New methods for developing (bi)cyclic peptides by phage display

PAR

Camille VILLEQUEY

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

Acknowledgements

Foremost, I would like to deeply thank my thesis director, Professor Christian Heinis. Thank you for accepting me as part of the LPPT and giving me the opportunity to take a deep dive into the field of peptide chemistry which was new to me. Discovering this area of science under your supervision has been a great learning experience that I have really enjoyed. Thank you for your guidance and for contributing to the scientific researcher I became.

I would like to thank Dr. Steven Dunn, Prof. Oliver Hantschel, Prof. Oliver Hartley and Dr. Horst Pick for being part of the jury and evaluate my work.

I am also very grateful to our secretary Beatrice for all the administrative help but also for her kindness and support when I was anxious. I always felt more relaxed after speaking with you.

Huge thanks to all past and present members of the group for making the lab such a fun place to work. In particular, thanks to the old generation (Davide, Charlotte, Inma, Lisa, Mike, Philippe, Raphael, Ranga, Silvia, Vanessa, Yuichi). You made me feel immediately comfortable in the lab and I appreaciate all the time you spent teaching me how to be a better scientist. From this period (2013-2014), I will always remember the epic events we had together such as the parties at Lisa's and Ranga's places and I am still laughing at some of our memories while writing these words. Then, I would like to thank the people who have spent most of the time with me in the lab (Christina, Jonas, Kaycie, Manh, Sangram, Simon, Vanessa, Xudong). A special mention goes to Sangram and Xudong who have greatly contributed to two of my projects. You are extremely talented scientists and I have been so lucky to work together with you. I am also very grateful to Vanessa and Jonas for taking the time to proof-read my thesis although you already had so much to do. I also had memorable days and weekends with this second generation of LPPT and I am very grateful to Davide and Vanessa for organizing one of the greatest weekend of my PhD (wine weekend in Piemont)! Now a new generation has arrived at LPPT (Carl, Joao, Ganesh, Patrick, Sevan, Yuteng) and the best I can wish to you guys is to have an as good experience as I had in this lab! Finally, I would like to thank the "permanent scientist" of LPPT, Ale, who could have appeared in any of the three categories. I know you will hate

me after reading this line but I also know you have a good sense of humour and that is why it is always so nice to speak and joke with you. I had four great years with you and you always gave me valuable suggestions and opinions on my projects.

Additional thanks go to the students that worked with me during my PhD: Adrian, Pierre, Isabel, Lisa and Peter. I hope you learned and enjoyed working with me as much as I learned and enjoyed working with you.

I also would like to thank members of our neighboring labs such as LIP and LCBIM for the fun BBQ-madness at the lake, the unforgettable raclette parties and other events we shared together.

This work would not have been possible without all my out-of-the-lab friends! A special thanks goes to (Professor) Ric: my time here would not have been the same without the amazing nickname you invented for me (hum hum)! I also would like to tell to my friends from ESBS (Agathe, Anaïs, Fleur, Katha, Laura, lea, Max, Virginie) how cool and nice it has been to travel around Europe and Cambodia to visit you. I had really great weekends with you and I am looking forward to the next ones.

Doing this PhD would also not have been possible without the support of my mum and dad as well as the rest of my family. I feel really lucky that you gave me the opportunity to study what I liked and always encouraged me to do my best. Although you have gotten a very pessimistic daughter, you were always supportive in the difficult moments and I knew I could count on you.

Finally, last but non least, thank you Jonas for making my life so much more exciting and beautiful everyday. Thank you for believing in me but even more in us considering our unusual situation. When everything was dark, you were the light that kept me going and I will miss that so deeply when I go to Paris. But now that we have learned how to work and live side by side, I hope it will be easy to cope with some few kilometers more and enjoy our future adventures in new places!

Lausanne, le 20 Septembre 2017

Abstract

Cyclic peptide ligands are a promising molecular format for the development of therapeutics. They combine some of the advantages of large protein therapeutics (high affinity and specificity) and of small molecule drugs (accessibility to chemical synthesis and low immunogenicity). With phage display, large combinatorial libraries of more than one billion mono- and polycyclic peptides can be developed and screened against virtually any target. During phage affinity selections, the physical link between phenotype (encoded peptide) and genotype (phage DNA) allows the identification of enriched peptide binders by sequencing of the phage vector DNA. Based on these concepts, our laboratory has established a robust chemistry in which phage displayed peptides containing two or three cysteines are cyclized or bi-cyclized with thiol-reactive linkers. This procedure has been generally applied for the isolation of potent and selective (bi)cyclic peptide ligands with therapeutic potential. The isolation of binders with optimal properties is highly dependent on the possibility to generate (bi)cyclic peptide libraries with high structural diversity and on the ability to thoroughly decode the phage selection output. In my PhD work, I developed new methods to generate large and structurally diverse libraries and to decode more efficiently phage selected peptides.

In my first project, I explore the cyclization of peptides with two chemical bridges as a method to provide rapid access to phage-encoded libraries with high scaffold diversity. To this end, a range of twelve structurally diverse chemicals were tested for thiol-reactivity and phage compatibility. Subsequently, peptide libraries of the format XCX_nCX_nCX (X = random amino acid, C = cysteine, n = 3 or 4) were cyclized with six of these chemical linkers and panned against the protease plasma kallikrein. The selection yielded inhibitors with remarkable affinity (sub-nM K_I), target selectivity (>1000-fold) and proteolytic stability ($t_{1/2}$ in plasma > 3 days) despite a relatively small molecular mass (~ 1200 Da).

In the second project, I developed phage-encoded libraries of small cyclic peptides in order to identify ligands that have the potential to be orally absorbed. A major challenge for the generation of such library is the usually low diversity of compounds due to the limited number of amino acids. In this work, we addressed this limitation by taking advantage of a recently discovered peptide

cyclization reaction in which the thiol of a cysteine is ligated with the N-terminal amino group using a chemical reagent. Unlike other cyclization strategies involving bridges between two amino acid side chains (e.g. two cysteines), this method maximizes the ratio between the number of target-interacting amino acid side chains and the number of peptide bonds. We applied this strategies to two phage-encoded peptide libraries of four and five amino acids (XXXC and XXXXC) that were cyclized with eight chemical linkers yielding a collection of more than 1.3 million compounds. Panning the libraries against the two disease targets plasma kallikrein and coagulation factor XIa yielded ligands with single-digit micromolar affinity.

In a third project, I developed a strategy to bypass the bacterial infection step in phage display. Bacterial infection is usually crucial for the propagation and identification of enriched clones by DNA sequencing. However, mutations or chemical modifications of the phage coat proteins, which are commonly used in phage-encoded cyclic peptide libraries, sometimes lead to decreased infection rates and therefore compromise the entire procedure. Consequently, there is an interest for developing methods in which this step could be completely omitted. Using next generation sequencing, I could prove that essentially all phage particles of a sample could be decoded by directly sequencing DNA from phage particles. In addition, the coverage of sequences was comparable to the one obtained when plasmid DNA from cells infected with wild-type phage was sequenced.

Key words: Phage display, peptide library, bicyclic peptides, macrocycles, next-generation sequencing.

Résumé

Les peptides cycliques sont un format prometteur de molécules thérapeutiques. Ils allient à la fois les avantages des protéines thérapeutiques, à savoir une excellente affinité et spécificité pour leurs cibles, et les propriétés favorables des petites molécules puisqu'ils sont facilement synthétisés et sont peu immunogéniques. En utilisant la technique du phage display, de larges banques contenant plusieurs milliards de peptides peuvent être créées et criblées sur des cibles d'intérêt. Pendant les sélections de phage, il existe un lien physique entre le phénotype (le peptide encodé) et le génotype (l'ADN codant du phage) ce qui permet l'identification des ligands peptidiques sélectionnés et enrichis grâce au séquençage de leur ADN. En se basant sur ces concepts, notre laboratoire a établi une réaction chimique avec laquelle les peptides exposés à la surface des phages peuvent être cyclisés. Ces peptides contiennent généralement deux ou trois cystéines à des positions fixes qui peuvent réagir avec un réactif de liaison bivalent ou trivalent. Cette procédure a été utilisée pour l'isolation de peptides (bi-)cycliques ayant une bonne affinité et un fort potentiel thérapeutique. La capacité de notre laboratoire à isoler des ligands ayant des propriétés optimales est fortement dépendante de notre faculté à produire des banques contenant des composés ayant une large diversité de structures. Un autre paramètre important consiste en notre habilité à parfaitement décoder les séquences d'ADN des peptides sélectionnés. Au cours de mon doctorat, j'ai travaillé sur ces deux aspects de la technologie du phage display au travers de trois projets qui sont brièvement décrit dans les paragraphes suivants.

Un premier objectif de ma thèse était d'évaluer une nouvelle façon de cycliser les peptides avec deux ponts chimiques. Cette méthode a le potentiel de générer rapidement de large banques de peptides cycliques ayant une grande diversité de structures. Pour ce faire, une sélection de douze réactifs de cyclisation ont été testés afin de déterminer leur réactivité avec les groupements thiols et leur compatibilité avec les phages. Par la suite, des banques de peptides, exposés sur des phages et disposant d'un format XCX_nCX_nCX (avec X = acide aminé randomisé, et n = 3 - 4), ont été cyclisées avec six de ces réactifs de liaison puis criblées sur une cible d'intérêt, la protéase plasma kallikrein. La sélection de phage a mis à jour des ligands avec une affinité excellente (jusqu'à moins d'un

nanomolaire), une bonne sélectivité pour la cible (> 1000 fois) et une borne résistance à la protéolyse ($t_{1/2}$ > 3 jour dans du plasma) bien qu'étant de faible poids moléculaire.

Dans mon deuxième projet, j'ai développé des banques de petits peptides cycliques encodés par des phages afin d'identifier des ligands ayant le potentiel d'être absorbés oralement. Un défi majeur pour la création de telles banques reste généralement leur faible diversité due au nombre limité d'acides aminés. Nous avons remédié à ce problème en profitant d'une nouvelle réaction de cyclisation récemment implémentée par notre laboratoire. Elle fonctionne via la connexion d'un réactif de liaison au groupement thiol d'une cystéine située en aval du peptide, suivie d'une deuxième réaction avec le groupement amine du N-terminus. A la différence d'autres stratégies de cyclisation basées généralement sur des ponts moléculaires entre deux chaines latérales d'acides aminés (par exemple deux cystéines), cette approche maximise le rapport entre le nombre d'acides aminés interagissant avec la cible et le nombre de liaisons peptidiques. Nous avons appliqué cette stratégie à deux bibliothèques de phage contenant quatre ou cinq acides aminés (XXXC ou XXXXC) qui ont été cyclisées avec huit réactifs de liaison. Par conséquent, une collection de plus d'1,3 milliard de composés a été créée. Des ligands ayant une affinité proche d'un micromolaire ont été identifiés lors de sélections de phage contre la plasma kallikréine et le facteur de coagulation XIa.

Dans mon dernier projet, j'ai développé une stratégie pour contourner l'étape d'infection bactérienne habituellement indispensable au phage display. Cette étape est normalement cruciale pour la propagation et l'identification des clones enrichis lors du processus de sélection par séquençage ADN. Cependant, certaines mutations ou modifications chimiques des protéines de capside du phage, qui sont communément utilisée dans les banques de peptides bicycliques, peuvent parfois diminuer fortement la capacité des phages à infecter leur hôte et par conséquent compromettre la totalité de la procédure. Ainsi, il serait intéressant de développer une méthode avec laquelle le matériel génétique des phages pourrait être séquencé sans nécessairement avoir besoin d'être injecté dans une bactérie. En utilisant le séquençage haut débit, nous avons pu prouver que la quasi totalité des virions présents dans un échantillon donné pouvait être décodée en séquençant directement l'ADN simple brin contenu dans les phages. D'autre part, la sensibilité de la méthode est comparable à celle obtenue avec les procédures standards où l'ADN plasmidique double brin est extrait des bactéries infectées par les phages.

Mots clés : phage display, peptides bicycliques, macrocycles, librairie de peptides, séquençage haut-débit.

Contents

List of abbreviations

ACN	Acetonitrile
aPTT	Activated partial thromboplastin time
BSA	Bovine serum albumin
bp	Base pairs
CDR	Complementarity-determining region
CsA	Cyclosporin A
Da	Dalton
DCM	Dichloromethane
DIEA/DIPEA	N,N-diisopropylethylamine
DMF	N,N-Diméthylformamide
DMSO	Dimethylsulfoxyde
ssDNA	Single-stranded deoxyribonucleic acid
EDT	Ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicine Agency
ESI-MS	Electrospray ionization mass spectometry
F	Oral availability
Fc region	Fragment crystallizable region
FDA	Food and Drug Administration
Fmoc	Fluorenylmethyloxycarbonyl
FXIa/FXIIa	Activated coagulation factor XI/XII

HBD	H bond donor
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt	Hydroxybenzotriazole
HPLC	High pressure liquid chromatography
IC_{50}	Half maximal inhibitory concentration
i.v.	Intra-veinous
K_d	Dissociation constant
K_I	Inhibition constant
K_M	Michaelis-Menten constant
LC-MS	Liquid chromatography mass spectrometry
mAb	Monoclonal antibody
MW	Molecular weight
NGS	Next generation sequencing
NMR	Nuclear magnetic resonance
pIII/pVIII	phage coat protein III/VIII
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PK	Plasma kallikrein
PT	Prothrombin time
(m or t)RNA	(messenger or transfer) ribonucleic acid
Rt	Room temperature
SAR	Structure activity relationship
SPPS	Solid-phase peptide synthesis
$t_{1/2}$	Half-life
TATA	1,3,5-triacryloyl-1,3,5-triazinane

TBAB   N,N′,N″-(benzene-1,3,5-triyl)-tris(2-bromoacetamide)

TBMB   1,3,5-tris(bromomethyl)benzene

TCEP   Tris(2-carboxyethyl)phosphine

TFA   Trifluoroacetic acid

uPa   Urokinase-type plasminogen activator

1. Introduction

1.1 – Cyclic peptides therapeutics

1.1.1   Properties of cyclic peptide therapeutics

Peptides are often defined as chains of approximately 3 to 50 amino acids, which is the commonly accepted boundary between proteins and peptides[1]. In recent years, peptide therapeutic have gained increasing attention from the pharmaceutical industry due to their potent and specific mode of action as well as their low inherent toxicity. Today, around one hundred peptide therapeutics are used in the treatment of a large variety of diseases. Interestingly, more than 40 of the approved peptide-based drugs are cyclic or polycyclic.

Several methodologies have been used to cyclize peptides such as head-to-tail, head-to-side chain, side chain-to-tail, or side chain-to-side chain[2] (Figure 1). In most cases, head-to-tail cycles are generated through amide bond or ester bond formation. By contrast, a wide range of chemistry can be applied for cyclizations involving side chains as for example disulfide bridging of cysteines, thioether bond formation, lactamization, etc. Moreover, innovative cyclization methodologies have also enabled the generation of polycyclic peptides[3].
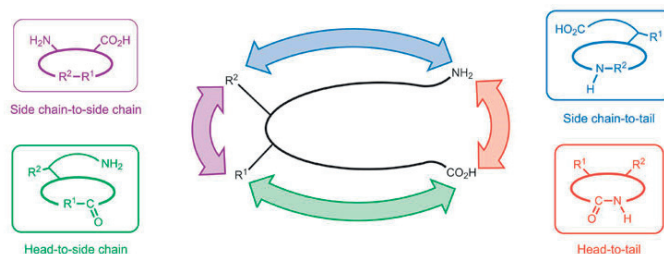
Figure 1. The four possible synthetic approaches used to constrain peptides into macrocycles: head-to-tail, head-to-side chain, side chain-to-tail or side chain-to-side chain. Adapted with permission from[2], reprinting license nr. 4047091488316, copyright © 2011 Nature Publishing Group.

Cyclic peptides offer numerous advantages in therapeutic applications compared to linear peptides[4]. First, their limited conformational flexibility reduces the entropic penalty upon binding which leads to higher binding affinity. Similar to proteins and antibodies, they can potentially bind to their targets with subnanomolar affinities. This can also be attributed to their intermediate size of 0.5 to 5 kDa that affords large surface of interaction with their target. These specific contacts also minimizes the risk of off-target side-effects, which is considered one of the greatest advantage of peptides over small molecules. Another advantage of cyclic peptides is that the scaffolding of

3

macrocycles can lock them into a particular shape or secondary structure that can favor the interactions with specific binding surfaces[5]. For example, peptide stapling is an approach that consists of locking short peptides in their bioactive alpha-helical conformation by covalently linking the side chain of two amino acids[6]. This has been particularly useful for targeting protein–protein interactions that are mediated by α-helices, enhance their cellular uptake and their resistance towards proteolytic degradation. Finally, macrocyclization renders peptides less sensitive to proteolytic cleavage by reducing the likelihood of cleavage by both endo- and exo-peptidases[7]. This has the effect of increasing their stability in systemic circulation as compared to linear peptides.

Apart from these differences, cyclic peptides also share all the favorable properties of their linear counterparts (Table I). They are associated with lower cost of production compared to antibodies and other protein-based therapeutics[8]. Over the second half of the 20th century, the generation of therapeutic peptides has been revolutionized by the emergence of solid phase chemical synthesis allowing the automated generation of peptides with high yields. Today, the method has evolved into a major technique and is compatible with thousands of amino acid building blocks, coupling reagents, etc., which facilitates complex peptide modifications, conjugations, or cyclizations. Furthermore, due to their small size and their biochemical nature, peptides display a lower risk of immunogenicity than large biologics and are usually not associated with the accumulation of harmful metabolites in the body as observed with small molecule drugs[9]. Finally, several studies have pointed to a correlation between smaller ligand size and greater tissue penetration, which offer peptides another advantage over protein-based therapeutics[8].

properties	small molecules	cyclic peptides	biologics/anti-bodies
size[10]	< 0.5 kDa	0.5 to 5 kDa	5 Da to 200 kDa
target specificity[11]	low	high	high
typical potency[11,12]	nM to μM range	pM to nM range	fM to nM range
membrane permeability[11,13]	good	rare	non-existent
oral availability[9,13]	usually good	usually poor	never observed
accumulation of harmful metabolites[14]	frequent	rare	rare
immunogenicity[9]	rare	rare	frequent
production costs[15]	rather low	rather low	high

Table I. Summary of the strength and weaknesses of cyclic peptides as compared to small molecules and biologics.

Despite these demonstrated qualities, several challenges remain in the development of cyclic peptide drugs. Poor oral bioavailability is certainly one of their main limitations[10]. This can be imputed to the fact that peptides often are not resistant to pancreatic proteases or membrane

permeable (Table I). Today, several approaches are being evaluated to increase pancreatic stability as well as the diffusion or active transport of peptides through the gut epithelium. These include the use of smaller and more constrained peptide formats, N-methylation of the peptide backbone, as well as innovative formulation strategies[13]. Another problem is that intraveinously (i.v)-injected peptides undergo a fast renal clearance due to their small size and usually have a short half-life[16]. Methods for increasing the *in vivo* half-life of peptides and proteins have been extensively studied in the past few decades. These strategies include fusion or conjugation to serum proteins[17–20], non-covalent binding to serum proteins[21–23], or conjugation to polymers with large hydrodynamic radii[24,25].

1.1.2   Examples of cyclic peptide therapeutics

Owing to their excellent properties, more than 40 cyclic peptides have made their way to the therapeutic drug market[26]. The vast majority of them are directed against extracellular targets because of the typically poor cell permeability[27]. Among these, nearly all are derivatives of natural products that are indicated for the treatment of a variety of pathologies including infections, cancer and metabolic diseases. Although most of the approved cyclic peptide therapeutics are monocyclic as for instance the oncology drug octreotide and the immunosuppressant cyclosporine, a handful of polycyclic peptides are also used in the clinic. Some successful examples of these two categories will be discussed in the following paragraphs.

The oncology drug octreotide and its derivatives (lanreotide, vapreotide and pasireotide) are a successful family of monocyclic peptide therapeutics. Indeed, octreotide (sandostatin, Novartis) is one of the best-selling cyclic peptide drug as it accounts for $1.6 billion of sales in 2016[28]. Octreotide, which was first launched in 1988, is indicated for the treatment of acromegaly, carcinoid tumors and other types of gastrointestinal and pancreatic tumours[29]. It is a disulfide cyclized octapeptide that acts by blocking the release of hormones such as the growth hormone via the inhibition of G-protein-coupled somatostatin receptors. As often experienced with cyclic peptide drugs, octreotide was shown to have a modest half-life of 90 to 120 min after subcutaneous administration. However, improved formulations of octreotide and its derivative lanreotide were recently introduced as for example octreotide long-acting release, lanreotide sustained release and lanreotide autogel[30]. Because of these innovations, a single administration every 14 or 28 days is required.

Cyclosporin A (CsA) is another famous example of monocyclic peptide-based drug with exceptional pharmacokinetic parameters. The molecule was first isolated in 1970 from the fungus *Tolypocladium inflatum*. The immunosuppressive effect of CsA was discovered one year later in a

screening test on immune suppression performed by Hartmann F. Stähelin at Sandoz[31]. Currently in clinical use for the treatment of various pathologies (organ transplants management, rheumatoid arthritis, keratoconjunctivitis sicca etc.), CsA is also one of the strongest segment of the peptide market with $1.9 billion annual sales in 2015[32]. It is well known for having a remarkable oral availability estimated between 10% and 60%[13]. Although the formulation clearly had an effect on its bioavailability, several structural properties of CsA were shown to affect its absorption as for example the cyclic backbone, the presence of N-methylated amino acids, the lipophilicity or the existence of four hydrogen-bonded amide NH protons. These features have considerably inspired the *de novo* design of orally available cyclic peptides.

Polycyclic peptides have also been evaluated as potential drug candidates. For instance, Vancomycin and its derivatives (telavancin, dalbavancin and oritavancin) are polycyclic antibacterials broadly applied for the treatment of gram positive infections[33]. By binding to the D-Ala-D-Ala target, these drugs block the transpeptidation of peptidoglycan precursors in the gram positive bacterial cell wall. In particular, the ability of vancomycin and its derivatives to interact with a molecule as small as a dipeptide underlined the importance of the high conformational constraint imposed by the three macrocyclic rings. Due to their efficacy against methicillin-resistant Staphylococcus Aureus (MRSA), they have been utilized for many years as "last resort" therapies for patients that were not responsive to other antibacterials. However, this trend is now changing because of the appearance of new resistant strains that necessitate novel therapeutic approaches such as the use of polymixins, which are another class of cyclic peptides.

While the structure of vancomycin is not exclusively peptidic, polycyclic peptides composed of only natural amino acids have been exploited in the clinic. For instance, linaclotide is a 14-amino acids peptide cyclized through three disulfide bonds that is currently used by patients with chronic constipation and irritable bowel syndrome. As a result of its constrained conformation, the peptide is highly resistant to proteolysis and can be orally administered while not orally available. Indeed the target of linaclotide is guanylate cyclase C, an enzyme located on the surface of intestinal epithelial cells. By activating its target, the drug triggers a signaling pathway that generate abundant fluid secretions into the lumen, thereby increasing colonic transit.

1.2 – In vitro evolution of cyclic peptides

1.2.1   In vitro evolution techniques

In the last ten years, several cyclic peptides that reached clinical trials were developed *de novo* using methods based on *in vitro* evolution[26]. This involves the generation of combinatorial libraries whose size and structural diversity are crucial in order to select lead candidates with desirable properties. Typically, *in vitro* evolution of cyclic peptide follows three different steps[34,35]. First, a combinatorial library of peptides is created and cyclized. For this step, the development of methods that maintain a one-to-one linkage between the phenotype (cyclic peptides) and the genotype (encoding sequence of DNA or mRNA) has greatly contributed to the development of large collections of compounds containing several trillion (10^{12}) of variants. Second, the encoded repertoire of peptides undergoes a Darwinian selection process in which molecules with the most desirable properties (in most cases, affinity, specificity or stability) are isolated. Ultimately, the sequence of the selected peptides are determined using DNA sequencing strategies such as Sanger or Next Generation Sequencing (NGS). In the following section, some examples of successful approaches for the *in vitro* evolution of cyclic peptide are presented. In the interest of brevity, the discussion will be restricted to technologies that have brought cyclic peptides into clinical trials; namely phage display and mRNA display.

Phage display is a technique that appeared in the mid-80's and was the first *in vitro* display system to be established[36]. A short introduction of the method is given in this paragraph. However, the principles of phage display as well as the associated selection of peptide macrocycles are further described in the section 1.2.2. The most commonly used procedure relies on filamentous phage display such as M13 or fd. The libraries are generated by means of plasmidic collections that are transformed into phage competent *E.coli*. Peptide-tagged phage are subsequently assembled, released from bacteria and subjected to affinity selection against an immobilized target (Figure 2). The size of the libraries is usually limited by the transformation efficiency, resulting in diversities comprised between 10^8 and 10^{10} variants[35]. In the case of cyclic peptides, the most common approach is to encode peptides bearing an even number of cysteines that are either oxidized[37] (Figure 3a) or cyclized with a chemical linker[38,39]. Peginesatide is the only cyclic peptide developed by phage display that has been approved by the Food and Drug Administration (FDA)[26]. It consists of a 45 kDa PEGylated erythropoietin mimicking peptide approved for the treatment of chronic kidney disorder associated anemia[40,34]. The drug was withdrawn from the market in 2012.
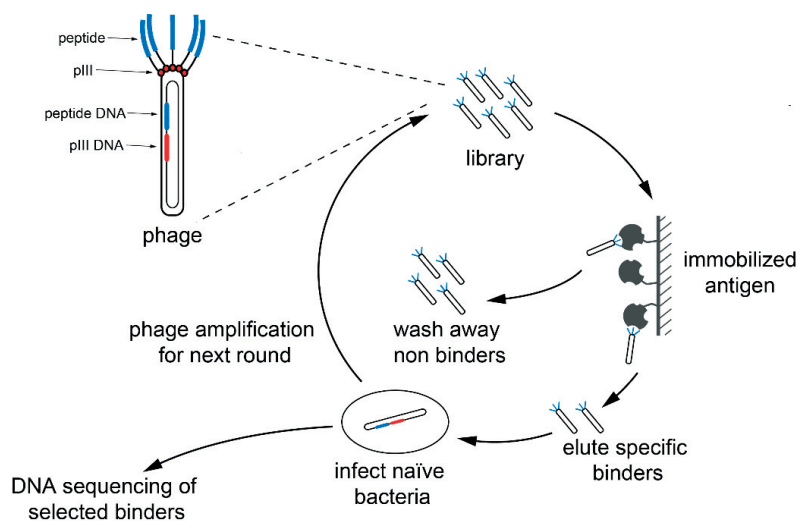
7

Figure 2. Schematic representation of a standard cycle of pIII phage display selection. DNA sequences encoding combinatorial libraries of peptides are cloned into a phage vector, in frame with phage coat protein III (pIII). Peptides are then displayed in 5 copies at the surface of the phage particles as pIII fusions. The phage-encoded library undergoes an affinity-selection against an immobilized target. Isolated binders are eluted and re-amplified in host *E.coli* cells. The procedure can be repeated several times before sequencing and identification of the selected clones.

As already mentioned, the production of large phage-encoded libraries is strongly dependent on cell transformation efficiency which limits their size to a few billion compounds. In comparison, translation of the peptides *in vitro* circumvents this problem and enables larger diversity of 10^{13} to 10^{14} molecules[35,41]. In order to maintain the nucleic acid encoding, the linkage between phenotype and genotype must then occur directly during translation. Such a strategy has been implemented for the display of a peptide on its coding mRNA (Figure 4). To do so, puromycin, a broadly used antibiotic that causes premature chain termination during ribosomal translation, is covalently attached to the 3'- end of an mRNA strand. The peptide library is then ribosomally synthesized using a reconstituted *in vitro* translation system. At the end of the translation, puromycin is able to react with the peptide at the stop codon position which establish the covalent linkage between the mRNA and displayed peptide. The generated library can then be subjected to several rounds of affinity selection against an immobilized target and hits are identified by reverse-transcription of the RNA and consecutive DNA sequencing.

Several cyclic peptide formats have been selected by mRNA display. For instance, Robert *et al.* efficiently cyclized mRNA-encoded peptide by reacting the two α-amino groups of a lysine and of the peptide N-terminus with disuccinimidyl glutarate (DSG)[42] (Figure 3b). The library was then used in a selection against the signaling protein Gαi1 in which a binder with low nanomolar dissociation constant (K_d = 2.1 nM) was identified[43]. In order to evolve ligands with better serum stability, they

also tried to apply a proteolytic pressure prior to affinity selection[44]. With this approach, they isolated a cyclic peptide with around 100-fold better stability in serum while retaining high affinity for Gαi1 (K_d = 9 nM).
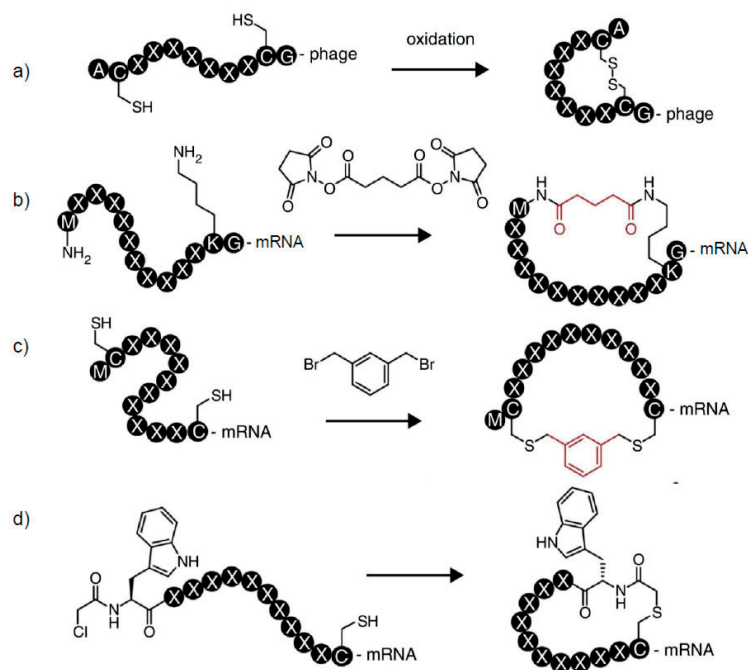
Figure 3. Examples of encoded cyclic libraries. a) Traditional cyclization of phage-encoded peptides by oxidation of two cysteines. b) Cyclization using the reaction of DSG with a mRNA-encoded peptide displaying two primary amines. c) Cyclization of mRNA-encoded peptides by crosslinking two cysteines with α,α'-dibromo-m-xylene. d) Cyclization of mRNA-encoded peptides by spontaneous reaction of a chloroacetamide group with the thiol of a cysteine. Reprinted with permission from[45], reprinting license n°4124691074190, copyright © 2015 Elsevier Ltd.

Another strategy for cyclizing mRNA-encoded peptide libraries is to react bivalent thiol-reactive compounds with peptides bearing two cysteines. This reaction was shown to be quantitative in aqueous buffers[46] and Szostak *et al.* proved that it was compatible with *in vitro* translation systems and mRNA display[47,48] (Figure 3c). They managed to create a library of more than 10^{13} peptides containing both natural and non-natural residues[49]. The incorporation of non-natural amino acids was made possible because of the tolerance of certain aminoacyl tRNA synthetases for analogues of proteinogenic residues. A selection against thrombin yielded an inhibitor with 1.5 nM affinity. The therapeutic peptide RA101495 was developed in a similar fashion. It inhibits the cleavage of the complement factor C5 into C5a and C5b and is currently under evaluation in a phase I trial for the treatment of paroxysmal nocturnal hemoglobinuria.

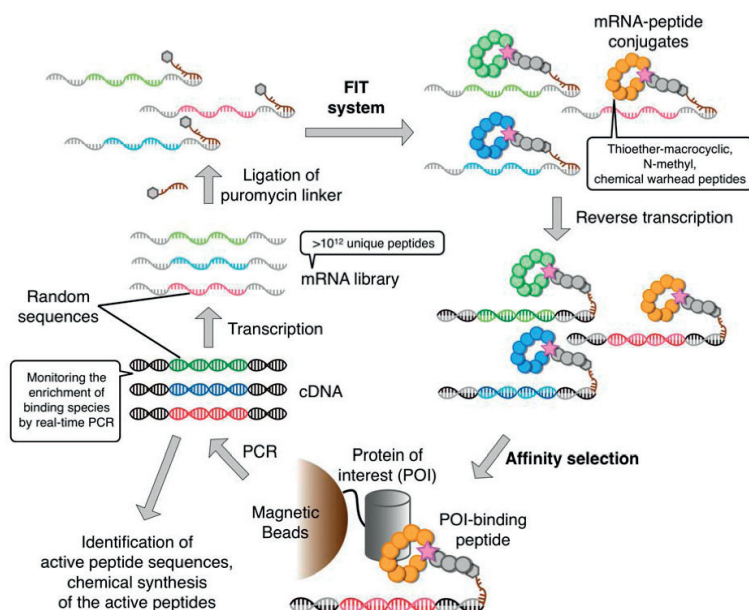**Figure 4. Schematic Illustration of a particular case of mRNA display.** The RaPID system developed by Suga and coworkers. With this strategy, multiple non-canonical amino acids can be incorporating using flexizyme or Flexible *in vitro* translation (FIT). The library of peptide is ribosomally synthesized from the mRNA pool conjugated to puromycin. The mRNA is then reverse transcribed and the library is subjected to affinity-selection against an immobilized target. cDNAs encoding the hit compounds are re-amplified by PCR and transcribed for the next round of panning. Ultimately, the enriched peptides are identified by NGS. Reprinted (open access) from[41], this article is available under the terms of CC-BY 3.0 license.

Cyclization of mRNA displayed peptides can also be achieved via the incorporation of unnatural amino acids acting as chemical handles for cyclization. In order to be able to load these residues on tRNA, Suga *et al.* evolved an RNA-based catalytic tool called Flexizyme[49] (Figure 4). With the flexible *in vitro* translation (FIT), a chloroacetamide functional group is conjugated to the α-amino group of tryptophan or tyrosine and can be displayed on a mRNA-encoded peptide. Immediately upon translation, this group reacts spontaneously and intramolecularly with a cysteine forming a thioether bond[50,51] (Figure 3d). With this method, Suga *et al.* identified several ligands to disease-relevant targets including ubiquitin ligase[52] ($K_d = 0.6$ nM), Akt2 kinase[53] ($IC_{50} = 92$ nM), and deacetylase SIRT2[54] ($K_d = 3.7$ nM). Subsequently, they commercialized their idea by founding the company Peptidream. Using their technology, Peptidream has successfully brought a cyclic peptide to clinical phase I for immune-oncology-related therapies[54].

1.2.2   Phage display

Phage display was described for the first time in 1985 by G.P. Smith[36]. In his study, he demonstrated that an exogenous protein can be expressed on the surface of a M13 phage coat protein III (pIII). To do so, Smith inserted a DNA sequence coding for a part of the EcoR1 enzyme on the 5' end of the gene III of the phage, creating an N-terminal fusion protein. After showing that the introduced sequence was properly exposed on the phage capsid, he could prove that the modified virions retained their infectivity and ability to propagate. Consequently, his work revealed one of the main characteristics of phage display which is the linkage between the phenotype (peptide displayed) and the genotype (DNA code).

The discovery of Smith was quickly associated with *in vitro* evolution of proteins and peptides. To this aim, DNA coding for randomized combinatorial libraries of peptides or proteins is inserted into the genome of a bacteriophage and expressed at the surface of a phage coat protein. The linkage between phenotype and genotype enables the isolation and identification of ligands through an iterative process of selection involving infection of bacteria and re-amplification of the phage (Figure 2). In 1990, McCafferty *et al.* successfully selected phage-displayed antibody fragments with desired properties[55]. Since then, monoclonal antibodies (mAbs) are the class of therapeutic molecules that has benefited the most by far from phage display technology developments[56,57]. Adalimumab was the first phage-selected antibody reaching the market in 2002. It is a fully human antibody developed by Cambridge Antibody Technology (Now MedImmune/AstraZeneca) to block the interaction between Tumor Necrosis Factor α and its receptors. Adalimumab is commercialized under the name of humira and is currently indicated for the treatment of Rheumatoid arthritis, psoriatic arthritis and Crohn's disease[58]. In 2016, the annual sales of humira reached the value of $16 billion [59].

Since the early developments of phage display, a variety of bacteriophage, such as filamentous phage, T7 and lambda have been exploited for *in vitro* selection. Among them, the filamentous phage (f1, fd and M13) are by far the most common because they represent excellent cloning vehicles[60]. Their genome is composed of genes coding for five different coat proteins that are located on a circular single-stranded DNA (ssDNA) of about 6400 base pairs. Almost all of these proteins have been used for phage display but pIII and pVIII are the most frequent. Approximately 2800 copies of pVIII, also called major phage coat protein, are arranged around the phage ssDNA and constitute its main body. Peptides displayed on the N-terminus of pVIII protein are limited to five or six amino acids because longer ones were shown to interfere with phage replication[61]. Unlike for pVIII, long polypeptides,

such as antibody fragments, can be displayed on pIII without impairing phage propagation. pIII is present in five copies at one end of the phage and plays a role in bacterial infection by interacting with the F pili and TolA receptor[62].
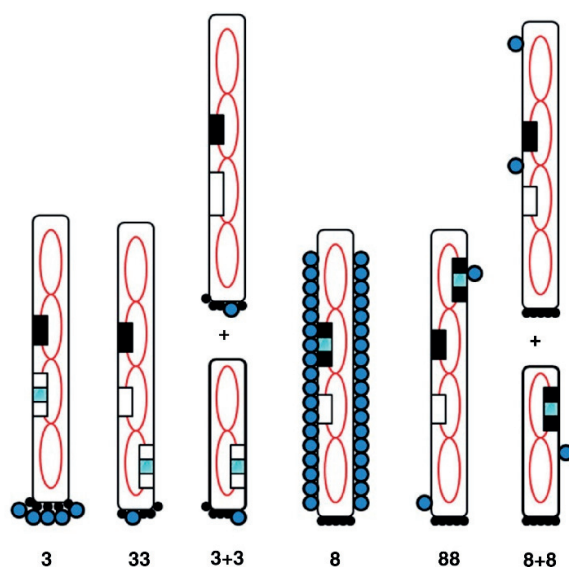

Figure 5. Schematic presentation of different types of phage display systems. In the type 3+3 and 8+8, a wild type version of pIII and of the other coat proteins is provided by a helper phage (not represented here). In the type 33 and 88, the wild type version of pIII is directly encoded on the phage vector. Reprinted with permission from[63], copyright © 2010 American Chemical Society.

Additionally, phage display systems can be categorized by the number of peptides exposed on phage surface[63]. This parameter can strongly influence avidity effects during selections and can be modulated accordingly. On the one hand, multivalent systems can help to select binders from naïve libraries, as it enhances the capture of phage with suboptimal binding affinities. However, on the other hand, it can be a drawback during affinity maturation procedures when strong binders are desired. This specificity depends on the chosen phage coat protein: more than 2800 peptides are exposed on the virion particles when using pVIII display but only 5 when using pIII display (Figure 5). Another strategy employed to decrease the valence is to co-express the engineered coat protein with the wild type counterpart. This can be achieved in two ways (Figure 5): by transfecting a phage vector co-expressing the wild type and engineered version of the coat protein (88 and 33 systems) or by using a phagemid system with helper phage (8+8 and 3+3 systems).

Practically, phage peptide libraries are created by inserting a DNA coding for a randomized peptide or protein fragment into a double stranded phage vector. This involves a PCR with primers containing degenerated codons, for example NNK (N = A,T,C,G ; K = T, G), to avoid the

overrepresentation of the most common amino acids. Alternatively, a more rational and controlled randomization can be adopted in order to display desirable or favored amino acids. These strategies can be applied using oligos containing define tri-nucleotide codon content. This step is usually followed by subsequent ligation into the phage vector and electroporation into phage competent *E.coli*. This last step limits the achievable diversity to $10^7 - 10^{10}$ transformants which is often much inferior to the theoretical diversity.

1.2.3   Phage selection of cyclic peptides

The first reported cyclic peptides identified during a phage selection were isolated from a 15-mer linear peptide library[64]. Peptides containing pairs of cysteines were enriched suggesting that they probably had been selected in a disulfide cyclized format. The disulfide bond formation most likely takes place in the oxidizing environment of the bacterial periplasm or after release of the phage particles in the oxygenated culture medium. These observations encouraged the design of a second generation of phage-encoded peptide libraries in which a central randomized sequence is flanked by two fixed cysteine residues[37]. The first studies conducted with these libraries confirmed the selective advantage offered by cyclization. For instance, a screen against platelet glycoproteins with a hexapeptide library (CX_6C, C = cysteine, X = any canonical amino acid), led to the identification of cyclic ligands with dissociation constants in the nanomolar range whereas their linear counterparts had a much weaker affinity[65]. In another study of Giebel *et al.*, a panning against streptavidin with cyclic peptide libraries resulted in the isolation of ligands having on average 100 to 1000-fold better binding affinities as compared to linear peptides selected against the same target[66].

A later approach consisted in diversifying the peptide formats using variations in the size of the loop and of newly added exocyclic regions[37]. With this aim, Fairbrother *et al.* generated phage-encoded libraries with the format $X_mCX_nCX_o$ (n = 4-10, m+n+o =18) and used them in a screen against vascular endothelial growth factor[67] (VEGF). Their results showed that a specific peptide format is usually preferred for a given target. This underlined the importance of screening multiple libraries having different peptide formats in order to increase the chances of isolating binders with desired affinity and specificity.

However, disulfide bridges are susceptible to reduction and disulfide exchange reactions. Therefore, other cyclization reactions leading to non-reducible cyclic peptides are desirable. The first phage-encoded peptides cyclized in such manner were the so-called bicyclic peptides developed by Heinis and Winter in 2009[68]. With their method, they could generate polycyclic compounds by

connecting the three cysteines of a displayed peptide with a trivalent thiol-reactive small molecule. Phage display and selection of bicyclic peptides is further described in the section 1.3 of this introduction. This pioneering strategy was then followed by many other examples, in which the nature and the function of the chemical linker was varied[39,69]. For instance, several groups have aimed at generating photoswitchable cyclic peptides circularized with an azobenzene moiety[38,70,71].

1.2.4   Phage selection of cyclic peptides for therapeutic applications

To date, peginesatide is the only cyclic peptide isolated by phage display that has been approved by FDA. However, the drug was withdrawn one year after its approval in 2012 due to toxicity issues. It consisted of a 45 kDa PEGylated peptide acting as an agonist of erythropoietin and was used for the treatment of chronic kidney disorder associated anemia[40]. Peginesatide was initially isolated from a cyclic peptide library having the CX_8C format (C = cysteine X = any canonical amino acid) and further affinity matured with the generation of two different phage libraries[72,73].

Another three peptides isolated by phage display are currently in clinical trials[26]. The pharmaceutical company Apellis has brought a PEGylated cyclic peptide inhibitor of complement factor C3 called APL-2 to the clinic. The molecule is currently undergoing evaluation in phase II for age-related macular degeneration (AMD) and is derived from a 27-mer peptide identified during phage selections against C3b[74]. The second one is ALRN-6924, a stapled peptide inhibitor of MDM2 and MDMX that restores normal p-53 mediated apoptosis activity in malignant cells[75]. ALRN-6924 was generated by optimizing the phage selected stapled peptide ATSP-7041. The molecule, developed by Aileron therapeutics, has reached a phase I/II clinical trial for the treatment of solid tumors and lymphomas.

1.3 – Bicyclic peptides

1.3.1   Phage selections of bicyclic peptides

Inspired by the favorable properties of phage-selected cyclic peptides, Heinis and Winter developed a new strategy to evolve bicyclic peptides on phage[68] (Figure 6). Their approach was aiming at obtaining even more constrained structures that would lead to a better affinity and stability. For their cyclization, they utilized an already existing chemistry that was introduced by Timmerman and co-workers in order to generate peptides mimicking non-linear epitopes of antibodies[46]. The idea was based on the reaction of a peptide containing three reduced cysteines with the trivalent thiol-reactive compound 1,3,5-tris(bromomethyl)benzene (TBMB). When performed in aqueous conditions, the reaction was shown to be quantitative, fast and selective for the thiol functional groups of cysteines.
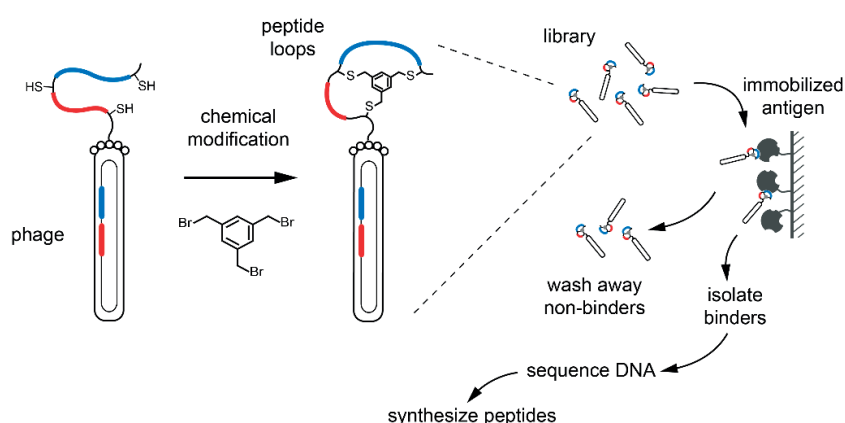
Figure 6. Phage selection of combinatorial bicyclic peptide libraries. Peptides containing three cysteines are exposed on pIII at the surface of a fd phage and chemically modified with TBMB. Random amino acid loops are encoded on the phage vector and indicated in blue and red. Libraries of bicyclic peptides are subjected to two or three rounds of affinity selection against an immobilized target. Weak and non binders are washed away and specific ligands are isolated and re-infected into *E.coli*. They can be either re-amplified for another round of panning of sequenced in order to identify peptide ligands.

In a proof of concept work published in 2009[68], Heinis and Winter showed that this reaction was compatible with phage and generated a large combinatorial library of more than 10^9 peptides containing two loops of six random amino acids flanked by three cysteines (format: ACX_6CX_6CG, named 6x6). In order to avoid side reactions with the cysteines of the D1 and D2 domains of pIII, the library was fused to the N-terminal end of a genetically-engineered disulfide-free pIII[76]. Phage carrying this mutated version of pIII are about hundred times less infective than the wild type phage but tolerate a great variety of alkylation and reduction reactions. After phage production, the cysteines

were reduced using TCEP and further modified using TBMB. Three rounds of iterative selections against Plasma Kallirein (PK) yielded bicyclic peptide inhibitors with IC_{50}s in the nanomolar range whereas their linear equivalents were nearly 250-fold less potent[68]. An additional affinity maturation selection allowed the identification a potent inhibitor of PK with a dissociation constant in the low nanomolar range (PK15; $K_I = 1.5$ nM).

This phage based strategy has been routinely applied to generate bicyclic peptide ligands to a variety of therapeutically relevant targets. As a result, bicyclic peptides with nanomolar affinities and good specificity could be identified and further characterized. A large effort has also been invested in the generation of combinatorial libraries of bicyclic peptides with higher structural diversity. The work done in these directions is discussed in the following subsections.

1.3.2   Combinatorial libraries of bicyclic peptides

Combinatorial libraries of peptides with large diversity are particularly interesting for the identification of ligands with high affinity for therapeutic targets. Although the randomized regions of phage-encoded peptide libraries offer a huge theoretical diversity, the achievable diversity is usually limited to $10^9 – 10^{10}$ clones considering the limited number of bacteria that can be transformed with DNA plasmid. However, other features of the bicyclic peptides can be modulated such as the number of amino acids in each of the two rings and the type of chemical linker used to cyclize the peptide. These two strategies induce major structural variations of the molecular backbone and are essential for the isolation of peptide ligands making compatible interactions with their targets. Both of them have been evaluated in our laboratory and are described in the following paragraphs.

As already discussed in the section 1.2.2, the number of amino acids in the loops of a cyclic peptide ligands can considerably influence its ability to bind to a specific target. The importance of screening multiple libraries of cyclic peptides had already been demonstrated by Fairbrother *et al.* in a work where they used libraries with the format $X_mCX_nCX_o$[67] (C= cysteine and X = any natural amino acid; n = 4-10, m+n+o = 18). A selection against VEGF with these libraries revealed that one particular format was preferentially enriched during the panning. Similarly, there was a need for phage libraries of differently sized bicyclic peptides. This was first illustrated with a problem encountered during the development of a plasma kallikrein inhibitor[77]. The bicyclic peptide PK15, isolated from the 6x6 library, displayed a good affinity for the human protease ($K_I = 1.5$ nM) but had little or no activity on mouse and rat orthologs. As a result, PK15 could not be tested in animal models. This strong specificity was attributed to the relatively large loops of 6 amino acids that were likely to interact with

regions of PK that are not conserved in the human and murine orthologs. For this reason, libraries of shorter macrocycles were tested in phage selections - the library 5x5 or 3x3 that contain two loops of 5 or 3 amino acids each, respectively. This resulted in the isolation of a clone from the 5x5 library that inhibited human, monkey and rat PK with K_I values in the low nanomolar range while sparing paralogous proteases. These observations motivated the creation of another three different bicyclic peptide phage libraries of the format ACX_mCX_nCG, where m and n can take values from 3 to 6[78] (Figure 7). These highly structurally diverse libraries, named A, B and C, were used in a selection against Urokinase-type plasminogen activator (uPA) and consensus sequences were found for all of the new formats. Although none of the new isolated ligands had a better affinity than the peptide UK18 initially identified with the 6x6 library, they revealed new binding motifs that were not found with this library.
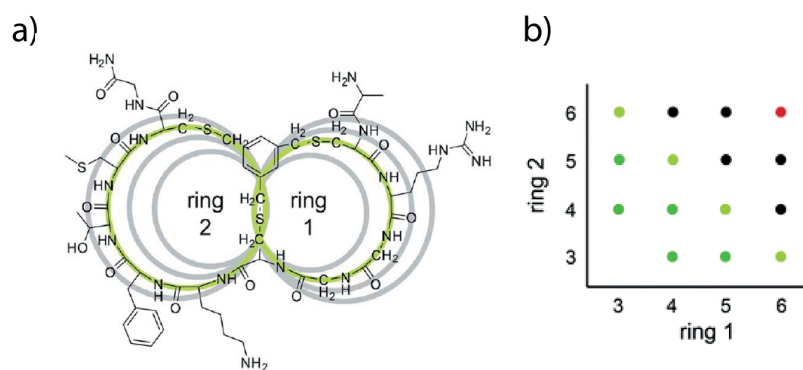

Figure 7. Phage display libraries of differently sized bicyclic peptides. a) The chemical structure of a representative bicyclic peptide (UK368) in green is represented together with peptide rings of 3, 4, 5 and 6 amino acids for size comparison (in grey) b) The libraries created by Rentero *et al.* combine loops of different lengths. The scheme shows the loop combinations in the library A (dark green), B (light green), C (black) and of the former 6x6 library (red). Reprinted with permission from[78], reprinting license n° 4124710316221, copyright © 2012 Royal Society of Chemistry.

In parallel, the use of alternative cyclization linker has also greatly contributed to improve the diversity of phage-encoded bicyclic peptide libraries. TBMB, the initial molecule used by Heinis and Winter, is a rather small molecule and does not display any H bond donor or acceptor[68]. Therefore, the linker itself can only form hydrophobic interactions with the target or with the amino acids present in the peptide loops. Additionally, NMR and crystal structures of two in-vitro evolved bicyclic peptides, PK15 and UK18, showed that the mesitylene core did not make any noncovalent bonding with the peptide loops[68,79]. In a study of Chen *et al.*, two other linkers were introduced with the aim of conferring new geometries to bicyclic peptides as well as favoring intramolecular H-bonding[80]. Based on TBMB structure, the choice was limited to trivalent thiol-reactive compounds with a threefold

rotational symmetry (Figure 8). Two molecules were proposed : 1,3,5-triacryloyl-1,3,5-triazinane (TATA) and N,N',N''-(benzene-1,3,5-triyl)tris(2-bromoacetamide) (TBAB). The calculation of the distance between sulfur atoms by molecular dynamics simulation confirmed that the three linkers would offer different loop geometry and flexibility (Figure 8).
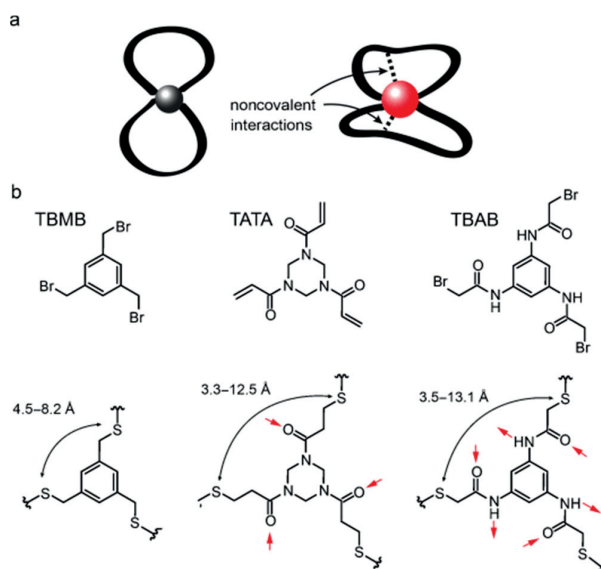

Figure 8. Structurally diverse small molecules for the cyclization of phage displayed bicyclic peptides. a) previous approaches have not favored interactions between the cyclization core (in grey) and the peptidic backbone. The use of small chemical linkers with polar groups would enable H-bond interactions with the peptide loops and would serves as nucleating scaffold. b) Structures of the three reagents chosen to generate bicyclic peptides (before and after reaction). All of them have a threefold rotational symmetry as well as three thiol-reactive groups. Potential H-bond donors and acceptors are indicated by red arrows. Computed distance ranges between the sulfur atoms of the cyclized peptides are indicated. Reprinted with permission from[81], reprinting license n° 4124710522727, copyright © 2014 John Wiley and Sons.

TBMB, TATA and TBAB, were successfully applied to phage selections against uPA using the 3x3 library[81]. This led to the identification of specific consensus motifs for each of them and showed the possibility to triplicate the diversity of the library. The best binder in this selection was a bicyclic peptide cyclized with TATA (K_I = 85 nM) whose inhibitory activity was not retained when cyclized with either TBAB or TBMB. Finally, the crystal structure of uPA in complex with one of the TBAB-selected bicyclic peptide showed that the linker itself was able to establish hydrogen bonds with the side chains and the backbone of the selected peptide.

Finally, the structural diversity of bicyclic peptide libraries could further be extended using a strategy based on the oxidation of peptides with four cysteines[82] (Figure 9). Indeed, the enrichment of peptides with a fourth cysteine was observed when phage libraries were subjected to panning without

chemical modification. Most probably, unpaired cysteines can form disulfide bridges with cysteines of neighboring displayed peptides, thus impairing the interaction of the pIII protein with the F-pili and TolA receptor during infection of host bacteria. Chen *et al.* took profit of this observation and screened oxydized phage displayed peptides against uPA. After two rounds of selection, 62% of the peptides had a fourth cysteine whereas none were found in the control where peptide were cyclized with TBMB. Sequencing results revealed the huge topological diversity offered by this method. The fourth cysteine appeared in any of the randomized amino acid positions. In addition, a single peptide sequence could lead to three different regioisomers which further expand the variety of conformation that can be obtained.
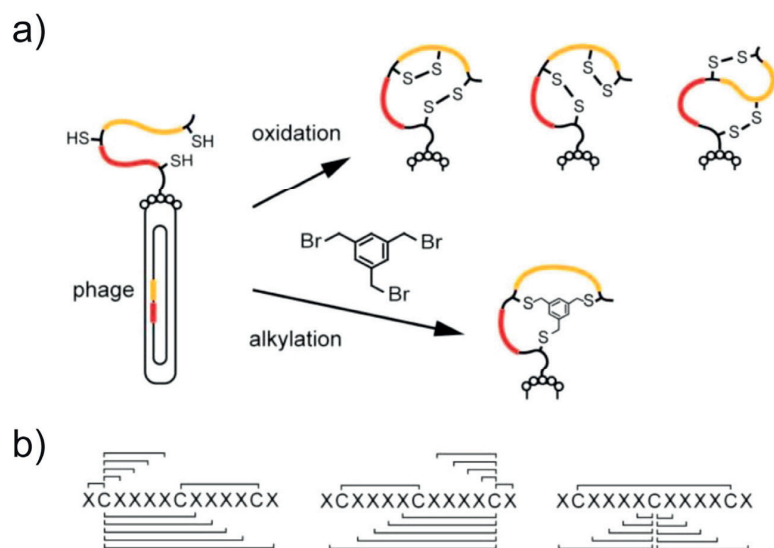
Figure 9. Generation of phage-encoded bicyclic peptides cyclized by two disulfide bridges. a) Peptides containing three cysteines in fixed positions and random sequences (red and yellow) are displayed on phage. A fourth cysteine can appear in the random sequence. Pairs of cysteines are air oxidized to form disulfide bridges and three different regioisomers can be formed. In the lower reaction, bicyclic peptides are formed by reacting the three cysteine residues with an alkylating reagent, such as TBMB. The single product isomer contains three thioether bonds. b) A large number of topologically diverse bicyclic structures can be generated by placing the fourth cysteine in any of the randomized positions and by forming disulfide bridges between two pairs of cysteines. Reprinted with permission from[82], copyright © 2013 American Chemical Society.

Nevertheless, the diversity offered by naïve bicyclic peptide phage libraries is sometimes not sufficient to isolate clones with desired properties. As a result, the creation of affinity maturation libraries is sometimes required. This involves the use of semi-randomized phage libraries based on the sequence of a lead bicyclic peptide isolated during naïve selections. To do so, the important residues of the lead peptide are identified using regions where a strong consensus was found or with the help of an alanine scan. After that, a new library is cloned in which the less important amino acids

are randomized with degenerated primers. The newly designed library is then panned against the target in one or two rounds of affinity selection with stringent conditions such as a reduced amount of immobilized target. This method has been successfully applied in our lab to increase the affinity of two bicyclic peptide inhibitors of PK[68] and coagulation factor XIIa[83] (FXIIa).

1.3.3   Examples of phage selected bicyclic peptides

Our laboratory has already identified bicyclic peptide ligands and inhibitors to a wide variety of targets including trypsin-like serine proteases[68,77,79,83], metalloproteases (unpublished data), sortase A[84], cancer associated cell receptors[85,86] and other intracellular proteins such as β-catenin[87] and actin (unpublished data). The affinity of the lead compounds identified during phage selections varies from the picomolar to the micromolar range and could be further improved by affinity maturation and medicinal chemistry (Figure 10).
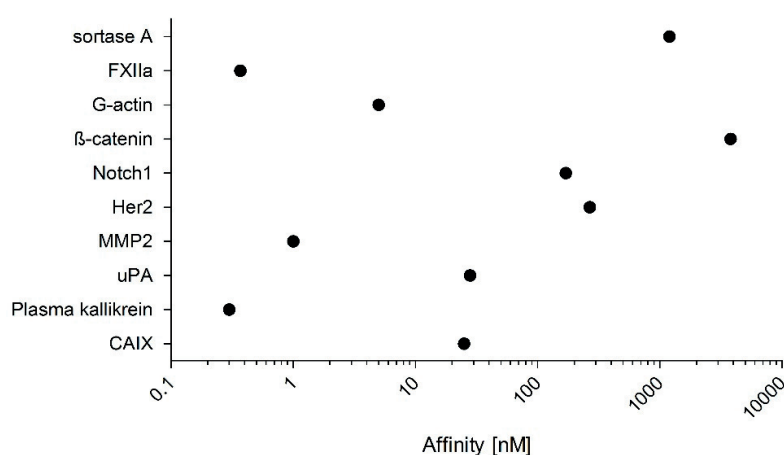
Figure 10. Affinity of bicyclic peptides isolated against a variety of targets.

Bicyclic peptide inhibitors of trypsin-like serine proteases have been the most extensively characterized. Serine proteases are components of signaling pathways that are tightly regulated and whose dysregulation can lead to important pathologies. Yet, targeting disease-related serine proteases remains challenging, mostly because of the difficulty to achieve selective binding to their active sites[88,89]. As already mentioned, bicyclic peptides bind to their targets with high affinity and specificity, which makes them a promising molecule format for the development of selective serine proteases inhibitors. With such strategy, our laboratory isolated potent inhibitors to the trypsin-like proteases

plasma kallikrein (PK)[68,77], urokinase-type plasminogen activator (uPA)[79] and coagulation factor XIIa (FXIIa)[83].

PK is involved in the intrinsic coagulation pathway as well as in the kallikrein-kininogen system. As a result dysregulation of PK activity can lead to thrombosis and inflammation and its inhibition would be beneficial to several pathologies. The only approved inhibitor of PK is a Kunitz domain-based polypeptide of 60 amino acids commercialized under the name of kalbitor (Dyax Corporation, now owned by Shire)[90]. It is indicated in the treatment of a rare inflammatory disease called Hereditary Angioedema (HAE). As kalbitor has a short *in vivo* half-life, it is rather used on demand during acute attacks in HAE patients[91]. For this reason, Dyax corporation/Shire is also working on the development of long acting inhibitor of PK[92]. They developed a monoclonal antibody called lanadelumab that is currently in phase III for HAE treatment and can be administered twice monthly.
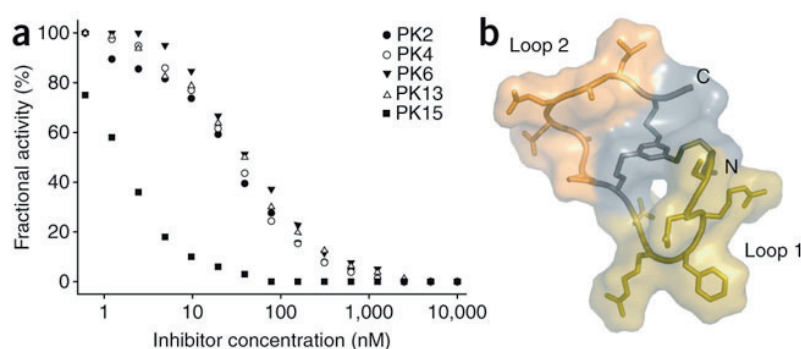
Figure 11. Inhibition of human plasma kallikrein by different bicyclic peptides and NMR solution structure of PK15. a) Clones PK2, PK4, PK6, PK13 and PK15 were isolated in phage selections using two different libraries. PK15 has an affinity of 1.5 nM as calculated with the enzyme inhibition assay. b) The peptide loops of conjugate PK15 are shown in yellow (loop 1) and orange (loop 2). The mesitylene core, the three Cys residues and the terminal Ala (N-terminus) and Gly (C-terminus) are shown in gray. Reprinted with permission from[68], license n° 4124750354452, copyright © 2009 Nature publishing group.

In 2009, the success of kalbitor inspired Heinis and Winter for the generation of smaller bicyclic peptide inhibitor of PK[68]. In their proof-of-concept study, they used their first generation of phage-encoded libraries in a selection against the serine protease and isolated PK15, a potent inhibitor with low nanomolar affinity ($K_I = 1.5$ nM) (Figure 11). In addition, the bicyclic peptide inhibitor was shown to have no activity on homologous proteases such as thrombin and Factor XIa. However, as already introduced in section 1.3.2, this strong specificity also resulted in a poor activity against the mouse and the rat orthologs, preventing the possibility to test its activity *in vivo*. This problem was solved with a second generation of bicyclic peptide inhibitors of PK that had smaller loop sizes[77].

Baeriswyl *et al.* performed a selection with the libraries 3x3 and 5x5 cyclized with TBMB and successfully isolated an inhibitor with high binding affinity to human PK ($K_I = 0.3$ nM) and retained activity on orthologous monkey and rat PK ($K_I = 0.4$ and 11 nM respectively).

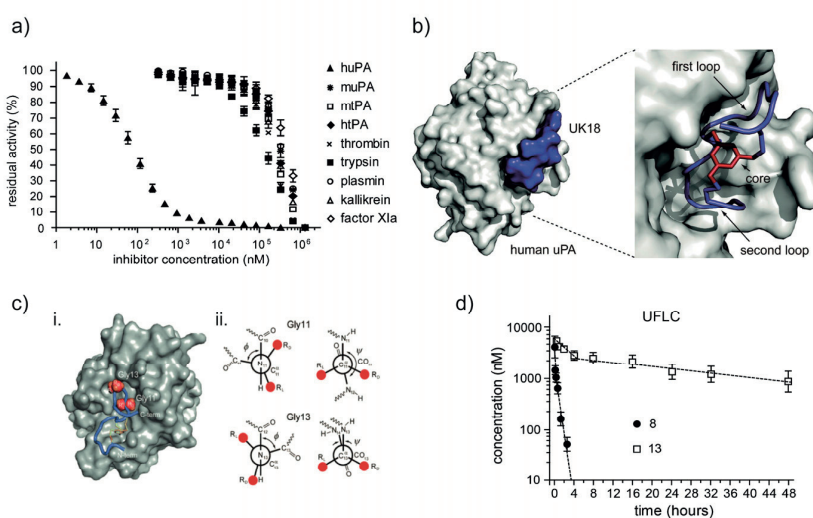**Figure 12. Summary of different experiments conducted with the uPA inhibitor UK18.** a) The affinity of UK18 for a range of homolous trypsin-like protease was measured in an enzyme activity assay. b) Crystallization of uPA (grey surface) in complex with UK18 (peptide ribbon in blue, mesitylene scaffold in red.) c) Prediction of the consequences of L to D-amino acid substitution in UK18 using Φ and Ψ torsion angles of Gly11 and Gly13. i. Crystal structure of uPA with bound bicyclic peptide UK18 showed potential steric clashes with surface residues of uPA when the two glycine residues were substituted with L-or D-amino acids. ii. Newman projections of Gly11 and Gly13 with the Φ and Ψ torsion angles found in the structure. D-amino acid substitutions in Gly11 create a clash between the side chain of the amino acid and the carbonyl group of Arg10; D-amino acid substitutions of Gly13 were not expected to create steric conflicts. d) Pharmacokinetic parameters of UK18-fluorescein and UK18-SA21-fluorescein in mice. Both compounds were applied intravenously (200 μL of 50 μM), blood samples were taken at the indicated time points, and the concentration of conjugate was measured by ELISA. The indicated values are averages of at least two mice. Reprinted with permission from[79,93,94], copyright © 2012 American Chemical Society, & reprinting license n°4124750799695, copyright © 2013 John Wiley and Sons.

The second bicyclic peptide isolated against a serine protease is UK18, an inhibitor of uPA. In several studies, this enzyme has been shown to favor tumor growth and metastasis by enhancing the degradation of the extracellular matrix and helping cancer cell evasion. Therefore, uPA was seen as a potential target for the preventive treatment of metastatic cancers[95]. UK18 was first identified by Angelini *et al.* in a phage selection with the 6x6 library[79]. As for PK15, It was active in the nanomolar range ($K_I = 53$ nM) and had more than 1000-fold reduced affinity for paralogous proteases (Figure 12a). In order to better understand the binding mechanism of UK18, the inhibitor was co-crystallized with its target (Figure 12b). The results underlined the large contact surface between the peptide and the protease and also provided useful information for its affinity maturation (Figure 12c.i).

Specifically, the analysis of the binding mechanism pointed two glycine residues that could potentially be replaced with D-amino acids[93] (Figure 12c.i and ii). These two positions were screened for a range of natural and non-natural D amino acids among whom two led to improvements of the peptide affinity and stability. UK18 has also been particularly important for understanding of the *in vivo* behavior of bicyclic peptides. A pharmacokinetic study showed that the peptide was undergoing rapid renal clearance (half-life = 30 min) after i.v injection in mice[94] (Figure 12d). After that, two strategies were evaluated to increase the circulation half-life: recombinant fusion to a Fc region[20] and conjugation to an albumin-binding peptide[94] (SA21, developed by Dennis and coworkers). These two strategies successfully prolonged the blood residence time to 36h and 24h, respectively (Figure 12d). Moreover, another mouse study demonstrated that the SA21-conjugate could efficiently diffuse into organ tissues and solid tumors[96].

The third example is a bicyclic peptide inhibitor of FXIIa that proved to be very potent both *in vitro* and *in vivo*. The protease target is the first serine protease of the intrinsic coagulation pathway and is also involved in other enzymatic cascades associated with inflammation and immune reaction[97]. As a consequence, FXIIa plays a role in coagulation and/or inflammation related pathologies by inducing unwanted blood clot formation[98,99]. For example, FXIIa dysregulation was associated with diseases such as thrombosis or hereditary angioedema (HAE). Moreover, the enzyme can also get activated during surgeries in which the blood comes in contact with medical devices, leading to life-threatening consequences. This is the case for coronary pulmonary bypass (CPB) and extracorporeal membrane oxygenation (ECMO), two medical procedures in which an instrument takes over the function of the heart and lungs. Currently, the addition of heparin to the blood circulation is the preferred solution to this problem despite strong side effects such as excessive bleeding[100]. For these reasons, the development of bicyclic peptide inhibitor of FXII was regarded as an interesting approach to generate safer anticoagulants. Up to now, Dyax[101,102] as well as several other laboratories[103,104] have developed mAbs blocking the activity of FXIIa.

Bicyclic peptide inhibitors of FXIIa have also been evolved by phage display. A first selection performed with the 4x4 library, in which the peptides were cyclized with TBMB, yielded only inhibitors with weak affinity ($K_I \geq 1.2$ μM)[105]. In a second trial, the structural diversity was increased by using several libraries (3x3, 4x4 and 6x6) as well as another linker (TATA)[83]. As a result, peptides with affinities in the nanomolar range were identified and an affinity maturation step enabled the isolation of peptide FXII618 with a K_I value of 22 nM (Figure 13a). This inhibitor was highly selective for FXIIa as compared to the already existing FXIIa corn trypsin inhibitor (CTI) (Figure 13b). It also

showed better prolongation of activated partial thromboplastin time (aPTT) which measures the time to coagulation upon initiation of coagulation via the intrinsic pathway (Figure 13c).
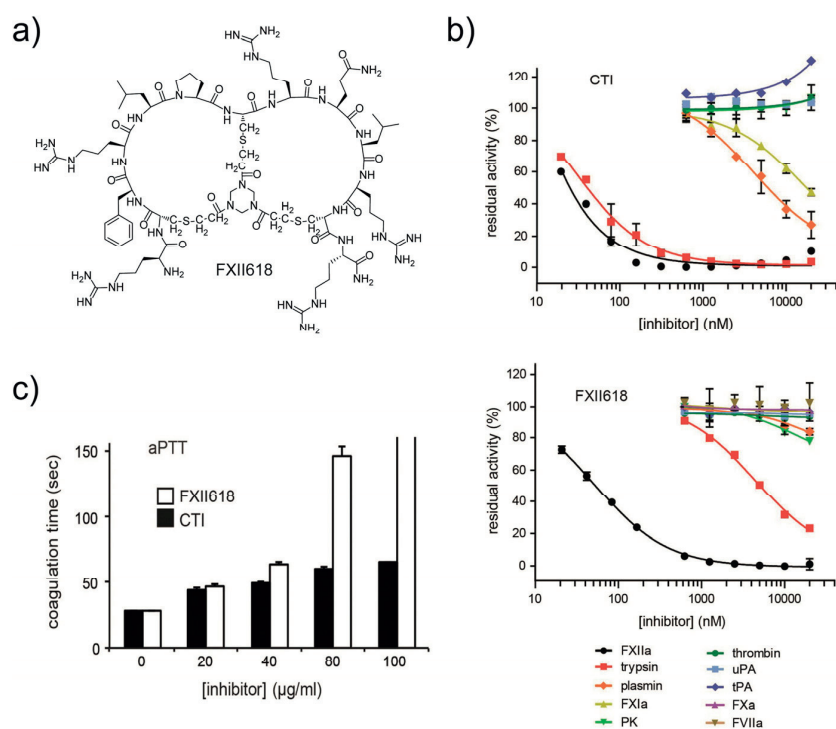
Figure 13. Comparison of FXII618 and CTI. a) Chemical structure of FXII618. b) Specificity profile and inhibitory activity of CTI (upper panel) and FXII618 (lower panel). The residual activity of several trypsin-like serine proteases is shown. c) Comparison of activated partial thromboplastin time (aPTT) in citrated platelet poor plasma in the presence of the same quantities of FXII618 and CTI. Reprinted with permission from[83], Copyright © 2015 American Chemical Society.

Later, the suboptimal affinity and plasma stability of FXII618 was again improved by generating small combinatorial libraries (10 to 100 compounds) of peptides by SPPS. In a first approach, residues important for stability and affinity were substituted with natural and non-natural analogues[106]. In a second approach, atoms were inserted into the peptide backbone using β-amino acids and cysteine analogues[107]. The amino acid sequence of best mutants were combined together and resulted into a 380 pM inhibitor, called JSM33-R that was stable for more than 100h in human plasma (unpublished data). In addition, the peptide had an enhanced activity on the mouse ortholog (K_I = 500 pM). Although JSM33-R has a really short *in vivo* half-life, it efficiently prevented blood coagulation in mouse models in which blood clotting is induced with ferric chloride (unpublished data).

1.4 – Next generation sequencing (NGS) and phage display

1.4.1   Application of NGS in phage display

In most cases, the analysis of phage panning outputs is done by Sanger sequencing. However, NGS has recently gained increasing attention from research groups performing phage selections. This technology has revolutionized the field of genomics and DNA sequencing by providing exceptional volumes of sequence data at an unprecedented speed and low cost[108]. In 2008, Mannocci *et al.* applied NGS for the first time to DNA-encoded libraries for the *in vitro* evolution of ligands[109]. Using 454 pyrosequencing[110], they successfully identified binders to streptavidin pulled out from a library of 4000 different compounds. After them many laboratories have worked with NGS to analyze naïve or matured phage libraries of peptides[111–113], antibodies[114–116] or protein domains[117–119]. In the majority of the cases, results were obtained with technologies[110] such as Illumina platform[115,116], 454 sequencing technology[114,117,118] or ion semiconductor sequencing[111,119] (Figure 14). In all these methods, the phage vector DNA is isolated by extracting plasmid dsDNA from infected *E.coli*. In a second step, the fragments of interest are amplified by one or two PCR resulting in amplicons flanked with adaptor sequences compatible with consecutive NGS (Figure 15).
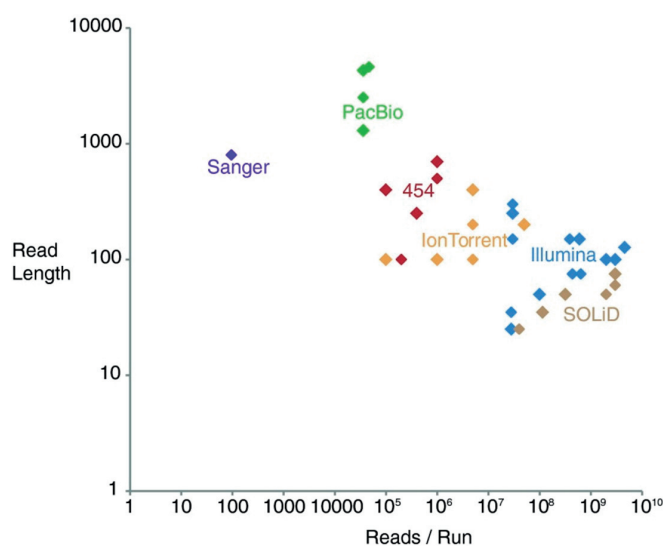

Figure 14. Specificities of different NGS technologies in terms of read length and throughput.
Points represent official platform/chemistry combination releases and are color-coded based on the platform family. Reprinted (open access) from[110], this article is available under the terms of CC-BY 3.0 license.
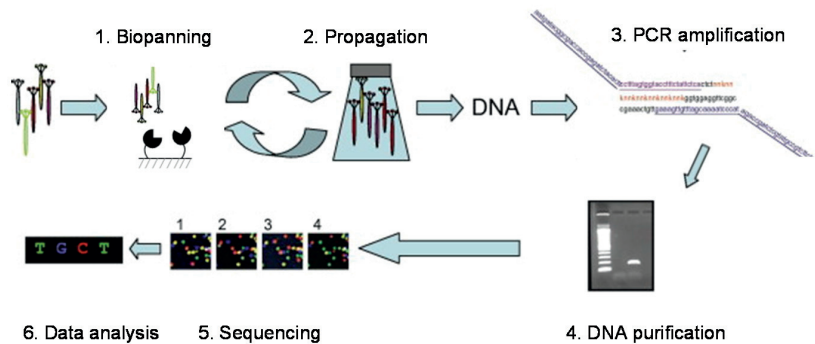
Figure 15. General procedure for NGS of phage-encoded polypeptides. (1) In the biopanning phase, the phage-encoded library is incubated together with the immobilized target. After washing away nonbound phages, binders are isolated and amplified in bacteria. The amplified libraries can be reused in a subsequent round of biopanning. (2) Alternatively, the DNA of the phages can be isolated usually after the propagation step (3) and amplified with PCR primers complementary to sequences flanking the region of interest. The PCR primers contain tails with the adapter sequences necessary for NGS. This step can involve more than one PCR (4) DNA fragments are purified (5) and sequenced (in this scheme, with Illumina technology). (6) Finally, the data are analyzed using bioinformatics tools. Adapted from, reprinting license n°4124760047073, copyright © 2012 Elsevier Ltd.

The first NGS of phage selected ligands was performed by Dias Nesto *et al.* using 454 pyrosequencing[108]. They obtained around 50 000 reads and analyzed them with the aim of identifying peptides selected against different organs. The same technology was applied in other phage panning and has been especially useful for the sequencing of antibody repertoires due to the long size of the reads (up to 700 bp)[114] (Figure 14). Nowadays, methods providing higher throughput are the most popular. In particular, Illumina platform is the most widely used because it can generate up to 4 billion different reads with low error rate as compared to 454 pyrosequencing and ion semiconductor sequencing[110]. However, the shorter read length has impaired the sequencing of entire single chain variable fragments (scFv) (Figure 14). As a result, NGS of antibody libraries has for a long time been restricted to the analysis of complementarity-determining region 3 (CDR3)[119,116]. Indeed, these domains were demonstrated to be the most relevant contributors to antibody specificity and are sometimes the only randomized regions in scFv libraries[116]. Last year, Lövgren *et al.* overcame for the first time these limitations and sequenced all the CDR of a framework antibody repertoire with Illumina technology[115]. This was achieved with the recently developed Illumina V3 2x300 bp paired end sequencing. As a result, they could easily replicate the enriched scFv by means of DNA synthesis, cloning and recombinant expression. Finally, another breakthrough in the analysis of phage selection data by NGS has been the use of barcoded primers for the preparation and multiplexing of the

sequenced DNA libraries. Sidhu and co-workers pioneered this approach and were able to analyze 22 independent panning experiments in a single run[117].

The application of NGS to phage displayed peptides provided valuable information about the diversity and abundance of selected clones and helped monitoring enrichment during the selection and amplification procedures. In most cases, peptides are ranked according to their copy number and the most frequent clones are further synthesized and characterized[115,119]. However, in some selections the average abundance is low and the isolated peptides diverse, which makes the identification of relevant clones difficult. As a consequence, other approaches based on sequence homology were developed[113,120]. For example, clustering of several thousands of peptides allowed the determination of target-specific consensus sequences and enabled to discriminate clones selected against M2 macrophages from the background[121]. NGS has also enhanced the recognition of target-unrelated peptides (TUPs). Indeed, there are several reasons that explain unintended selection of TUPs in phage pannings[112,122,123]. First of all, proliferation advantages during the phage amplification steps can have a profound impact on the output of the selection. Second, peptides can be systematically isolated because of their affinity to constant parts of the screening platform such as plastic or beads. One way to avoid the enrichment of TUPs is to reduce the number of iterative rounds of panning as illustrated with the work of 't Hoen *et al*[112]. Other procedures consist of including several biological replicates as well as an unselected but amplified library[120,121,123]. Actual target binders should appear with similar frequency in both replicates but should be much less abundant or absent in the negative control.

1.4.2   NGS software developed for of phage-encoded libraries

In the past, our laboratory was working with Sanger sequencing, resulting in labor and cost-intensive procedures that were rarely used to analyze more than a hundred clones per selection. Members of the laboratory speculated that NGS would help to identify target-specific binding motifs and afford a better coverage of all the diversity of isolated peptides. In line with this objective, Rentero *et al.* developed a method based on Ion Torrent PGM™ platform given the high throughput (10^6 to 10^7 reads) and adequate read length (approximately 200 bp length on the Ion 316TM Chips they used) offered by this technology. For their experiment, they isolated plasmid DNA from phage infected bacteria and amplified the region of interest by PCR. During this step, a six-letter barcode was inserted via the forward primer, allowing the multiplexing and sequencing of several phage selections in a single run. After production of the DNA libraries, samples were run on an Ion316™ Chip yielding a

total of more than 10^6 sequences that could be separated into different files according to their barcode[120].

Consecutively, Rentero *et al.* developed a tool to analyze NGS data of phage selected peptides based on both abundance and sequence homology. The different scripts were written with the MATLAB software and were used to (i) filter reads with suboptimal quality, (ii) eliminate erroneous clones appearing during sequencing or PCR, (iii) classify peptides according to the number of cysteine residues and ring size, (iv) order peptides into consensus sequences and finally compare different datasets with each other (Figure 16).
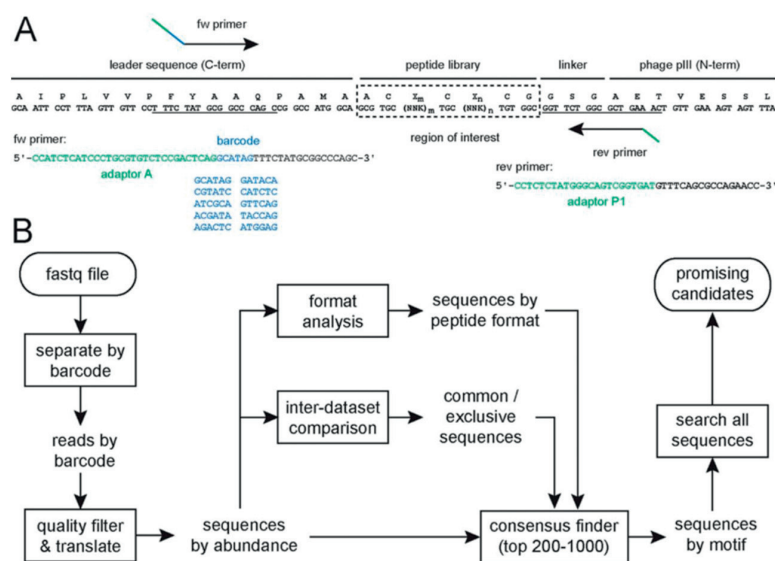
Figure 16. NGS and sequence analysis strategy of Rentero *et al.* (A) Primer design for Ion Torrent sequencing of bicyclic peptide libraries. The insertion of barcodes enable the multiplexing of several phage selection outputs on the same chip (B) Procedure for the analysis of sequencing data applying MatLab scripts. First, reads are separated into several files according to their barcode/selection. Second, low-quality sequences are removed from the dataset, and remaining sequences are translated and sorted by abundance. Additionally, peptides can be distributed based on their format (e.g. peptide ring size in the case of bicyclic peptides) and datasets compared with each other. Then, the sequences of the most abundant peptides (e.g. top 200) are compared and clustered in consensus groups, allowing the identification of specific motifs. Finally, the entire pool of sequences is searched for other less abundant sequences sharing such motifs in order to identify promising candidates. Reprinted (open access) from[111], this article is available under the terms of CC-BY 4.0 license.

Filtering sequences with substandard quality was a critical step in the data processing. Indeed, all sequencing technologies including ion semiconductor sequencing can lead to insertion/deletion errors that are often caused by inacurate flow calls occuring during the sequencing of homopolymeric regions. In order to detect these events, sequencing data are provided with Q-values, which indicate the confidence of each scanned nucleotide. Q-values are indicated with a single-character Phred-

based quality score. Rentero et al used this information to develop a function in which reads containing low confidence basecalls can be removed. Practically, the filter excludes all sequences with more than x (X ≥ 1) bases having a quality score below a certain value, often chosen as Q18. Both parameters can be modulated according to the need of the user. The remaining sequences are translated and sorted by abundance.

In a second script, reads arising from sequencing or PCR errors could be eliminated. These reads overrule the quality filtering and are still easily spotted because they have identical nucleotide sequences except for one or two positions (Figure 17). Indeed, with the size and the diversity of our standard libraries, the probability to have two clones with almost the exact same nucleotide sequence is close to zero. Peptides present in high copy number will engender more sequencing errors that can be misleading and result in the generation of incorrect consensus group. The fixingerrors.m function enables one to pool together all DNA sequences differing only by a specified number of bases determined by the user (Figure 17). The procedure is iterative and time consuming when large data set are analyzed. For this reason, the number of sequences to correct can also be selected and the most abundant clones are processed first.

Figure 17. Example for the identification of false consensus sequences due to sequencing errors. In this selection, the most abundant sequence (present 4592 times) was clustered with sequences differing in only 1 nucleotide and being present at much lower frequency. These sequences likely resulted from sequencing errors. The fixingerrors.m function fount 467 erroneous reads arising from this parental sequence. Reprinted (open access) from[111], this article is available under the terms of CC-BY 4.0 license.

The number of cysteines found in phage-selected bicyclic peptides can often indicate how they have been cyclized. For example, oxidized peptides often are selected with four cysteines. In addition, some formats are preferentially enriched when performing panning with pooled libraries containing different ring sizes. As a result, a script was written in order to be able to classify and separate peptides according to these parameters. This allows the identification of peptide topologies favored during selections against specific targets.

One of the most important achievements was the creation of a function allowing the clustering of peptides into consensus groups. The script is based on the calculation of pair-wise distances between peptides. Subsequently, it generates a phylogenetic tree and finally organizes clones according to particular sequence motifs. Several parameters can be varied by the user such as the number of sequences to cluster, the stringency, the minimal size of a cluster etc. However, the number of clones that can be analyzed for similarity is usually inferior to 1000 due to the long processing time.

Finally, the *CommonSeq.m* function allows the comparison of two to three data set with each other. This was used to compare two parallel independent selections against the same target and facilitated the identification of target binding motifs from noisy dataset.

Since its development, NGS of phage-selected bicyclic peptides has been particularly valuable for the identification of rare target binding peptide motifs and choose appropriate peptide leads. This has been the case for selections against β-catenin[87], sortase A[84] and other unpublished projects.

2. Aim of the thesis

The goal of my thesis was to develop new methods for the phage display selection of (bi)cyclic peptides. Specifically, I aimed at the following three objectives: (i) increasing the scaffold diversity of phage-encoded bicyclic peptide libraries by cyclizing peptides with two chemical bridges, (ii) selecting peptide macrocycle ligands using phage libraries with unprecedented small size and (iii) bypassing bacterial infection in phage display by sequencing DNA released from phage particles. These three aims are described in details in the following paragraphs.

(i) Small structural diversity is a major limitation of combinatorial peptide macrocycle libraries. This problem can result in the isolation of peptides with suboptimal affinity, specificity, and/or stability. In most cases, our laboratory has successfully solved these problems by varying the backbone of the macrocycle. This has been achieved by cloning libraries with different ring size[78] as well as by using different chemical linkers[81]. At the beginning of my PhD, our laboratory was in possession of seven different libraries of bicyclic peptides, each of them having different combinations of ring sizes. In addition, two commercially available trivalent linkers – TBMB and TATA – could be employed to cyclize phage exposed peptides. All together, these libraries and linkers could be applied to generate phage-encoded libraries of bicyclic peptides containing around 40 different peptide backbones. We hypothesized that a much larger scaffold diversity would enable the identification of ligands with improved binding properties. Consequently, the first aim of my PhD was to address this problem by developing a new strategy that provides rapid access to thousands of different bicyclic peptide formats. Together with Dr S. Kale, we established a method in which linear peptides containing four cysteines were easily cyclized with two bivalent linkers. As a result, an enormous scaffold diversity could be obtained because the two bridges divide a peptide into three segments that all can be varied in length in a combinatorial fashion. In addition, the cyclization of peptides with four cysteines leads to three regioisomers which would triplicate the diversity. Finally, at least 20 symmetrical bivalent linkers, having two potential thiol-reactive groups, are commercially available. Following this strategy, I performed phage display against plasma kallikrein using libraries cyclized with six of the structurally diverse linkers, tested and validated by Dr. Kale.

(ii) The development of peptide therapeutics that are orally available has been a long standing goal in the pharmaceutical industry. Most cyclic peptides are not orally available because they are too large and/or polar to passively diffuse through cell membranes. Among all factors influencing cyclic peptide passive permeability, the molecular size and the number of H bond donors (HBD) are the most important. Peptides with small molecular mass (500 < MW < 1300) and fewer HBD are more likely orally absorbed[124]. Nevertheless, developing cyclic peptide ligands having a small size is

challenging as they have fewer amino acids that can potentially interact with the target. In addition, generating large combinatorial libraries of small cyclic peptides is difficult, because cyclization strategies often require bridges between two amino acid side chains (e.g two cysteines) which signifies that not all amino acids are free for diversification. In this project, my aim was to develop a strategy allowing the creation of large phage-encoded libraries of small cyclic peptides that promises intestinal absorption. Large combinatorial diversity was achieved using a newly established cyclization reaction occurring between the side chain of a C-terminal cysteine and the N-terminal amino group. Unlike other cyclization strategies, this method maximizes the ratio between the number of target-interacting amino acid side chains and the total number of residues. In order to validate this approach, phage-encoded libraries of the format XXXC and XXXXC were cyclized with eight structurally different linkers and subjected to affinity selection against the two disease targets plasma kallikrein and coagulation factor XIa.

(iii) In a third project, I intended to verify which fraction of phage clones could be decoded by directly sequencing DNA from phage particles isolated after panning. This was motivated by the fact that the identification of phage-selected clones invariably requires re-infection into competent *E.coli* before extraction and sequencing of plasmid dsDNA. This can lead to three major complications. First, some phage selection procedures depend on phage with mutated coat protein that can lead to the generation of poorly infective virions. Second, this restricts the type of post-translational chemical modifications that can be applied to phage. Finally, re-infection can create a bias for certain peptide sequences and affect the read out of the sequencing results. The two first points are particularly important in our laboratory since we are reacting a broad range of chemical linkers with peptides displayed on mutated pIII coat proteins of phage and we often observe an impact on phage infectivity. Therefore, I tested a new strategy for bypassing this bacterial infection step via direct sequencing of the ssDNA contained in phage particles. In this work, high-throughput sequencing was used to quantify the fraction of phage clones that could be identified using this approach.

3. Cyclization of peptides with two chemical bridges affords large scaffold diversities

3.1 – Abstract

Successful screening campaigns depend on large and structurally diverse collections of compounds. In macrocycle screening, variation of the molecular scaffold is important for structural diversity, but to date it is challenging to diversify this aspect in large combinatorial libraries. Herein, we report the cyclization of peptides with two chemical bridges to provide rapid access to thousands of different macrocyclic scaffolds in libraries that are easy to synthesize, screen, and decode. Application of this strategy to phage-encoded libraries allowed for the screening of an unprecedented structural diversity of macrocycles against plasma kallikrein, important in the swelling disorder hereditary angioedema. These libraries yielded inhibitors with remarkable binding properties (sub-nM Ki, > 1,000-fold selectivity) despite the small molecular mass (~ 1200 Da). An interlaced bridge format characteristic of this strategy provided a high proteolytic stability ($t_{1/2}$ in plasma > 3 days), making double-bridged peptides potentially amenable to topical or oral delivery.

3.2 – Introduction

Macrocycles have emerged as an interesting therapeutic class because they can bind to challenging targets that are not easily accessible to traditional small molecule compounds[4,125]. They owe their favorable binding properties to a larger size that enables interactions with extended surfaces where small molecules cannot normally bind and a ring-shaped structure that limits the conformational flexibility, and thus the entropic penalty, upon target binding[126]. The vast majority of macrocycle drugs are natural products or derivatives thereof[127]. For many targets, there are no available natural product-derived macrocycles, or the high complexity of the natural product hampers their synthetic modification and optimization. Thus, there is a need for synthetic macrocycles that are developed in a combinatorial fashion, which can greatly reduce the overall complexity. The necessary library diversity required to make these combinatorial macrocycles can be acquired by incorporating variable backbone structures (macrocyclic scaffold), by decorating the scaffolds with diverse functional groups, and by attaching peripheral substituents outside of the main ring. The variation of the macrocyclic scaffold is considered to be the most important determinant of structural diversity[128]. Therefore, a key step for achieving a high structural diversity in combinatorial macrocyclic libraries is to find a way to incorporate these variable scaffolds into the screening system. To this end, a range of elegant diversity-oriented synthesis strategies were published in recent years[129]. However, the

development of strategies for diversifying the scaffolds in large combinatorial libraries comprising ideally millions of macrocycles remains an important challenge.

Display technologies such as phage and mRNA display have enabled the generation and screening of enormously large numbers of peptide macrocycles, typically reaching several billion compounds at a time[48,35,130,131]. Efficient chemical reactions allow for the transformation of genetically encoded linear peptides into cyclic structures, which can provide access to variably cyclized peptides while retaining the benefits of the genetically encoded library[45,132]. Typically, two or three amino acids are ligated to generate mono- or bicyclic peptide libraries. By combining orthogonal reactions, more complex macrocycle structures have also been generated. Recent technology innovations have enabled the incorporation of unnatural amino acids into mRNA display, expanding the diversity of chemical groups that can be used to build these types of libraries[133]. To this point, however, a major limitation of phage or mRNA display macrocyclic libraries is the low scaffold diversity. Due to the dependence on ribosomal translation, the backbones of the macrocycles are required to be polypeptidic in nature. In most of the reported libraries, the macrocycles contain a single polypeptide scaffold, and the diversity is based only on the variation of the amino acid side chains. By varying the number of amino acids in the random peptides[53,78,134] or the chemical cyclization linkers[81], libraries comprising several different scaffolds were generated, but none of the reported libraries contains more than a dozen different scaffolds.

Here we report an efficient and robust strategy for generating phage-encoded macrocyclic libraries that contain thousands of different scaffolds. We cyclize peptides through two chemical bridges that connect two pairs of cysteines (Figure 18a). The two bridges subdivide the peptides into three segments containing m, n and o random amino acids, respectively (highlighted in red, orange and green in Figure 18a). Altering the number of random amino acids in each segment in a combinatorial fashion allows for the generation of a much larger scaffold diversity than when peptides are divided by only two or three cysteines into one or two segments, respectively. Additional structural diversity stems from the connectivity between the cysteines, due to the fact that four cysteines can be connected by two chemical linkers three different ways. The greater number of different bicyclic peptide scaffolds that can be generated with two chemical bridges as opposed to only one is exemplified for example 9-amino acid peptides in Figure 18b. While cyclizing nonapeptides with one bridge (3 cysteines, 6 random amino acids) yields only 7 different bicyclic scaffolds, two bridges (4 cysteines, 5 random amino acids) yields 63 different bicyclic scaffolds. This already provides a nine-

fold improvement in diversity that will increase with an increasing peptide size. For peptide macrocycles between 8 and 14 amino acids in length (including the four cysteines), a total of 798 different di-bridged bicyclic peptide formats can be obtained versus the 63 that are possible with a mono-bridge connecting three cysteines. The scaffold diversity can be further increased by cyclizing the same peptide library in parallel with many different chemical bridges that can further vary the ring size as well as include functional groups that can participate in binding. Dozens of reagents with two thiol-reactive groups are commercially available[135], in direct contrast to the only two commercially available reagents with three symmetrical thiol-reactive groups[81]. This calculation based on the cyclization of 8- to 14-mer peptides with 10 different bridges has the potential to yield > 7000 different macrocyclic scaffolds.
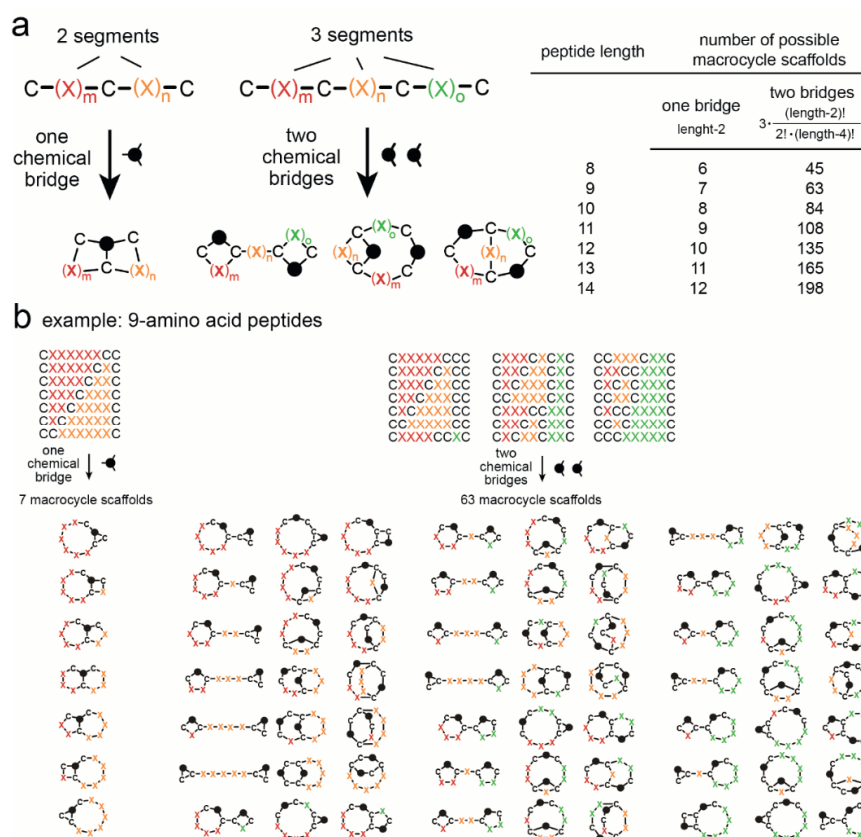
Figure 18. Cyclization of peptides with two chemical bridges. a) Cyclization of peptides with two chemical bridges (connecting four cysteines) yields a much larger number of bicyclic peptide scaffolds than cyclization with one bridge (connecting three cysteines). The number of different macrocycle scaffolds is indicated for peptides of different lengths. b) Illustration of the large macrocycle diversity with a double-bridged nonapeptide as an example.

An important goal of our laboratory is the development of small, highly stable peptide macrocycles that can be applied topically or orally. Developing such small and compact ligands can

be challenging because they have fewer amino acids that can interact with targets. We speculated that this handicap may be overcome by a molecular shape that is perfectly complementary to a target binding site, and that such molecules could potentially be isolated from macrocycle libraries with high scaffold diversities. We thus applied the "double-bridge" cyclization strategy to small-sized peptides and used these to isolate inhibitors of plasma kallikrein. This protein is an important target of the swelling disorder hereditary angioedema to which injectable, protein-based inhibitors are already in clinical use[136,137] or development[138] but no small molecule drug is currently available.

3.2 – Results

3.2.1 Reactivity of the cyclization reagents
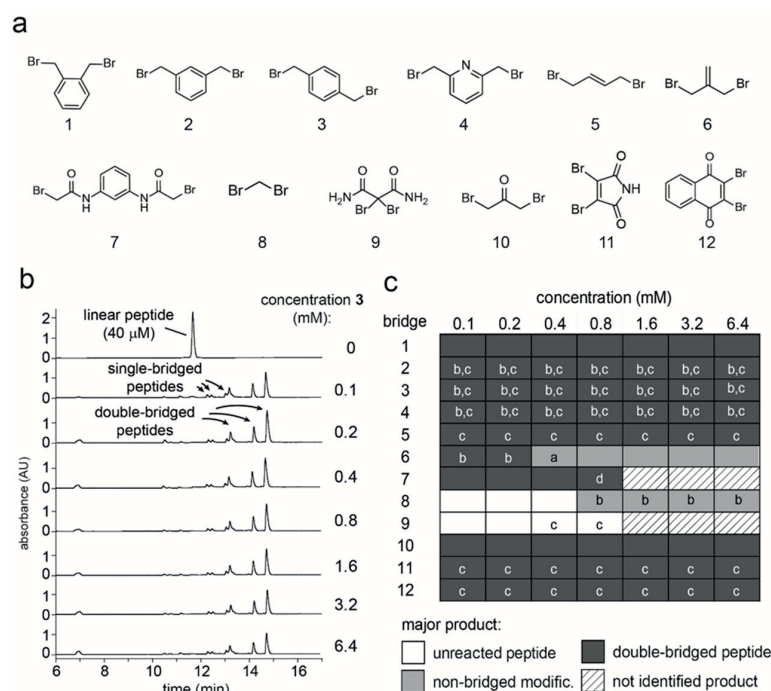
Figure 19. Cyclization reagents and reaction efficiency. a) Chemical linkers each containing two thiol-reactive groups. b) Reaction of the model peptide ACSRCVECGWCG-NH2 with the indicated concentrations of chemical linker 3. Products were analyzed by analytical HPLC. At all reagent concentrations, the major products were the three desired double-bridged peptides. c) Products formed upon incubation of the model peptide with twelve different chemical reagents at the indicated concentrations. Fields shaded with dark gray indicate that the major product was the double-bridged peptide (sum of three isomers > 50% of total product). Minor products are indicated with small letters: a. bicyclic peptide, b. non-bridged modification, c. single bridge, d. not identified. – Data produced by Dr S. Kale

A number of reagents containing thiol-reactive groups were reported to efficiently cross-link or cyclize peptides and proteins via cysteines in aqueous buffer and under mild conditions[139–142]. Of those, we chose 12 structurally diverse compounds and tested their ability to quantitatively and selectively link two pairs of cysteines in a model peptide in conditions that are compatible with phage display (ACSRCVECGWCG-NH$_2$; Figure 19a). Eight out of the 12 reagents produced the double-bridged peptide as the main product at a wide range of concentrations, suggesting that the reactions are robust and would work efficiently with peptides of variable sequences when displayed on phage (Figure. 19b and c). The observed side products were mostly peptides with only one bridge and, at higher reagent concentrations, peptides that were modified with more than two linkers such that the cysteines were not bridged.

3.2.2   Impact of cyclization reagents on phage infectivity
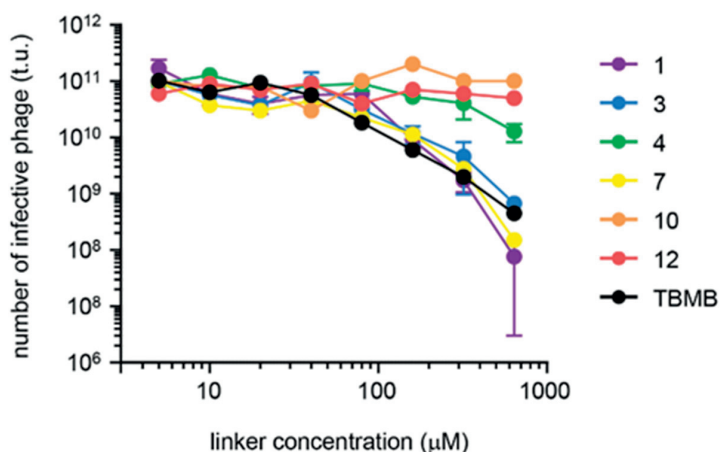
Figure 20. Influence of chemical modification on phage infectivity. 10^{11} t.u. of phage from library 1 were incubated with different concentrations of the linkers in 80% aqueous buffer (20 mM NH$_4$HCO$_3$, pH 8) and 20% ACN for 1 h and the number of infective phages was determined.

Most of the thiol-reactive compounds did not inactivate filamentous phage or only interfered at concentrations far above those needed to fully cyclize the model peptide (Figure 20).

3.2.3   Phage selections with various cyclization linkers

We applied the compounds **1**, **3**, **4**, **7**, **10** and **12** to two peptide phage display libraries each comprising more than 100 million random peptides of the formats XCX$_3$CX$_3$CX-phage (library 1) and XCX$_4$CX$_4$CX-phage (library 2). In these libraries, there was a 20% and 23% chance, respectively, for the peptides to contain four cysteines due to the probability of a cysteine occurring in the random
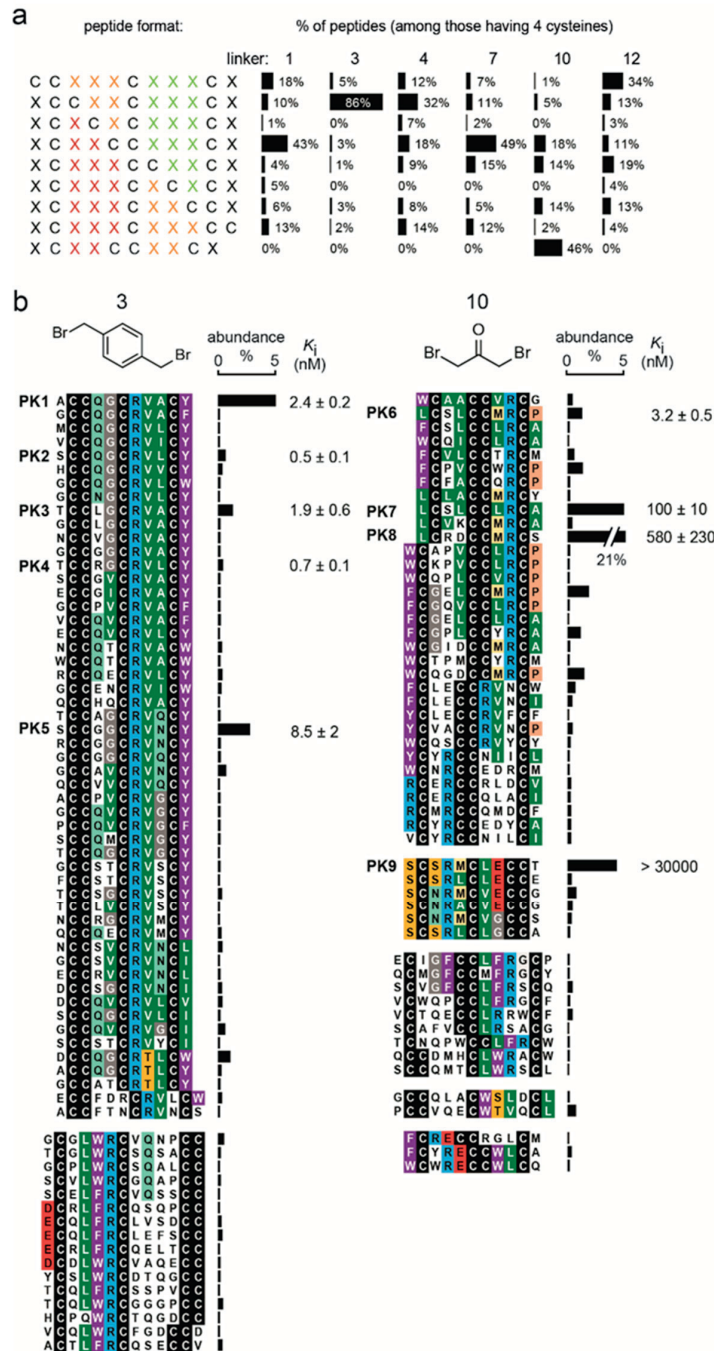
Figure 21. **Phage selection of double-bridged peptides.** a) Abundance of peptide formats after two rounds of panning library 1 cyclized with the indicated linkers against plasma kallikrein. Abundance is indicated in % of those peptides containing four cysteines. Abundance of peptide formats enriched from library 2 is shown in Supplementary Figure 3.2. b) Peptides isolated from libraries 1 and 2 after two rounds of panning using linkers 3 and 10. Peptides with similar sequences are aligned in groups and amino acid similarities highlighted in color. Peptide formats isolated with four other linkers are shown in Supplementary Figure 3.3.Ten peptides were synthesized and characterized. The K_I of the most active isomer of each peptide is indicated. Standard deviations of at least three K_I measurements are shown.

positions. Cyclization of the peptides yielded 54 different backbone formats for each linker (Supplementary Figure 3.1), producing 324 different macrocyclic scaffolds in total. The two libraries each cyclized with the six different linkers were individually panned against immobilized plasma kallikrein. High-throughput sequencing of phage isolated after two rounds of selection showed an enrichment for peptides containing four cysteines, that the fourth cysteine was localized to certain positions depending on the thiol-reactive reagent used (Figure 21a for library 1; Supplementary Figure 3.2 for library 2), and that peptides shared strong consensus sequences (Figure 21b and Supplementary Figure 3.3). These findings suggested that the library peptides were efficiently cyclized on phage, due to the strong prevalence of four cysteines, and that target-selective peptides were isolated because of the consensus. In phage selections with linkers **3** and **10**, certain peptide formats were particularly enriched (Figure 21b). For example, most peptides cyclized with linker **3** contained the fourth cysteine in amino acid position 3. In the selection with linker **10**, many of the peptides isolated had the truncated format XCX_2CCX_2CX, which contained 10 instead of 11 amino acids. Such truncated peptides are generated during library cloning through erroneous DNA primers and occur only rarely in the finished library. The strong enrichment of some peptide formats, especially such rare formats, suggested that certain molecular scaffolds are particularly suited for target binding. This finding was the first hint that the high scaffold diversity obtained through the double-bridge strategy was key for the isolation of binders.

3.2.4   Binding affinity of selected bicyclic peptides

We synthesized several peptides and cyclized them by randomly bridging two pairs of cysteines through all three possible combinations, and each isomer was separately isolated and tested for binding. We picked the short peptides of 10 or 11 amino acids isolated from library 1 cyclized with linkers **3** and **10** due to our interest in small macrocycles (~1200 Da). The K_Is of the most active isomer for each peptide are indicated in Figure 21.b. Most of the peptides had activities in the nanomolar range, with some reaching sub-nanomolar K_Is. For each peptide, there was obviously one isomer that was the selected format, as one of the three tested isomers proved to be the best inhibitor for plasma kallikrein. Because it was so important for the activity, the connectivity of two peptides, PK4 and PK6, was deciphered by synthesizing the three isomers using orthogonal cysteine protecting groups Mmt and Dpm (Supplementary Figure 3.4 and 3.5). An efficient procedure including one on-resin cyclization was established that enabled the synthesis of double-bridged peptides in two days. It was determined that for PK4, isomer 3 (Cys1/Cys3, Cys2/Cys4) inhibited the protease with a K_I of 0.7 ±

0.1 nM while the other two isomers showed at least a 200-fold weaker inhibition (Figure 22a). In PK6, isomer 1 (Cys1/Cys2, Cys3/Cys4) was by far the most active $(K_I = 3.2 \pm 0.5$ nM) (Figure 22a). Not surprisingly, the most active inhibitors of a consensus group had the same cysteine connectivity and thus the same macrocyclic scaffold.
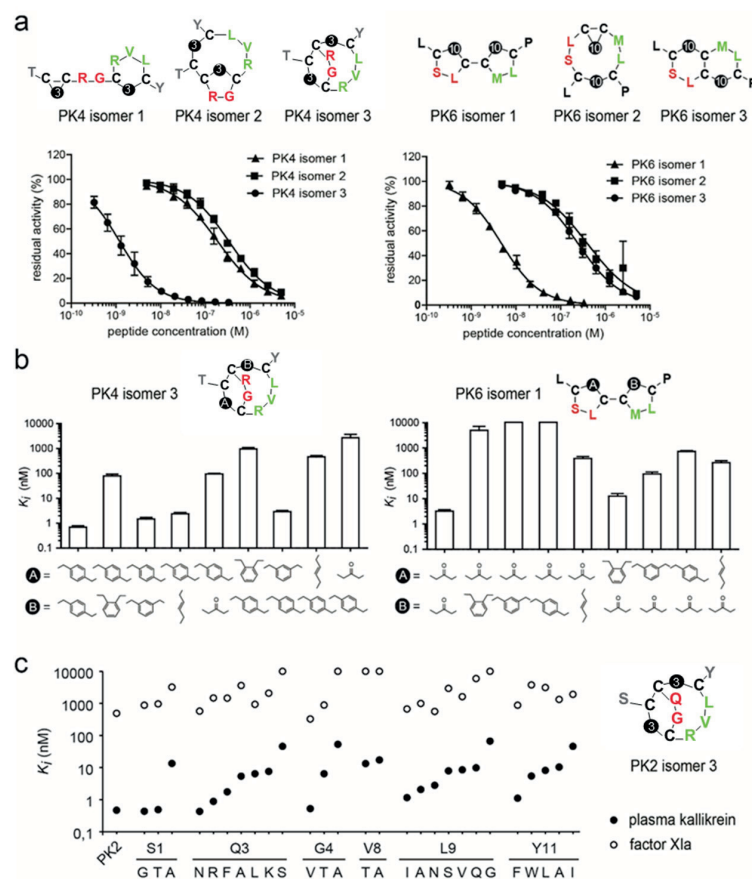
Figure 22. Role of chemical bridges. a) Structures of the three isomers of PK4 and PK6. One out of the three isomers efficiently inhibits plasma kallikrein. HPLC traces of the active isomers are shown in Supplementary Figure 3.5. b) Replacement of the chemical bridges in PK4 and PK6 reduced the inhibitory activity to various extents. c) SAR study of PK2 (isomer 3). Inhibition of plasma kallikrein and FXIa by PK2 variants with the indicated amino acid substitutions. – Data shown in b and c were produced by Dr S. Kale.

3.2.5   Structural and functional role of the cyclization linker

The convergence of peptides to certain consensus groups based on the use of different linkers indicated an important structural and functional role for the chemical bridges. We assessed the role of the two linkers in the double-bridged peptides PK4 (isomer 3) and PK6 (isomer 1) in a structure-activity relationship (SAR) study in which we replaced one linker at a time to a set of diverse linkers

(Figure 22b, Supplementary Figures 3.6, to 3.8), some of which rendered the peptide macrocycles entirely inactive at the highest concentration tested (1 μM). This linker swapping experiment showed that both of the two bridges are important. We then wondered if linker substitution could be a strategy for enhancing the macrocycles' inhibitory activities. To test this, we substituted linker **3** in the most active macrocycle PK2 (isomer 3; $K_I = 0.5 \pm 0.1$ nM) with *para*-dibromomethyl-benzene linkers **13**-**19** carrying diverse groups attached to the benzyl ring (Figure 23). While none of the macrocycles had an enhanced activity, we discovered that some of them again had a dramatically reduced activity, showing that even minor structural changes in the linker, such as the addition of small substituents, substantially impact the macrocycle activity. This suggested that linkers **13**-**19**, despite the similarity to **3**, should all be applied in parallel in future phage selections in order to generate even larger diversities.
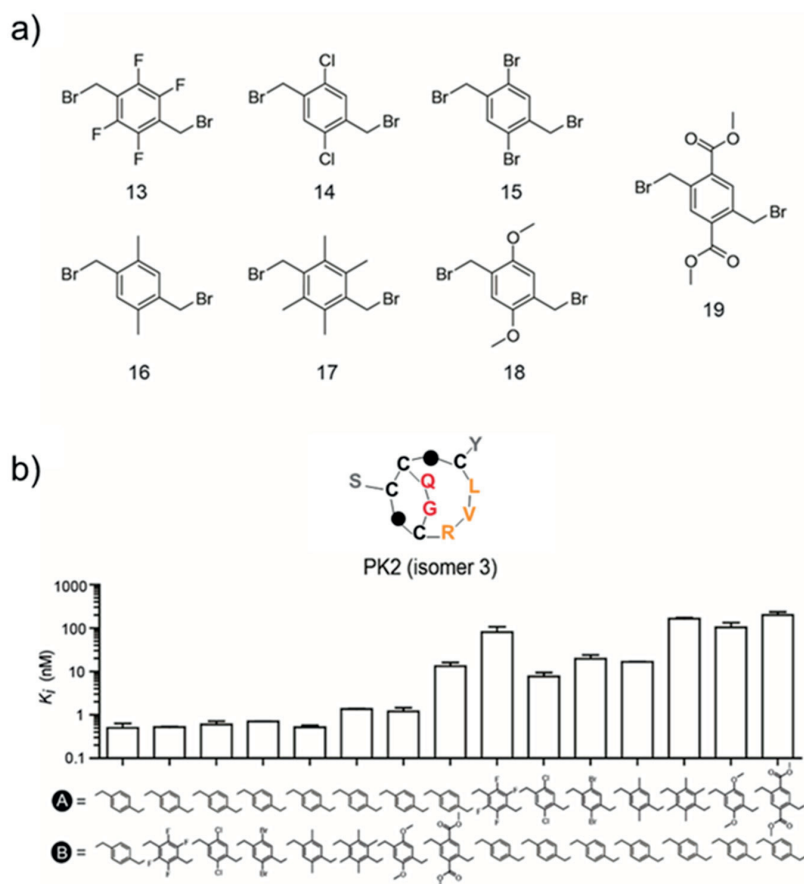
Figure 23. Cyclization of peptides with linkers having similar structures. a) Chemical reagents resembling linker 3. b) Inhibition of plasma kallikrein by bicyclic peptide PK2 (isomer 3) in which one of the two linkers was substituted.

3.2.6   Target selectivity of bicyclic peptides

The macrocycles PK2, PK4, and PK6 showed high target selectivity in a specificity profile performed with a panel of eight homologous trypsin-like serine proteases (Table II). Six of the proteases were not inhibited at all. Even coagulation factor XIa (FXIa), sharing the highest sequence identity with plasma kallikrein (69% amino acid homology in the catalytic domain), was only weakly inhibited, and the macrocycles still demonstrated around 1000-fold selectivity for plasma kallikrein over FXIa. We subsequently investigated the structural determinants of the target specificity for one of the macrocycles, PK2. Specifically, we tested if PK2 could be turned into a FXIa inhibitor by mutating different amino acids (Figure 22c), but none of the substitutions improved the inhibition of FXIa. Detrimental mutations reduced the inhibition of both proteases to similar extents, indicating that the macrocycle binds both proteases through similar key contacts. Apparently, the backbone formed by a $XCCX_2CX_3CX$ peptide double-bridged by linker **3** fits perfectly to the active site of plasma kallikrein and contributes to the target selectivity.

	K_I (nM)			
	PK2	PK4	PK6	PK10
plasma kallikrein	0.5 ± 0.1	0.7 ± 0.1	3.2 ± 0.5	3.6 ± 0.5
factor XIa	580 ± 180	1300 ± 500	2700 ± 200	2500 ± 100
factor XIIa	> 30000	> 30000	> 30000	> 30000
thrombin	> 30000	> 30000	> 30000	> 30000
uPA	> 30000	> 30000	> 30000	> 30000
tPA	> 30000	> 30000	> 30000	> 30000
plasmin	> 30000	> 30000	> 30000	> 30000
factor Xa	> 30000	> 30000	> 30000	> 30000
factor VIIa	> 30000	> 30000	> 30000	> 30000

Table II. Target specificity. Inhibition of plasma kallikrein and a panel of structurally homologous or physiologically important paralogous proteases by isolated inhibitors. Average values and standard deviations of at least three measurements are indicated. – Data produced by Dr S.Kale.

3.2.7   Proteolytic stability of selected bicyclic peptides

Short peptides constrained by two linkers promised a high proteolytic stability due to the inaccessibility of the peptide backbone to proteases. Indeed, some of the double-bridged peptides, like PK2 (isomer 3), showed an impressive stability in human plasma at 37°C (Figure 24a). While the two

exocyclic amino acids were cleaved by proteases, first the N-terminal serine ($t_{1/2}$ = 29 min) followed by the C-terminal tyrosine ($t_{1/2}$ = 6.9 hours), the double-bridged macrocycle core remained intact after two days.
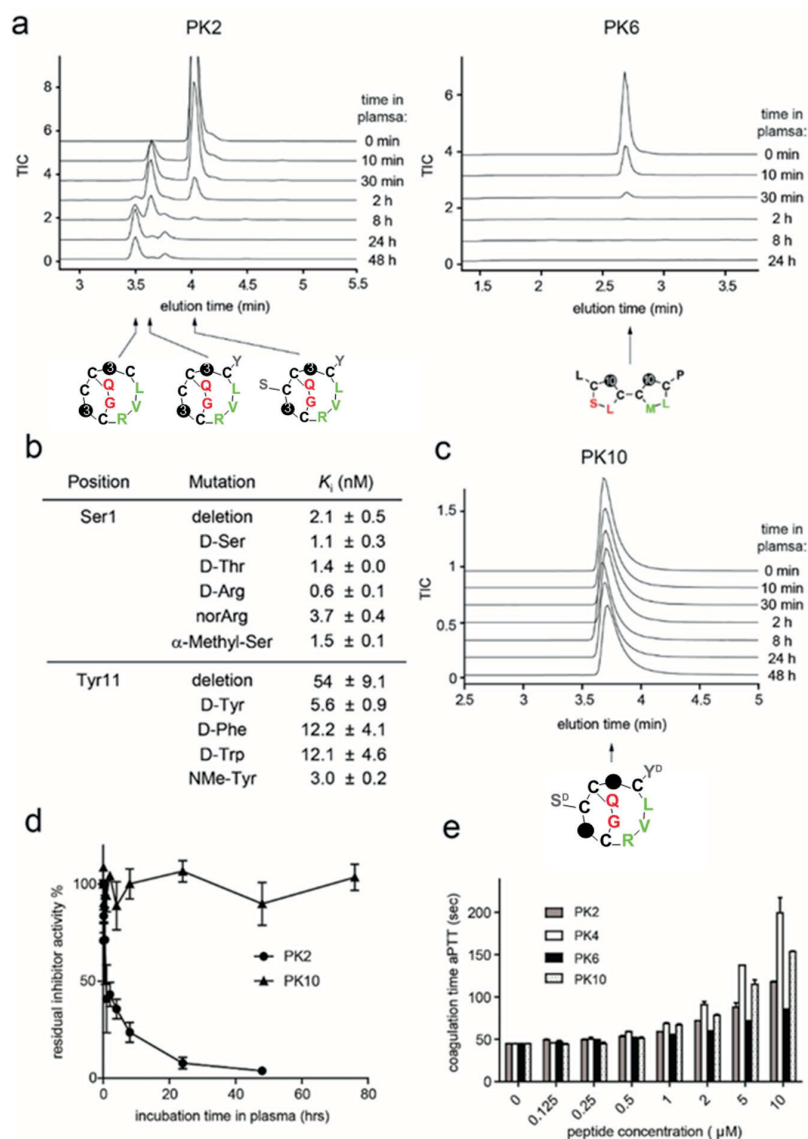
Figure 24. **Stability of double-bridged peptides.** a) Proteolytic degradation of PK2 (isomer 3) and PK6 (isomer 1) in human plasma analyzed by LC-MS. Total ion count (TIC) in SIM mode of identified species is shown. b) Inhibitory activities of N- or C-terminally modified PK2 (isomer 3). Average values and standard deviations of K_is were calculated from at least three measurements. c) LC-MS analysis of PK10 after incubation in human blood plasma. d) Fraction of functional PK10 after incubation in human blood plasma as determined in a FXIIa inhibition assay. e) Inhibition of the intrinsic coagulation pathway. Coagulation time (aPTT) was measured in human blood ex vivo at different concentrations of the indicated plasma kallikrein inhibitors. – Data produced by Dr S. Kale.

The high stability of PK2 most likely results from interlaced bridging that tightly connects the two macrocycles, reducing their conformational flexibility and/or rendering them less accessible to proteases. This was confirmed by testing a peptide, PK6 (isomer 1), lacking the interlaced configuration due to bridges between Cys1/Cys2 and Cys3/Cys4, which was rapidly degraded ($t_{1/2} \sim 5$ minutes; Figure 24b). For PK2, the loss of the terminal amino acids reduced the activity of the peptide 4-fold (Ser1) and 108-fold (Tyr11), respectively (Figure 24c). To mitigate this loss, we tested a small set of peptides in which the vulnerable exocyclic residues were replaced by D-amino acids (Figure 24c). Appending D-Arg at the N-terminus and D-Tyr at the C-terminus increased the inhibitory activity 1.9- and 9.6-fold, respectively. The resulting macrocycle PK10 had a K_I of 3.6 ± 0.5 nM, compared to 0.5 ± 0.1 nM for the PK2 precursor, though remained fully intact upon incubation in plasma ($t_{1/2} > 3$ days; Figure 24c and d).

3.2.8   Inhibition of the intrinsic coagulation pathway

Plasma kallikrein plays a role in various diseases including hereditary angioedema (HEA), diabetic macular edema, and thrombosis, and contributes to contact activation in extracorporeal circulation during procedures such as coronary artery bypass grafting. Two protein-based inhibitors, C1 inhibitor and ecallantide, are approved for the treatment of acute HEA attacks[136,137]. Several drug development programs are ongoing to generate antibody (e.g. lanadelumab) or small molecule inhibitors of plasma kallikrein[138,143]. We tested if the double-bridged peptides could block the activation of the intrinsic coagulation pathway in human plasma, which is dependent on the reciprocal activation of plasma kallikrein and factor XII (FXII). Human plasma was incubated with inhibitor, the activation of FXII was triggered by Pathromtin®, and the time to coagulation was measured to determine the activated partial thromboplastin time (aPTT). The peptides PK2, PK4, PK6, and PK10 all efficiently inhibited activation of the intrinsic coagulation pathway in human plasma ex vivo (Figure 24e); PK10 doubled the aPTT (EC$_{2x}$) at a concentration of 3 μM.

3.4 – Discussion

We show that macrocyclic peptide libraries with a large structural diversity can be generated by cyclizing peptides with two chemical bridges that each connect a pair of cysteines. This approach yielded libraries comprising far more different macrocyclic scaffolds than any previously developed libraries, which tend to be based on one or, at the most, a handful of different scaffolds. High-

throughput sequencing of isolated peptides revealed a strong preference for certain peptide formats, clearly showing the importance of the large scaffold diversity. Even though the new strategy generates three macrocyclic isomers that are all encoded by the same phage DNA, we show that multiple products do not impair the phage panning procedure, that the three isomers can be efficiently synthesized with an orthogonal protection strategy, and that the active isomer can easily be identified.

As a direct comparison, similarly sized bicyclic peptides (11 amino acids) that were cyclized by only one chemical linker (connecting three cysteines) were previously developed to the same target by screening the same type of library (XCX_3CX_3CX)[77]. The best bicyclic peptide plasma kallikrein inhibitor developed through these pans had a K_I of 5.2 nM and thus a 10-fold weaker binding affinity than the di-bridged peptides. It is likely that the affinity improvement can be attributed to the larger scaffold diversity that was generated and sampled with the new approach. A plasma kallikrein inhibitor with a comparable affinity ($K_I = 0.3$ nM) was previously obtained only by screening a library of much longer bicyclic peptides that form a larger binding interface with the target (15 amino acids of the form XCX_5CX_5CX cyclized by one linker)[77]. The development of smaller macrocycles with a molecular weight approaching 1 kDa is of interest for the development of drugs amenable to topical or oral administration. Several orally available peptide macrocycles, including vacomycin (1.5 kDa) and cyclosporin (1.2 kDa), have molecular weights in this range, suggesting that phage selected double-bridged peptides might be applied as leads for the development of topical or oral drugs.

The SAR study revealed that both of the two chemical bridges in the bicyclic peptides are essential for binding, wherein one bridge was slightly more important in all of the examples studied. Substitution to structurally similar bridges reduced the binding affinity, suggesting that cyclizing peptide phage libraries in parallel with similar reagents such as compounds **13-19** will allow for the generation of even larger macrocycle diversities. A SAR study altering the amino acid sequence of double-bridged peptide PK2 showed that its high target selectivity was based, at least to some extent, on the specific peptide backbone architecture. None of the numerous amino acid mutations tested increased the affinity for the homologous protease FXIa. This is in stark contrast to the previously developed single-bridged bicyclic peptide inhibitors of plasma kallikrein that also inhibited FXIa when specific amino acid positions were mutated[77]. This observation is an additional indication that the larger scaffold diversity allowed for the identification of ligands that are perfectly complementary in shape and polarity to the target binding site.

An interesting feature of the ligands with interlaced bridges is the high proteolytic stability. The core structure of PK2 resisted protease degradation when it was incubated for several days in human

plasma at 37°C. This high stability is likely due to the close connection of the two macrocyclic rings, resulting from interlacing bridges Cys1/Cys3 and Cys2/Cys4. This hypothesis is further supported by the low proteolytic stability of the double-bridged PK6 that contains two independent monocycles connected via a flexible linker. The enormous stability of the interlaced peptide macrocycle format might be exploited for the development of oral drugs that must survive the proteolytic pressure in the gastrointestinal tract. This could include either drugs that are absorbed into the bloodstream or that act on targets in the gastrointestinal tract. The feasibility of peptide-based drugs acting in the lumen of the gastrointestinal tract is demonstrated with the recently approved blockbuster drug linaclotide. This 14-amino acid peptide derived from nature and stabilized by three disulfide bridges is orally applied. The low stability for exocyclic amino acids, such as the Ser1 and Tyr11 in PK2, should therefore not be a limitation as they can be substituted to non-natural building blocks that resist proteolytic cleavage, as demonstrated in this work. Alternatively, exocyclic amino acids may be completely omitted and the loss in affinity compensated by substituting amino acids in the core to unnatural ones that can more efficiently bind the target. The latter strategy would allow a further reduction of the molecular weight of the ligands, and would reduce the size of PK2 to below 1 kDa.

In summary, we have developed an efficient strategy for the generation and screening of macrocyclic peptide libraries comprising an unprecedented scaffold diversity. We show that ligands with a better affinity and target selectivity can be isolated from such libraries despite a rather small molecular mass of the compounds, and that a scaffold format with interlaced macrocyclic rings has a particularly high stability. The combined properties of the generated ligands including high affinity and selectivity, small size, and high proteolytic stability make them interesting as drug development leads.

3.5 – Materials and methods

Cyclization of model peptide with thiol-reactive reagents

10 μL of thiol-reactive reagents at concentrations ranging from 500 μM to 32 mM in acetonitrile (ACN) were added to 40 μL of 50 μM of the peptide ACSRCVECGWCG-NH$_2$ in NH$_4$CO$_3$ buffer (60 mM, pH 8.0) and incubated for 1 h at 30°C. All linkers were commercially available except for linker **7** which was synthesized as previously described[77]. Reactions were quenched with 40% formic acid (10 μL) and analyzed by LC-MS (LCMS-2020, Shimadzu). Injection volume: 1 μL, column: Kinetix®C18, 2.6 μm, 100 Å, 50x2.1 mm (Phenomenex), gradient: solvent A (99.95% H$_2$O, 0.05%

formic acid) and 10-60% of solvent B (99.95% ACN, 0.05% formic acid) over 5 min, flow: 1 mL min-1, MS analysis: positive mode. Potential detrimental effects of the thiol-reactive reagents on phage were assessed by incubating 10^{11} phage (transforming units, t.u.) with the reagents and measuring the number of infective phage. 20 μL of thiol-reactive reagents at concentrations ranging from 20 μM to 3 mM in ACN were added to 80 μL of 10^{11} t.u phage in 60 mM NH_4CO_3 buffer pH 8.0, and incubated for 1 h at 30°C. The number of functional phage was quantified by infecting 20 μL of exponentially growing *E. coli* cells with 10-fold dilutions of the phage and plating the cells on selective agar plates (30 μg.ml^{-1} chloramphenicol).

Phage selection of double-bridged peptides

Phage of library 1 and 2 were produced in separate 1 L *E. coli* cultures, pooled, purified by PEG-precipitation, and the cysteines in the peptides reduced as previously described[144]. Phage were dissolved in 50 mL aqueous buffer containing 20 mM NH_4CO_3, 5 mM EDTA, pH 8 and distributed into 6 tubes. The peptides were chemically modified by adding 2 mL of the cyclization reagents in ACN to 8 mL phage, and incubating for 1 h at 30 °C in a water bath. Modified phage were precipitated in 20% PEG, 500 mM NaCl, and dissolved in 3 mL washing buffer (10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.4). Human plasma kallikrein (Molecular Innovations) was biotinylated by incubating the protease at a concentration of 10 μM with a 15-fold molar excess of EZ-Link® Sulfo-NHS-LC-Biotin (ThermoFisher Scientific) for 1 h at RT and the excess of biotinylation reagent was removed by size exclusion chromatography. 2.5 μg of biotinylated plasma kallikrein was incubated with 50 μL magnetic streptavidin beads (Dynabeads® M-280 Streptavidin, ThermoFisher Scientific) in 500 μL of washing buffer for 15 minutes and non-immobilized protein was removed by washing three times. Both, phage and the beads were separately blocked by the addition of 500 μl of washing buffer containing 1% w/v BSA and 0.1% v/v Tween 20, and incubated for 30 minutes. Phage and magnetic beads were mixed and incubated for 1 h on a rotating wheel at RT. The beads were washed eight times with washing buffer containing 0.1% v/v Tween 20 and twice with washing buffer only. Phage were eluted by incubating the beads with 100 μL of 50 mM glycine, pH 2.2 for 5 min. The solution containing the eluted phage was neutralized by the addition of 50 μL 1 M Tris-Cl, pH 8.0 and added to 30 mL of exponentially growing TG1 *E. coli* cells (OD_{600} = 0.4). After a 90 min incubation, the freshly infected bacteria were plated on 2YT/chloramphenicol plates (30 μg.ml^{-1}) and incubated overnight at 37°C. Bacterial cells of the colonies were recovered the next day in a small volume of 2YT media and stored as glycerol stocks at -80°C. The second round of selection was

performed using neutravidin-coated magnetic beads to prevent the enrichment of streptavidin-specific peptides. Neutravidin beads were generated by coupling neutravidin to tosyl-activated magnetic beads according to the manufacturer's recommended protocol (Dynabeads M-280 Tosylactivated, Invitrogen Dynal Biotech AS).

High-throughput DNA sequencing of enriched phage

Phage DNA was extracted from E. coli cell pellets (around 50 mg) using a kit wherein the DNA was eluted with 5 mM Tris-HCl, pH 8.5. The DNA region encoding the peptides was amplified by PCR using 100 ng phage DNA as template, one of the following forward primers, and the reverse primer 5'-CCTCTCTATGGGCAGTCGGTGATTTTCAACAGTTTCAGCGGAGTG-3'. Each primer contains a sequence for annealing to the phage vector (italics) and an adapter sequence for ion semiconductor sequencing. The forward primers contain a barcode that encodes the cyclization reagent used in the phage selection (underlined), allowing for pooled sequencing of all isolated phage and deconvolution.

Bridge 1	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG<u>CTCTGA</u>*CGCAATTCCTTTAGTTGTTC*-3'
Bridge 3	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG<u>CGTCAT</u>*CGCAATTCCTTTAGTTGTTC*-3'
Bridge 4	5'-CCATCTCATCCCTGCGTGTCTCCGACTCA<u>GATACT</u>*CCGCAATTCCTTTAGTTGTTC*-3'
Bridge 7	5'-CCATCTCATCCCTGCGTGTCTCCGACTCA<u>GGACTTA</u>*CGCAATTCCTTTAGTTGTTC*-3'
Bridge 10	5'-CCATCTCATCCCTGCGTGTCTCCGACTCA<u>GTTCAGT</u>*CGCAATTCCTTTAGTTGTTC*-3'
Bridge 12	5'-CCATCTCATCCCTGCGTGTCTCCGACTCA<u>GCAGTT</u>C*CGCAATTCCTTTAGTTGTTC*-3'

A standard PCR reaction was performed (25 cycles, 30 s 95°C, 30 s 55°C, 30 s 72°C) in a volume of 50 µL containing Taq polymerase buffer, 250 µM dNTP, 500 nM primer, and 1 unit Taq polymerase. Three additional PCR cycles were performed with fresh reagents wherein 10 µL of the first PCR reaction served as a template, in order to limit the amount of DNA heteroduplex. PCR products were separated by agarose gel electrophoresis (2.5% gel UltraPure agarose, Invitrogen), extracted using a kit, and DNA elution with 5 mM Tris-HCl, pH 8.5. Ion semiconductor sequencing was performed by the Centre for Research in Agricultural Genomics (Barcelona, Spain) on an Ion

Personal Genome Machine (PGM™) sequencer. Sequence data was processed and sequences were aligned using MATLAB scripts[120].

Peptide synthesis

Peptides were synthesized on an Advanced ChemTech 348Ω parallel peptide synthesizer (AAPPTec) by standard Fmoc solid-phase chemistry on Rink Amide MBHA resin (0.26 mmol/g resin, 0.03 mmol scale). The coupling was carried out twice for each natural amino acid (4 eq.) using HBTU (4 eq.), HOBt (4 eq.) and DIEA (6 eq.) in 1.3 mL DMF with shaking at 400 rpm for 30 min. The coupling of unnatural amino acids (2 eq.) was performed once using HATU (2 eq.) and DIEA (4 eq.) in 1.5 mL DMF and shaking at 400 rpm for 1 h. Fmoc groups were removed by incubating the resin twice with 20% v/v piperidine in DMF with shaking 400 rpm for 5 min. Washing steps were performed with 3 mL DMF. The resin was washed four times before Fmoc deprotection and five times after deprotection.

Peptide cleavage and deprotection

Final peptides were deprotected and cleaved from the resin under reducing conditions by incubation in 90% TFA, 2.5% H_2O, 2.5% thioanisole, 2.5% phenol, 2.5% EDT with shaking for 4 h at room temperature. The resin was removed by vacuum filtration, peptides precipitated with cold diethyl ether (50 mL) and a 30 min incubation at -20°C, and pelleted by centrifugation at 4400 g for 5 min. The peptide pellets were washed twice with cold diethyl ether (30 mL) and lyophilized or directly purified by HPLC.

Peptide cyclization by bridging random pairs of cysteines

Purified peptide (9 mL, 1.2 mM) was dissolved in 40% v/v ACN and 60% v/v aqueous buffer (60 mM NH_4HCO_3, pH 8.0) and the cyclization reagent was added (1 mL, 40 mM in ACN). The final reagent and solvent concentrations in the reaction mixture were 1 mM peptide, 4 mM cyclization reagent, 50% v/v aqueous buffer, and 50% v/v ACN. The reaction mixture was incubated at 30°C in a water bath for 1 h and then lyophilized.

Peptide cyclization by bridging specific pairs of cysteines

Peptides with defined pairs of cysteines bridged by chemical linkers were synthesized as shown in Supplementary Figure 3.4 and described in the following. Peptides were synthesized on Rink amide resin as described above wherein the N-terminal Fmoc group was not removed. Two of the cysteines were protected at the side chain with Mmt and two with Dpm. After washing the resin with 5 mL DCM, the two cysteines with the Mmt groups were deprotected by shaking the resin in 5 mL of TFA:TIS:DCM (1:5:94) in a fritted syringe for 3 min. The solution was removed and the deprotection repeated 9 times. The resin was washed 3 times with DCM for 1 min and swollen by washing 4 times with DMF. 2 eq. of cyclization reagent in 3 mL DMF and 3 eq. of DIPEA in 1 mL were added together stepwise and the reaction shaken for 1 h at RT. The reaction solution was removed and the resin washed 4 times with DMF. The N-terminal amino group was deprotected by adding twice 3 mL of 20% piperidine in DMF for 5 min and shaking. The resin was washed 4 times with DMF and subjected to global deprotection in 5 mL of the cleavage mixture TFA:TIS:EDT:H_2O (97:1:1:2) for 12 h. In the case of linker 10 EDT was avoided in the cleavage mixture because of a reaction of EDT with the ketone group. The resin was filtered and peptide was precipitated by addition of 45 mL cold ether and incubation at -20 °C for 20 min. The peptide was pelleted by centrifugation at 4400 g for 5 min, washed twice with 30 mL ether and lyophilized. The crude peptide was dissolved in around 2 mL in a 1:1 mixture of H_2O and ACN to reach a concentration of 10 mM. 1.5 eq of cyclization reagent in 5 mL ACN and 9 mL of 60 mM NH_4HCO_3 pH 8 were added stepwise. If the solution was turbid, 4 mL of ACN were added. The reaction was incubated at 30 °C for 1 h. The reaction was quenched by addition of 1 mL of 40% HCOOH. The peptide was lyophilized and purified by RP-HPLC.

Peptide purification

Peptides were purified with a HPLC system (Prep LC 2535 HPLC, Waters) using a preparative C18 reversed-phase column (SunfireTM prep C18 OBD 10 μm, 100 Å, 19x250 mm, Waters) or a semi-preparative C18 reversed-phase column (X-bridge peptide BEH C18 5 μm, 300 Å, 10x250 mm, Waters) applying a flow rate of 20 or 6 mL.min^{-1} and a linear gradient of 10-50% v/v solvent B in 40 min (A: 99.9% v/v H_2O and 0.1% v/v TFA; B: 99.9% v/v ACN and 0.1% v/v TFA). Fractions containing the desired peptide were lyophilized.

Analysis of peptide purity and mass

The purity was assessed by analyzing around 10 μg of peptide by RP-HPLC (1260 HPLC system, Agilent) using a C18 column (ZORBAX 300SB-C18, 5 μm, 300 Å, 4.6x250 mm, Agilent). Peptides were run at a flow rate of 1 mL.min^{-1} with a linear gradient of 0-100% or 0-50% v/v of solvent B over 15 min (A: 94.9% v/v H$_2$O, 5% v/v ACN and 0.1% v/v TFA; B: 94.9% v/v ACN, 5% v/v H$_2$O and 0.1% v/v TFA). The mass of purified peptides was determined by electrospray ionization mass spectrometry (ESI-MS) in positive ion mode on a single quadrupole liquid chromatography mass spectrometer (LCMS-2020, Shimadzu).

Protease inhibition assay

Inhibition constants (K$_I$) were determined by measuring the residual activities of protease in the presence of different dilutions of inhibitor (2-fold dilutions, ranging from 3 μM to 0.2 nM final concentration) using fluorogenic substrates. Activities were measured at 25°C in buffer containing 10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10 mM MgCl$_2$, 1 mM CaCl$_2$, 0.1% w/v BSA and 0.01% v/v Triton-X100 by monitoring the change in fluorescence intensity over one hour using a Tecan infinite M200 Pro plate reader (excitation at 367 nm, emission recorded at 468 nm). All proteases were human in origin and were purchased from Molecular Innovations or Innovative Research and the following concentrations were used: 0.1 nM plasma kallikrein, 0.25 nM factor XIa, 2 nM factor XIIa, 2 nM thrombin, 1.5 nM uPA, 7.5 nM tPA, 2.5 nM plasmin, 6 nM factor Xa, 50 nM factor VIIa. The following fluorogenic substrates were used at 50 μM final concentration: Z-Phe-Arg-AMC for plasma kallikrein, Boc-Phe-Ser-Arg-AMC for factor XIa, Boc-Gln-Gly-Arg-AMC for factor XIIa, Z-Gly-Gly-Arg-AMC for thrombin, uPA and tPA, H-D-Val-Leu-Lys-AMC for plasmin, D-Phe-Pro-Arg-ANSNH-C$_4$H$_9$ for factor Xa and factor VIIa. The IC$_{50}$ values were determined by fitting sigmoidal curves to the data using the following four-parameter logistic equation used for dose response curves:

$$y = \frac{100}{1 + 10^{(\log IC_{50} - x)p}}$$

The *K$_I$* values were calculated based on the IC$_{50}$s using the Cheng-Prusoff equation.

Plasma stability measurement by quantifying functional inhibitor

20 μM of peptide was incubated in human plasma (Innovative Research) at 37 °C in a water bath and 30 μL probes were taken at different time points (5 min, 0.5, 1, 2, 4, 8, 24, and 48 h). The samples were diluted by addition of 770 μL assay buffer containing 0.1% (w/v) BSA and 0.01% (v/v) Triton-X100 and the proteases were heat-inactivated by incubation at 69 °C for 20 min. The samples were stored at -20 °C. Residual inhibitor activity was measured using the protease activity assay described above. The samples were centrifuged at 16,000 g for 5 min and the supernatant was transferred to new tubes. Two-fold dilutions of the samples were prepared and the residual activity of 0.1 nM human plasma kallikrein was measured with 50 μM Z-Phe-Arg-AMC substrate as described above. Residual inhibition in % was calculated using the equation $IC_{50,0h}/IC_{50,xh}*100$ wherein $IC_{50,0h}$ is the functional strength of the inhibitor at time point 0 and $IC_{50,xh}$ the functional strength of inhibitor after one of the different plasma incubation periods mentioned above. The resulting percentages of inhibition were plotted versus time and analyzed using a one-phase exponential decay model.

Plasma stability measurement by LC-MS analysis of bicyclic peptide

40 μM of peptide was incubated in human plasma (Innovative Research) at 37 °C in a water bath. 5 μL probes were taken at different time points (5 min, 0.5, 1, 2, 4, 8, 24, and 48 h), diluted by addition of 5 μL 6 M guanidine hydrochloride pH 2 and incubated at RT for 30 min. The plasma proteins were precipitated by addition of cold ethanol (200 μL) and incubated for 30 min at -20 °C. The samples were centrifuged at 9000 g for 20 min, the supernatant was transferred to new tubes and the solvent was evaporated under vacuum (speed vac). The samples were dissolved in 20 μL of H_2O and analyzed by LC-MS as described above.

aPTT and PT experiment

aPTT was determined in human plasma using a STAGO STart4 Coagulation analyzer (Diagnostica). Human single donor plasma was used (Innovative Research). 100 μL of plasma was incubated with 100 μL of Pathromtin® SL (silicon dioxide particles, plant phospholipids in HEPES buffer system, Siemens) for 2 min at 37°C. The coagulation was triggered by addition of 100 μL $CaCl_2$ solution (25 mM, Siemens). Upon addition of this reagent the electromagnetically induced movement of a steel

ball in the plasma was monitored. The time until the ball stops moving was recorded as coagulation time.

3.6 – Supplementary figures and tables
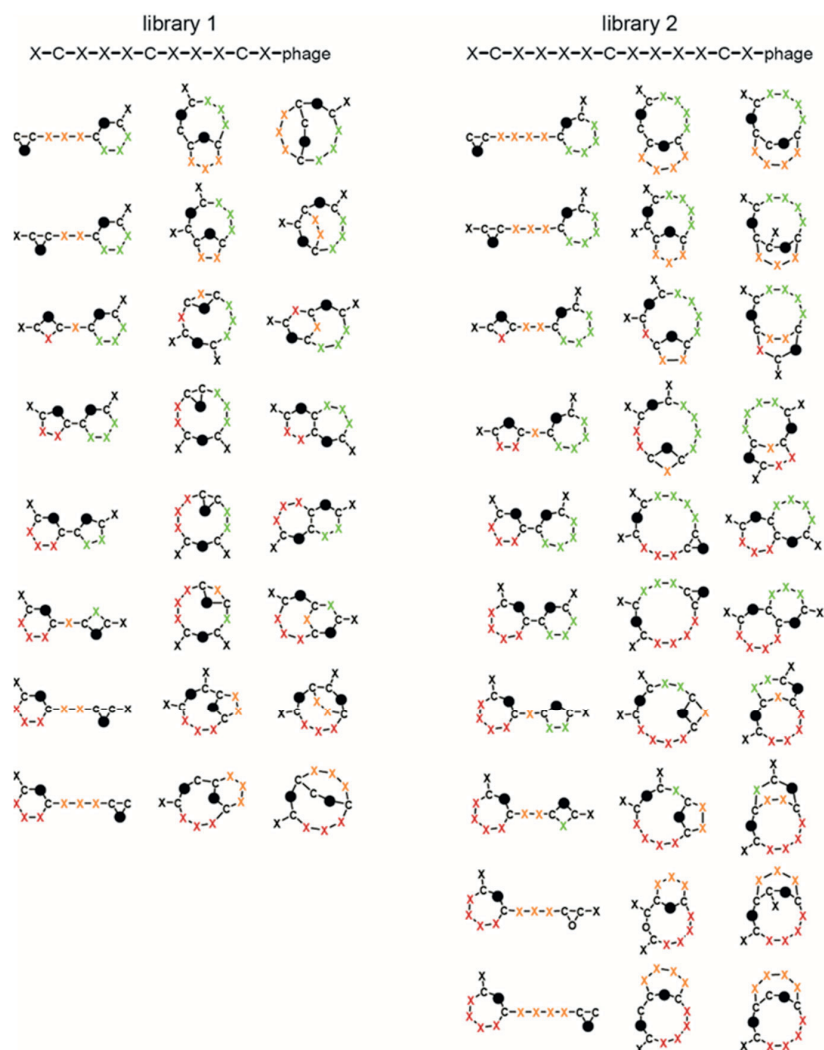
Supplementary Figure 3.1. Scaffold diversity in libraries 1 and 2. Around 20% of peptides in library 1 (XCX$_3$CX$_3$CX-phage) and 23% of peptides in library 2 (XCX$_4$CX$_4$CX-phage) contain an additional cysteine in one of the random positions indicated with 'X'. Cyclization with a chemical linker (indicated as a black ball) yields the indicated scaffolds. Random amino acids in the three segments are shown in red, orange and green.

peptide format:	% of peptides (among those having 4 cysteines)					
linker:	1	3	4	7	10	12
C C X X X X C X X X X C X	7%	8%	2%	2%	4%	7%
X C C X X X C X X X X C X	16%	20%	11%	11%	18%	3%
X C X C X X C X X X X C X	1%	2%	2%	1%	2%	2%
X C X X C X C X X X X C X	7%	5%	29%	3%	5%	2%
X C X X X C C X X X X C X	22%	9%	28%	4%	25%	3%
X C X X X X C C X X X C X	4%	6%	4%	19%	9%	2%
X C X X X X C X C X X C X	13%	3%	14%	11%	8%	2%
X C X X X X C X X C X C X	6%	2%	4%	10%	6%	1%
X C X X X X C X X X C C X	7%	12%	2%	15%	6%	2%
X C X X X X C X X X X C C	17%	34%	4%	24%	16%	76%

Supplementary Figure 3.2. Abundance of peptide formats after panning library 2 cyclized with the indicated linkers against plasma kallikrein. Abundance is indicated in % of those peptides containing four cysteines. Abundance of peptide formats from library 1 is shown in Figure 21a.
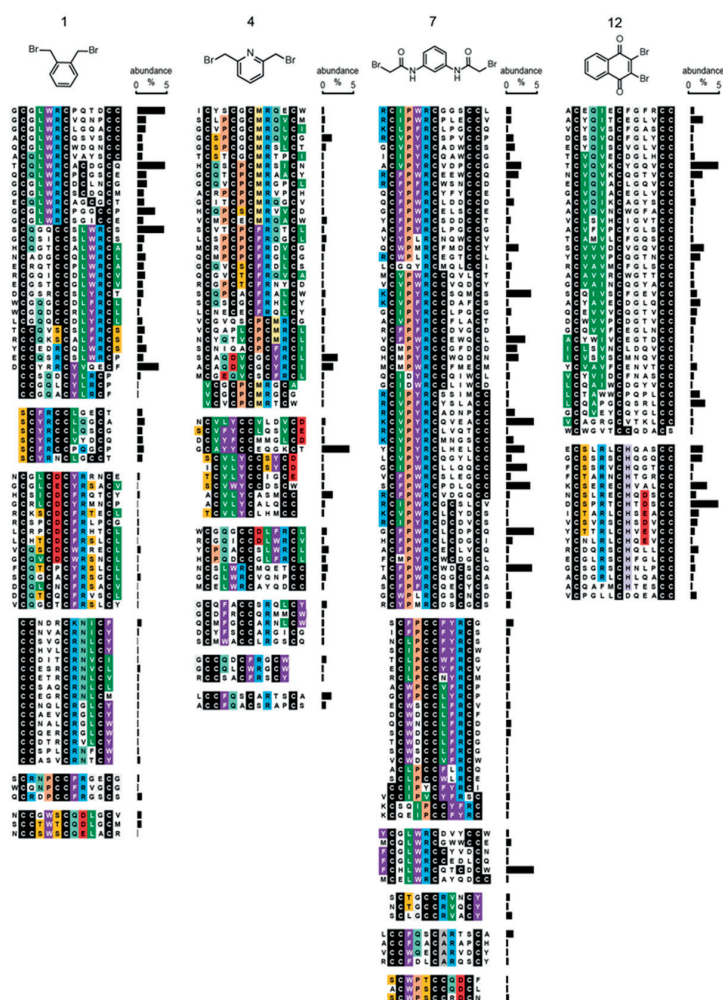
Supplementary Figure 3.3. Consensus sequences. Peptides isolated from libraries 1 and 2 cyclized with the chemical linkers **1**, **4**, **7** and **12** after two rounds of selection. Peptides isolated with the chemical linkers **3** and **10** are shown in Figure 21b.
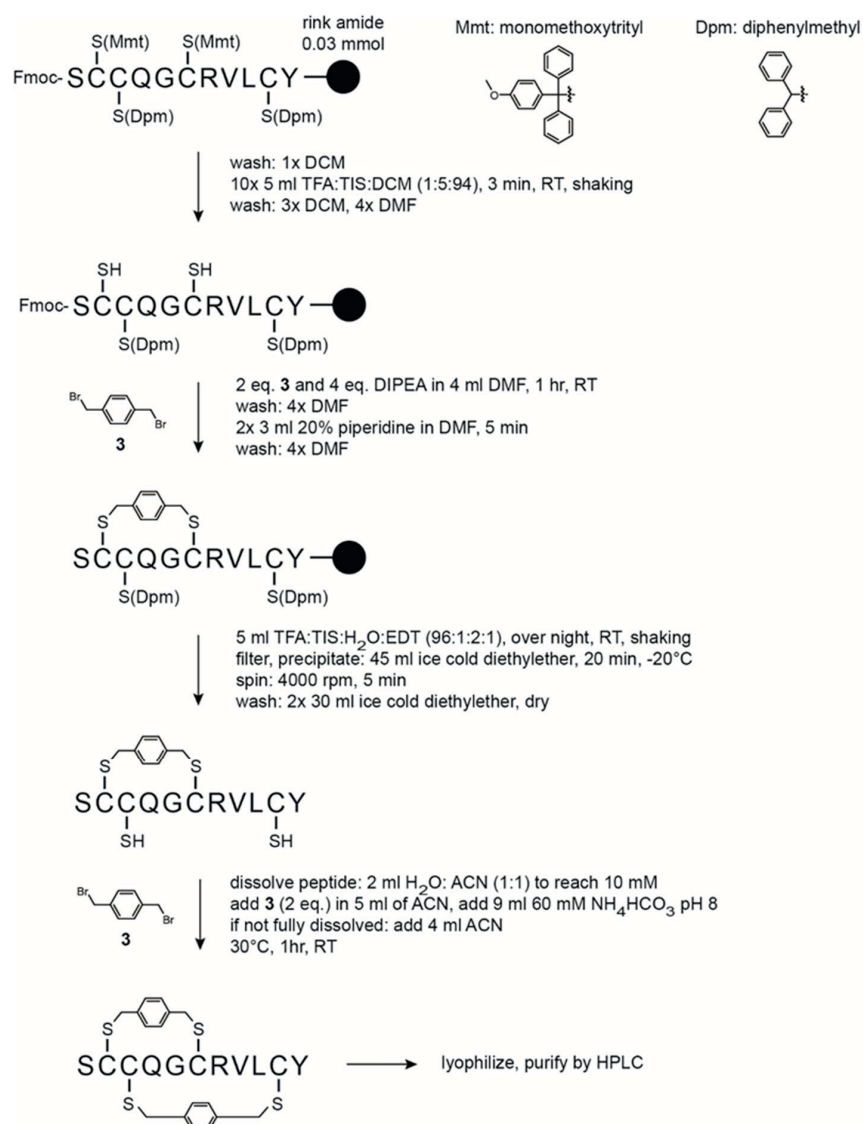
Supplementary Figure 3.4. Strategy for the synthesis of individual regioisomers of double-bridged peptides. The first pair of cysteines is reacted on resin and the second pair in solution.
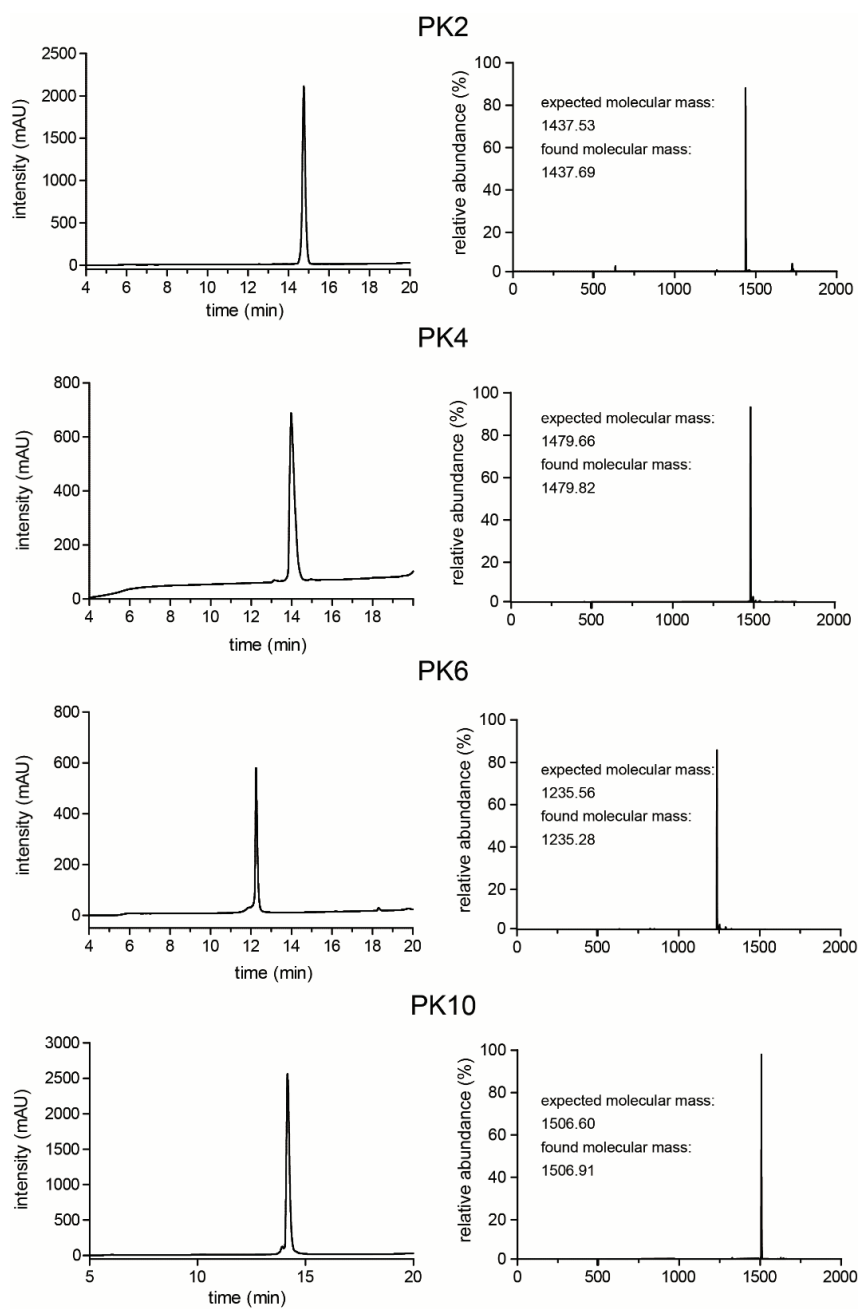
Supplementary Figure 3.5. Analytical RP-HPLC chromatograms of the plasma kallikrein inhibitors PK2 (isomer 3), PK4 (isomer 3), PK6 (isomer 1) and PK10 (isomer 3). Expected and measured molecular masses are indicated.
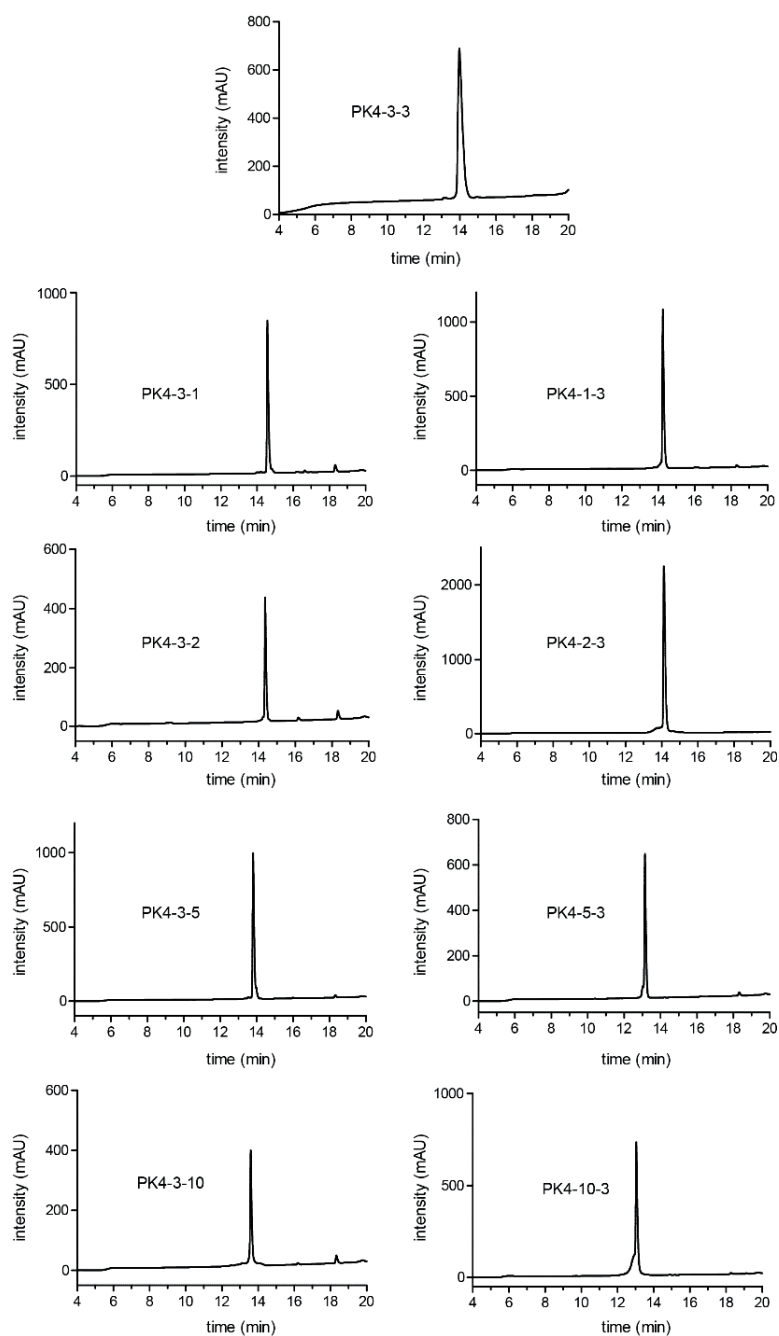
Supplementary Figure 3.6. Analytical RP-HPLC chromatograms of bicyclic peptide PK4 (isomer 3) in which one of the two linkers was exchanged. The two numbers following the name "PK4" indicate the first and second linker, respectively.
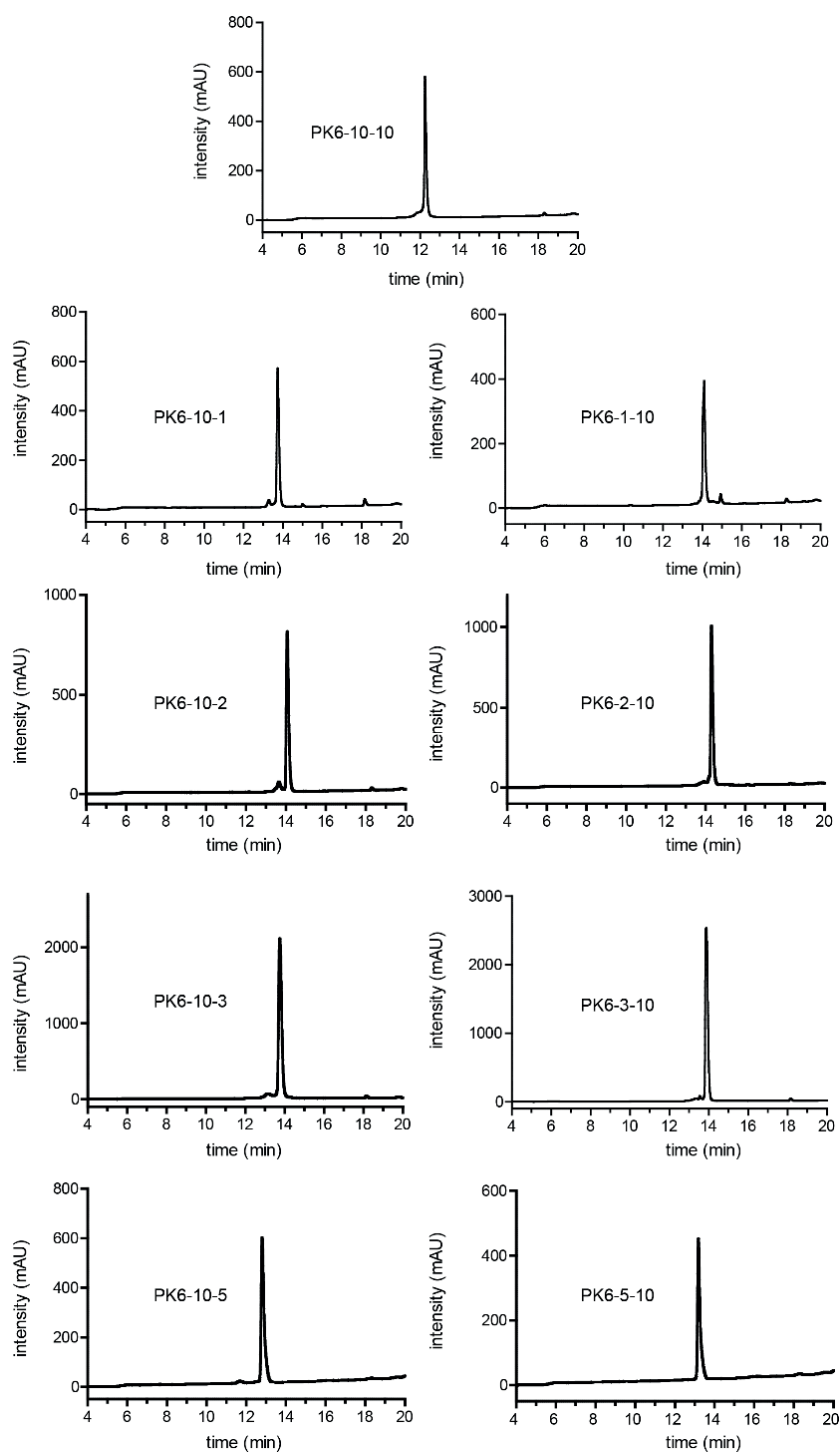
Supplementary Figure 3.7. Analytical RP-HPLC chromatograms of bicyclic peptide PK6 (isomer 1) in which one of the two linkers was exchanged. The two numbers following the name "PK6" indicate the first and second linker, respectively.
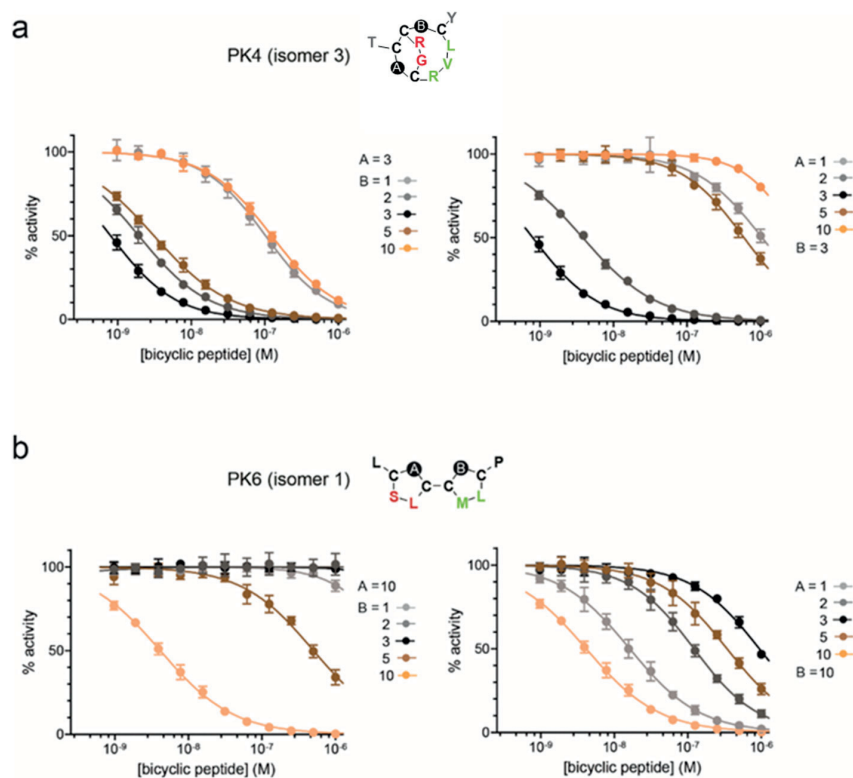
Supplementary Figure 3.8. Linker swapping. a) One of the two linkers of the bicyclic peptides PK4 (isomer 3) cyclized with linker **3** was replaced to either of five different linkers. b) One of the two linkers of the bicyclic peptides PK6 (isomer 1) cyclized with linker **3** was replaced to either of five different linkers. The residual activity of plasma kallikrein is shown. Values and standard deviations were calculated based on three measurements.
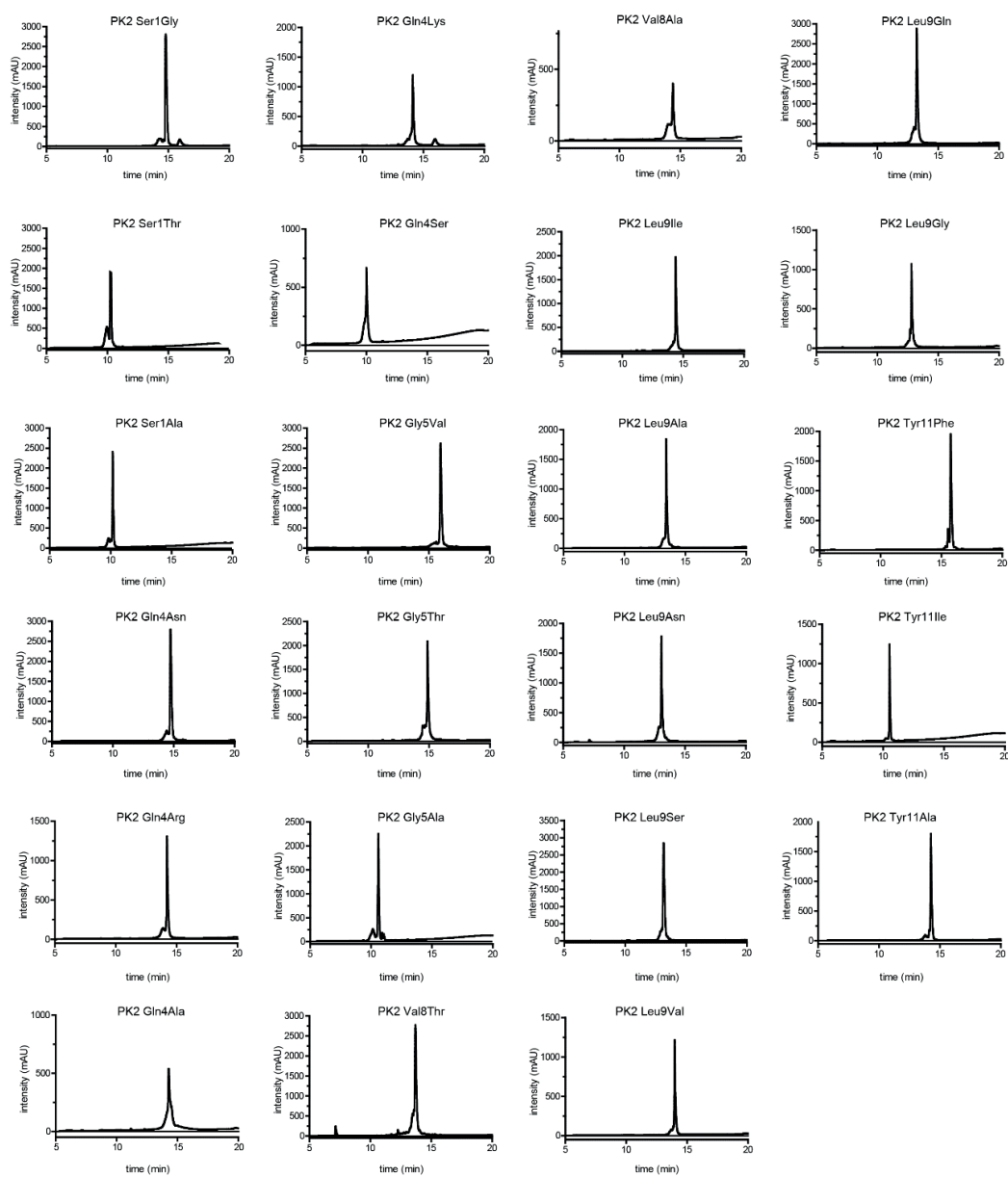
Supplementary Figure 3.9. Analytical RP-HPLC chromatograms of bicyclic peptide PK2 (isomer 3) variants in which one amino acid was substituted as indicated in the names.

4. Phage selection of head-to-side chain cyclized peptides allows for the generation of small cyclic peptide ligands

4.1 – Abstract

The development of orally available peptide drugs has been a longstanding goal in the pharmaceutical industry. A major challenge towards this aim is the poor gastrointestinal absorption of the relatively large peptides. In this work, we aimed at developing cyclic peptide ligands with the smallest possible molecular weight; ideally well below 1 kDa to facilitate intestinal absorption. We applied a recently discovered peptide macrocyclization reaction in which the side chain of a cysteine is ligated with the N-terminal amino group to cyclize phage-encoded peptides of the form XXXC and XXXXC. Unlike other cyclization strategies that rely on the cyclization of two amino acid side chains (e.g. two cysteines to form a disulfide bridge), this approach requires only one constant amino acid - the cysteine, reducing the size of the peptides by one amino acid. Panning the libraries against the two disease targets plasma kallikrein and coagulation factor XIa yielded ligands with low micromolar affinities despite the small molecular weights around 750 Da. The approach should be applicable to many other disease targets and could potentially yield leads for the generation of orally available macrocycle drugs.

4.2 – Introduction

Cyclic peptides can bind with high affinity and selectivity to protein targets, making them an attractive modality for the development of therapeutics[145,146]. Cyclic peptide ligands can be identified for "challenging" targets that are not tractable to small molecules such as globular protein with featureless surfaces or protein-protein interactions. The good binding properties arise from their relatively large binding interface and the limited conformational flexibility imposed by the circular structure. Other strengths of cyclic peptides are the low toxicity of metabolic products, being mostly amino acids, as well as the ease of production by chemical synthesis. Cyclic peptides binding to virtually any targets of interest can be developed thanks to powerful *in vitro* display techniques such as phage display or mRNA display[147–149]. Several *in vitro* evolved cyclic peptides are currently tested in clinical trials[26].

A limitation of cyclic peptides is that they are typically not orally available. All the *in vitro* evolved cyclic peptides in clinical studies were developed for administration by injection. Given the obvious advantages of needle-free application, there is a high interest in developing orally available cyclic peptides. The reasons for the poor oral availability are degradation by proteases in the

gastrointestinal tract and/or poor absorption across the epithelium of the gastrointestinal tract. While proteolytic stability can be achieved relatively easily by modifying the peptide around the cleavable bonds through the introduction of unnatural amino acids or peptide bond surrogates, the poor absorption across the intestinal epithelium is more difficult to overcome. Key parameters are the molecular mass and the size of the polar surface[13,126]. The limited oral availability of several cyclic peptide drugs with a molecular weight around 1 kDa, as for example oxytocin (1007 Da, F = 0.9% in rat), actinomycin (1255 Da, F = 5% in human) or caspofungin (1093 Da, F < 0.2 %), illustrates that cyclic peptides should be smaller to achieve good uptake via the oral route.[124] Omission of amino acids with polar side chains or N-methylation of peptide bonds can facilitate oral delivery. Studies with partially N-methylated cyclic pentapeptides show that good oral availability can be reached for molecules with a molecular weight around 600 Da, but that bioavailability depends on the specific peptide sequence.[150]

In this work, we aimed at using phage display to generate cyclic peptide ligands that have the smallest possible molecular weight, ideally well below 1 kDa. A challenge in developing such small cyclic peptides is that they contain only a few amino acids, limiting the number of contacts that can be formed between cyclic peptide and protein target. Most peptide ligands developed by phage display or other in vitro display techniques contain more than ten amino acids. The smallest cyclic peptide ligand isolated by phage display is a six-amino acid peptide cyclized by a disulfide bridge that was isolated against streptavidin[66]. The peptide cyclo-Ac[CHPQFC]NH2 binds streptavidin with a Kd of 270 nM. Slightly larger disulfide-cyclized peptides, containing seven or eight amino acids, were isolated against Neutravidin[151], the paratope of an anti-somatostatin antibody[152] and integrins[153]. A limitation of disulfide-cyclized peptides is that they require two constant amino acids - the cysteines – that cannot be varied in a combinatorial fashion but contribute together 240 Da to the molecular weight of the ligand. For instance, if the molecular weight of cyclic peptides is limited to 700 Da, the number of amino acids in a cyclic peptide should not exceed four or five. Given that two cysteines are required, the number of variable amino acids is thus limited to two or three and the maximal diversity of the libraries is thus only 400 for peptides of the format CXXC and 8,000 for peptides of the form CXXXC.

Slightly smaller cyclic peptides can be generated if the peptides are cyclized by connecting a side chain with one of the peptide termini. We have recently discovered a chemical reaction in which peptides are efficiently cyclized by connecting the thiol side chain of a C-terminal cysteine with the amino group of the N-terminal amino acid (Kale, S. et al., unpublished). With this reaction, only one

of the amino acids is required for the cyclization instead of two. In short cyclic peptides of four or five amino acids, three or four amino acid positions remain for diversification, allowing a diversity of 8,000 and 160,000 cyclic peptides, respectively. In this work, we chose to exploit the head-to-side chain cyclization to generate and screen large libraries of small cyclic peptides (Figure 25a). We reasoned that such peptides, even if they have only moderate binding affinity, could be used as lead structures for further optimization using unnatural amino acids to improve binding affinity, proteolytic stability and bioavailability.

4.3 – Results

4.3.1   Design and cloning of the libraries

We have cloned two phage display libraries encoding peptides of the form XXXC (library 1) and XXXXC (library 2; Figure 25a) as fusion with phage pIII. A mutated pIII was employed that lacks six cysteines present in the wild-type phage coat protein[76], preventing reaction of other cysteines with the cyclization reagents. Degenerate primers containing NNK codons were used for the library construction to allow all 20 amino acids in the randomized positions. The theoretical numbers of different peptides in these libraries are 8,000 and 160,000, respectively. The number of bacterial cells that were transformed with a plasmid of the library exceeded 4.9×10^7 (library 1) and 3.2×10^7 (library 2), ensuring that nearly all the possible sequences are represented in the two phage display libraries.

We incubated phage of the two peptide libraries in separate reactions with eight different reagents that all contain two thiol-reactive groups (Figure 25b). These reagents react in a first step rapidly with the constant cysteine present in each peptide, and in a second step intra-molecularly with another nucleophilic group within the randomized region of the peptide[154] (Figure 25a). If a cysteine is present in any of the random amino acids, the second reactive group of the linker reacts with the sulfhydryl group of this additional cysteine (Figure 25a, dashed arrow). Depending on the position of this cysteine in the peptide, three (library 1) or four (library 2) different cyclic peptide formats are generated. The number of different macrocycles that can be generated in this way is 3 x 400 (library 1) and 4 x 8,000 (library 2) for one cyclization reagent, and a total of 9,400 macrocycles (library 1) and 256,000 macrocycles (library 2), if the peptides are cyclized in parallel with the eight different reagents.

If no cysteine is present in the random region, the second electrohilic group of the chemical linker can react with another, less nucleophilic group. As described above, we have recently found

that peptides containing one cysteine are efficiently macrocyclized by bivalent thiol-reactive reagents shown in Figure 25b by reacting first with the cysteine side chain and second with the N-terminal amino group (Kale, S. et al., unpublished). In this case, the number of different molecules generated is much larger, namely around 7000 (library 1) and 130 000 (library 2) for one cyclization reagent, and a total of more than 54,000 macrocycles (library 1) and 1,000,000 macrocycles (library 2), if the peptides are cyclized in parallel with the eight different reagents. Alternatively, the peptides may also be cyclized by connecting the side chains of the constant cysteine with a lysine, or with a methionine[154] (Figure 25a; dashed arrows).
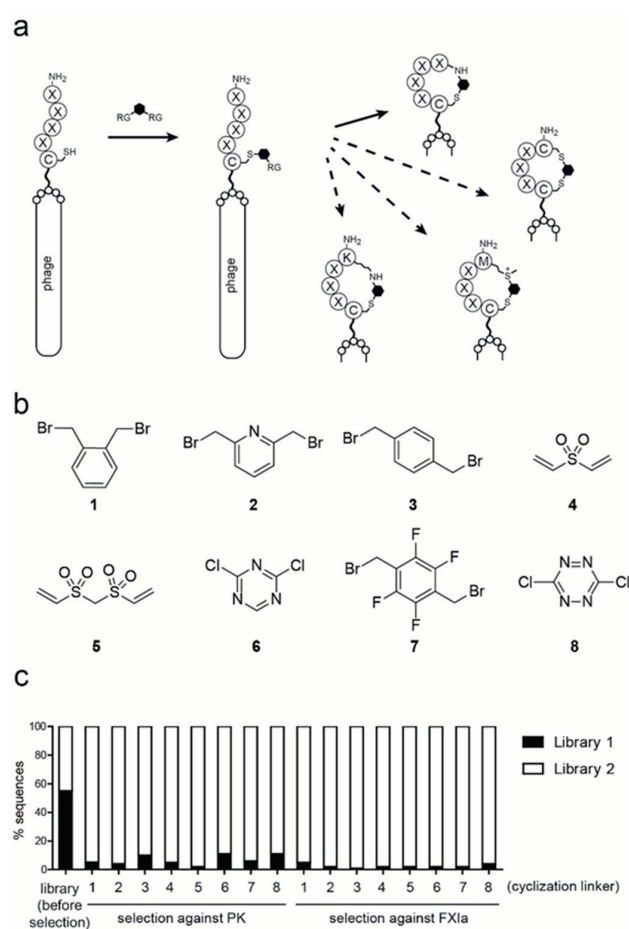
Figure 25. Phage selection of head-to-side chain cyclized short peptides. (a) Linear penta-peptides displayed on phage (library 2), each containing one constant cysteine and four random amino acids are cyclized by a reagent having two thiol-reactive groups (RG). The reagent reacts first inter-molecularly with the thiol group of the cysteine and then intra-molecularly with a second electrophilic group such as the N-terminal amino group or the side chains of a second cysteine, a lysine or methionine. (b) Chemical reagents applied for the cyclization phage-displayed peptides. (c) Percentage of clones belonging to library 1 and 2 before the phage selection and after two rounds of phage selection against PK or FXIa. The numbers indicate the reagents used for the cyclization of the peptides on phage.

4.3.2   Phage selections and NGS

We performed phage selections against the two serine proteases plasma kallikrein (PK) and coagulation factor XI (FXIa). Uncontrolled activity of plasma kallikrein causes the swelling disorder hereditary angioedema (HEA). While a protein-based inhibitor of PK, ecallantide, is approved for the treatment of HEA, no oral drug is yet available[155]. FXIa plays an important role in various thrombotic disorders and is an attractive target as it promises a safe anti-thrombotic therapy. As for PK, no small molecule drugs are available for FXIa. Phage of the two libraries were produced in separate *E.coli* cultures, purified by PEG precipitation and equal numbers of phage from each library were pooled. Around 10^{14} phage of the mixture were cyclized in eight separate reactions with the reagents shown in Figure 25b and panned against PK and FXIa in two iterative rounds of selection. Phage of the libraries 1 and 2 before selection and cyclization, and phage isolated in the 16 different selections, were identified by next generation sequencing (NGS; > 1 million reads for most samples). Peptides of the format XXXXC were strongly enriched over peptides of the format XXXC (Figure 25c). The enrichment of peptides from library 2 suggested that larger cyclic peptides bind substantially better than smaller ones. This most likely arises from the larger combinatorial diversity of library 2 and the larger interaction surface of these peptides.

Using previously published MATLAB scripts, we analyzed the 500 most abundant clones of each selection by aligning them in groups of peptides that share sequence similarities[111]. Peptides containing an amber stop codon were strongly enriched in the selections. Sequencing of the library immediately after transformation of the plasmids into bacterial cells showed an equal distribution of all codons, including the UAG codon, showing that no bias was introduced during the library cloning (Supplementary figure 4.1). Sequencing of the library after phage production and before affinity panning revealed that the amber stop codon was enriched at this stage around 3-fold over other codons (Supplementary figure 4.1). The amber codon UAG is read as stop codon in around half of the translation events in the E.coli strain used. We speculate that cells harboring a phage clone with amber codon grow faster and that these clones were enriched during phage production but not in the phage affinity selection. We excluded peptides containing amber stop codons from the analysis. In most of the 16 selections, several consensus groups were identified (Figures 26 and 27). Given the redundancy of the genetic code, we expected abundant peptides to be encoded by several different DNA sequences. Indeed, for most peptides, different DNA sequences were identified, as indicated in the column "number of different DNA sequences" in Figures 26 and 27. The consensus sequences were different for peptides cyclized with the eight structurally diverse chemical linkers, suggesting that

the peptides were affinity-selected in a linker-modified form and thus most likely in a cyclic form. The consensus sequences found for the selections against PK and FXIa differed too, suggesting the isolation of target-specific ligands.

Clone	Sequence	Abundance	Number of different DNA sequences	K_i (µM) Linear peptide	Cyclic peptide
Linker 1 (1.6 million reads)					
PK1	C P W R C	139709	6	110 ± 40	10.5 ± 1.1
	C P Y R C	9988	6		
	C P F R C	957	2		
PK2	C L W R C	121809	8	300 ± 6	13.5 ± 1.5
	C M W R C	3577	1		
	C Q W R C	645	2		
PK3	L C G R C	21716	17	570 ± 90	180 ± 50
	F C G R C	1432	3		
	M C G R C	456	2		
	C C G R C	645	1		
	W Y R C C	34123	3		
Linker 2 (1.9 million reads)					
	R C V H C	710	2		
Linker 3 (1.1 million reads)					
PK4	C P F R C	156621	6	900 ± 200	75 ± 12
	C P W R C	27986	6		
	C D W R C	1792	1		
	C D F R C	140	1		
	C E W R C	120	1		
	C M W R C	1797	1		
	C G W R C	800	1		
	C Q W R C	659	2		
	C T W R C	451	2		
	C L F R C	334	2		
	C F R C	139	1		
Linker 6 (1.2 million reads)					
	S R S C C	370	1		
	D R S C C	276	1		
	R S C C	316	2		
	R L A C	424	1		
	R V L C	200	1		
	R L L C	194	1		
Linker 7 (1.5 million reads)					
	R P R I C	601	1		
	R P R L C	222	1		
	R P R T C	369	1		
	W C W M C	545	1		
	W C W L C	247	1		
	W C W I C	149	1		
	W H C C	334	1		
	Y H C C	166	1		
	F H C L C	459	2		
	F H C M C	162	1		
Linker 8 (1.4 million reads)					
	L L W E C	4210	8		
	W L W E C	3226	3		
	C L W E C	3126	3		
	M L W E C	2842	3		
	F L W E C	307	2		
	Q L W E C	125	1		
	W F W E C	94	1		

Figure 26. Peptides selected against PK. For each selection, around one million of clones were sequenced and the most abundant 500 peptides aligned. Sequences are shown for those selections for which a consensus sequence was identified, or for which one clone was strongly enriched over all other library members. Sequence similarities are highlighted in color. The abundance of each sequence found by NGS is shown. For most peptides, multiple different DNA sequences were found due to codon degeneracy. Kis for linear and cyclic peptides were determined for selected clones. Indicated average values and standard deviations were calculated based on at least three measurements.

78

Clone	Sequence	Abundance	Number of different DNA sequences	K_i (µM) Linear peptide	K_i (µM) Cyclic peptide
Linker 1 (0.1 million reads)					
	L I W Y C	576	16		
	L V W F C	102	5		
	L I W F C	29	2		
	L L W F C	27	2		
	C L W F C	12	1		
	F V W L C	22	1		
	F L W C C	18	1		
	K L C C	1774	3		
	K I C C	326	1		
	K V C C	15	1		
Linker 2 (1.3 million reads)					
FXI1	R C Q W C	73280	3	175 ± 10	> 1000
Linker 3 (1.6 million reads)					
	W R G C C	257020	6		
FXI2	M R G E C	17667	6	> 1000	210 ± 10
	R P P P C	950	4		
	R D P P C	284	1		
	R A P P C	191	1		
	Q P P P C	194	1		
Linker 6 (1.6 million reads)					
FXI3	R P T V C	13160	22	> 1000	700 ± 200
FXI4	R P T I C	8236	12	> 1000	680 ± 110
FXI5	R P T L C	948	3	> 1000	> 1000
	R P S V C	803	2		
	R P G V C	222	1		
	R P T C C	242	1		
FXI6	R T R L C	539	1	> 1000	> 1000
	R T R V C	225	1		
	R T K A C	225	1		
FXI7	M A D R C	1926	5	> 1000	8.8 ± 1.5
FXI8	M A N R C	1494	4	> 1000	7.1 ± 0.3
Linker 7 (1.7 million reads)					
FXI9	R H P W C	15977	6	> 1000	1.7 ± 0.4
	K H P W C	1391	2		
FXI10	S H P W C	12257	6	> 1000	2.6 ± 1.1
	T H P W C	3508	4		
FXI11	V H P W C	8641	4	> 1000	7.5 ± 0.2
	A H P W C	1104	3		
	I H P W C	957	2		
FXI12	Q H P W C	10401	4	> 1000	6.5 ± 0.8
	E H P W C	667	2		
	H H P W C	922	2		
	W H P W C	179	1		
FXI13	M R G E C	8274	6	> 1000	270 ± 5
Linker 8 (1.6 million reads)					
FXI14	R P F S C	4728	17	> 1000	35 ± 5
	R P Y W C	2775	6		
FXI15	R P F W C	1120	5	> 1000	> 1000
	R P F G C	116	1		
	R P L W C	920	5		
	M A W E C	14397	2		
	M V W E C	555	2		

Figure 27. Peptides selected against FXIa. For each selection, around one million of clones were sequenced and the most abundant 500 peptides aligned. Sequences are shown for those selections for which a consensus sequence was identified, or for which one clone was strongly enriched over all other library members. Sequence similarities are highlighted in color. The abundance of each sequence found by NGS is shown. For most peptides, multiple different DNA sequences were found due to codon degeneracy. Kis for linear and cyclic peptides were determined for selected clones. Indicated average values and standard deviations were calculated based on at least three measurements.

In several consensus sequences, particularly those identified in selections against PK, a second cysteine was present, indicating that the peptides were cyclized via two cysteine side chains. The additional cysteine was mostly found at the N-terminal end of the peptides, but also in the other

positions that were randomized. Peptides containing two cysteines were strongly enriched in the selections (see "abundance" in Figures 26 and 27), possibly because the chemical cyclization via two cysteines is more efficient than the cyclization via another pair of nucleophiles, and more phage carry the desired product. Some of the consensus sequences without a second cysteine contained a methionine at the N-terminus that had potentially reacted with the chemical linkers to form a sulfonium ion. Most consensus sequences did not contain an amino acids with a nucleophilic group in the side chain, suggesting that they were either isolated as cyclic peptides in which the C-terminal cysteine was linked by a chemical bridge to the N-terminal amino group, or they were isolated as linear peptides.

4.3.3   Characterization of small cyclic peptides

We synthesized peptides from different formats and consensus groups with the aim of testing if the peptides bind in their cyclic or linear form, and to determine their binding affinity. A range of peptides selected based on abundance and consensus group were synthesized by solid phase peptide chemistry and purified. A fraction of the linear peptides was saved for testing the binding of the linear forms while the rest of the compounds was reacted with the appropriate cyclization reagent and finally purified by reversed phase HPLC (Supplementary figures 4.2 to 4.6). Assuming that most of the peptides were binding to the active sites of the two proteases, as found previously for selections with bicyclic peptides[77,83], we tested the binding in inhibition assays using fluorogenic and chromogenic substrates of PK and FXIa (Supplementary figure 4.7). Among the peptides isolated against PK, all five peptides synthesized inhibited the protease, with K_Is in the micromolar range. The peptides belong to three different consensus groups, all containing two cysteines cyclized via the linker 1 or 3 and an arginine residue that most likely binds to the S1 pocket of the trypsin-like serine protease. All four cyclic peptides inhibited PK at least 10-fold better than their linear counterparts, suggesting that they were isolated in their cyclized form during the phage panning. The best inhibitors, PK1 and PK2 cyclized with compound 1 inhibited PK with K_Is of 11 ± 1 μM and 13 ± 1 μM, respectively.

Among peptides isolated against FXIa, six out of the 15 tested cyclic peptides inhibited the protease with single-digit micromolar inhibition constants. The same peptides being not cyclized did not inhibit FXIa at the highest concentration tested (1 mM), clearly indicating that the peptides had been cyclized on phage and isolated in their cyclic form. The six peptides belong to two consensus groups, both containing only the constant C-terminal cysteine. Peptides of one of the consensus groups (MAXRC) contain a methionine at the N-terminus, and are potentially cyclized via the

methionine side chain. The two cyclic peptides of this group, FXI7 and FXI8 cyclized with linker 6 inhibited FXIa with K_Is of 8.8 ± 1.5 μM and 7.4 ± 0.3 μM. Peptides of the other consensus group (XHPWC) are cyclized via the side chain of the constant cysteine and the N-terminal amino group. The best inhibitor, FXI9 blocked FXIa with a K_I of 1.7 ± 0.4 μM while its linear analogue did not inhibit the protease at a 1000-fold higher concentration.

4.4 – Conclusions

To our knowledge, the cyclic peptide ligands developed here are smaller than any cyclic peptide previously evolved by phage display. The small size was achieved by limiting the number of randomized amino acids to four and by applying a new head-to-side chain macrocyclization reaction that requires only a single constant amino acid, the cysteine. With the new strategy, we generated large libraries comprising more than a million different cyclic peptides that have an average molecular weight smaller than 750 Da. We show that the thiol-to-amine ligation reaction is compatible with phage display by isolating ligands that i) converge to consensus sequences possessing only one cysteine and ii) bind only when cyclized by a linker via the cysteine side chain and the N-terminus. The phage selections against two important disease targets, PK and FXIa yielded ligands with decent binding affinities in the low micromolar range. The approach and libraries should be applicable to other protein targets in order to identify small cyclic peptide leads to a broad range of targets. The affinity, stability, oral availability and other properties may be optimized in iterative rounds of improvement following established strategies that are routinely used to translate linear peptide leads into peptidomimetic drugs.

4.4 – Materials and methods

Cloning peptide phage display libraries

The phage peptide libraries 1 and 2 were cloned by ligating DNA coding for the variant peptides, a GSG linker and the cysteine-free N-terminal domains D1 and D2 of p3 into the phage vector fd0D12[144], between the leader peptide sequence and the C-terminal domain of pIII. The DNA coding for the N-terminal domains of pIII was amplified in a PCR reaction using the following pair of forward and reverse primers and the vector fdg3p0ss21[76] as template:

Prepcr1Clib: 5'-TGTGGTTCTGGCGCTGAAACTGTTGAAAGTAG-3'

sfinotfo: 5'-CCATGGCCCCCGAGGCCGCGGCCGCATTGACAGG-3'

The PCR product was separated by electrophoresis on an agarose gel and the DNA extracted with a gel purification kit. DNA coding for diverse peptides of the formats XXXC and XXXXC was appended to the PCR product in a second PCR using the following degenerate forward primers, reverse primer and the product of the first PCR reaction as template:

PCRlib1: 5'-TATGCGGCCCAGCCGGCCATGGCANNKNNKNNKTGTGGTTCTGGCGCTG-3' (library 1)

PCRlib2: 5'-TATGCGGCCCAGCCGGCCATGGCANNKNNKNNKNNKTGTGGTTCTGGCGCTG-3' (library 2)

sfinotfo: 5'- CCATGGCCCCCGAGGCCGCGGCCGCATTGACAGG-3'

The final PCR products were digested with the restriction enzyme SfiI and ligated into the vector fd0D12 that was linearized by sequential digestion with EcoRI and SfiI. The ligated vector was electroporated into *E.coli* TG1 electrocompetent cells (Lucigen). After the transformation, the cells were rescued with SOC medium, incubated 1 h at 37°C and plated on large chloramphenicol plates (30 μg/ml). The bacteria were detached from the plates by suspension in 2xYT medium containing 20% glycerol and stored in stocks of 1 mL at -80°C. The number of transformants was estimated based on the number of colonies growing on selective plates after plating various dilutions of transformed cells.

Chemical cyclization of peptides on phage

Phage were produced and modified as described before (Kale,S. et al., unpublished). Four 2 L flasks containing each 0.5 L of 2xYT medium supplemented with 30 μg/mL chloramphenicol were inoculated with TG1 *E.coli* cells harboring library 1 and 2 (2 flasks each) to reach an OD_{600} of 0.1. After overnight incubation for phage production (30°C, 200 rpm), phage were PEG-precipitated, dissolved in reaction buffer (20 mM NH_4CO_3, 5 mM EDTA, pH 8) and the two libraries pooled. The cysteines of the peptides were reduced by incubating the phage with 1 mM TCEP at 42°C for 1 h. The phage were concentrated in centrifugation tubes at 4°C and diluted with ice cold reaction buffer to lower the TCEP concentration below 10 μM. The phage were split into eight equal fractions and each fraction complemented with reaction buffer to a volume of 800 μL. 200 μL of the 0.2 mM cyclization reagents in ACN were added to reach a reagent concentration of 40 μM. After incubation at 30°C for 1 h, excess reagent was removed by precipitating the phage with PEG. The pellet was dissolved in washing buffer (10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10 mM MgCl2, 1 mM CaCl2, pH 7.4)

containing 1% bovine serum albumin (BSA) and 0.01% tween 20. For the second round of panning, phage were produced in volumes of 25 mL cultures.

Affinity selection against PK and FXIIa

Human PK (activated protease; HPKA, Molecular Innovations) and FXIa (activated protease; HFXIA, Molecular Innovations) were biotinylated by incubating 0.2 mg protein (10 μM) in PBS buffer pH 7.4 with a 10 -fold molar excess of sulfo-NHS-LC-biotin (ThermoFisher) at RT. Unreacted biotinylation reagent was removed by size exclusion chromatography using a disposable PD10 desalting column (GE Healthcare) and PBS buffer. For each selection, 1 μg biotinylated protease was immobilized on 50 μL of streptavidin-coated magnetic beads (streptavidin Dynabeads M-280°; ThermoFischer). The beads and free biotin binding sites were blocked with washing buffer containing 1% BSA, 0.01% Tween 20 and 1 μM biotin in order to prevent the isolation of binders to streptavidin or the bead surface. For each target and linker, 200 μL of phage and 50 μL of blocked beads were mixed together and incubated 1 h at RT. Unbound phage were removed by washing the beads six times with washing buffer containing 0.01% Tween 20 (v/v) and twice with washing buffer. The remaining phage were eluted by incubating the beads 5 min with 100 μL of 50 mM glycine, pH 2.2. The solution containing the eluted phage was neutralized with 100 μL of 1 M Tris-Cl, pH 8.0, added to 25 mL of exponentially growing TG1 *E.coli* cells ($OD_{600} = 0.4$) and incubated for 90 min at 37°C. The cells were plated on chloramphenicol plates (30 μg/mL) and incubated overnight at 30°C. Cells of the colonies were harvested the next day with 2xYT containing 20% glycerol. The second round of selection was performed following the same procedure but using neutravidin-coated magnetic beads (25 μL) to prevent enrichment of streptavidin-binding peptides. Magnetic neutravidin beads were generated by coupling neutravidin to tosyl-activated magnetic beads (Dynabeads M-280 Tosylactivated; Thermofisher) as described before[144].

High-throughput DNA sequencing

DNA released from phage particles or DNA isolated from infected bacterial cells was sequenced using NGS Illumina technology as described in the dedicated chapter 5. Double-stranded DNA from

infected bacterial cells was obtained by centrifugating 300 μL glycerol stock and extracting plasmid DNA from the cell pellet using a purification kit (NucleoSpin Plasmid, Macherey-Nagel). The phage DNA was amplified in two consecutive PCR reactions as described in the following. In the first reaction, the forward and reverse primers below were used. A 30 μL PCR reaction was performed adding reagents in the following order to achieve the final concentrations indicated in brackets. 20.4 μL H_2O, 6 μL 5x standard HF buffer, 0.9 μL DMSO, 0.6 μL dNTP (250 μM), 0.6 μL primer mixture (the 10 primers were pre-mixed, 40 nM of each primer; 400 nM in total), 0.5 μL Phusion® high-fidelity DNA polymerase (1 unit; NEB), and 1 μL phage (10^{10} t.u.) or plasmid DNA (100 ng). 25 PCR cycles were performed (98°C for 15 sec, 55°C for 30 sec and 72°C for 15 sec). Formation of product was assessed by agarose gel electrophoresis (2.5%).

Primer for 1st PCR reaction:

NGS Forward1   5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTCTATGCGGCCCAGC-3'

NGS Forward2   5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNATTCTATGCGGCCCAGC-3'

NGS Forward3   5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNATTCTATGCGGCCCAGC-3'

NGS Forward4   5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNATTCTATGCGGCCCAGC-3'

NGS Forward5   5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNATTCTATGCGGCCCAGC-3'

NGS Reverse1   5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTTCAGCGCCAGAACC-3'

NGS Reverse2   5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGNTTTCAGCGCCAGAACC-3'

NGS Reverse3   5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGNNTTTCAGCGCCAGAACC-3'

NGS Reverse4   5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGNNNTTTCAGCGCCAGAACC-3'

NGS Reverse5   5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGNNNNTTTCAGCGCCAGAACC-3'

A second PCR was performed to append barcodes to DNA from different samples and to append DNA sequences required for Illumina sequencing. In the primers shown below, the regions of the barcodes are underlined. A 60 μL PCR reaction was performed adding reagents in the following order to achieve the final concentrations indicated in brackets. 41.2 μL H_2O, 12 μL 5x standard HF buffer, 1.8 μL DMSO, 1.2 μL dNTP (250 μM), 0.6 μL Phusion® high-fidelity DNA polymerase (0.6 unit; NEB), 0.6 μL forward and reverse primer (400 nM) and 2 μL product of the 1st PCR. PCR products were purified from a 2.5% agarose gel (UltraPure agarose, Invitrogen) using a commercial agarose gel purification kit (QIAquick Gel Extraction Kit, Qiagen). DNA was eluted with 10 mM Tris-HCl, pH

8.5. Libraries were subsequently loaded at 1.5 pM on a Mid Output flow cell (Illumina) and sequenced with a NextSeq 500 instrument (Illumina) according to manufacturer instructions, yielding paired-end reads of 75 nucleotides. Overlapping paired-end reads were merged using CLC Genomics Workbench v9.5 (Qiagen). The DNA was sequenced along with at least 30% of unrelated non-amplicon-based libraries to avoid low-diversity problems that could occur on the Illumina NextSeq 500 instrument.

Primer for 2nd PCR reaction:

NGS 2nd Forward S507   5'-AATGATACGGCGACCACCGAGATCTACAC<u>AAGGAGTA</u>TCGTCGGCAGCGTC-3'

NGS 2nd Forward S520   5'-AATGATACGGCGACCACCGAGATCTACAC<u>AAGGCTATT</u>CGTCGGCAGCGTC-3'

NGS 2nd Reverse N704   5'-CAAGCAGAAGACGGCATACGAGAT<u>GCTCAGGA</u>GTCTCGTGGGCTCGG-3'

NGS 2nd Reverse N705   5'-CAAGCAGAAGACGGCATACGAGAT<u>AGGAGTCC</u>GTCTCGTGGGCTCGG-3'

NGS 2nd Reverse N706   5'-CAAGCAGAAGACGGCATACGAGAT<u>CATGCCTA</u>GTCTCGTGGGCTCGG-3'

NGS 2nd Reverse N707   5'-CAAGCAGAAGACGGCATACGAGAT<u>GTAGAGAG</u>GTCTCGTGGGCTCGG-3'

NGS 2nd Reverse N714   5'-CAAGCAGAAGACGGCATACGAGAT<u>TCATGAGC</u>GTCTCGTGGGCTCGG-3'

NGS 2nd Reverse N715   5'-CAAGCAGAAGACGGCATACGAGAT<u>CCTGAGAT</u>GTCTCGTGGGCTCGG-3'

NGS 2nd Reverse N716   5'-CAAGCAGAAGACGGCATACGAGAT<u>TAGCGAGT</u>GTCTCGTGGGCTCGG-3'

NGS 2nd Reverse N719   5'- CAAGCAGAAGACGGCATACGAGAT<u>TACTACGC</u>GTCTCGTGGGCTCGG-3'

The sequencing data were analyzed using MATLAB scripts described in the dedicated chapter 5. The 500 most abundant clones were compared to identify consensus sequences using the script *clustering.m*.

Peptide synthesis

Peptides were synthesized on an Intavis Multipep RSI synthesizer using a standard protocol for Fmoc solid phase chemistry. Each amino acid (4 equivalents, 0.5 M final concentration) was coupled twice for 30 min on 0.05 mmol of Rink amide resin using HBtU/HOBt (4 equivalent each) and N-methylmorpholine (8 equivalents) as a base in a volume of 0.85 mL DMF. The resin was washed twice with 6 mL DMF after each coupling reaction. Unreacted amino groups were capped by addition of 5% (v/v) acetic anhydride and 6% (v/v) 2,6-lutidine in 0.8 mL DMF for 5 min. The resin was washed 7 times with 6 mL DMF. Fmoc protecting groups were removed by incubating the resin twice in 0.8

ml 20% (v/v) piperidine in DMF for 30 min. The resin was washed 7 times with 6 mL DMF. At the end of the synthesis, the resin was washed twice with 6 mL DCM.

Peptides were cleaved from the resin and deprotected by incubation of the resin with 5 ml of cleavage mixture (TFA/phenol/water/thioanisole/ethanedithiol = 90/2.5/2.5/2.5/2.5) for 2 h with shaking. The resin was removed by filtration and the peptides were precipitated by addition of 50 mL of ice cold Et2O and incubation for 1 h at -20°C. The precipitated peptides were pelleted by centrifugation at 2700 g for 10 min. The Et2O supernatant was discarded and the remaining solvent was evaporated at RT.

Crude peptide (typically 10-15 mg) was dissolved in 10 ml H_2O containing 0.1 % TFA and purified on a preparative C18 column (Vydac C18 TP1022 250 × 22mm, 10 mm) using a linear gradient of solvent B (99.9% v/v MeCN and 0.1% v/vTFA) over solvent A (99.9% v/v H_2O, 0.1% v/v TFA) (28 min, 0-20%, flow rate of 20 mL/min). Fractions containing the desired product were identified by ESI-MS in positive ion mode on a single quadrupole liquid chromatography mass spectrometer (LCMS-2020, Shimadzu) and lyophilized. Typically, 5-10 mg of purified peptide was obtained.

Peptide cyclization

Purified linear peptides (typically 4-5 mg) were cyclized by incubation with an excess of linker in aqueous buffer (60 mM NH_4HCO_3, pH 8) containing 20% MeCN for 1h at 30°C. Peptides containing two cysteines were cyclized at a concentration of 1 mM with 1.2 equivalents of linker. Peptides containing one cysteine were cyclized at a concentration of 100 μM with 8 equivalents of linker. After verification of the cyclization reaction by LC-MS (LCMS-2020, Shimadzu), the solvent was evaporated using lyophilization, the peptide dissolved in 10 mL of 99.9 % H_2O/0.1 % TFA and purified by HPLC using a preparative C18 column (Vydac C18 TP1022 250 × 22mm, 10 mm) and a linear gradient of solvent B over solvent A (28 min, 0-25%, flow rate of 20 mL/min). The purity was assessed by analyzing around 20 μg of peptide by RP-HPLC (Agilent 1260 HPLC system) using a C18 column (Agilent ZORBAX 300SB-C18, 4.6 × 250 mm, 5 μm) and a linear gradient (15 min, 0-30%, flow rate of 1 mL/min) of solvent B over solvent A' (94.9% v/v MECN, 5% v/v H_2O and 0.1% v/v TFA).

Protease inhibition assay

Inhibitory activities were determined by incubation of PK or FXIa with different concentrations of the linear and cyclic peptides and quantification of the residual activities. The activity assays were

performed in washing buffer containing 0.1% w/v BSA and 0.01% v/v Triton-X100 at 25°C. Activity of PK (0.1 nM) was measured with the fluorogenic substrate Z-Phe-Arg-AMC at a final concentration 50 μM) and recording of the fluorescence intensity over 30 min (excitation at 367 nm, emission at 468 nm) on a Tecan infinite M200 Pro plate reader. Activity of FXIa (1 nM) was measured with the chromogenic substrate Pyr-Pro-Arg-pNA at a final concentration 400 μM and recording of the absorbance at 405 nm on a Tecan infinite M200 Pro plate reader. IC50 values were calculated by fitting the Hill equation to the data using GraphPad Prism 5 software. The inhibitory constant (K_I) was calculated according to the Cheng–Prusoff equation: $K_I = \text{IC}_{50}/(1+([S]_0/Km))$, where $[S]_0$ is the total substrate concentration, and Km is the Michaelis–Menten constant (165 μM for PK and 260 μM for FXIa).

4.5 – Supplementary figures
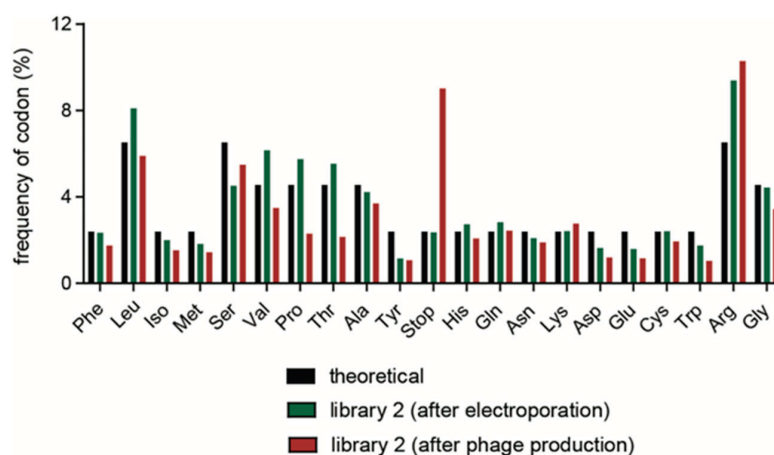
Supplementary figure 4.1. Amino acids in the random region of peptides of library 2. The frequency of codons was determined by sequencing thousands of plasmids after electroporation of library 2 into bacterial cells (double-stranded phage vector isolated from bacterial cells grown as colonies) or after phage production (single-strated phage DNA from phage particles), and compared to the theoretically expected frequency based on codon degeneracy.
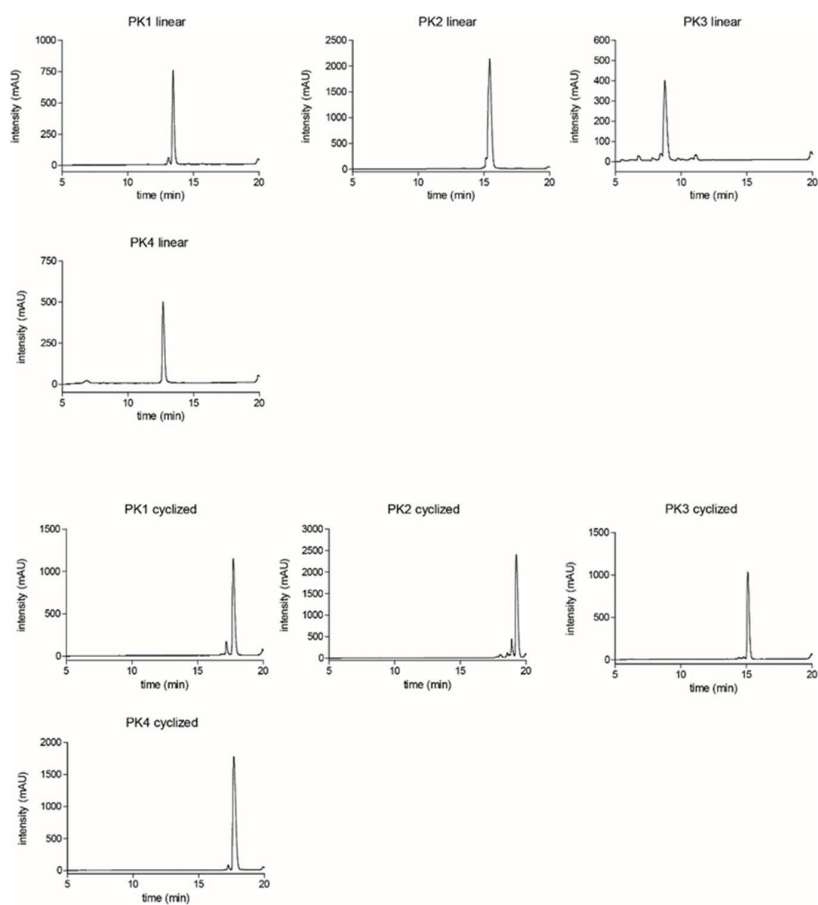
Supplementary figure 4.2. Cyclic peptides selected against PK. Purified linear and cyclic peptides were analyzed by HPLC.
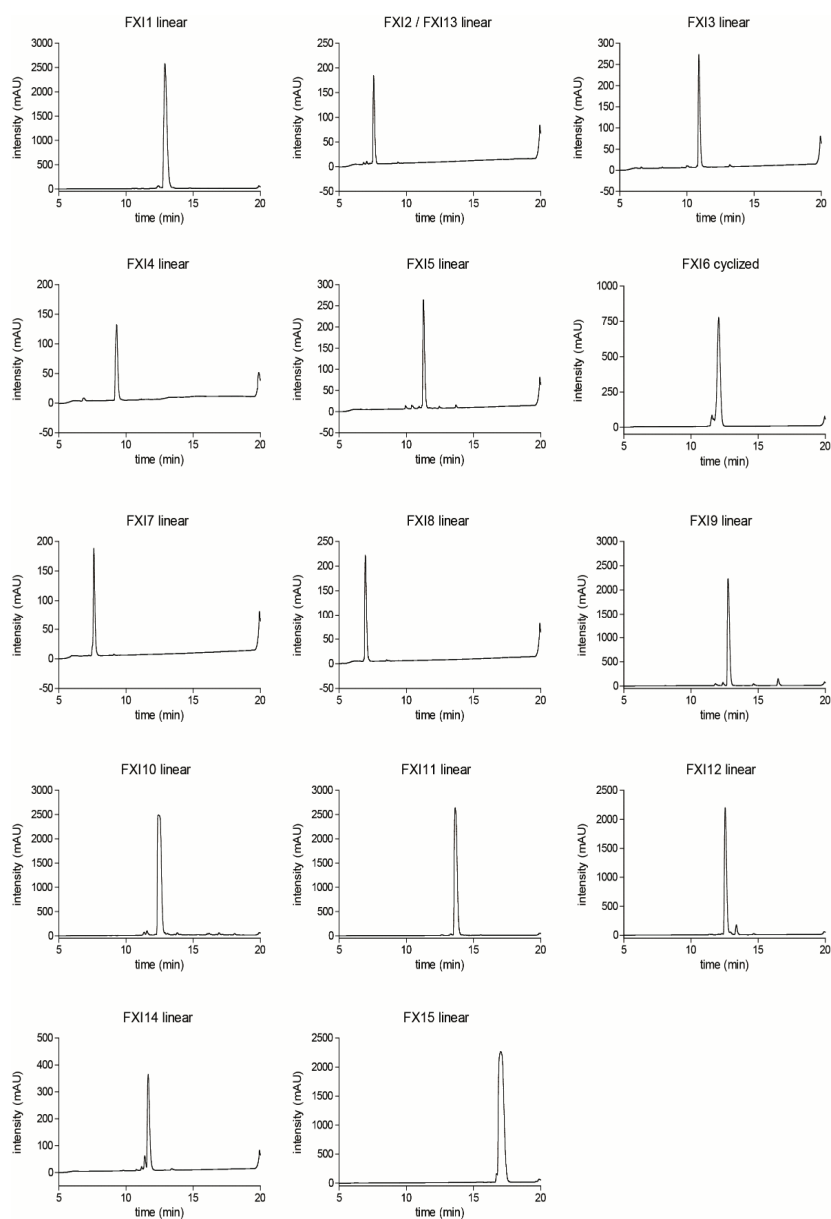
Supplementary figure 4.3. Cyclic peptides selected against FXIa. Purified linear peptides were analyzed by HPLC.
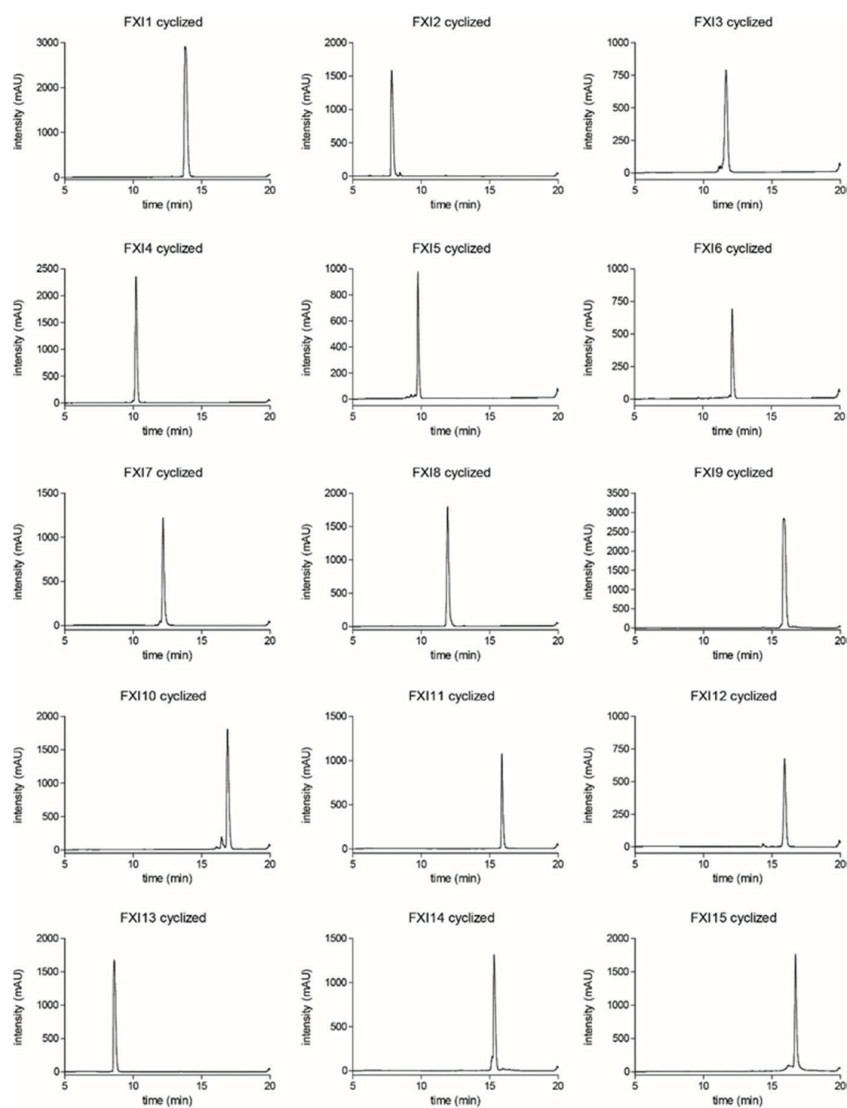
Supplementary figure 4.4. Cyclic peptides selected against FXIa. Purified cyclic peptides were analyzed by HPLC.
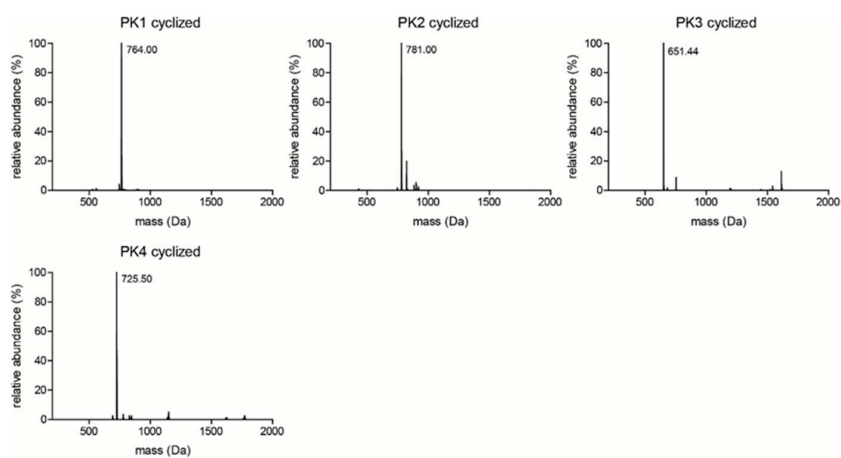
Supplementary figure 4.5. Cyclic peptides selected against PK. Purified cyclic peptides were analyzed by mass spectrometry.
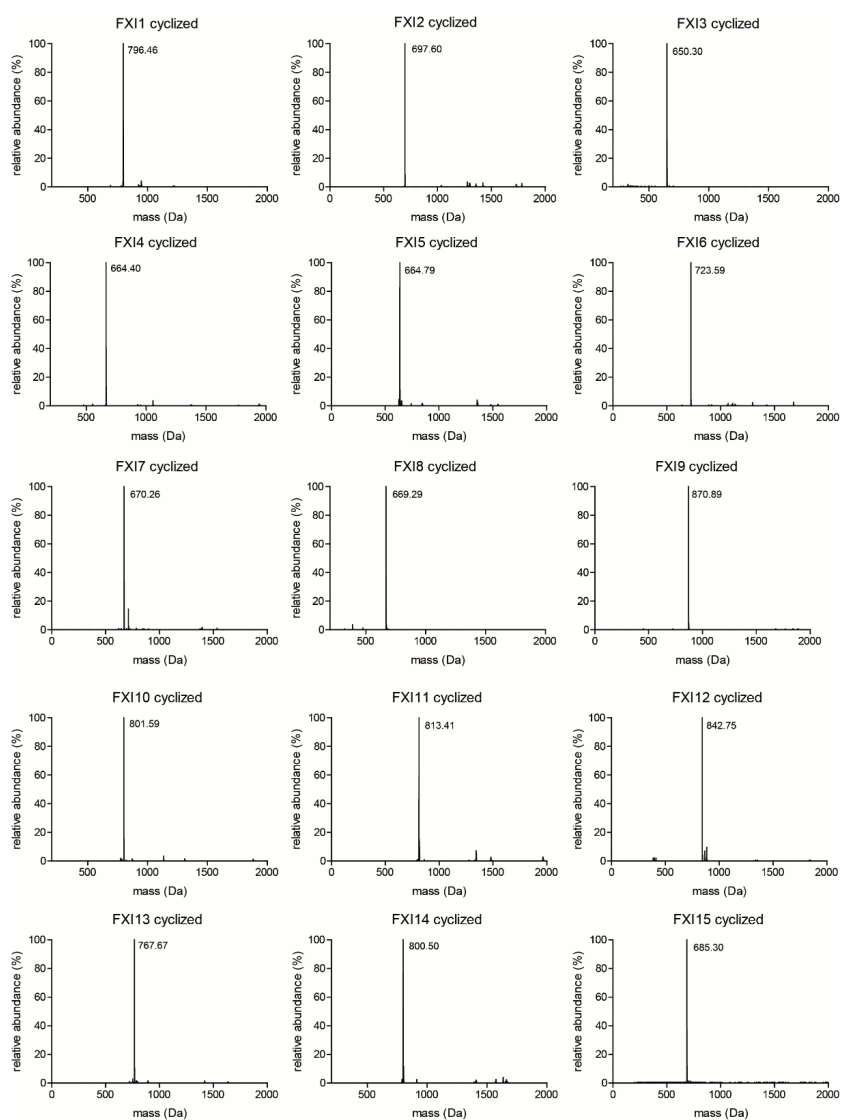
Supplementary figure 4.6. Cyclic peptides selected against FXIa. Purified cyclic peptides were analyzed by mass spectrometry.
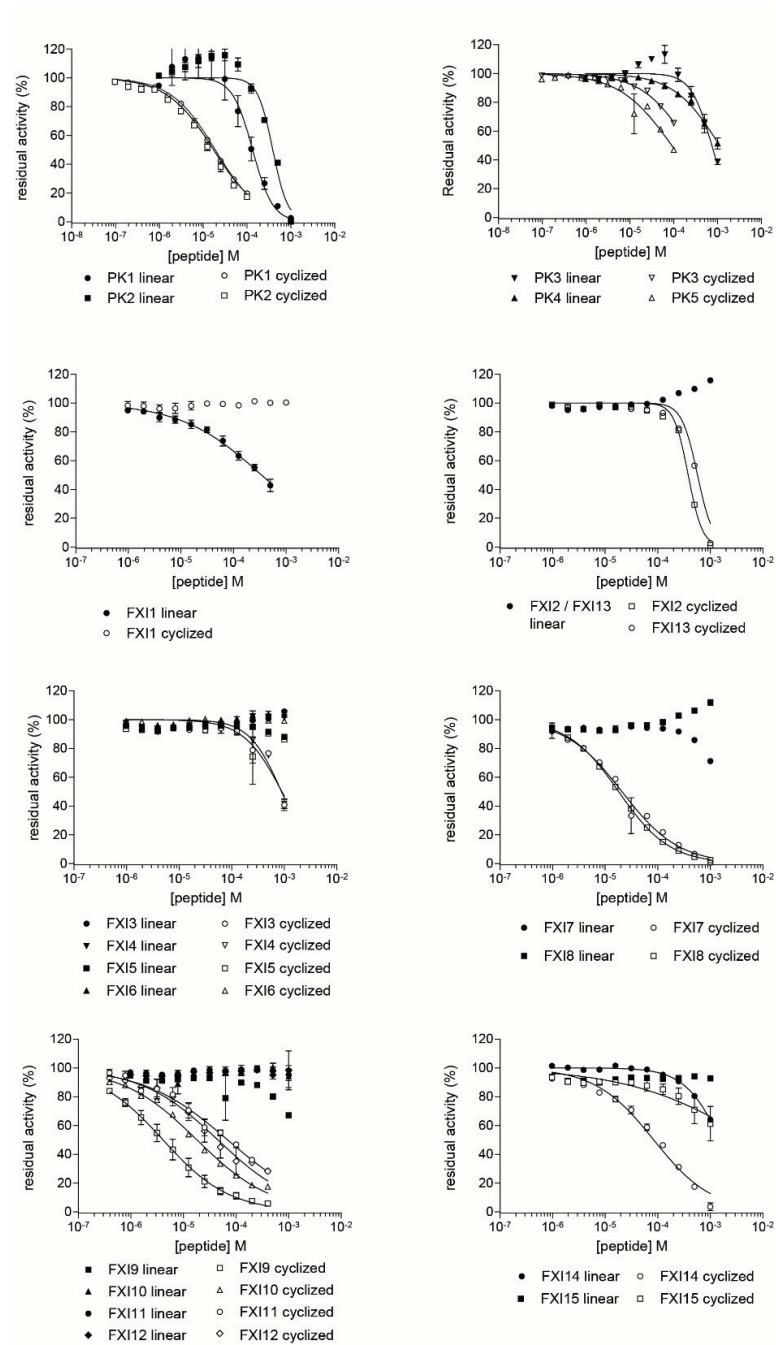
Supplementary figure 4.7. Inhibition of PK and FXI by linear and cyclic peptides. For peptides showing > 50% inhibition at a concentration of 1 mM, the Hill equation was fit to the data.

5. Bypassing bacterial infection in phage display by sequencing DNA released from phage particles

This chapter is based on a manuscript that has been submitted for publication and is currently under review. I will be the only first author in this publication.

5.1 – Abstract

Phage display relies on a bacterial infection step in which the phage particles are replicated to perform multiple affinity selection rounds and to enable the identification of isolated clones by DNA sequencing. While this process is efficient for wild-type phage, the bacterial infection rate of phage with mutant or chemically modified coat proteins can be low. For example, a phage mutant with a disulfide-free pIII coat protein, used for the selection of bicyclic peptides, has a more than 100-fold reduced infection rate compared to the wild-type. A potential strategy for bypassing this bacterial infection step is to directly sequence DNA extracted from phage particles. In this work, high-throughput sequencing was used to quantify the fraction of phage clones that can be identified by directly sequencing DNA from phage particles. The results show that the DNA of essentially all of the phage particles can be "decoded", and that the sequence coverage for mutants equals that of plasmid DNA extracted from cells infected with wild-type phage. This procedure is particularly attractive for selections with phage that have a compromised infection capacity, and it may allow phage display to be performed with particles that are not infective at all.

5.2 – Introduction

Phage display is a powerful *in vitro* screening technique that allows for the identification of ligands from large libraries of peptides or proteins. Each member of a library is displayed on the surface of a phage as a fusion with a coat protein, and the encoding gene is enclosed in the phage particle. Ligands are isolated from the libraries in iterative rounds of *in vitro* panning and amplification and are subsequently identified by sequencing the phage DNA. The phage amplification and sequencing steps rely on bacterial infection in which the phage binds to a cell and injects its DNA into the cytoplasm. For phage propagation, the cells synthesize copies of single-stranded DNA (ssDNA) and phage proteins that are assembled into new phage particles. For sequencing, typically circular phage DNA replicated by the bacterial cells is isolated. For filamentous phage, more than half of the particles of a phage preparation can infect bacterial cells[156], indicating that the infection is highly efficient. This high efficiency ensures that essentially all ligands isolated in a phage panning experiment can be amplified or identified, making phage display a highly effective method for ligand identification.

The bacterial infection step can be compromised in some phage selection systems for different reasons. First, the infectivity of filamentous phage can be reduced when large guest proteins are fused to the pIII and impair the interaction of the pIII with the F-pilus and/or TolA of bacterial cells[156] or when the displayed proteins are not properly folded and the fusion proteins, including the N-terminal domains of pIII, are proteolytically degraded[157]. Second, some phage selection procedures require a mutated coat protein. For example, a filamentous phage developed by Schmidt and co-workers that is lacking disulfide bridges in the pIII has a substantially reduced infectivity[76], with only around one out of 100 phages able to infect bacterial cells. This phage clone is applied for the selection of bicyclic peptides that are generated by chemically cyclizing linear peptides via three cysteines contained in the ligand sequence prior to affinity selection[68]. Third, the infectivity of phage can be reduced after exposure to reagents that are applied to modify peptides or proteins displayed on phage. For example, in the generation of bicyclic peptides, the reagent 1,3,5-tris(bromomethyl)benzene (TBMB) applied at a concentration of 10 μM typically reduces the number of infective phage 3 to 10-fold[78]. Other chemical transformations using harsher reaction conditions may reduce the infectivity by larger factors[158–160]. A compromised infection rate can negatively affect the phage display selection process. First, not all clones will be propagated during the phage selection. Second, several studies have reported biases that are introduced in the step of phage infection and/or propagation[122,161,162], and these biases may be more pronounced if the efficiency of the infection step is reduced.

The development of high-throughput sequencing techniques has enabled the identification of millions of clones after a phage selection. These techniques have been applied for identifying antibodies, protein domains, and peptides that were isolated by phage display[163–165]. Sequencing such a large number of clones allowed for the identification of binders immediately after a single round of phage display, omitting the infection step necessary for phage amplification for subsequent rounds[112,120]. This in turn has reduced the number of bacterial infections needed in a phage display selection process to a single one – the one required for the sequencing step.

Sequencing the DNA extracted from phage particles could in principle omit this step of bacterial infection completely. ssDNA purified from phage particles has been used routinely as a template for sequencing phage DNA[166]. The DNA in intact phage can also be efficiently amplified by incubating whole phage particles in a PCR reaction, as the capsids disassemble during the denaturation step to expose the DNA[167]. This method, termed "whole phage PCR", was applied for the direct sequencing of peptides in phage selections[168]. While a large number of clones could be sequenced with this method, it was unclear which fraction of the total phage clones could actually be

identified. An idea about the efficiency of the PCR amplification of DNA from phage particles can be obtained from quantitative PCR (qPCR) using phage particles as a template. Dias-Neto et al. showed that the DNA of samples containing as few as 10 phage particles can be identified with a readable signal in qPCR[108].

In our work, we measured what fraction of phage clones can be decoded by directly sequencing DNA from phage particles. This procedure was compared with a strategy in which plasmid DNA, obtained from cells infected with wild-type phage, was sequenced. It was shown that the sequencing of phage particle DNA is highly efficient and can identify approximately the same number of clones as when a bacterial infection step is used before the sequencing.

5.3 – Results

6.3.1   Quantification of phage particles

For assessing the fraction of phage that can be identified by sequencing the DNA released from phage particles, the absolute number of phage particles in a sample had to be determined. Wild-type fd phage displaying a library of random peptides was produced and purified by PEG precipitation (peptides of 14 amino acids fused to pIII, diversity > 10^9). The t.u. was $7.3 \times 10^{13} \pm 1.4 \times 10^{13}$ per milliliter as determined by infecting exponentially growing *E. coli* TG1 cells with serial dilutions of the sample and counting colonies grown on selective plates. The number of phage particles was determined by measuring the absorbance using two related protocols, one based on the absorbance of intact phage particles at 269 nm[169] and one based on the absorbance of extracted ssDNA at 260 nm. In the latter method, phage were incubated at 95°C and precipitated coat proteins were removed by centrifugation. The two methods gave similar phage particle concentrations of $7.2 \times 10^{13} \pm 0.3 \times 10^{13}$ and $6.5 \times 10^{13} \pm 0.2 \times 10^{13}$ per milliliter, respectively. These numbers of physical phage particles were close to the number of infective phage.

6.3.2   Quantification of ssDNA released from phage particles

The quantity of ssDNA that is released from the phage particles upon dissociation of the coat protein by heating was determined by qPCR. Purified phage plasmid DNA was used as a standard. Threefold serial dilutions of plasmid DNA and phage were subjected to 40 cycles of qPCR using a Taqman probe for signal detection (Figure 28a). The reactions with the highest and lowest

concentrations were prepared to contain 100,000 and 5 molecules of plasmid DNA and phage particles, respectively. The qPCR measurements were performed in duplicate. Cycle threshold (Ct) values were plotted against the number of molecules (plasmid DNA or phage particles) on a logarithmic scale (Figure 28b). For both, plasmid DNA and phage particles, a correlation coefficient of 0.997 was found. Amplification slopes of -3.65 and -3.30 indicated an amplification efficiency of more than 88%. The Ct values for identical dilutions of plasmid DNA and phage were similar, indicating that DNA was quantitatively released from the phage particles and that it was efficiently used as a template in the PCR reaction. Even in the reaction containing five phage particles, DNA could be released and amplified. This important step suggested that bacterial replication of DNA is not a requirement for phage display sequencing, and that DNA amplification can be efficiently accomplished solely by PCR using eluted phage particles as template.
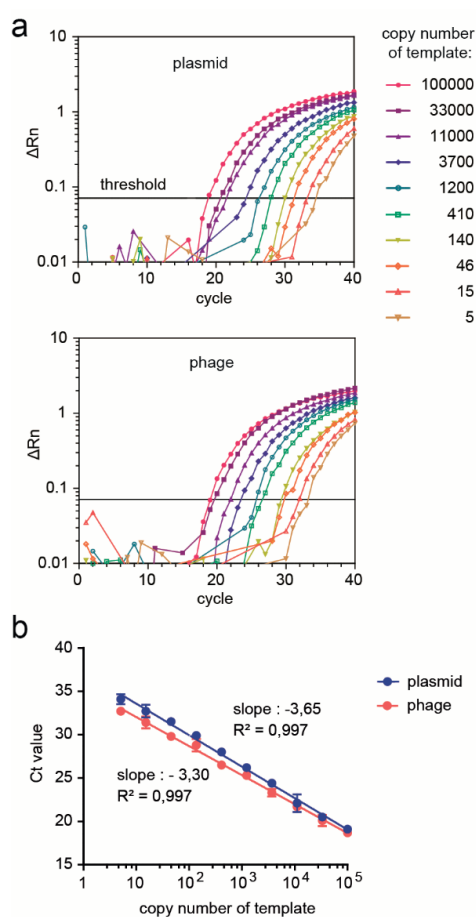
Figure 28. Quantification of ssDNA released from phage particles using qPCR. a) qPCR amplification of plasmid DNA and DNA from phage particle. Threefold dilutions of plasmid DNA and phage were prepared and qPCR reactions were performed in duplicate. Curves from duplicates were overlapping and only one dataset is shown for clarity. b) Cycle threshold (Ct) values are

plotted against the number of molecules (plasmid DNA or phage particles) on a logarithmic scale. Correlation coefficients and slopes are indicated.

6.3.3 Sequencing DNA derived from phage particles

The number of phage clones that could be identified by sequencing DNA released from phage particles was quantified by NGS using Illumina technology. Dilutions of the peptide phage library described above were prepared so that samples contained approximately 7, 70, 700, 7000, or 70,000 phage particles. The phage peptide library used contains a diversity greater than 10^9 so that essentially all samples, even the one with as many as 70,000 phage, contains clones that are all different. For comparison, equivalent dilutions of plasmid DNA were prepared.
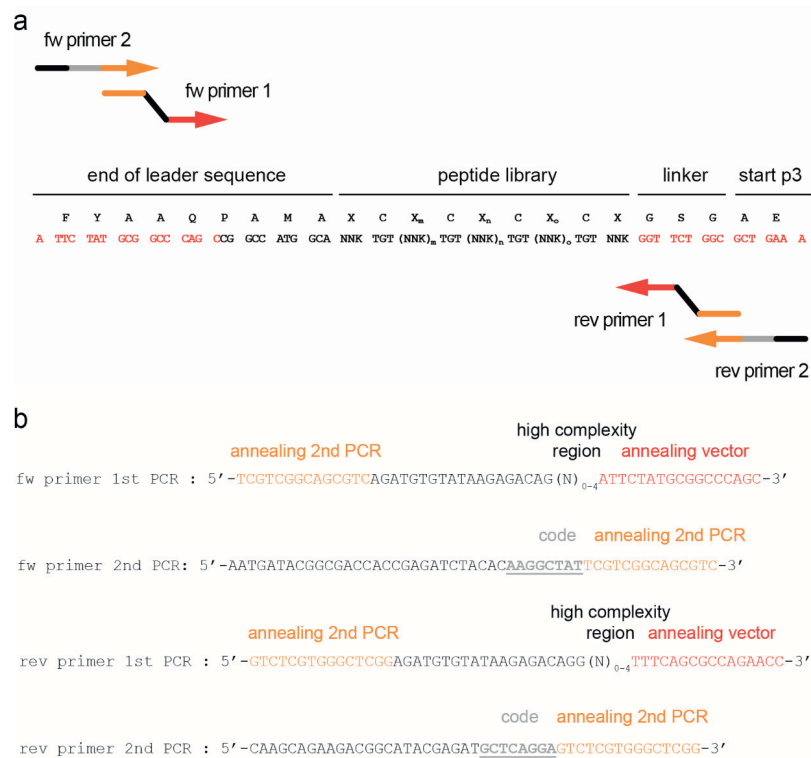
Figure 29. DNA primers for the sequencing of phage clones by Illumina NextSeq technology.
a) The region of the phage DNA covering the leader sequence, the peptide, and the N-terminal end of pIII is shown. Library peptides contain 14 amino acids (m + n + o = 8). The annealing regions of the first primer pair are shown in red. b) Primer pairs used for the two consecutive PCR reactions. Primers of the first PCR contain a high complexity region to avoid problems of low-diversity sequencing. Primers of the second PCR contain barcodes for multiplex sequencing. Both primer pairs contain regions required for Illumina NextSeq sequencing.

The plasmid DNA was obtained by infecting *E. coli* TG1 cells with the same peptide phage library, incubating the cells overnight in shaking cultures, and extracting the plasmid DNA. Two

consecutive PCR reactions were performed to prepare the DNA for NGS (Figure 29). In the first PCR reaction, the DNA coding for the random peptides in the library were amplified. In the second PCR reaction, primers were used that append a barcode to each different sample, allowing for multiplexed sequencing. The PCR products of all of the reactions were pooled, sequenced by the Illumina NextSeq 500 instrument, and the sequences were sorted based on the barcode. Between 0.5 and 4 million reads were obtained for each data set, which was suitable for the analysis of samples with a diversity estimated to be smaller than 100,000 (Tables III and IV). Quality filters were applied to eliminate reads of poor quality and wrong length as described previously[120].

	Number of plasmids subjected to sequencing[1]				
	7	70	700	7000	70000
Number of NGS reads[2]	1243589	653981	2269552	1416812	2049766
Number of different peptides	75	6398	7597	50376	104325
Number of different peptides with high abundance	12	55	450	4750	33000
Calculated number of different peptides (a of eq. 2)	11	43	453	4109	33384

Table III. NGS of plasmid DNA. [1]Calculated based on absorption at 260 nm and the dilution factor. [2]After applying a quality filter to remove reads with more than two bases with a quality score lower than Q18 and too-short or too-long peptides.

	Number of phage subjected to sequencing[1]				
	7	70	700	7000	70000
Number of NGS reads[2]	724779	1730575	2465070	1762376	481201
Number of different peptides	64	7015	8365	60822	802011
Number of different peptides with high abundance	9	90	650	5600	55000
Calculated number of different peptides (a of eq. 2)	10	103	614	5569	67439

Table IV. NGS of phage DNA. [1]Calculated based on absorption at 260 nm and the dilution factor. [2]After applying a quality filter to remove reads with more than two bases with a quality score lower than Q18 and too-short or too-long peptides.

6.3.4 Number of phage clones identified by NGS

The NGS data was analyzed to quantify the number of different clones that were identified in each sample. The number of different clones in each sample was much larger than the number of clones that were prepared in the various dilutions (plasmid DNA or phage particle). Comparison of the sequences showed that for the most abundant peptides, a group of similar peptides was found that

differed in only one or a few amino acids (or bases). Such sequences are due to sequencing errors that occur in the NGS. A software tool was applied to correct sequences that differ by 6 or fewer bases[120] which substantially reduced the number of different clones. The peptides were then ordered based on their abundance and the copy number was plotted in graphs for each dilution of plasmid DNA (Figure 30, left panels) and phage particle (Figure 31, left panels).
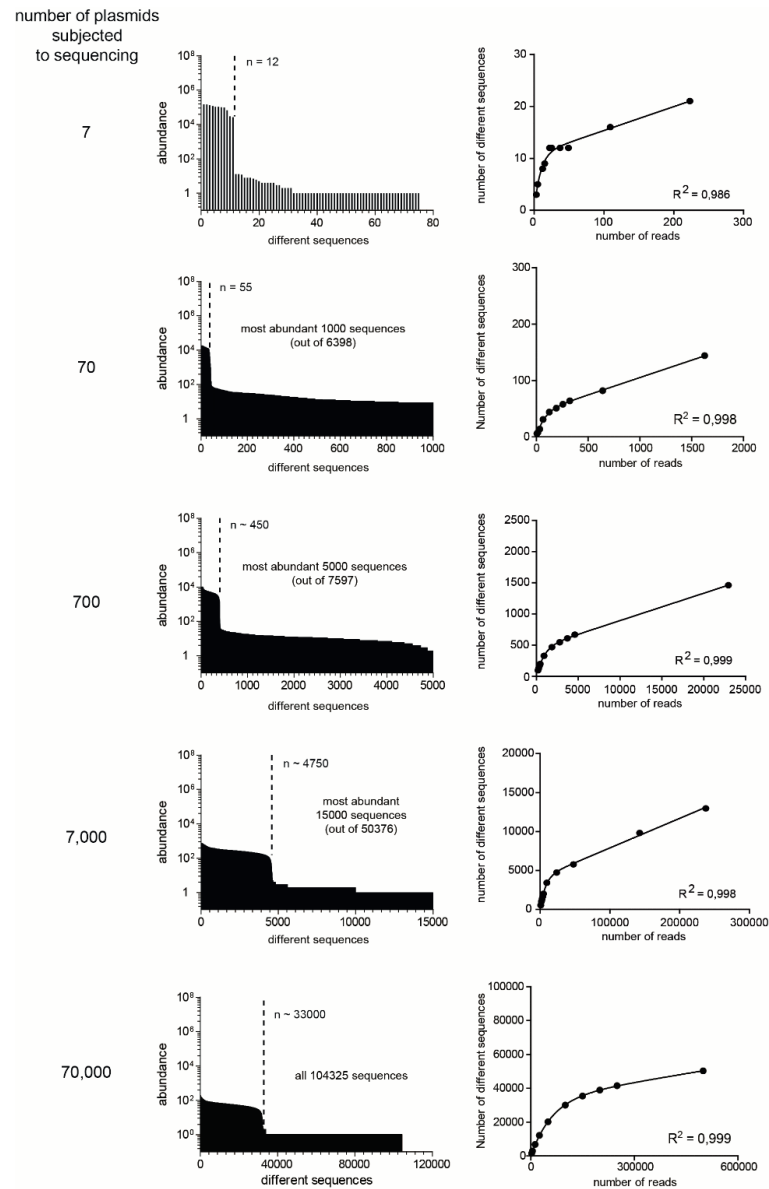
Figure 30. Sequencing of plasmid DNA. Left panels: The graphs indicate how often each peptide was identified by NGS in a sample containing the indicated number of plasmids. The peptides are arranged according to their abundance. Right panels: Number of different sequences found when increasing numbers of reads were analyzed. Saturation plots were used for the calculation of the total number of different sequences, a of equation 2, shown in Table III.
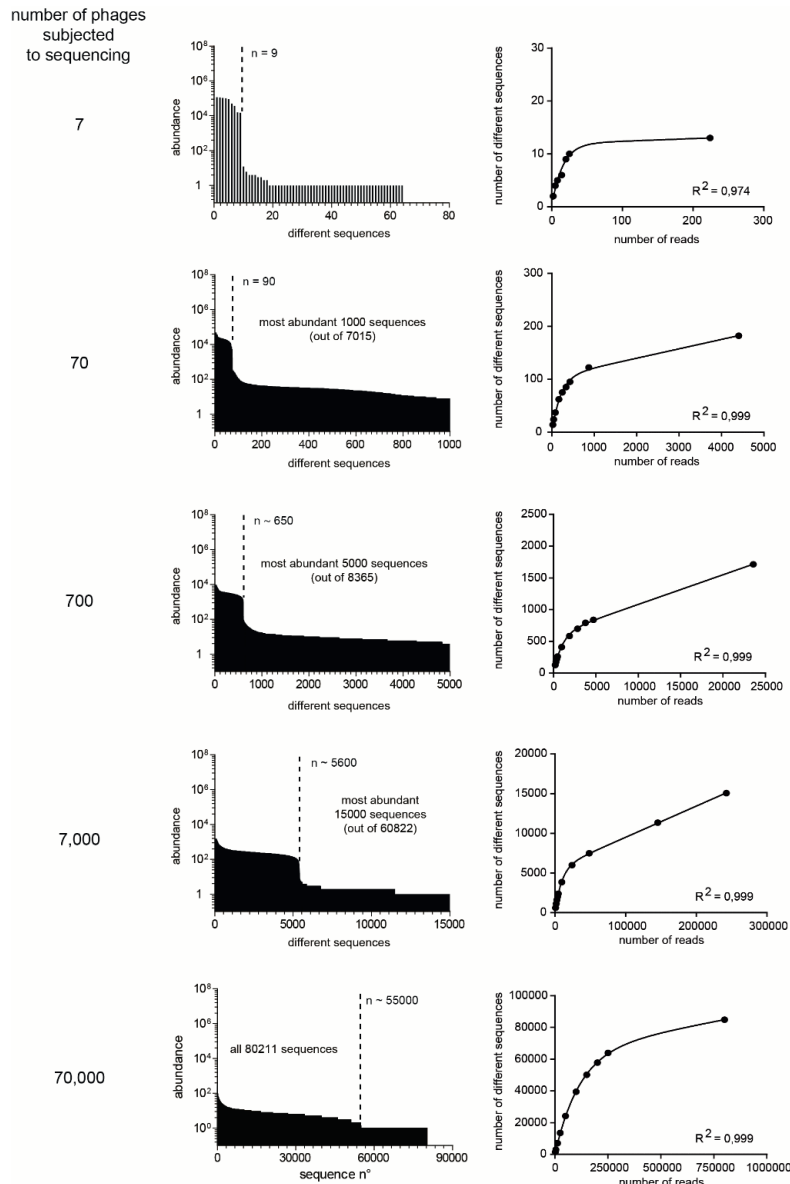
Figure 31. Sequencing of ssDNA from phage particles. Left panels: The graphs indicate how often each peptide was identified by NGS in a sample containing the indicated number of plasmids. The peptides are arranged according to their abundance. Right panels: Number of different sequences found when increasing numbers of reads were analyzed. Saturation plots were used for the calculation of the total number of different sequences, a of equation 2, shown in Table IV.

For all dilutions, an abundance distribution was found in which a small fraction of all identified clones showed a much higher copy number than the remaining clones. The two populations of high- and low-abundance clones can be distinguished visually by a sharp transition of the copy number. This transition is indicated in the graphs with a dashed line and the number of high-abundance clones is indicated as a number. Analysis of the low-abundance clones showed that they are mostly artefacts

arising from the multiplex sequencing where errors in barcode reading led to incorrect assignments. The number of high-abundance peptides thus provided a good estimate for the number of different clones that were identified within a sample.

The number of clones identified in a sample can potentially be overestimated if clones containing sequencing errors were not filtered out completely and, consequently, appear as additional clones. If the number of distinct sequences (y) is plotted as a function of the total number of reads analyzed (x), the graph should show a curve that approaches the total number of different clones (a). The data can be fitted to equation 1 where k is a constant that depends on the abundance distribution of the sample.

$$\text{Equation 1:} \qquad y = a\left(1 - e^{-\frac{x}{k}}\right)$$

If the sequencing errors are not fully filtered out, the number of distinct sequences is expected to continuously grow in a linear manner at large numbers of analyzed reads. In this case, the data can be fitted to equation 2 where b is the frequency of sequencing errors.

$$\text{Equation 2:} \qquad y = a\left(1 - e^{-\frac{x}{k}}\right) + bx$$

Fitting the data from the different samples to equation 2 (Figure 30 and 31, right panels) yielded values for the parameter b and revealed relatively low sequencing error rates between 1.7% and 6.2%. The number of different clones identified for the various samples, found by fitting the data to equation 2, is shown in Tables III and IV. Due to the low error rate after filtering, these numbers do not deviate largely from those identified by counting the high-abundance clones in the graphs shown in the left panels of Figures 30 and 31.

A comparison of the number of phage subjected to NGS sequencing and the number of identified peptide sequences shows that essentially all phage clones could be identified. The sequencing of DNA released from phage was as efficient as that of purified plasmid DNA, meaning that this technique could equivalently replace traditional phage display sequencing and has the advantage of being able to be used in situations in which bacterial infection is inefficient and sequencing from bacterial infections would not identify all clones.

5.4 – Conclusions

In this study we show that essentially all clones in a sample of phage can be identified by sequencing the ssDNA that is released upon heating the phage particles in a PCR reaction. While ssDNA from phage particles has been used as a template for sequencing phage clones for a long time, the efficiency of the process was never accessed, and the number of clones that are actually able to be identified has never been counted. Our data show that the coverage of sequences is as good as the normal phage display sequencing technique – when phage are first transformed into bacterial cells, the DNA is extracted from cells or amplified phage, and the DNA is sequenced. Based on these results, the bacterial infection for DNA sequencing may be completely replaced by direct sequencing of ssDNA from phage particles, even in experiments with wild-type phage for which the infection step is efficient. The procedure of directly sequencing the ssDNA from phage particles is particularly attractive for selections with phage that have a compromised infection capacity where only a fraction of phage-selected clones are able to infect bacterial cells in order to be identified in the sequencing step. This is, for example, the case for selections performed with phage that have a disulfide-free pIII or for selections in which chemical reagents are applied for peptide modification that impair the infectivity. This direct sequencing approach may even allow phage display with phage particles that have completely lost their ability to infect bacteria, greatly increasing the number of interesting modifications that can be performed using a phage library.

5.5 – Materials and methods

Peptide phage display library

A peptide phage display library based on the vector fd-tet[170] was used for all experiments. Phage of this library display random linear peptides of 14 amino acids encoded by NNK codons. The library comprises around 2.5×10^9 different peptides as estimated based on the number of colonies formed after transforming the library plasmid DNA into *E. coli* TG1 cells. Phage were produced by inoculating 0.5 L 2YT media containing 10 µg/mL tetracycline with library glycerol stock and growing the culture overnight at 30°C. Phage were purified by PEG precipitation as described previously and dissolved in water[144].

Quantification of infective phage

Ten-fold dilutions of phage were incubated with exponentially growing *E. coli* TG1 cells for 30 minutes at 37°C without shaking, plated on 2YT agar plates containing tetracycline (10 μg/mL), and incubated at 37°C overnight. Colonies were counted and the transforming units (t.u.) were calculated.

Quantification of phage particles

The number of phage particles was quantified based on the optical density as described previously[169]. Dilutions of a phage solution containing 2×10^{13} t.u./mL were prepared, and the absorption at 269 nm and 320 nm was measured. The number of phage particles was calculated using the following equation 3. The size of the fd-tet vector is 9.2 kb.

$$\text{Equation 3:} \qquad \text{virions/mL} = \frac{(A_{269} - A_{320}) \times 6 \times 10^{16}}{\text{number of bases / virion}}$$

In parallel, the number of phage particles was quantified by extracting the single-stranded phage DNA and measuring its optical density. A solution of phage containing 2×10^{13} t.u./mL was heated at 95°C for 10 min. The denatured capsid protein was removed by centrifugation at 10,000 g for 10 min at 4°C. The absorbance at 260 nm was measured, and the number of phage particles was calculated using an extinction coefficient of 0.027 $(\mu g/mL)^{-1}.cm^{-1}$ for ssDNA.

Phage and plasmid DNA quantification by qPCR

Plasmid DNA was purified from E. coli glycerol stocks of the peptide phage display library (around 50 mg of cell pellet) using a standard plasmid miniprep kit. The DNA was eluted from the columns with 50 μL of 5 mM Tris-Cl pH 8.5. The absorbance of the plasmid DNA at a wavelength of 260 nm was measured and the concentration calculated using an extinction coefficient of 0.02 $(\mu g/mL)^{-1}.cm^{-1}$.

Phage (7×10^{13} t.u./mL) and plasmid DNA (10^9 plasmids/mL) were diluted with water to obtain samples containing 10^5, 3.3×10^4, 1.1×10^4, 3.7×10^3, 1.2×10^3, 410, 140, 46, 15, and 5 molecules per μL (phage or plasmid DNA). The primer pair qPCRxd14f1 (5'-ATGGCAAACGCTAATAAG-3') / qPCRxd14r1 (5'-GTCACCAATGAAACCATC-3'), amplifying a region of 130 base pairs in the gene

of the phage pIII, and the TaqMan probe qPCRxd14p1 (FAM-ACCGTAATCAG-TAGCGACAGAATCAA-BHQ1), annealing within the amplified region, were used for the detection of PCR DNA amplification by fluorescence. qPCR reactions were performed in a total volume of 10 µL in MicroAmpTM optical 96-well reaction plates (ThermoFisher). The reagents were added in the following order to achieve the final concentrations indicated in brackets: 5 µL TaqMan Fast Advanced Master Mix, 0.8 µL of each primer (800 nM), 0.25 µL TaqMan probe (250 nM), 1 µL template (phage or plasmid DNA). A control reaction without template was included in each run. The reactions were run on a 7900HT Fast qPCR instrument. The thermal cycle conditions started with a 10 minute-long denaturation step at 95°C followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min each. The reactions were performed in duplicate.

Next-generation sequencing

Dilutions of phage and plasmid DNA were prepared by diluting phage 10-fold with water. PCR reactions were prepared that contained 7, 70, 700, 7000, or 7×10^4 molecules of plasmid DNA or phage particles. The primers for the two PCR reactions were designed following the guidelines of the Nextera DNA Library Preparation Kit (Illumina). In the first reaction, the forward and reverse primers listed below were used. The part of the sequence that is underlined anneals with the template DNA. The region with the "N" repeats is a high complexity region that was inserted to avoid problems of low-diversity sequencing (in Illumina sequencing, the presence of many identical sequences yields sequencing images that pose a considerable challenge to the image recognition algorithm). The 5'-regions of the primers contain sequences that are needed for Illumina sequencing. A 30 µL PCR reaction was performed by adding reagents in the following order to achieve the final concentrations indicated in brackets: 20.6 µL H_2O, 6 µL 5× standard HF buffer, 0.9 µL DMSO, 0.6 µL dNTP (250 µM), 0.6 µL primer mixture (the 10 primers were pre-mixed, 40 nM of each primer; 400 nM in total), 0.3 µL Phusion® high-fidelity DNA polymerase (0.6 unit; New England Biolabs), and 1 µL phage or plasmid DNA. 25 PCR cycles were performed (98°C for 15 sec, 55°C for 30 sec, and 72°C for 15 sec). Formation of product was assessed by agarose gel electrophoresis (2.5%).

Primers for 1st PCR reaction:

NGS Forward1	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG<u>ATTCTATGCGGCCCAGC</u>-3'
NGS Forward2	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGN<u>ATTCTATGCGGCCCAGC</u>-3'
NGS Forward3	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNN<u>ATTCTATGCGGCCCAGC</u>-3'

NGS Forward4	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNN<u>ATTCTATGCGGCCCAGC</u>-3'
NGS Forward5	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNN<u>ATTCTATGCGGCCCAGC</u>-3'
NGS Reverse1	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGG<u>TTTCAGCGCCAGAACC</u>-3'
NGS Reverse2	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGN<u>TTTCAGCGCCAGAACC</u>-3'
NGS Reverse3	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGNN<u>TTTCAGCGCCAGAACC</u>-3'
NGS Reverse4	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGNNN<u>TTTCAGCGCCAGAACC</u>-3'
NGS Reverse5	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGNNNN<u>TTTCAGCGCCAGAACC</u>-3'

A second PCR was performed to append barcodes to the DNA to code for different samples and to append the DNA sequences required for Illumina sequencing. In the primers shown below, the regions of the barcodes are underlined. A 60 μL PCR reaction was performed by adding reagents in the following order to achieve the final concentrations indicated in brackets: 41.2 μL H$_2$O, 12 μL 5x standard HF buffer, 1.8 μL DMSO, 1.2 μL dNTP (250 μM), 0.6 μL Phusion high-fidelity DNA polymerase (0.6 unit; New England Biolabs), 0.6 μL forward and reverse primer (400 nM), and 2 μL product of the 1st PCR. PCR products were purified from a 2.5% agarose gel (UltraPure agarose, Invitrogen) using a commercial agarose gel purification kit (QIAquick Gel Extraction Kit, Qiagen). DNA was eluted with 10 mM Tris-HCl, pH 8.5. Libraries were subsequently loaded at 1.5 pM on a Mid Output flow cell (Illumina) and sequenced with a NextSeq 500 instrument (Illumina) according to the manufacturer's instructions, yielding paired-end reads of 75 nucleotides. Overlapping paired-end reads were merged using CLC Genomics Workbench v9.5 (Qiagen). The DNA was sequenced along with at least 30% of unrelated non-amplicon-based libraries to avoid the low-diversity problems.

Primer for 2nd PCR reaction:

NGS 2nd Forward S507	5'-AATGATACGGCGACCACCGAGATCTACAC<u>AAGGAGTA</u>TCGTCGGCAGCGTC-3'
NGS 2nd Forward S508	5'-AATGATACGGCGACCACCGAGATCTACAC<u>CTAAGCCT</u>TCGTCGGCAGCGTC-3'
NGS 2nd Forward S510	5'-AATGATACGGCGACCACCGAGATCTACAC<u>CGTCTAAT</u>TCGTCGGCAGCGTC-3'
NGS 2nd Reverse N704	5'-CAAGCAGAAGACGGCATACGAGAT<u>GCTCAGGA</u>GTCTCGTGGGCTCGG-3'
NGS 2nd Reverse N705	5'-CAAGCAGAAGACGGCATACGAGAT<u>AGGAGTCC</u>GTCTCGTGGGCTCGG-3'
NGS 2nd Reverse N706	5'-CAAGCAGAAGACGGCATACGAGAT<u>CATGCCTA</u>GTCTCGTGGGCTCGG-3'
NGS 2nd Reverse N707	5'-CAAGCAGAAGACGGCATACGAGAT<u>GTAGAGAG</u>GTCTCGTGGGCTCGG-3'
NGS 2nd Reverse N714	5'-CAAGCAGAAGACGGCATACGAGAT<u>TCATGAGC</u>GTCTCGTGGGCTCGG-3'

Processing of sequencing data

Sequencing results were analyzed with previously developed MATLAB scripts[120]. In a first step, DNA reads and information about their quality were extracted from the .fastq file and saved in a *.txt* file using the MATLAB script *Step1.m*. In a second step, sequences were filtered to keep those with sufficient sequencing quality and the desired length, and translated into peptide sequences using the MATLAB script *Step2.m* that was modified as follows: the *ACGTOnly* parameter of the nt2aa function was set to "false" because Illumina sequencing returns ambiguous nucleotide characters (R, Y, K, M, S, W, B, D, H, V, and N). The script was executed using parameters to eliminate reads having more than two bases with a quality score lower than Q18 and reads coding for peptides that are too long or too short. In a third step, some of the remaining sequencing errors we corrected using the MATLAB script fixingerrors.m. Parameters were chosen such that this script recognized sequences that differ by 6 or fewer bases from the 1000 most abundant sequences. Such high sequence similarities were statistically unlikely and were interpreted as sequencing errors.

Analysis of sequencing data

The absolute number of different peptide sequences a was estimated by subtracting the number of sequences that were added through sequencing errors which could not be filtered or corrected by the procedures described above. The average error rate b was calculated by plotting the number of different peptide sequences y as a function of the number of analyzed reads x and fitting the data to equation 3 shown in the results section[120]. The different curves were plotted with GraphPad Prism 5 (GraphPad software), and the coefficients a, b, and k were determined by non-linear regression.

6. General conclusions and outlook

The first objective of my PhD was to set up a new method to increase the combinatorial diversity of bicyclic peptide libraries. In particular, we aimed to generate bicyclic peptide libraries that contain a high diversity of scaffolds. Although variation of the scaffold is the most important determinant of structural diversity, this has been prohibitively difficult to achieve in large combinatorial libraries up to now. Unfortunately, libraries with limited scaffold variation often result in poor screening outcomes due to incompatible interactions with their protein targets. To address this problem, we have conceived and developed a new strategy that provides rapid access to thousands of different macrocyclic scaffolds in libraries easy to synthesize, screen, and decode. The chosen approach was based on the use of bivalent linkers to cyclize peptides comprising four cysteines. To this end, my colleague, S. Kale, tested the ability of 12 structurally different chemical linkers to quantitatively and selectively link two pairs of cysteines in a model peptide. Nine linkers out of twelve reacted efficiently at concentrations that would not affect phage infectivity and propagation. In a proof of concept study, I applied six of these linkers to two phage-encoded libraries of bicyclic peptides (library 1: XCX_3CX_3CX and library 2: XCX_4CX_4CX) that were panned against the serine protease PK. As observed in prior studies, variation of the chemical linker had a strong effect on the sequence of the selected peptides[81]. In addition, our new approach led to the isolation of a peptide with 10-fold higher affinity than inhibitors developed by screening library 1 using the previous format with tri-functional linkers[77].

The great improvement of theoretical scaffold diversity has to be mitigated by the fact that not all double-bridged peptides will form on phage and strong isomer bias might occur. As seen on Figure 19, not all isomer are generated in the same proportion when cyclizing synthetic peptides. A similar result is expected for cyclization on phage. In addition, swapping the linker (Figure 22 and 23) often decreased the cyclization efficiency, peptide solubility and synthesis yield (data not shown). This implies that not all scaffolds can be easily formed and/or lead to well-behaved peptides in solution. However, even if only half of the possible scaffolds were forming, the library would still contain 37 times more scaffolds than the former libraries with 3 cysteines. Finally, the bias for a specific isomer might lead to a more difficult identification of hits after phage selection. As observed on Figure 21, peptides with the best affinities are not always the most enriched in our selection. The fact that the most active isomer might not be the most represented on the five copies of pIII could be an explanation for this poor correlation between enrichment and K_I. This limitation might be overcome by increasing the stringency of our affinity selections.

Target	Associated disease/ application	Affinity	Project done by
Factor XI	Thrombotic diseases	$K_I = 3$ nM	Dr. X Kong, V. Carle
Interleukin 17	Chronic inflammatory and autoimmune diseases	$K_d = 32$ nM	Dr. X Kong
Kallikrein-related peptidase 5	Netherton Syndrome (NS)	$K_I = 5$ nM	P. Gonschorek
Kallikrein-related peptidase 7	Netherton Syndrome (NS)	$K_I = 50$ nM	P. Gonschorek
FKBP12	Point of care detection of FK506	$K_d = 18$ nM	C. Berti

Table V. Affinity of double bridge bicyclic peptides isolated with our new method and libraries

Though this work proved the feasibility of this method, our concept was not fully exploited because we used two pre-existing phage libraries that were tailored for cyclization with tri-functional linkers. Only around 20% of the clones in library 1 and 2 contained a fourth cysteine in one of the randomized positions. In addition, the fact that one of the central cysteines was fixed limited the overall scaffold diversity that could be generated with the double bridge strategy. Although 324 peptide scaffolds were obtained after cyclization with the six chosen linkers, we anticipated that we could easily generate more than 7000 different scaffolds by varying the size of the peptides, changing the position of the central cysteines, and employing different linkers for cyclization. With this objective, two researchers of our laboratory, Dr Xudong Kong and Vanessa Carle, cloned new phage libraries comprising four cysteines ($XCX_mCX_nCX_oCX$, with m+n+o = 3-8 in X. Kong's library, and m+n+o = 12 in V. Carle's library). The number of peptide sequences in each library was estimated around 10^{10}, meaning that more than 10^{11} bicyclic peptides could potentially be generated by cyclizing with ten different linkers. These two different formats provide the peptides with different properties. In the library of V. Carle, the large size of the peptides is expected to lead to better interaction contacts between the ligand and its targets. Combined with the large peptide backbone diversity, this library has the potential to deliver ligands with high affinity and selectivity to difficult target classes, such as those involving protein-protein interactions. By contrast, X. Kong designed peptide libraries of shorter length that should theoretically be more resistant to proteolytic degradation. Our method and the new libraries have had an important impact on the future developments in our laboratory. More than seven different researchers are currently developing double bridged peptides by phage display. Binders have been isolated against a variety of targets (Table V). In all cases, phage selections with

these libraries have yielded peptides with nanomolar affinities to their targets that show great promise as therapeutic or diagnostic agents.

Finally, the double bridge macrocycle format presented in this work is of particular interest to a larger research audience not only because it allows for the generation of better ligands to challenging targets but also because it is not limited to phage-encoded libraries and can be applied to any synthetic peptide, one-bead-one-compound or mRNA display library.

In my second project, my aim was to identify cyclic peptide ligands, using phage display technology, whose size would be small enough to promise potential intestinal absorption. Indeed, the poor absorption of cyclic peptides in the gastronintestinal tract is a major limitation for their oral availability. Their large size and the numerous HBD present in amide bonds and in side chains typically prevent cell permeation. On the other hand, combining the properties of small size and high affinity in a cyclic peptide is difficult because fewer amino acids signifies fewer interactions with the target. In addition, phage displayed peptides are often cyclized via cysteine side chains which implies that two amino acid positions are not free for diversification and target interaction while contributing to the size and number of HBD. This can have a significant impact on the diversity of short cyclic peptide libraries. For example, in cyclic peptides of four or five amino acids, the number of variable amino acids is limited to two or three resulting in diversities of only 400 for peptides of the format CXXC and 8,000 for peptides of the form CXXXC. Such libraries are thus not likely to deliver ligands to any target.

Recently, two new developments in our laboratory have enabled the creation of phage-encoded libraries of small cyclic peptides that retain a relatively high diversity. The first is the establishment of a highly efficient peptide macrocyclization reaction in which a chemical linker reacts with the thiol group of a C-terminal cysteine and subsequently with the N-terminal amino group. With this reaction, only one amino acid side chain is absolutely required for the cyclization. As a result, in peptides containing four or five residues, three or four amino acid positions are free for diversification, which allows a diversity of 8,000 (XXXC) and 160,000 cyclic peptides (XXXXC), respectively. The second is the discovery of several linkers compatible with this thiol-to-amine cyclization, of which most were identified during the development of the double bridge strategy. Thus, more than 1.3 million different cyclic peptides of the format XXXC and XXXXC can be obtained if they are cyclized in parallel with the eight different reagents. I applied this strategy to selections against the two disease targets plasma kallikrein and coagulation factor XIa. This led to the identification of small thiol-to-amine cyclized pentapeptides with single-digit micromolar affinity.

As expected, the affinity of the selected peptides was more than one order of magnitude weaker than the affinity of other macrocyclic peptide inhibitors identified during phage selections with libraries containing longer peptides (PK6 = LCSLCCLRCA, K_I = 3.2 ± 0.5 nM; and FXI-X1 = ECSNIFCCRCP , K_I = 3 nM). These results highlight one limitation of using small cyclic peptide libraries. Nevertheless, this project gave valuable insight on what can be achieved with phage display technology. First, we proved that the thiol-to-amine cyclization was compatible with phage displayed peptides. This further expands the type of already known reactions that can be applied to cyclize phage displayed peptides. Second, we have isolated the smallest phage selected peptides known to date. However, the fact that peptides containing five amino acids were strongly enriched over those with four indicates that at least four side chains are needed for target engagement. Finally, the suboptimal affinities suggest that the chemical diversity conferred by the 20 canonical amino acids is too low to generate more potent binders. Thus, the thiol-to-amine strategy likely has greater potential to deliver peptides with high affinity when used in combinatorial strategies allowing the incorporation of non-natural building blocks.

As stated previously, cyclic peptides that are small, less polar, and resistant to proteases are more likely to be orally available. In this project, we addressed only the problem of the size. For instance, the two single digit micromolar inhibitors FXI7 and FXI8 (K_I = 8.8 ± 1.5 μM and 7.4 ± 0.3 μM, respectively) have a molecular mass below 700 Da. Although this is in line with our objectives, it is not sufficient to achieve oral availability. In particular, Nielsen *et al.* showed that the number of HBD is one of the most crucial parameter for oral availability of cyclic peptides[124]. Among the 125 cyclic peptides reported in their review, only one compound with HBD > 6 shows an oral availability F > 10%. Consequently, FXI7 and FXI8 contain too many HBD (HBD = 11 or 12) and the arginine alone accounts for 6 of them. A medicinal chemistry study aiming at reducing the polarity and mumber of HBD would be necessary to develop an orally available cyclic peptide. This may include modification of the side chains, N-methylation of the backbone etc. This problem has often been observed during the development of inhibitors for other trypsin-like serine proteases such as thrombin for which arginine mimetics or prodrug strategies were devised[171,172]. A similar approach could be applied to our peptides.

In conclusion, further improvement of our peptides in terms of affinity, polarity and possibly stability would be required in order to develop a drug that has the potential to be orally administered. This represents a large medicinal chemistry effort that would take several months or years. For this reason, we decided to discontinue the development of these inhibitors and rather apply the thiol-to-

amine cyclization to other combinatorial strategies that allow the incorporation of non-natural amino acids.

In my last project, I tested whether directly sequencing DNA from phage particles isolated after panning would be suitable to bypass the bacterial infection step in phage display. Therefore, we designed a study in which we showed that essentially all clones in a sample of phage could be identified by next generation sequencing of the ssDNA released upon heating the phage particles in a PCR reaction. Even a low amount of phage (< 100 particles) could be identified with this method, which is far below the typical amount of particles harvested after panning. This new method has not yet been employed for phage selection in our laboratory because standard procedures have shown to be robust enough for the type of coat protein mutations and post-translational modifications that we are currently using. In addition, applying only one round of selection would lead to a difficult isolation and identification of binders with desired characteristics. When large libraries are used, and therefore only few copies of any given binding phage clones are present, no obvious enrichment may be observed after a single round. Moreover, as stringency during the first round is generally low, relatively weak binders might be overrepresented. Therefore, our method is more likely to be useful for libraries having a small diversity and in cases where harsher genetic or chemical modifications disrupting the function of coat proteins are contemplated.

7. References

(1) Uhlig, T., Kyprianou, T., Giancarlo, F., Alberto, C., Heiligers, D., Hills, D., Ribes, X., and Verhaert, P. (2014) The emergence of peptides in the pharmaceutical business : From exploration to exploitation. *EUPROT 4*, 58–69.

(2) White, C. J., and Yudin, A. K. (2011) Contemporary strategies for peptide macrocyclization. *Nat. Chem. 3*, 509–524.

(3) Baeriswyl, V., and Heinis, C. (2013) Polycyclic Peptide Therapeutics 377–384.

(4) Driggers, E. M., Hale, S. P., Lee, J., and Terrett, N. K. (2008) The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat. Rev. Drug Discov. 7*, 608–624.

(5) Pelay-Gimeno, M., Glas, A., Koch, O., and Grossmann, T. N. (2015) Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew. Chemie - Int. Ed. 54*, 8896–8927.

(6) Fairlie, D. P., and Dantas de Araujo, A. (2016) Review stapling peptides using cysteine crosslinking. *Biopolymers 106*, 843–852.

(7) Di, L. (2015) Strategic approaches to optimizing peptide ADME properties. *AAPS J. 17*, 134–43.

(8) Zahnd, C., Kawe, M., Stumpp, M. T., De Pasquale, C., Tamaskovic, R., Nagy-Davidescu, G., Dreier, B., Schibli, R., Binz, H. K., Waibel, R., and Plückthun, A. (2010) Efficient tumor targeting with high-affinity designed ankyrin repeat proteins: Effects of affinity and molecular size. *Cancer Res. 70*, 1595–1605.

(9) Diao, L., and Meibohm, B. (2013) Pharmacokinetics and Pharmacokinetic--Pharmacodynamic Correlations of Therapeutic Peptides. *Clin. Pharmacokinet. 52*, 855–868.

(10) Craik, D. J., Fairlie, D. P., Liras, S., and Price, D. (2013) The Future of Peptide-based Drugs. *Chem. Biol. Drug Des. 81*, 136–147.

(11) Góngora-Benítez, M., Tulla-Puche, J., and Albericio, F. (2014) Multifaceted roles of disulfide bonds. peptides as therapeutics. *Chem. Rev. 114*, 901–926.

(12) Boder, E. T., Midelfort, K. S., and Wittrup, K. D. (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. U. S. A. 97*, 10701–5.

(13) Wang, C. K., and Craik, D. J. (2016) Cyclic peptide oral bioavailability: Lessons from the past. *Biopolymers 106*, 901–909.

(14) Tsomaia, N. (2015) Peptide therapeutics: Targeting the undruggable space. *Eur. J. Med. Chem. 94*, 459–470.

(15) Mäde, V., Els-Heindl, S., and Beck-Sickinger, A. G. (2014) Automated solid-phase peptide synthesis to obtain therapeutic peptides. *Beilstein J. Org. Chem. 10*, 1197–1212.

(16) Pollaroa, L., and Heinis, C. (2010) Strategies to prolong the plasma residence time of peptide drugs. *Med. Chem. Commun. 1*, 319–324.

(17) Wang, B., Nichol, J. L., and Sullivan, J. T. (2004) Pharmacodynamics and pharmacokinetics of AMG 531, a novel thrombopoietin receptor ligand. *Clin. Pharmacol. Ther. 76*, 628–638.

(18) Kratz, F. (2008) Albumin as a drug carrier: Design of prodrugs, drug conjugates and nanoparticles. *J. Control. Release 132*, 171–183.

(19) Baggio, L. L., Huang, Q., Cao, X., and Drucker, D. J. (2008) An Albumin-Exendin-4 Conjugate Engages Central and Peripheral Circuits Regulating Murine Energy and Glucose Homeostasis. *Gastroenterology 134*, 1137–1147.

(20) Angelini, A., Diderich, P., Morales-Sanfrutos, J., Thurnheer, S., Hacker, D., Menin, L., and Heinis, C. (2012) Chemical macrocyclization of peptides fused to antibody Fc fragments. *Bioconjug. Chem. 23*, 1856–1863.

(21) Parker, L. (2002) Research Design and Methods. *Pediatr. Hematol. 19*, 287–287.

(22) Zobel, K., Koehler, M. F. T., Beresini, M. H., Caris, L. D., and Combs, D. (2003) Phosphate ester serum albumin affinity tags greatly improve peptide half-life in vivo. *Bioorganic Med. Chem. Lett. 13*, 1513–1515.

(23) Dennis, M. S., Zhang, M., Meng, Y. G., Kadkhodayan, M., Kirchhofer, D., Combs, D., and Damico, L. A. (2002) Albumin binding as a general strategy for improving the pharmacokinetics of proteins. *J. Biol. Chem. 277*, 35035–35043.

(24) Veronese, F. M., and Pasut, G. (2005) PEGylation, successful approach to drug delivery. *Drug Discov. Today 10*, 1451–1458.

(25) Schellenberger, V., Wang, C.-W., Geething, N. C., Spink, B. J., Campbell, A., To, W., Scholle, M. D., Yin, Y., Yao, Y., Bogin, O., Cleland, J. L., Silverman, J., and Stemmer, W. P. C. (2009) A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol. 27*, 1186–1190.

(26) Zorzi, A., Deyle, K., and Heinis, C. (2017) Cyclic peptide therapeutics: past, present and future. *Curr. Opin. Chem. Biol. 38*, 24–29.

(27) Morgan, B., Reid, P., Sawyer, T., and Verdine, G. (2012) Bringing macrocycles full circle.

(28) https://www.novartis.com/investors/financial-data/product-sales.

(29) Stunkard, A. J. (2009) NIH Public Access. *Psychiatry Interpers. Biol. Process. 162*, 214–220.

(30) Barbieri, F., Bajetto, A., Pattarozzi, A., Gatti, M., Würth, R., Thellung, S., Corsaro, A., Villa, V., Nizzari, M., and Florio, T. (2013) Peptide Receptor Targeting in Cancer : The Somatostatin Paradigm *2013*.

(31) Borel, J.F, Z.L. Kis, T. beveridge. (1995) The history of the dicovery and development of cyclosporine (Sandimmune), in *The Search for Anti-Inflammatory Drugs: Case Histories from Concept to Clinic*, pp 27–30.

(32) http://www.drug-dev.com/Main/Back-Issues/PROTEIN-THERAPEUTICS-MARKET-Technology-Advances-Sp-1211.aspx.

(33) Levine, D. P. (2006) Vancomycin : A History. *Clin. Infect. Dis. 42*, S5-12.

(34) Goldflam, M., and Ullman, C. G. (2015) Recent Advances Toward the Discovery of Drug-Like Peptides De novo. *Front. Chem. 3*, 1–8.

(35) Obexer, R., Walport, L. J., and Suga, H. (2017) Exploring sequence space: harnessing chemical and biological diversity towards new peptide leads. *Curr. Opin. Chem. Biol. 38*, 52–61.

(36) Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (80-.). 228*, 1315 LP-1317.

(37) Chen, S., and Heinis, C. (2013) Phage Selection of Mono- and Bicyclic Peptide Ligands.

(38) Bellotto, S., Chen, S., Rentero Rebollo, I., Wegner, H. A., and Heinis, C. (2014) Phage selection of photoswitchable peptide ligands. *J. Am. Chem. Soc. 136*, 5880–5883.

(39) Kalhor-Monfared, S., Jafari, M. R., Patterson, J. T., Kitov, P. I., Dwyer, J. J., Nuss, J. M., and Derda, R. (2016) Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone. *Chem. Sci. 7*, 3785–3790.

(40) Macdougall, I. C., Provenzano, R., Sharma, A., Spinowitz, B. S., Schmidt, R. J., Pergola, P. E., Zabaneh, R. I., Tong-Starksen, S., Mayo, M. R., Tang, H., Polu, K. R., Duliege, A.-M., and Fishbane, S. (2013) Peginesatide for Anemia in Patients with Chronic Kidney Disease Not Receiving Dialysis. *N. Engl. J. Med. 368*, 320–332.

(41) Ito, K., Passioura, T., and Suga, H. (2013) Technologies for the synthesis of mRNA-encoding libraries and discovery of bioactive natural product-inspired non-traditional macrocyclic peptides. *Molecules 18*, 3502–3528.

(42) Millward, S. W., Takahashi, T. T., and Roberts, R. W. (2005) A general route for post-translational cyclization of mRNA display libraries. *J. Am. Chem. Soc. 127*, 14142–14143.

(43) Millward, S. W., Fiacco, S., Austin, R. J., and Roberts, R. W. (2007) Design of cyclic peptides that bind protein surfaces with antibody-like affinity. *ACS Chem. Biol. 2*, 625–634.

(44) Howell, S. M., Fiacco, S. V., Takahashi, T. T., Jalali-Yazdi, F., Millward, S. W., Hu, B., Wang, P., and Roberts, R. W. (2015) Serum Stable Natural Peptides Designed by mRNA Display. *Sci. Rep. 4*, 6008.

(45) Heinis, C., and Winter, G. (2015) Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol. 26*, 89–98.

(46) Timmerman, P., Beld, J., Puijk, W. C., and Meloen, R. H. (2005) Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces. *ChemBioChem 6*, 821–824.

(47) Seebeck, F. P., and Szostak, J. W. (2006) Ribosomal synthesis of dehydroalanine-containing peptides. *J. Am. Chem. Soc. 128*, 7150–7151.

(48) Josephson, K., Ricardo, A., and Szostak, J. W. (2014) MRNA display: From basic principles to macrocycle drug discovery. *Drug Discov. Today 19*, 388–399.

(49) Guillen Schlippe, Y. V., Hartman, M. C. T., Josephson, K., and Szostak, J. W. (2012) In vitro selection of highly modified cyclic peptides that act as tight binding inhibitors. *J. Am. Chem. Soc. 134*, 10469–10477.

(50) Sako, Y., Goto, Y., Murakami, H., and Suga, H. (2008) Ribosomal synthesis of peptidase-resistant peptides closed by a nonreducible inter-side-chain bond. *ACS Chem. Biol. 3*, 241–249.

(51) Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. (2008) Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol. 3*, 120–129.

(52) Yamagishi, Y., Shoji, I., Miyagawa, S., Kawakami, T., Katoh, T., Goto, Y., and Suga, H. (2011) Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library. *Chem. Biol. 18*, 1562–1570.

(53) Hayashi, Y., Morimoto, J., and Suga, H. (2012) In vitro selection of anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors. *ACS Chem. Biol. 7*, 607–613.

(54) Morimoto, J., Hayashi, Y., and Suga, H. (2012) Discovery of macrocyclic peptides armed with a mechanism-based warhead: Isoform-selective inhibition of human deacetylase SIRT2. *Angew. Chemie - Int. Ed. 51*, 3423–3427.

(55) McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*.

(56) Manuscript, A. (2008) NIH Public Access. *Growth (Lakeland) 23*, 1–7.

(57) Bruin, R. de, Spelt, K., Mol, J., Koes, R., and Quattrocchio, F. (1999) Selection of high-affinity phage antibodies from phage display libraries. *Nat. Biotechnol. 17*, 397–399.

(58) Nelson, A. L., and Reichert, J. M. (2009) Development trends for therapeutic antibody fragments. *Nat. Biotechnol. 27*, 331–337.

(59) https://www.statista.com/statistics/318206/revenue-of-humira/.

(60) Barbas, C. F. (2001) Phage Display: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

(61) Clackson, T., and Lowman, H. B. (2004) Phage Display: A Practical Approach. OUP Oxford.

(62) Nilsson, N., Malmborg, A.-C., and Borrebaeck, C. A. K. (2000) The Phage Infection Process: a Functional Role for the Distal Linker Region of Bacteriophage Protein 3. *J. Virol. 74*, 4229–4235.

(63) Smith, G. P., and Petrenko, V. A. (1997) Phage Display. *Chem. Rev. 97*, 391–410.

(64) Devlin, J. J., Panganiban, L. C., and Devlin, P. E. (1990) Random peptide libraries: a source of specific protein binding molecules. *Science (80-.). 249*, 404 LP-406.

(65) O'Neil, K. T., Hoess, R. H., Jackson, S. A., Ramachandran, N. S., Mousa, S. A., and DeGrado, W. F. (1992) Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library. *Proteins Struct. Funct. Bioinforma. 14*, 509–515.

(66) Giebel, L. B., Cass, R. T., Milligan, D. L., Young, D. C., Arze, R., and Johnson, C. R. (1995) Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities. *Biochemistry 34*, 15430–15435.

(67) Fairbrother, W. J., Christinger, H. W., Cochran, A. G., Fuh, G., Keenan, C. J., Quan, C., Shriver, S. K., Tom, J. Y. K., Wells, J. A., and Cunningham, B. C. (1998) Novel peptides selected to bind vascular endothelial growth factor target the receptor-binding site. *Biochemistry 37*, 17754–17764.

(68) Heinis, C., Rutherford, T., Freund, S., and Winter, G. (2009) Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol. 5*, 502–507.

(69) Fukunaga, K., Hatanaka, T., Ito, Y., Minami, M., and Taki, M. (2014) Construction of a crown ether-like supramolecular library by conjugation of genetically-encoded peptide linkers displayed on bacteriophage T7. *Chem. Commun. 50*, 3921.

(70) Jafari, M. R., Lakusta, J., Lundgren, R. J., and Derda, R. (2016) Allene Functionalized Azobenzene Linker Enables Rapid and Light-Responsive Peptide Macrocyclization. *Bioconjug. Chem. 27*, 509–514.

(71) Jafari, M. R., Deng, L., Kitov, P. I., Ng, S., Matochko, W. L., Tjhung, K. F., Zeberoff, A., Elias, A., Klassen, J. S., and Derda, R. (2014) Discovery of light-responsive ligands through screening of a light-responsive genetically encoded library. *ACS Chem. Biol. 9*, 443–450.

(72) Livnah, O., Stura, E. A., Johnson, D. L., Middleton, S. A., Mulcahy, L. S., Wrighton, N. C., Dower, W. J., Jolliffe, L. K., and Wilson, I. A. (1996) Functional Mimicry of a Protein Hormone by a Peptide Agonist: The EPO Receptor Complex at 2.8 Å. *Science (80-.). 273*, 464 LP-471.

(73) Wrighton, N. C., Farrell, F. X., Chang, R., Kashyap, A. K., Barbone, F. P., Mulcahy, L. S., Johnson, D. L., Barrett, R. W., Jolliffe, L. K., and Dower, W. J. (1996) Small Peptides as Potent Mimetics of the Protein Hormone Erythropoietin. *Science (80-.). 273*, 458 LP-463.

(74) Sahu, A., Kay, B. K., Lambris, J. D., Sahu, A., Kay, B. K., and Lambris2, J. D. (1996) phage-displayed random peptide library. C3-binding peptide isolated from a Inhibition of human complement by a Inhibition of Human Complement by a C3-Binding Peptide Isolated from a Phage-Displayed Random Peptide Library'. *J Immunol J. Immunol. 157*, 884–891.

(75) Chang, Y. S., Graves, B., Guerlavais, V., Tovar, C., Packman, K., To, K.-H., Olson, K. A., Kesavan, K., Gangurde, P., Mukherjee, A., Baker, T., Darlak, K., Elkin, C., Filipovic, Z., Qureshi, F. Z., Cai, H., Berry, P., Feyfant, E., Shi, X. E., Horstick, J., Annis, D. A., Manning, A. M., Fotouhi, N., Nash, H., Vassilev, L. T., and Sawyer, T. K. (2013) Stapled α–helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. 110*, E3445–E3454.

(76) Kather, I., Bippes, C. A., and Schmid, F. X. (2005) A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution. *J. Mol. Biol. 354*, 666–678.

(77) Baeriswyl, V., Rapley, H., Pollaro, L., Stace, C., Teufel, D., Walker, E., Chen, S., Winter, G., Tite, J., and Heinis, C. (2012) Bicyclic peptides with optimized ring size inhibit human plasma kallikrein and its orthologues while sparing paralogous proteases. *ChemMedChem 7*, 1173–1176.

(78) Rebollo, I. R., Angelini, A., and Heinis, C. (2012) Phage display libraries of differently sized bicyclic peptides. *Medchemcomm* 145–150.

(79) Angelini, A., Cendron, L., Chen, S., Touati, J., Winter, G., Zanotti, G., and Heinis, C. (2012) Bicyclic peptide inhibitor reveals large contact interface with a protease target. *ACS Chem. Biol. 7*, 817–821.

(80) Chen, S., Morales-Sanfrutos, J., Angelini, A., Cutting, B., and Heinis, C. (2012) Structurally Diverse Cyclisation Linkers Impose Different Backbone Conformations in Bicyclic Peptides. *ChemBioChem 13*, 1032–1038.

(81) Chen, S., Bertoldo, D., Angelini, A., Pojer, F., and Heinis, C. (2014) Peptide ligands stabilized by small molecules. *Angew. Chemie - Int. Ed. 53*, 1602–1606.

(82) Chen, S., Rentero Rebollo, I., Buth, S. A., Morales-Sanfrutos, J., Touati, J., Leiman, P. G., and Heinis, C. (2013) Bicyclic peptide ligands pulled out of cysteine-rich peptide libraries. *J. Am. Chem. Soc. 135*, 6562–6569.

(83) Baeriswyl, V., Calzavarini, S., Chen, S., Zorzi, A., Bologna, L., Angelillo-Scherrer, A., and Heinis, C. (2015) A Synthetic Factor XIIa Inhibitor Blocks Selectively Intrinsic Coagulation Initiation. *ACS Chem. Biol. 10*, 1861–1870.

(84) Rentero Rebollo, I., McCallin, S., Bertoldo, D., Entenza, J. M., Moreillon, P., and Heinis, C. (2016) Development of Potent and Selective S. aureus Sortase A Inhibitors Based on Peptide Macrocycles. *ACS Med. Chem. Lett. 7*, 606–611.

(85) Diderich, P., and Heinis, C. (2014) Phage selection of bicyclic peptides binding Her2. *Tetrahedron 70*, 7733–7739.

(86) Urech-Varenne, C., Radtke, F., and Heinis, C. (2015) Phage Selection of Bicyclic Peptide Ligands of the Notch1 Receptor. *ChemMedChem 10*, 1754–1761.

(87) Bertoldo, D., Khan, M. M. G., Dessen, P., Held, W., Huelsken, J., and Heinis, C. (2016) Phage Selection of Peptide Macrocycles against β-Catenin to Interfere with Wnt Signaling. *ChemMedChem 11*, 834–839.

(88) Walker, B., and Lynas, J. F. (2001) Strategies for the inhibition of serine proteases. *Cell. Mol. Life Sci. 58*, 596–624.

(89) Turk, B. (2006) Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov. 5*, 785–799.

(90) Lehmann, A. (2008) Ecallantide (DX-88), a plasma kallikrein inhibitor for the treatment of hereditary angioedema and the prevention of blood loss in on-pump cardiothoracic surgery. *Expert Opin. Biol. Ther. 8*, 1187–1199.

(91) Duffey, H., and Firszt, R. (2015) Management of acute attacks of hereditary angioedema: role of ecallantide. *J. Blood Med. 6*, 115–23.

(92) Kenniston, J. A., Faucette, R. R., Martik, D., Comeau, S. R., Lindberg, A. P., Kopacz, K. J., Conley, G. P., Chen, J., Viswanathan, M., Kastrapeli, N., Cosic, J., Mason, S., DiLeo, M., Abendroth, J., Kuzmic, P., Ladner, R. C., Edwards, T. E., TenHoor, C., Adelman, B. A., Nixon, A. E., and Sexton, D. J. (2014) Inhibition of plasma kallikrein by a highly specific active site blocking antibody. *J. Biol. Chem. 289*, 23596–23608.

(93) Chen, S., Gfeller, D., Buth, S. A., Michielin, O., Leiman, P. G., and Heinis, C. (2013) Improving binding affinity and stability of peptide ligands by substituting glycines with D-amino acids. *ChemBioChem 14*, 1316–1322.

(94) Angelini, A., Morales-Sanfrutos, J., Diderich, P., Chen, S., and Heinis, C. (2012) Bicyclization and tethering to albumin yields long-acting peptide antagonists. *J. Med. Chem. 55*, 10187–10197.

(95) Andreasen, P. a, Egelund, R., and Petersen, H. H. (2000) The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell. Mol. Life Sci. 57*, 25–40.

(96) Pollaro, L., Raghunathan, S., Morales-sanfrutos, J., Angelini, A., Kontos, S., and Heinis, C. (2015) Bicyclic Peptides Conjugated to an Albumin- Binding Tag Diffuse Ef fi ciently into Solid Tumors 151–162.

(97) Long, A. T., Kenne, E., Jung, R., Fuchs, T. A., and Rennöe, T. (2016) Contact system revisited: An interface between inflammation, coagulation, and innate immunity. *J. Thromb. Haemost. 14*, 427–437.

(98) Kenne, E., Nickel, K. F., Long, A. T., Fuchs, T. A., Stavrou, E. X., Stahl, F. R., and Renné, T. (2015) Factor XII: A novel target for safe prevention of thrombosis and inflammation. *J. Intern. Med. 278*, 571–585.

(99) Björkqvist, J., Sala-Cunill, A., and Renné, T. (2013) Hereditary angioedema: A bradykinin-mediated swelling disorder. *Thromb. Haemost. 109*, 368–374.

(100) Dobrovolskaia, M. A., and McNeil, S. E. (2015) Safe anticoagulation when heart and lungs are "on vacation". *Ann. Transl. Med. 3*, S11.

(101) http://www.businesswire.com/news/home/20151206005017/en/Dyax-Corp.-Presents-DX-4012-Data-2015-American.

(102) https://ash.confex.com/ash/2015/webprogramscheduler/Paper83101.html.

(103) Worm, M., Köhler, E. C., Panda, R., Long, A., Butler, L. M., Stavrou, E. X., Nickel, K. F., Fuchs, T. A., and Renné, T. (2015) The factor XIIa blocking antibody 3F7: a safe anticoagulant with anti-inflammatory activities. *Ann. Transl. Med. 3*, 247.

(104) Kokoye, Y., Ivanov, I., Cheng, Q., Matafonov, A., Dickeson, S. K., Mason, S., Sexton, D. J., Rennée, T., McCrae, K., Feener, E. P., and Gailani, D. (2016) A comparison of the effects of factor XII deficiency and prekallikrein deficiency on thrombus formation. *Thromb. Res. 140*, 118–124.

(105) Baeriswyl, V., Calzavarini, S., Gerschheimer, C., Diderich, P., Angelillo-Scherrer, A., and Heinis, C. (2013) Development of a selective peptide macrocycle inhibitor of coagulation factor XII toward the generation of a safe

antithrombotic therapy. *J. Med. Chem. 56*, 3742–3746.

(106) Middendorp, S. J., Wilbs, J., Quarroz, C., Calzavarini, S., Angelillo-Scherrer, A., and Heinis, C. (2017) Peptide Macrocycle Inhibitor of Coagulation Factor XII with Subnanomolar Affinity and High Target Selectivity. *J. Med. Chem. 60*, 1151–1158.

(107) Wilbs, J., Middendorp, S. J., and Heinis, C. (2016) Improving the binding affinity of *in-vitro*-evolved cyclic peptides by inserting atoms into the macrocycle backbone. *ChemBioChem 17*, 2299–2303.

(108) Dias-Neto, E., Nunes, D. N., Giordano, R. J., Sun, J., Botz, G. H., Yang, K., Setubal, J. C., Pasqualini, R., and Arap, W. (2009) Next-generation phage display: Integrating and comparing available molecular tools to enable costeffective high-throughput analysis. *PLoS One 4*.

(109) Mannocci, L., Zhang, Y., Scheuermann, J., Leimbacher, M., De Bellis, G., Rizzi, E., Dumelin, C., Melkko, S., and Neri, D. (2008) High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc. Natl. Acad. Sci. 105*, 17670–17675.

(110) Hodkinson, B. P., and Grice, E. A. (2015) Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. *Adv. Wound Care 4*, 50–58.

(111) Rentero Rebollo, I., Sabisz, M., Baeriswyl, V., and Heinis, C. (2014) Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res. 42*, e169.

(112) 'T Hoen, P. A. C., Jirka, S. M. G., Ten Broeke, B. R., Schultes, E. A., Aguilera, B., Pang, K. H., Heemskerk, H., Aartsma-Rus, A., Van Ommen, G. J., and Den Dunnen, J. T. (2012) Phage display screening without repetitious selection rounds. *Anal. Biochem. 421*, 622–631.

(113) Matochko, W. L., Chu, K., Jin, B., Lee, S. W., Whitesides, G. M., and Derda, R. (2012) Deep sequencing analysis of phage libraries using Illumina platform. *Methods 58*, 47–55.

(114) Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G. R., Ni, I., Mei, L., Sundar, P. D., Day, G. M. R., Cox, D., Rajpal, A., and Pons, J. (2009) Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc. Natl. Acad. Sci. 106*, 20216–20221.

(115) Lövgren, J., Pursiheimo, J. P., Pyykkö, M., Salmi, J., and Lamminmäki, U. (2016) Next generation sequencing of all variable loops of synthetic single framework scFv-Application in anti-HDL antibody selections. *N. Biotechnol. 33*, 790–796.

(116) Ravn, U., Gueneau, F., Baerlocher, L., Osteras, M., Desmurs, M., Malinge, P., Magistrelli, G., Farinelli, L., Kosco-Vilbois, M. H., and Fischer, N. (2010) By-passing in vitro screening - Next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Res. 38*.

(117) Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P. M., Bader, G. D., and Sidhu, S. S. (2010) Coevolution of PDZ domain–ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol. Biosyst. 6*, 1782.

(118) Di Niro, R., Sulic, A. M., Mignone, F., D'Angelo, S., Bordoni, R., Iacono, M., Marzari, R., Gaiotto, T., Lavric, M., Bradbury, A. R. M., Biancone, L., Zevin-Sonkin, D., De Bellis, G., Santoro, C., and Sblattero, D. (2010) Rapid interactome profiling by massive sequencing. *Nucleic Acids Res. 38*, 1–10.

(119) D'Angelo, S., Kumar, S., Naranjo, L., Ferrara, F., Kiss, C., Bradbury, A. R. M., and Daggett, V. (2014) From deep sequencing to actual clones. *Protein Eng. Des. Sel. 27*, 301–307.

(120) Rebollo, I. R., Sabisz, M., Baeriswyl, V., and Heinis, C. (2014) Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res. 42*, 1–12.

(121) Liu, G. W., Livesay, B. R., Kacherovsky, N. A., Cieslewicz, M., Lutz, E., Waalkes, A., Jensen, M. C., Salipante, S. J., and Pun, S. H. (2015) Efficient Identification of Murine M2 Macrophage Peptide Targeting Ligands by Phage Display and Next-Generation Sequencing. *Bioconjug. Chem. 26*, 1811–1817.

(122) Matochko, W. L., Cory Li, S., Tang, S. K. Y., and Derda, R. (2014) Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res. 42*, 1784–98.

(123) Christiansen, A., Kringelum, J. V, Hansen, C. S., Bøgh, K. L., Sullivan, E., Patel, J., Rigby, N. M., Eiwegger, T., Szépfalusi,

Z., Masi, F. De, Nielsen, M., Lund, O., and Dufva, M. (2015) High-throughput sequencing enhanced phage display enables the identification of patient- specific epitope motifs in serum. *Sci. Rep.* 1–13.

(124) Nielsen, D. S., Shepherd, N. E., Xu, W., Lucke, A. J., Stoermer, M. J., and Fairlie, D. P. (2017) Orally Absorbed Cyclic Peptides. *Chem. Rev.* acs.chemrev.6b00838.

(125) Levin, J. (2015) Macrocycles in Drug Discovery. The Royal Society of Chemistry.

(126) Villar, E. A., Beglov, D., Chennamadhavuni, S., Jr, J. A. P., Kozakov, D., Vajda, S., and Whitty, A. (2014) How proteins bind macrocycles *10*.

(127) Giordanetto, F., and Kihlberg, J. (2014) Macrocyclic drugs and clinical candidates: What can medicinal chemists learn from their properties *J. Med. Chem. 57*, 278–295.

(128) Dow, M., Fisher, M., James, T., Marchetti, F., and Nelson, A. (2012) Towards the systematic exploration of chemical space. *Org. Biomol. Chem. 10*, 17–28.

(129) Collins, S., Bartlett, S., Nie, F., Sore, H. F., and Spring, D. R. (2016) Diversity-Oriented Synthesis of Macrocycle Libraries for Drug Discovery and Chemical Biology. *Synth. 48*, 1457–1473.

(130) Jones, L., and McKnight, A. J. (2013) Biotherapeutics. The Royal Society of Chemistry.

(131) Ladner, R. C., Sato, A. K., Gorzelany, J., and De Souza, M. (2004) Phage display-derived peptides as therapeutic alternatives to antibodies. *Drug Discov. Today 9*, 525–529.

(132) Ng, S., Jafari, M. R., and Derda, R. (2012) Bacteriophages and viruses as a support for organic synthesis and combinatorial chemistry. *ACS Chem. Biol. 7*, 123–138.

(133) Bashiruddin, N. K., Nagano, M., and Suga, H. (2015) Synthesis of fused tricyclic peptides using a reprogrammed translation system and chemical modification. *Bioorg. Chem. 61*, 45–50.

(134) Hacker, D. E., Hoinka, J., Iqbal, E. S., Przytycka, T. M., and Hartman, M. C. T. (2017) Highly Constrained Bicyclic Scaffolds for the Discovery of Protease-Stable Peptides via mRNA Display. *ACS Chem. Biol. 12*, 795–804.

(135) Jo, H., Meinhardt, N., Wu, Y., Kulkarni, S., Hu, X., Low, K. E., Davies, P. L., Degrado, W. F., and Greenbaum, D. C. (2012) Development of α-helical calpain probes by mimicking a natural protein-protein interaction. *J. Am. Chem. Soc. 134*, 17704–17713.

(136) Cicardi, M., Levy, R. J., Mcneil, D. L., Li, H. H., Ph, D., Sheffer, A. L., Campion, M., Horn, P. T., Ph, D., Pullman, W. E., and Ph, D. (2010) Ecallantide for the Treatment of Acute Attacks in Hereditary Angioedema — NEJM. *N. Engl. J. Med. 363*, 523–531.

(137) Plosker, G. L. (2012) Recombinant Human C1 Inhibitor (Conestat Alfa). *BioDrugs 26*, 315–323.

(138) Banerji, A., Busse, P., Shennak, M., Lumry, W., Davis-Lorton, M., Wedner, H. J., Jacobs, J., Baker, J., Bernstein, J. A., Lockey, R., Li, H. H., Craig, T., Cicardi, M., Riedl, M., Al-Ghazawi, A., Soo, C., Iarrobino, R., Sexton, D. J., TenHoor, C., Kenniston, J. A., Faucette, R., Still, J. G., Kushner, H., Mensah, R., Stevens, C., Biedenkapp, J. C., Chyung, Y., and Adelman, B. (2017) Inhibiting Plasma Kallikrein for Hereditary Angioedema Prophylaxis. *N. Engl. J. Med. 376*, 717–728.

(139) Wang, Y., and Chou, D. H. C. (2015) A Thiol-Ene Coupling Approach to Native Peptide Stapling and Macrocyclization. *Angew. Chemie - Int. Ed. 54*, 10931–10934.

(140) Smeenk, L. E. J., Dailly, N., Hiemstra, H., Van Maarseveen, J. H., and Timmerman, P. (2012) Synthesis of water-soluble scaffolds for peptide cyclization, labeling, and ligation. *Org. Lett. 14*, 1194–1197.

(141) Kowalczyk, R., Harris, P. W. R., Brimble, M. A., Callon, K. E., Watson, M., and Cornish, J. (2012) Synthesis and evaluation of disulfide bond mimetics of amylin-(1-8) as agents to treat osteoporosis. *Bioorganic Med. Chem. 20*, 2661–2668.

(142) Chua, K., Fung, E., Micewicz, E. D., Ganz, T., Nemeth, E., and Ruchala, P. (2015) Small cyclic agonists of iron regulatory hormone hepcidin. *Bioorganic Med. Chem. Lett. 25*, 4961–4969.

(143) Li, Z., Partridge, J., Silva-Garcia, A., Rademacher, P., Betz, A., Xu, Q., Sham, H., Hu, Y., Shan, Y., Liu, B., Zhang, Y., Shi, H., Xu, Q., Ma, X., and Zhang, L. (2017) Structure-Guided Design of Novel, Potent, and Selective Macrocyclic Plasma Kallikrein Inhibitors. *ACS Med. Chem. Lett. 8*, 185–190.

(144) Rentero Rebollo, I., and Heinis, C. (2013) Phage selection of bicyclic peptides. *Methods 60*, 46–54.

(145) Waldmann, H., Valeur, E., Guéret, S. M., Adihou, H., Gopalakrishnan, R., Lemurell, M., Grossmann, T. N., and Plowright, A. T. (2017) New Modalities for Challenging Targets in Drug Discovery. *Angew. Chemie Int. Ed.*

(146) Driggers, E. M., Hale, S. P., Lee, J., and Terrett, N. K. (2008) The exploration of macrocycles for drug discovery - an underexploited structural class. *Nat. Rev. Drug Discov. 7*, 608–624.

(147) Josephson, K., Ricardo, A., and Szostak, J. W. (2014) MRNA display: From basic principles to macrocycle drug discovery. *Drug Discov. Today*.

(148) Morioka, T., Loik, N. D., Hipolito, C. J., Goto, Y., and Suga, H. (2015) Selection-based discovery of macrocyclic peptides for the next generation therapeutics. *Curr. Opin. Chem. Biol.*

(149) Deyle, K., Kong, X.-D., and Heinis, C. (2017) Phage selection of cyclic peptides for application in research and drug development. *Acc. Chem. Res.*

(150) White, T. R., Renzelman, C. M., Rand, A. C., Rezai, T., McEwen, C. M., Gelev, V. M., Turner, R. A., Linington, R. G., Leung, S. S. F., Kalgutkar, A. S., Bauman, J. N., Zhang, Y., Liras, S., Price, D. A., Mathiowetz, A. M., Jacobson, M. P., and Lokey, R. S. (2011) On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat. Chem. Biol. 7*, 810–7.

(151) Meyer, S. C., Gaj, T., and Ghosh, I. (2006) Highly selective cyclic peptide ligands for neutravidin and avidin identified by phage display. *Chem. Biol. Drug Des. 68*, 3–10.

(152) Wright, R. M., Gram, H., Vattay, A., Byme, S., Lake, P., and Dottavio, D. (1995) Binding epitope of somatostatin defined by phage-displayed peptide libraries. *Biotechnology. (N. Y). 13*, 165–9.

(153) Koivunen, E., Wang, B., and Ruoslahti, E. (1995) Phage libraries displaying cyclic peptides with different ring sizes: ligand specificities of the RGD-directed integrins. *Biotechnology. (N. Y). 13*, 265–270.

(154) Peraro, L., Siegert, T. R., and Kritzer, J. A. (2016) Conformational Restriction of Peptides Using Dithiol. *Pept. Protein Enzym. Des.* 1st ed. Elsevier Inc.

(155) Bork, K. (2016) A Decade of Change: Recent Developments in Pharmacotherapy of Hereditary Angioedema (HAE). *Clin. Rev. Allergy Immunol.*

(156) Nilsson, N., Malmborg, a C., and Borrebaeck, C. a. (2000) The phage infection process: a functional role for the distal linker region of bacteriophage protein 3. *J. Virol. 74*, 4229–4235.

(157) Heinis, C., Huber, a, Demartis, S., Bertschinger, J., Melkko, S., Lozzi, L., Neri, P., and Neri, D. (2001) Selection of catalytically active biotin ligase and trypsin mutants by phage display. *Protein Eng. 14*, 1043–1052.

(158) Ng, S., Jafari, M. R., and Derda, R. (2012) Bacteriophages and viruses as a support for organic synthesis and combinatorial chemistry. *ACS Chem. Biol.*

(159) Heinis, C., and Winter, G. (2015) Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol.*

(160) Bernard, J. M. L., and Francis, M. B. (2014) Chemical strategies for the covalent modification of filamentous phage. *Front. Microbiol.*

(161) Derda, R., Tang, S. K. Y., Li, S. C., Ng, S., Matochko, W., and Jafari, M. R. (2011) Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules*.

(162) Rodi, D. J., Soares, A. S., and Makowski, L. (2002) Quantitative assessment of peptide sequence diversity in M13 combinatorial peptide phage display libraries. *J. Mol. Biol. 322*, 1039–1052.

(163) Ravn, U., Gueneau, F., Baerlocher, L., Osteras, M., Desmurs, M., Malinge, P., Magistrelli, G., Farinelli, L., Kosco-

Vilbois, M. H., and Fischer, N. (2010) By-passing in vitro screening--next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Res. 38*, e193.

(164) Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P. M., Bader, G. D., and Sidhu, S. S. (2010) Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol. Biosyst. 6*, 1782–90.

(165) Christiansen, A., Kringelum, J. V, Hansen, C. S., Bøgh, K. L., Sullivan, E., Patel, J., Rigby, N. M., Eiwegger, T., Szépfalusi, Z., Masi, F. De, Nielsen, M., Lund, O., and Dufva, M. (2015) High-throughput sequencing enhanced phage display enables the identification of patient- specific epitope motifs in serum. *Sci. Rep.* 1–13.

(166) Scott, J. K., and Smith, G. P. (1990) Searching for peptide ligands with an epitope library. *Science (80-.). 249*, 386 LP-390.

(167) Kingshury, G. A., and Junghans, R. P. (1995) Screening of phage display immunoglobulin libraries by anti-M13 ELISA and whole phage PCR. *Nucleic Acids Res. 23*, 2563–2564.

(168) Ngubane, N. A. C., Gresh, L., Ioerger, T. R., Sacchettini, J. C., Zhang, Y. J., Rubin, E. J., Pym, A., and Khati, M. (2013) High-throughput sequencing enhanced phage display identifies peptides that bind mycobacteria. *PLoS One 8*.

(169) Smith, G. P., and Scott, J. K. (1993) Libraries of Peptides and Proteins Displayed on Filamentous Phage. *Methods Enzymol. 217*, 228–257.

(170) Zacher, A. N., Stock, C. A., Golden, J. W., and Smith, G. P. (1980) A new filamentous phage cloning vector: fd-tet. *Gene 9*, 127–140.

(171) Gustafsson, D., Nyström, J. E., Carlsson, S., Bredberg, U., Eriksson, U., Gyzander, E., Elg, M., Antonsson, T., Hoffmann, K. J., Ungell, A. L., Sörensen, H., Nagard, S., Abrahamsson, A., and Bylund, R. (2001) The direct thrombin inhibitor melagatran and its oral prodrug H 376/95: Intestinal absorption properties, biochemical and pharmacodynamic effects. *Thromb. Res. 101*, 171–181.

(172) Peterlin-Mavic, L., and Kikelj, D. (2001) Arginine mimetics. *Tetrahedron 57*, 7073–7105.

Camille Villequey

PhD student in chemical biology

CONTACT

Rue des crêtes,15
1018 Lausanne (CH)

+41 78 941 32 14

camille.villequey@epfl.ch

PERSONAL DETAILS

- 27 years old
- French
- Driving licence
- Swiss B permit (Oct 2018)

KEY STRENGTHS

- Phage display
- Drug discovery
- Protein & peptide analytics
- Protein & peptide engineering

LANGUAGES

French
Native language

English
Full professional proficiency

German
Currently learning
Limited working proficiency (B1)

Spanish
Listening & reading (B2)

PROFILE

Protein & peptide biochemist with 4 years PhD experience and one experience (8 months) in a biopharmaceutical start-up. Extensive knowledge of in-vitro evolution of ligands. Excellent social & organizational skills obtained from my safety coordinator & teaching responsibilities.

Available from October 2017

PROFESSIONAL EXPERIENCE

October 2013 - present
Doctoral scientist, EPFL – Lausanne (CH)

PhD supervisor : Christian Heinis

- Established a method affording the generation of libraries with high structural diversity (> 10^{12} variants, > 7500 backbone scaffolds) that is now used by most of the researchers in the laboratory
- Phage selections of binders/inhibitors to serine proteases & carbonic anhydrase IX with low nanomolar or picomolar affinity
- Peptide SAR and affinity/stability maturation
- Evaluation of new strategies to cyclize phage displayed peptides resulting in the isolation of ligands with unprecedented small size

January – August 2013
Research intern at Covagen (J&J) – Schlieren (CH)

- Cloning, selection and screening of libraries of scaffolded proteins
- Identification of binders to cancer targets with high affinity (< 20 nM)
- Characterization of ligands (affinity, specificity, epitope mapping)
- Strategic design of bispecific antibodies
- Establishment of stable cell lines and cell culture maintenance

July – August 2012
Research intern LOMIC (CNRS) – Banyuls-sur-mer (FR)

- Establishment of efficient DNA extraction from cyanobacteria
- PCR on toxin genes for subsequent RT-qPCR arrays
- Sequencing of 16S RNA

June – August 2011
Research intern at APEX Biosolutions, Besancon (FR)

- Development of strip-tests for detection of Hepatitis B virus
- Characterization immunoblot tests (sensitivity, specificity)
- Other: cell culture, bactericidal tests (NF 1040/ NF 1276)

Teaching assistant
Chemistry lab practice, chemistry and exam supervision and correction

Lab safety coordinator
Inform and advise collaborators about safety in the laboratory, organize waste management, report incidents – official swiss certificate

Stocks management
Take inventories, manage stocks and ordering

Organization of group events
For scientific and non-scientific purposes (between 20 and 30 people). Advertising, accounting, logistic organization etc

Communication & sponsoring
In student association (2010-2012). Advertise events (poster creation, emails and social media), contact sponsors for obtaining funds or goods, staffing during events welcoming up to 200 people

SCIENTIFIC ACHIEVEMENTS

Poster Prize
2016 – Swiss Summer School of Chemical biology

Oral Communication
2016 – Swiss Summer School of Chemical biology

Poster & Oral Communication
2017 – Gordon Research Conference/Seminar of High Throughput Chemistry and Chemical Biology

SKILLS

Molecular biology
- Cloning (especially of phage libraries), gene design (fusion constructs)
- Directed mutagenesis
- qPCR
- Next generation sequencing (Illumina & ion torrent)

Protein & peptide biochemistry
- SPPS, recombinant expression (prokaryotic systems)
- Purification (His-Tag, Strep-tag, SEC, RP-HPLC)
- Characterization (electrophoresis, HPLC, MS)
- Post-translational modifications

Screening assays
- Enzymatic assays, binding assays (FACS, SPR, ELISA, FP)
- Phage display

EDUCATION

2010 – 2013
Engineering Diploma (equivalent to master degree)
École supérieure de biotechnologie de Strasbourg, Illkirch (FR)
Tri-national school with Basel (CH) and Freiburg (DE) universities
Specialization: synthetic biology

2008 – 2010
Post-secondary preparatory school
Lycée Victor Hugo, Besançon (FR)
Classes preparing for entrance examinations to the French "Grandes Écoles" – BCPST division (Maths, Chemistry, Physics, Life science and Geology)

2008
Baccalauréat (scientific)
Lycée Aragon, Héricourt (FR)
Obtained with highest honors

LIST OF PUBLICATIONS

Sangram Kale, Camille Villequey, Alessandro Zorzi, Kaycie Deyle, Christian Heinis. "**Cyclizing peptides with two chemical bridges affords large macrocycle scaffold diversity**," *Manuscripts submitted, to be published in 2017.Co-authorship.*

Camille Villequey, Xudong Kong, Christian Heinis. "**Analyzing phage panning outputs by direct sequencing of ssDNA from selected virions**," *Advanced manuscript, to be published in 2017.*

Camille Villequey, Sangram S.Kale, Peter Chen, Christian Heinis. "**Phage selection of macrocyclic peptide inhibitors of serine proteases with unprecedented small size**," *Manuscript in progression.*