing immunostimulatory liposomes as a genetically programmable synthetic vaccine. Systems and Synthetic Biology 2011;5(1):21-31.

Teunissen EA, **De Raad M**, Mastrobattista E. Production and biomedical applications of virus-like particles derived from polyomaviruses. J Controlled Release 2013;172(1):305-321.

van der Heijden I, **De Raad M**, van den Berg JH, Mastrobattista E, Engbersen JFJ, Haanen JBAG, Beijnen JH, Nuijen B. Histidine enhances transfection efficiency of poly(amido amine)s polyplexes for intradermal DNA tattoo vaccination. *Submitted for publication.*

Selected abstracts

de Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E., High-content screening of peptide-based non-viral gene delivery systems. Podium presentation at the Annual meeting of The Netherlands Society of Gene and Cell Therapy (NVGCT), Heeze, the Netherlands, March 2012.

de Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E., High-content screening of peptide-based non-viral gene delivery systems. Poster presentation at the European Sumposium on Controlled Drug Delivery(ESCD), Egmond aan Zee, the Netherlands, April 2012.

de Raad M, Teunissen EA, Lelieveld D, Egan DA, Mastrobattista E., High-content screening of peptide-based non-viral gene delivery systems. Podium presentation at the Annual meeting of the Controlled Release Society (CRS), Québec, Canada, July 2012.

Dankwoord | *Acknowledgements*

"'Cause I don't shine, if you don't shine"
Read my mind - The Killers

Dit is het dan. Het zit er op. Hoe vaak beginnen dankwoorden niet met "wat ging die 4 jaar toch snel voorbij" of "de afgelopen 4 jaar zijn voorbij gevlogen". En dat klopt ook. De afgelopen 4 jaar zijn gewoon snel voorbij gegaan. En het was allemaal niet gelukt zonder de hulp en steun van al die bijzondere mensen, de collega's, familie en vrienden, die betrokken waren bij mijn project. Daarom een woord van dank (en hopen dat ik niemand vergeet…)

Beste Daan, geachte promotor, ik ben zeer dankbaar dat je betrokken was bij mijn project. Je bent eerlijk, kritisch en direct en daar heeft het project (zeker als de resultaten weer tegenvielen) alleen maar baat bij gehad. Ontzettend bedankt voor alle nuttige discussies en 'moeilijke vragen' tijdens de werkbesprekingen en het kritisch doornemen van mijn manuscripten. En de aanbevelingsbrieven van Prof. Dr. Crommelin hebben zeker geholpen tijdens mijn sollicitatie voor een Postdoc positie in Berkeley!

Beste Peter, geachte promotor, ondanks dat je meer van de virussen dan van de peptiden bent, was je input in mijn project altijd zeer waardevol. Jij ook bedankt voor alle discussies tijdens de werkbesprekingen en het kritisch doornemen van mijn manuscripten.

Beste Enrico, geachte copromotor, je bent al sinds het begin van mijn Master Drug Innovation nauw betrokken bij het begin van mijn wetenschappelijke carrière. Eerst als stagebegeleider en nu als co-promotor. Bedankt voor al je enthousiasme, optimisme, creativiteit, al je hulp bij het schrijven en de gezelligheid! Ik kon altijd bij je binnen lopen voor advies of een hart onder de riem. En ook jij bedankt voor je hulp met het vinden van mijn Postdoc!

Tijdens de afgelopen 4 jaar heb ik met verschillende mensen mogen samenwerken. Zonder jullie had dit boekje er zeker anders uitgezien.

Beste Erik T., op papier werkten we op hetzelfde project (jij van boven en ik van beneden;), maar in de praktijk waren het eigenlijk twee verschillende projecten.

Ondanks dat hebben we elkaar zeker kunnen helpen. Bedankt voor al je hulp, input en gezelligheid op het ML-I lab, tijdens de werkbesprekingen en buiten het lab. Veel succes met de laatste lootjes!

Dear David and Daphne , thank you for all the help setting up the HCS assay. David, thanks for all the hours behind the computer staring at cells and programming the algorithm. En Daphne, bedankt voor het meer dan eens redden van mijn platen uit de klauwen van de robot;).

Beste Paul en Raimond, ik wou ook aan de nanobodies en dat heeft geresulteerd in een leuke samenwerking en Hoofdstuk 6. Bedankt voor al jullie input en hulp!

Tijdens mijn AIO periode heb ik een aantal studenten mogen begeleiden: Honayda, bedankt voor al je inzet tijdens de eerste paar maanden van mijn project. Sander, dank voor al je hulp met het op poten zetten van Hoofdstuk 3. Jouw 'gouden handjes' zijn goed van pas gekomen en hebben geresulteerd in een mooi co-auteurschap. Ben ook benieuwd naar jouw boekje! Nattida, dank je voor al positivisme op het lab! Natasa, thanks for all your hard work and effort on the BiPar project. And for all the Greek lessons;). Good luck in Crete with your Ph.D. and ευχαριστώ πολύ!

En dan de paranimfen… Beste Roy, al waren onze projecten totaal verschillend, ik heb toch een hoop van je kunnen leren op het lab. Ik bewonder hoe jij iedereen altijd graag helpt, voor iedereen klaar staat en hoeveel je voor anderen doet/hebt gedaan. Om je eigen woorden te gebruiken, je bent een ontzettend goede gast! Dank voor alle gezellige sappies/biertjes binnen en buiten het lab. En onze trip naar Quebec en New York zal ik niet snel vergeten! We gaan elkaar zeker tegenkomen aan de andere kant van de Atlantische oceaan!

Beste Jo, klein broertje, super dat je mijn paranimf wil zijn! Heb ik je toch een beetje nageaapt , door het pad van de farmacie te kiezen. Zonder jou had ik waarschijnlijk niet bij Biofarmacie stage gelopen en was ik dus waarschijnlijk ook nooit gepromoveerd. Ik bewonder hoe je eigen weg hebt gevonden in de farmaceutische wereld en hoe goed je aan de weg timmert bij Janssen. En een kleine update betreft onze 'publicatiestrijd', we staan momenteel gelijk maar ik ga daar zeker verandering in brengen;).

Ik blijf steeds maar weer terugkeren bij Biofarmacie. Na meerdere stages vond ik het nodig om daar nog eens 4 jaar aan toe te voegen door te gaan promoveren (het is jullie niet gelukt om mij buiten de deur te houden;). Het is bij Biofarmacie gewoon super gezellig, er is een top werksfeer en voor het menig vertier wordt goed gezorgd!

Beste Wim, ik wist niet dat 020-suporters ook aardige mensen kunnen zijn;). Ik heb zeer veel respect voor je harde werken en interesse in al de AIO's die rond lopen. Bedankt voor de nodige gezelligheid op de borrels, de CRS en natuurlijke alle filosofielessen…

En wat kan een afdeling zonder een topteam van analisten? Helemaal niks. Mies, bedankt voor je hulp met de SEC en je gezelligheid! Roel, dank voor al je hulp in de ML-I en ML-II labs. Je hebt me geïntroduceerd in de wondere wereld die recombinante eiwit productie heet. Maar ook voor de celkweek tips & trics en alle biertjes in Z700 (en daar buiten)! Ebel (en Willemiek), dank voor de *in vivo* lessen. En ook Georgi, Louis, Kim en Joep, bedankt voor alle hulp op de labs.

Dear Isil, thanks for the great times we had in and outside the lab. I admire your focus and how you manage to do all the things your own way. And maybe you made me a little more punctual with your lunch regime;). Sorry for destroying your view on dolphins and hopefully I will see you in San Francisco one day!

Beste Kristel, jij bedankt voor al je gezelligheid in het Went en DdW! Of het nou op het lab, tijdens het bierproeven of op 't Bokbierfestival is, je bent altijd vrolijk, dank daarvoor! Ik ben bewonder jouw positivisme en hoe je al je samenwerkingen en projecten managet. En straks ben jij aan de beurt, ik ben zeer benieuwd naar jouw boekje.

En dan het fameuze UMC lunch team, Roy, isil, (frontcook) Kirstel, Sietze, Niels, Erik O., Jos en Meriem. Mijn lunch ervaring is voor altijd veranderd. Dank voor alle levendige discussies en lunch pret!

Thank you '2009 people': Roy, Peter, Isil, Afrouz, Negar, Erik T., Audrey, Filis, Neda and Kimberly.
Thank you to all my predecessors, who showed how to defend a thesis and how to party: Albert, Marina, Rolf, Amir G, Pieter, Holger, Niels, Joost (ook bedankt voor je hulp/begeleiding en de prettige tijd bij B3), Ellen, Joris, Ethlinn, Bart, Maria (thank you both for a great time in NYC!), Marcel, Karlijn (bedankt en alvast bedankt voor de geweldige tijd in SF!), Twan, Frank, Hajar, Roberta, Sabrina, Inge, Laura, Melody and Emmy.
And thank you to all the others: Yang, Erik O., Iris, Sima, Steffen, Maryam, Barbara, Raymond, Susan, Herre, Tina, Robbert Jan, Gert, Huub, Andhyk, Mazda, Martin, Paul, Rachel, Farshad, Anne, Vera, Anastasia, Herman, Oil, Orn, Mehrnoosh, Neda, Louann, Dandan, Yinan, Shima, Anna, Reka, Merel, Negar, Burcin, Jan Jaap and Eduardo.

Zeker in het begin van mijn AIO carrière was het iedere dinsdag weer stressen… stipt om half zes op de fiets richting Starsound… oefenen met de band. Een mislukt experiment was snel weer vergeten in de oefenruimte. Marijn en Lars, bedankt voor alle mooie momenten en jullie vriendschap!

Mijn allereerste stapjes in Utrecht en in het lab heb ik gezet in de F.C. Donderstraat. Alle vrienden van Uranymus, bedankt voor alle gezellige biertjes tijdens de kampen, in de Poli, in Neudezicht/'t College/de Witte Ballons/de Stad, LowLands en overige feestjes in het verleden en in de toekomst!

Bovenste Beste Boykes (jullie weten wie ik bedoel), we zijn druk met van alles en nog wat en zien elkaar misschien wat te weinig, maar als we bij elkaar zijn, is het als een warm bad en is alles één groot feest. Ontstressen gaat zeer goed bij jullie. Bedankt voor alle steun, gezelligheid, pinte pakke en het vele lachen! De Snor heeft gesproken.

Lieve familie, dank voor alle support door de jaren heen. Hopelijk is het nu na al die jaren wat duidelijker wat ik heb uitgespookt. En anders leg ik het graag nog een keer uit. And for a change, it's all in English;).

Lieve ouders, jullie wisten waarschijnlijk zo'n 29 jaar geleden al dat ik ging promoveren bij Biofarmacie. Dat krijg je als je een zoon bent van een Bioloog en een Apotheker. Hoe cliché ook, woorden kunnen niet beschrijven hoe dankbaar ik ben voor alle steun en liefde door de jaren heen. Zonder jullie was dit zeker niet gelukt. Bedankt voor alle adviezen, support en wijze raad die ik kreeg tijdens mijn onderzoek.

Last but zeker not least, Claire! Ik moet nog steeds wennen aan het feit dat je mijn vrouw bent en niet meer mijn vriendin, maar 14 september was geweldig en om nooit te vergeten! Ook bij jou is van toepassing dat woorden te kort schieten… Dank voor al je begrip, steun en zorgzaamheid tijdens mijn Ph.D.! Zonder jouw had ik het veel zwaarder gehad. Bedankt voor de momenten dat je er was als ik je nodig had. En nu gaan we samen aan een nieuw avontuur beginnen. Maar zoals altijd, met ons tweeën komt alles goed! Dikke kus en Love you!!!