Phage-encoded combinatorial chemical libraries based on bicyclic peptides

Christian Heinis*[1,3], Trevor Rutherford[2], Stephan Freund[2] and Greg Winter*[1]

[1] Laboratory of Molecular Biology, and [2] Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, UK

[3] Present address: Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

*Corresponding authors

E-mail addresses: christian.heinis@epfl.ch, winter@mrc-lmb.cam.ac.uk

Abstract

Here we describe a phage strategy for the selection of ligands based on bicyclic or linear peptides attached covalently to an organic core. Peptide repertoires were designed with three reactive cysteine residues, each spaced apart by several random amino acid residues, and fused to the phage gene-3-protein. Conjugation with tris-(bromomethyl)benzene via the reactive cysteines generated repertoires of peptide conjugates with two peptide loops anchored to a mesitylene core. Iterative affinity selections yielded several enzyme inhibitors; after further mutagenesis and selection, we were able to chemically synthesize a lead inhibitor (PK15) (K_i = 1.5 nM) specific to human plasma kallikrein that efficiently interrupted the intrinsic coagulation pathway in human plasma tested *ex vivo*. This approach offers a powerful means of generating and selecting bicyclic macrocycles (or if cleaved linear derivatives thereof) as ligands poised at the interface of small molecule drugs and biologics.

Introduction

The discovery of novel ligands to receptor, enzyme and nucleic acid targets represents the first stage in the development of therapeutic drugs. For drugs based on small organic ligands, high throughput screening (HTS) has proved a popular strategy; large libraries of compounds are synthesized (or purchased) and each compound assayed for binding to the targets. With the use of robots it is possible to screen 10^5-10^6 compounds per day, but the hits usually require further chemistry to improve their binding affinity and target specificity[1, 2].

For drugs based on nucleic acids, peptides or proteins, biological selection methods offer an alternative strategy. These methods (such as phage display, ribosome display, mRNA display or RNA/DNA aptamer technologies) rely on (a) creating a diverse genetic library wherein the phenotype (binding to target) of each member of the library is linked to its genotype (the encoding DNA or RNA)[3, 4], and (b) an interative cycle in which library members are selected for binding to target, and then amplified (by replication in a host cell, or by copying of the encoded nucleic acid *in vitro*)[5, 6]. At each round of selection the binders are thereby enriched over the non-binders. Very large libraries (10^9-10^{13} members) can be efficiently screened by a few rounds of selection and lead hits can be refined by mutation and further selection[7-9]. The approach is very powerful and has been used to fashion ligand binding sites in antibodies[10-12] and other protein scaffolds[13], leading to the first human therapeutic antibody Humira™ to be approved by the FDA.

Several attempts have been made to develop selection methods for the isolation of small organic ligands. Typically DNA is used as a tag that can be readily synthesized,

sequenced, amplified and/or hybridized. For example, small molecules can each be conjugated to a unique DNA[14-16] (or bacteriophage[17]) tag, and the conjugates mixed together to create a tagged small molecule library. After selection of the library against the target, the small molecule "hits" can be identified by the sequences of their (amplified) tags. Alternatively the DNA tags can be introduced during the synthesis of combinatorial chemical libraries. For example, small molecules and a corresponding tag are synthesised in parallel on the same bead[18], or hybridisation of the tag is used to govern the route of chemical synthesis[19-21]. From such libraries the synthetic route (and thereby structure) of the selected hits can be deduced from the sequence of the tag. Notwithstanding their ingenuity, these methods suffer from common disadvantages; either the small molecule is linked to the DNA tag only during the first round of selection, or the method requires a high efficiency of chemical translation[20], rendering iterative cycles either impossible (and limiting application to small libraries) or technically demanding.

Methods have been described for tethering peptides through reactive cysteine side chains to the functional groups of an organic scaffold[22]. More recently others have envisaged that these methods could be used for the generation of variant peptide conjugates that mimic discontinuous epitopes of folded proteins, and with binding activities assayed by HTS[23-25]. As an alternative, we wondered if we could chemically modify the cysteine residues of the peptides displayed on phage while maintaining the phage infectivity[26, 27], and by harnessing the power of iterative selection have access to much greater chemical diversity than possible by HTS (**Fig. 1a**). In a first embodiment, these conjugates are reminiscent of the peptide macrocyclic drugs[28, 29] (**Fig. 1b**), but whereas many peptide macrocycles (from fungus, yeast and bacteria)

are made *in vivo* by non-ribosomal peptide synthases, our strategy uses ribosomal synthesis. In a second embodiment, cleavage of the cyclic peptide repertoires generates an organic core decorated with highly diverse linear peptide side chains (**Fig. 1a**).

Results

Conjugation of an organic scaffold to peptides on phage

We used the small organic compound tris-(bromomethyl)benzene (TBMB; 1) as a scaffold to anchor peptides containing three cysteine residues[22, 24] (**Fig. 1a**). The reaction occurs in aqueous solvents at room temperature, and the three-fold rotational symmetry of the TBMB molecule ensures the formation of a unique structural and spatial isomer.

We first elaborated the reaction conditions for conjugation of the peptide ^{N}GCGSGCGSGCG^{C} fused to the soluble D1-D2 domains of the phage pIII, analysing the molecular weight of the products by mass spectrometry. However, we were unable to selectively conjugate the three cysteine residues of the peptide with TBMB while sparing the disulphide bridges of D1 and D2 (C7-C36, C46-C53, C188-C201). This prompted us to take advantage of a disulfide-free gene-3-protein recently developed by Schmidt F. X. and co-workers[30]. The peptide-D1-D2 (disulfide free) fusion protein was reduced with tris-(carboxyethyl)phosphine (TCEP), the TCEP removed and TBMB added. A concentration of 10 μM TBMB was sufficient for quantitative reaction with of peptide-fusion protein at 30°C in one hour, giving predominantly one product with the expected molecular mass (Δ mass expected = 114 Da; **Fig. 2a**). No product was detected with the (disulfide-free) D1-D2 protein. Reaction of TBMB with peptide-D1-D2 (disulfide-free) fusions containing only two cysteine residues (^{N}AGSGCGSGCG^{C}-D1-D2) yielded a product with a molecular mass consistent with reaction of both cysteines and the α-amino group at the peptide N-terminus (**Supplementary Fig. 1a,b**). Similarly, the reaction of TBMB with a peptide-D1-D2 (disulfide free) fusion having one cysteine and a lysine (^{N}AGSGKGSGCG^{C}-D1-D2)

yielded a molecular mass consistent with the reaction of the cysteine, the α-amino group of the N-terminus and the ε-amino group of the lysine (**Supplementary Fig. 1c,d**). Thus the functional groups of TBMB preferentially react with the thiol groups of cysteines but can also react intramolecularly with primary amines in the absence of cysteine. Having identified suitable conditions, we reacted TBMB with (disulfide-free p3) phage bearing the peptide ^{N}G<u>C</u>GSG<u>C</u>GSG<u>C</u>G^{C}. This led to a small loss (5-fold) of phage infectivity at 10 μM TBMB, but much greater losses at higher concentrations (**Fig. 2b**); we suggest that this may be due to cross-linking of the phage coat proteins through lysine residues.

Creation of polycyclic peptide library and affinity selection

We designed a library of peptides comprising two sequences of six random amino acids flanked by three cysteines (Cys-$(Xaa)_6$-Cys-$(Xaa)_6$-Cys; **Fig. 3a**) for display on the (disulphide-free p3) phage. An alanine residue was added to the N-terminus of the peptide to ensure a correct processing of the signal sequence. A Gly-Gly-Ser-Gly linker was placed between the third cysteine and the gene-3-protein. As the (disulfide-free p3) phage had a 100-fold reduced infectivity compared to wild-type phage, we grew up a 1-litre volume of culture (incubated over night at 30°C this typically yielded 10^{11}-10^{12} infective particles).

We tested the library (estimated $4.4x10^9$ variants) of polycyclic peptides for binding and inhibition of the human proteases plasma kallikrein and cathepsin G. About 10^{12} purified infective phage particles were chemically modified with TBMB and then incubated with the biotinylated target proteins. After capture on magnetic streptavidin or avidin beads, the enriched phage were treated to two further rounds of selection,

each round comprising amplification (by infection of bacteria), chemical conjugation and capture with the biotinylated targets. The phage titre increased after the second and third rounds suggesting enrichment of specific binders. DNA encoding the peptides was PCR-amplified from the selected population of phage in the third round, and recloned for periplasmic expression as peptide-D1-D2(disulfide-free) fusion proteins and sequenced. This revealed consensus sequences in one or both of the peptide loops (**Fig. 3b,c**) and several of the fusion proteins were expressed, purified, conjugated with TBMB and tested for their inhibitory activity to protease. The best plasma kallikrein and cathepsin G inhibitors had an IC_{50} of 400 nM (PK2 and PK4) and 100 nM (CG2 and CG4) respectively when tested as a D1-D2 fusion. Since we screened the phage selected clones for inhibition (rather than binding) we do not know whether peptide conjugates were selected that bind but not inhibit the proteases. However, the finding that the vast majority of clones tested after the three rounds of phage selection displayed inhibitory activities suggests that inhibitors were predominantly selected.

Affinity maturation of human plasma kallikrein inhibitors

Most of the sequences of the kallikrein binders revealed consensus sequences in one or other of the peptide loops. Three new libraries were created with each one of the three consensus regions in one loop and six random amino acids in the other loop (**Fig. 4a**). The libraries were mixed and phage panned under stringent conditions (1 nM to 200 pM biotinylated kallikrein). The random sequence converged to a new consensus, yielding clones with consensus sequences in both loops (**Fig. 4b**). Inhibition assays revealed that the IC_{50} of the best inhibitor (PK15) was 20 nM when tested as a D1-D2 fusion protein.

Activity and specificity of chemically synthesized inhibitors

Peptides corresponding to four kallikrein inhibitors from the primary selection (PK2, PK4, PK6 and PK13) and the best inhibitor from the affinity maturation selection (PK15) were chemically synthesized on a solid phase. The peptides had an alanine residue at the N-terminus and an amidated glycine at the C-terminus to represent the charge and chemical environment of the phage displayed peptides. The TBMB conjugated synthetic peptides proved at least 250-fold more potent inhibitors of kallikrein activity than the unconjugated peptides (Table 1). They were also more potent than the D1-D2 conjugates by a factor of more than ten (**Fig. 5a and Table 1**); this may be due to steric blocking of the conjugated peptide moiety by the D1-D2 moiety. The apparent inhibition constant (K_i) of the peptide conjugate PK15 (**Fig. 1b**) was calculated to be 1.5 nM using the equation of Cheng and Prusoff[31].

Incubation of the conjugate PK15 with kallikrein led to hydrolysis of a peptide bond after prolonged incubation (90% cleavage after 24h at 37 °C), as shown by a mass gain of 18 Da, but the inhibitory activities of cleaved and uncleaved samples proved similar (IC_{50} 2.2 nM and 1.6 nM respectively). Investigation of the cleavage site through incubation of the cleaved PK15 with leucyl aminopeptidase from *Aeromonas Proteolytica* and mass-spectrometric analysis of the formed products revealed that PK15 is cleaved by plasma kallikrein between Arg7 and Asn8 (**Supplementary Fig. 2**). The aminopeptidase also cleaved off the N-terminal alanine of PK15 which confirmed that the N-terminal amine had not reacted with TBMB.

The five inhibitors were also tested against mouse plasma kallikrein (79% sequence identity) or the homologous human serine proteases factor XIa (63% sequence identity) and thrombin (36% sequence identity). None inhibited these enzymes at the highest concentration tested (10 μM).

Interruption of the intrinsic coagulation pathway

Human plasma kallikrein plays a key role in the first events of the intrinsic coagulation pathway by converting factor XII to factor XIIa which then acts on the next protease in the pathway. We tested whether conjugate PK15 could inhibit the activation of factor XIIa in human plasma samples. The pathway was triggered with caolin and the activity of factor XIIa was measured with a colorimetric substrate. The activity of XIIa was halved in the presence of 160 nM conjugate PK15 (**Supplementary Fig. 3**). By comparison 5 μM of aprotinin, a 6 kDa bovine serine protease inhibitor also used clinically as a plasma kallikrein inhibitor (K_i = 30 nM), was required for the same effect.

Structure determination of TBMB modified peptide PK15

2D ^{1}H NMR spectra of the conjugate PK15 were recorded (**Supplementary Fig. 4**) and a sequence specific assignment of the chemical shifts of the TOCSY and NOESY spectra was possible. From the NOE data there was no evidence of interactions between the loops, no NOEs across the loops, and no evidence of short segments with regular secondary structure. Only one NOE (a 'weak' NOE between Arg5 HB and Cys2 HN) was between protons separated by more than two residues in the sequence. The absence of long range and medium range NOEs precluded a restraint-driven minimisation, and hinders the definition of a single defined structure with confidence.

A structural model of PK15 was generated using a standard restrained simulated annealing protocol in version 1.1 of CNS (**Fig. 5b**). Of the 50 accepted structures, none violated the NOE restraints and the closest H-H contact in the model that did not have a corresponding NOE correlation was 3.7 A (**Supplementary Table 1**). The model was therefore consistent with the available experimental data. However, we have no evidence of the extent of flexibility in the loops, and the range of conformers that may contribute to the conformational ensemble that is present in solution. In the model, the two peptide loops do not interact with each other and are arranged (but not closely packed) around the mesitylene core to which they are covalently tethered.

Discussion

We have shown how the reaction of tris-(bromomethyl)benzene (TBMB)[22, 24] with libraries of cysteine-rich peptides displayed on filamentous bacteriophage generates conjugates amenable to iterative selection. It was a challenge to conjugate the displayed peptide while sparing the phage, and we had to vary reagent concentrations, solvent composition and reaction temperature, and also use phage lacking disulfides in the gene-3-protein. From a library of $>10^9$ members and iterative selections we succeeded in isolating potent human plasma kallikrein inhibitors (<2000 Da). Our lead inhibitor (PK15) with K_i = 1.5 nM efficiently interrupted the intrinsic coagulation pathway in human plasma tested *ex vivo*, and was highly specific: it did not inhibit mouse plasma kallikrein or the homologous human plasma proteases factor XIa and thrombin. Inhibitors of human plasma kallikrein are being developed clinically for treatment of hereditary angioedema and coronary bypass surgery, but it has proved difficult to make small molecules that are kallikrein-specific (reviewed in[32, 33]). It is therefore promising that our strategy yielded a high affinity and specific inhibitor.

Our repertoire was built from 17 residue peptides with three cysteines, each spaced apart by six random amino acids. These conjugates should have advantages over other peptide forms. For example, compared to monocyclic peptides[34, 35], the bicyclic conjugates should be more capable of extensive interactions with globular proteins and perhaps thereby more able to block protein-protein interactions; compared to disulfide-bonded cyclic peptides, the cross-links should be inert to exchange and stable in reducing environments[36]; and compared to linear peptides the conjugates should be constrained and bind more tightly to targets (due to the smaller loss of conformational entropy on binding). Indeed our literature review of the most potent

peptide inhibitors isolated by phage display (**Supplementary Table 2**) shows that the majority are constrained by disulfide bonds. Constrained peptides are also expected to be more resistant to protease cleavage and/or inactivation than linear peptides, as is consistent with our observations that PK15 was cleaved in one of the loops only after prolonged incubation with human plasma kallikrein (and even then remained active).

After conjugation with TBMB the peptides are expected to form two six-residue loops covalently attached to a mesitylene core, as confirmed by the structure of the PK15 kallikrein inhibitor solved by NMR (**Fig. 5b**). The N-terminal peptide loop of the PK15 conjugate has a tight turn that is similar to those found in other peptide macrocycles, including the hormones oxytocin (**2**)[37, 38] and octreotide (**3**)[39, 40] (they both contain hexapeptides that are cyclized by disulfide bonds). However in contrast to most peptide macrocycles the PK15 conjugate has more polar side chains and is more soluble in aqueous solution. As hydrophobic side chains appear to be important for the interaction of antibacterial or antifungal peptide macrocyclic drugs with membranes, it may be necessary to increase the hydrophobic character of the side chains or core in our macrocycles to create drug leads.

The peptide macrocycles have some structural similarities with zinc finger proteins. In both types of molecules polypeptide chains are coordinated via cysteine residues (in zinc finger proteins also via histidine residues) to a central molecule to gain a tertiary structure. While in zinc finger proteins cysteine and histidine residues form non-covalent bonds to Zn^{2+}, the macrocycles isolated in this work form covalent bonds to the mesitylene core. The zinc finger proteins have a slightly larger molecular mass (around 30 amino acids), have obligate secondary structure elements (an anti-parallel

β-sheet and an α-helix) and are likely to be more tightly folded due to these elements.

The molecular weight of PK15 (1939.4 Da; both ring sizes are 31 atoms) is higher than several peptide macrocycles with biological activity such as oxytocin, octreotide, polymyxin B (**4**), daptomycin (**5**), cyclosporine (**6**), and caspofungin (**7**) (with molecular weights of 1000 to 1600 and 20 – 32 atoms in the ring), but it would be possible to use shorter loops. For example by altering the spacing of the cysteines, the loop length is readily varied, or even extra segments added to the peptide termini. Further variations could include mutagenesis of the loops (as with the affinity maturation of PK15); proteolytic cleavage in one or both loops to generate a mesitylene core with three peptide side-chains each linked to the core through cysteines; chemical conjugation to the nascent peptide N- or C-termini after loop cleavage[41]; use of other reactive amino acids (e.g. selenocysteine[42, 43]) to conjugate the peptide to the core; or use of variant organic cores. In particular a larger organic core, or one with more functional groups could help stabilize the peptide loops or interact directly with the target. In this respect we note that the two most potent peptide inhibitors in Supplementary Table 2 (with inhibition constants comparable to PK15) both comprise at least two tryptophan residues[44, 45] which may act in similar manner in these peptides. New functionalities such as fluorescence might also be introduced via the chemical core. The ability to isolate and engineer variant conjugates by both genetic and chemical engineering on the phage, and to make further variations by routine chemical synthesis of free conjugates, makes this strategy a remarkably flexible and attractive means for generating and developing drug leads. The small size and simplicity of the chemistry may also facilitate the manufacture of products by total synthesis.

Materials and Methods

The protocols for the chemical modification of peptides on phage, the phage panning, the inhibition screening and the chemical synthesis of the inhibitors are described in this section. The protocols for the cloning of the phage libraries, the cloning and expression of peptide fusion proteins, the activity measurement of the inhibitors and the structure determination can be found online in the Supplementary Methods.

Chemical modification of peptide repertoires with TBMB on phage

We cloned and produced phage peptide libraries that are based on the plasmid fdg3p0ss21[30] as described in the Supplementary Methods. Typically 10^{11}-10^{12} t.u. of PEG purified phage were reduced in 20 ml of 20 mM NH_4HCO_3, pH 8 with 1 mM TCEP at 42°C for 1hr. The phage were concentrated at 4000 rpm in a Vivaspin-20 filter (MWCO of 10,000) to 1 ml, washed twice with 10 ml ice cold reaction buffer (20 mM NH_4HCO_3, 5 mM EDTA, pH 8), taken up in 32 ml reaction buffer and 8 ml of 50 μM TBMB in ACN were added to obtain a final TBMB concentration of 10 μM. The reaction was incubated at 30°C for 1 hr before non-reacted TBMB was removed by precipitation of the phage with 1/5 volume of 20% PEG, 2.5 M NaCl on ice and centrifugation at 4000 rpm for 30 minutes.

Phage selections with human plasma kallikrein and cathepsin G

We incubated biotinylated human plasma kallikrein and cathepsin G (5 to 20 μg; biotinylation as in the Supplementary Methods) in 0.5 ml washing buffer (10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1mM $CaCl_2$) containing 1% BSA and 0.1 % Tween 20 for 30 minutes. To the chemically modified phage (typically 10^{10}-10^{11} t.u. dissolved in 2 ml washing buffer) was added 1 ml of washing buffer

containing 3% BSA and 0.3% Tween 20, and after 30 min 0.5 ml antigen was added and incubated for 30 minutes on a rotating wheel at RT. 50 μl magnetic steptavidin beads (Dynal, M-280 from Invitrogen, Paisley, UK), pre-incubated in 0.5 ml of washing buffer containing 1% BSA and 0.1 % Tween 20 for 30 minutes, was added to the phage/antigen mixture and incubated for 5 minutes at RT with rotation. The beads were washed 8 times with washing buffer containing 0.1% Tween 20 and twice with washing buffer before incubation with 100 μl of 50 mM glycine, pH 2.2 for 5 minutes. Eluted phage were transferred to 50 μl of 1 M Tris-Cl, pH 8 for neutralization, incubated with 50 ml TG1 cells at OD_{600}=0.4 for 90 minutes at 37°C and the cells were plated on large 2YT/chloramphenicol plates. We performed two additional rounds of panning using the same procedures. In the second round of selection, neutravidin-coated magnetic beads were used to prevent the enrichment of streptavidin-specific peptides. The neutravidin beads were prepared by reacting 0.8 mg neutravidin (Pierce, Rockford, IL, USA) with 0.5 ml tosyl-activated magnetic beads (Dynal, M-280 from Invitrogen, Paisley, UK) according to the supplier's instructions.

Screening selected clones for inhibitory activity

We cloned the genes that encode the peptides selected in the second and third round of biopanning into a pUC119 based vector for expression of the peptide-D1-D2 fusion proteins (disulfide-free D1-D2 protein; the cloning and expression procedures are described in the Supplementary Methods). Oxidized sulfhydryl groups of the peptides were reduced by incubation of the protein (1-10 μM) with 1 mM TCEP in 20 mM NH_4HCO_3, pH 8 at 42°C for 1 hr. The reducing agent was removed by size exclusion chromatography with a PD-10 column (Amersham Pharmacia, Uppsala, Sweden)

using 20 mM NH_4HCO_3, 5 mM EDTA, pH 8 buffer. The thiol groups of the proteins were reacted by incubation with 10 μM TBMB in reaction buffer (20 mM NH_4HCO_3, 5 mM EDTA, pH 8, 20% ACN) at 30°C for 1 hr. For removal of non-reacted TBMB and concentration the protein was filtered with a Microcon YM-30 (Millipore, Bedford, MA). The concentrations of the products were determined by measuring the optical absorption at 280 nm. The IC_{50} was measured by incubating various concentrations of the modified peptide fusion proteins (2-fold dilutions) with human plasma kallikrein (0.1 nM) or cathepsin G (20 nM) and determining the residual activity in 10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1mM $CaCl_2$, 0.1% BSA, 0.01% Triton-X100. Human plasma kallikrein activity was measured with the fluorogenic substrate Z-Phe-Arg-AMC (Bachem, Bubendorf, Switzerland) at a concentration of 100 μM on a Spectramax Gemini fluorescence plate reader (excitation at 355 nm, emission recording at 460 nm; Molecular Devices, Sunnyvale, CA, USA). Human cathepsin G activity was measured with the colorimetric substrate N-Suc-Ala-Ala-Phe-Pro-pNA (Bachem, Bubendorf, Switzerland) at a concentration of 1 mM with a Spectramax absorption plate reader (recording at 410 nm; Molecular Devices, Sunnyvale, CA, USA).

Chemical synthesis of bicyclic peptides

Peptides with a free amine at the N-terminus and an amide at the C-terminus were chemically synthesized on a 25 mg scale by solid phase chemistry (JPT Peptide Technologies, Berlin, Germany). The crude peptides in 1 ml 70% NH_4HCO_3, pH 8 and 30% ACN (1 mM) were reacted with TBMB (1.2 mM) for 1 hr at RT. The reaction product was purified by reversed-phase high-performance liquid chromatography (HPLC) using a C18 column and gradient elution with a mobile

phase composed of ACN and 0.1% aqueous trifluoroacetic acid (TFA) solution at a flow rate of 2 ml/min. The purified peptides were freeze-dried and dissolved in DMSO or a buffer of 50 mM Tris-Cl pH 7.8, 150 mM NaCl for activity measurements.

Acknowledgments

We thank Peter Jones, Linda Judd, Agnes Jaulent, and Farida Begum and Sew-Yeu Peak-Chew respectively for expert laboratory advice, media preparation, peptide purification, and mass spectrometric analysis. We also thank Insa Kather and Franz Xaver Schmid from the University of Bayreuth for the engineered phage with disulfide-free gene-3-protein. C. H. was supported by the Swiss National Science Foundation (SNSF) and the Novartis Foundation (formerly Ciba-Geigy Jubilee Foundation).

Competing Interests Statement

The authors declare that they have no competing financial interests.

References

1. Bleicher, K.H., Bohm, H.J., Muller, K. & Alanine, A.I. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2**, 369-378 (2003).
2. Hüser, J. High-Throughput Screening in Drug Discovery, Vol. 35. (Wiley-VCH, Weinheim; 2006).
3. Huse, W.D. et al. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* **246**, 1275-1281 (1989).
4. Ward, E.S., Gussow, D., Griffiths, A.D., Jones, P.T. & Winter, G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. *Nature* **341**, 544-546 (1989).
5. McCafferty, J., Griffiths, A.D., Winter, G. & Chiswell, D.J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554 (1990).
6. Scott, J.K. & Smith, G.P. Searching for peptide ligands with an epitope library. *Science* **249**, 386-390 (1990).
7. Marks, J.D., Hoogenboom, H.R., Griffiths, A.D. & Winter, G. Molecular evolution of proteins on filamentous phage. Mimicking the strategy of the immune system. *J Biol Chem* **267**, 16007-16010 (1992).
8. Lipovsek, D. & Pluckthun, A. In-vitro protein evolution by ribosome display and mRNA display. *J Immunol Methods* **290**, 51-67 (2004).
9. Ulrich, H. DNA and RNA aptamers as modulators of protein function. *Med Chem* **1**, 199-208 (2005).
10. Barbas, C.F., 3rd Synthetic human antibodies. *Nat Med* **1**, 837-839 (1995).

11. Holliger, P. & Hudson, P.J. Engineered antibody fragments and the rise of single domains. *Nat Biotechnol* **23**, 1126-1136 (2005).

12. Lerner, R.A. Manufacturing immunity to disease in a test tube: the magic bullet realized. *Angew Chem Int Ed Engl* **45**, 8106-8125 (2006).

13. Nygren, P.A. & Skerra, A. Binding proteins from alternative scaffolds. *J Immunol Methods* **290**, 3-28 (2004).

14. Doyon, J.B., Snyder, T.M. & Liu, D.R. Highly sensitive in vitro selections for DNA-linked synthetic small molecules with protein binding affinity and specificity. *J Am Chem Soc* **125**, 12372-12373 (2003).

15. Melkko, S., Scheuermann, J., Dumelin, C.E. & Neri, D. Encoded self-assembling chemical libraries. *Nat Biotechnol* **22**, 568-574 (2004).

16. Mannocci, L. et al. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc Natl Acad Sci U S A* **105**, 17670-17675 (2008).

17. Woiwode, T.F. et al. Synthetic compound libraries displayed on the surface of encoded bacteriophage. *Chem Biol* **10**, 847-858 (2003).

18. Brenner, S. & Lerner, R.A. Encoded combinatorial chemistry. *Proc Natl Acad Sci U S A* **89**, 5381-5383 (1992).

19. Gartner, Z.J. et al. DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601-1605 (2004).

20. Halpin, D.R. & Harbury, P.B. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biol* **2**, E174 (2004).

21. Tse, B.N., Snyder, T.M., Shen, Y. & Liu, D.R. Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for in vitro selection. *J Am Chem Soc* **130**, 15611-15626 (2008).

22. Kemp, D.S. & McNamara, P.E. Conformationally restricted cyclic nonapeptides derived from L-cysteine and LL-3-amino-2-piperidino-6-carboxylic acid (LL-acp), a potent b-turn-inducing dipeptide analogue. *Journal of Organic Chemistry* **50**, 5834-5838 (1985).

23. Timmerman, P., Beld, J., Meloen, R.H. & Puijk, W.C. (WO 2004077062.

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

25. Timmerman, P., Puijk, W.C., Slootstra, J.W., Van Dijk, E. & Meloen, R.H. (WO 2006078161.

26. Jespers, L.S.A., Winter, G.P., Bonnert, T.P. & Simon, T.M. (WO9501438.

27. Jespers, L., Bonnert, T.P. & Winter, G. Selection of optical biosensors from chemisynthetic antibody libraries. *Protein Eng Des Sel* **17**, 709-713 (2004).

28. Wessjohann, L.A., Ruijter, E., Garcia-Rivera, D. & Brandt, W. What can a chemist learn from nature's macrocycles?--a brief, conceptual view. *Mol Divers* **9**, 171-186 (2005).

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

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

31. Cheng, Y. & Prusoff, W.H. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol* **22**, 3099-3108 (1973).

32. Abbenante, G. & Fairlie, D.P. Protease inhibitors in the clinic. *Med Chem* **1**, 71-104 (2005).

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

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

35. Litovchick, A. & Szostak, J.W. Selection of cyclic peptide aptamers to HCV IRES RNA using mRNA display. *Proc Natl Acad Sci U S A* **105**, 15293-15298 (2008).

36. Cremlyn, R.J. An introduction to organosulfur chemistry, Edn. 1. (Wiley, 1996).

37. Wood, S.P. et al. Crystal structure analysis of deamino-oxytocin: conformational flexibility and receptor binding. *Science* **232**, 633-636 (1986).

38. Bhaskaran, R., Chuang, L.C. & Yu, C. Conformational properties of oxytocin in dimethyl sulfoxide solution: NMR and restrained molecular dynamics studies. *Biopolymers* **32**, 1599-1608 (1992).

39. Pohl, E. et al. Structure of octreotide, a somatostatin analogue. *Acta Crystallogr D Biol Crystallogr* **51**, 48-59 (1995).

40. Melacini, G., Zhu, Q. & Goodman, M. Multiconformational NMR analysis of sandostatin (octreotide): equilibrium between beta-sheet and partially helical structures. *Biochemistry* **36**, 1233-1241 (1997).

41. Jackson, D.Y. et al. A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **266**, 243-247 (1994).

42. Sandman, K.E. & Noren, C.J. The efficiency of Escherichia coli selenocysteine insertion is influenced by the immediate downstream nucleotide. *Nucleic Acids Res* **28**, 755-761 (2000).

43. Sandman, K.E., Benner, J.S. & Noren, C.J. Phage display of selenopeptides. *J Am Chem Soc* **122**, 960-961 (2003).

44. Dennis, M.S. et al. Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* **404**, 465-470 (2000).

45. Huang, L. et al. Novel peptide inhibitors of angiotensin-converting enzyme 2. *J Biol Chem* **278**, 15532-15540 (2003).

Figure Legends

Figure 1

(**a**) Generation of phage encoded combinatorial chemical libraries. A phage encoded peptide with three cysteine residues is tethered to the tri-functional compound tris-(bromomethyl)benzene in a nucleophilic substitution reaction (see ref. 18-20 for the chemical reaction). The resulting chemical entities could optionally be further modified through enzymatic reactions such a proteolysis. (**b**) Chemical structure of a macrocyclic plasma kallikrein inhibitor isolated by phage display (PK15).

Figure 2

Conjugation of peptide fusions with tris-(bromomethyl)benzene (TBMB). (**a**) Molecular spectrometry of GCGSGCGSGCG-D1-D2 (disulphide-free) fusion protein before and after reaction with 10 μM TBMB as in Methods. The mass difference corresponds to the mesitylene core. (**b**) Titres (transducing units) of phage treated with various concentrations of TBMB as in Methods.

Figure 3

Sequences of selected conjugates. (**a**) Designed sequence of peptides attached to fusion proteins as expressed by phage Library 1. Processing of the leader sequence upon secretion of the protein is expected to give a peptide with an N-terminal alanine, two random 6-amino acid sequences flanked by three cysteines and a Gly-Gly-Ser-Gly linker that connects the peptide to the gene-3-protein**.** (**b** and **c**) Amino acid sequences of conjugates selected with human plasma kallikrein (**b**) and cathepsin G (**c**) with corresponding inhibitory activity. Only those clones with sequence

similarities or that were isolated multiple times are displayed. Residues of similar character are highlighted in colour. ND = Not determined.

Figure 4

Affinity maturation of human plasma kallikrein inhibitors. (**a**) Design of library 2, 3 and 4. In each library, one of the peptide loops has the sequence of a consensus motif identified in the first selections and the other contains six random amino acids. (**b**) Amino acid sequences of clones selected with human plasma kallikrein. All clones derive from library 2. The inhibitory activities of TBMB modified peptide-D1-D2 fusion proteins are indicated. Only clones with sequence similarities or that were isolated multiple times are displayed. The colours highlight sequence similarities in the second binding loop. ND = Not determined.

Figure 5

Inhibition of human plasma kallikrein by conjugates and NMR solution structure of conjugate PK15. (**a**) The inhibitory activity is expressed as the fractional activity (inhibited rate/uninhibited rate) at varying inhibitor concentrations (one measurement per inhibitor concentration). Clones PK2, PK4, PK6 and PK13 were isolated in phage selections using library 1. PK15 derives from library 2 and is an affinity matured inhibitor. (**b**) The peptide loops of conjugate PK15 are shown in yellow (loop 1) and orange (loop 2). The mesitylene core, the three cysteine residues and the terminal alanine (N-terminus) and glycine (C-terminus) are shown in grey. The backbone atoms of the peptide are represented as a sausage and the side chains of the amino acids as sticks.

Tables

Table 1

Parental clones	Amino acid sequence	Mass (Da)		IC_{50} (nM)	
		Linear peptide	Bi-cyclic conjugate	Linear peptide	Bi-cyclic conjugate
PK2	H-ACSDRFRNCPLWSGTCG-NH_2	1871.2	1985.3	> 10’000	28.6
PK4	H-ACSTERRYCPIEIFPCG-NH_2	1942.9	2055.9	7181	33
PK6	H-ACAPWRTACYEDLMWCG-NH_2	1974.8	2088.7	5707	21.2
PK13	H-ACGTGEGRCRVNWTPCG-NH_2	1764.8	1879.1	> 10’000	39.1
PK15	H-ACSDRFRNCPADEALCG-NH_2	1825	1939.4	> 10’000	1.7

Table 1 Chemically synthesized peptide inhibitors. The amino acid sequences of five plasma kallikrein inhibitors (17-mers) are shown. The sequences of the synthetic peptides derive from the clones PK2, PK4, PK6, PK13 (isolated in phage selections using library 1) and from clone PK15 (an affinity matured clone isolated from library 2). Indicated are the molecular masses and the inhibitory activities before and after the modification of the peptides with TBMB. The reduced linear peptides were incubated with plasma kallikrein and the inhibitory activity was measured immediately to minimize the risk of peptide oxidation.

Figure 1

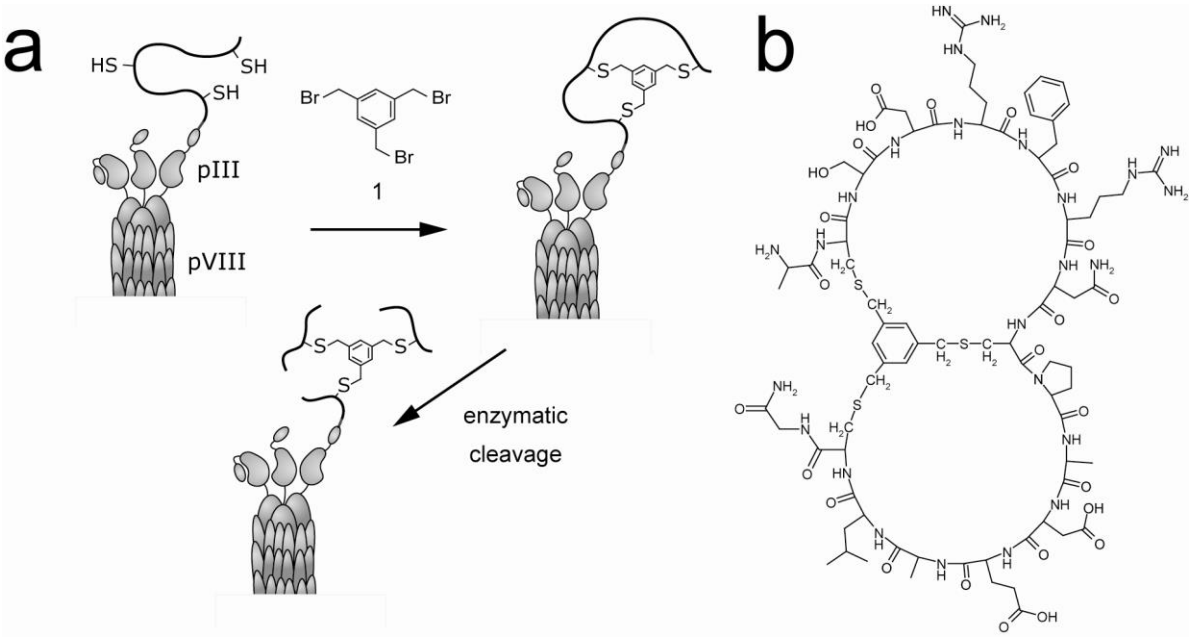


Figure 2

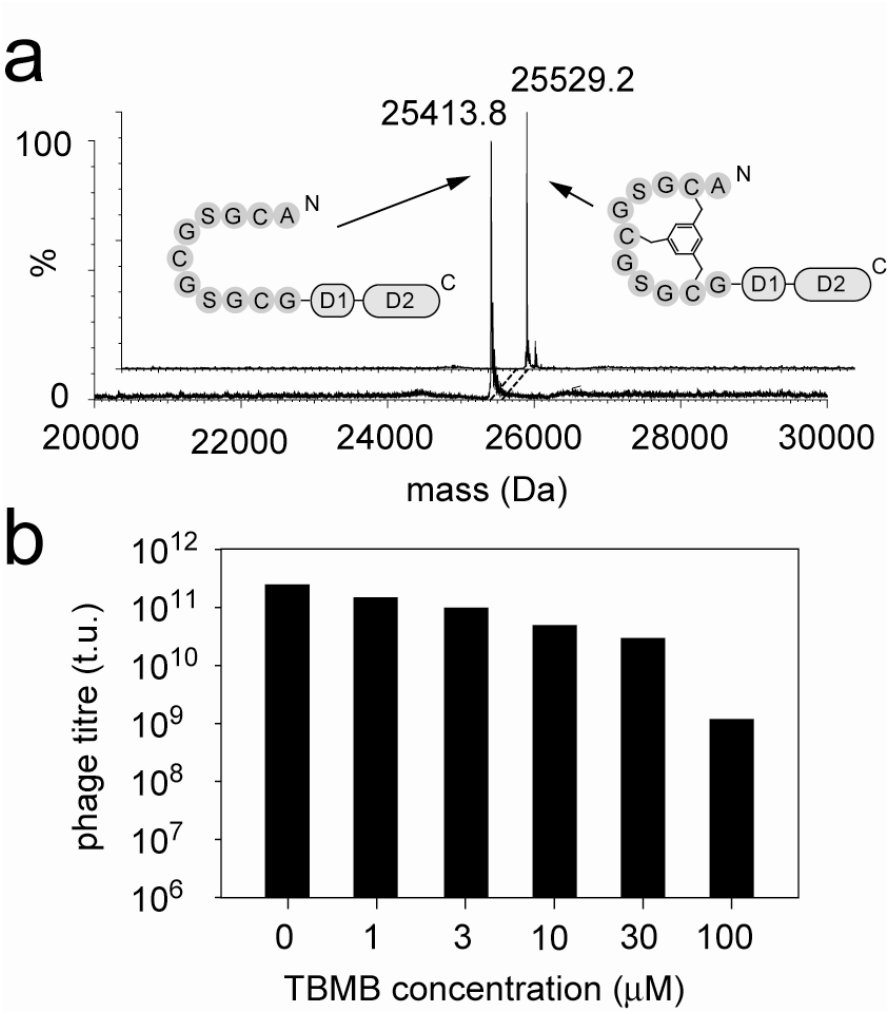

a
25529.2
25413.8
100
%
0
20000
22000
24000
26000
28000
30000
mass (Da)
S G C A N
G
C
G
S G C G
D1
D2
C
b
phage titre (t.u.)
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
0
1
3
10
30
100
TBMB concentration (μM)

Figure 3

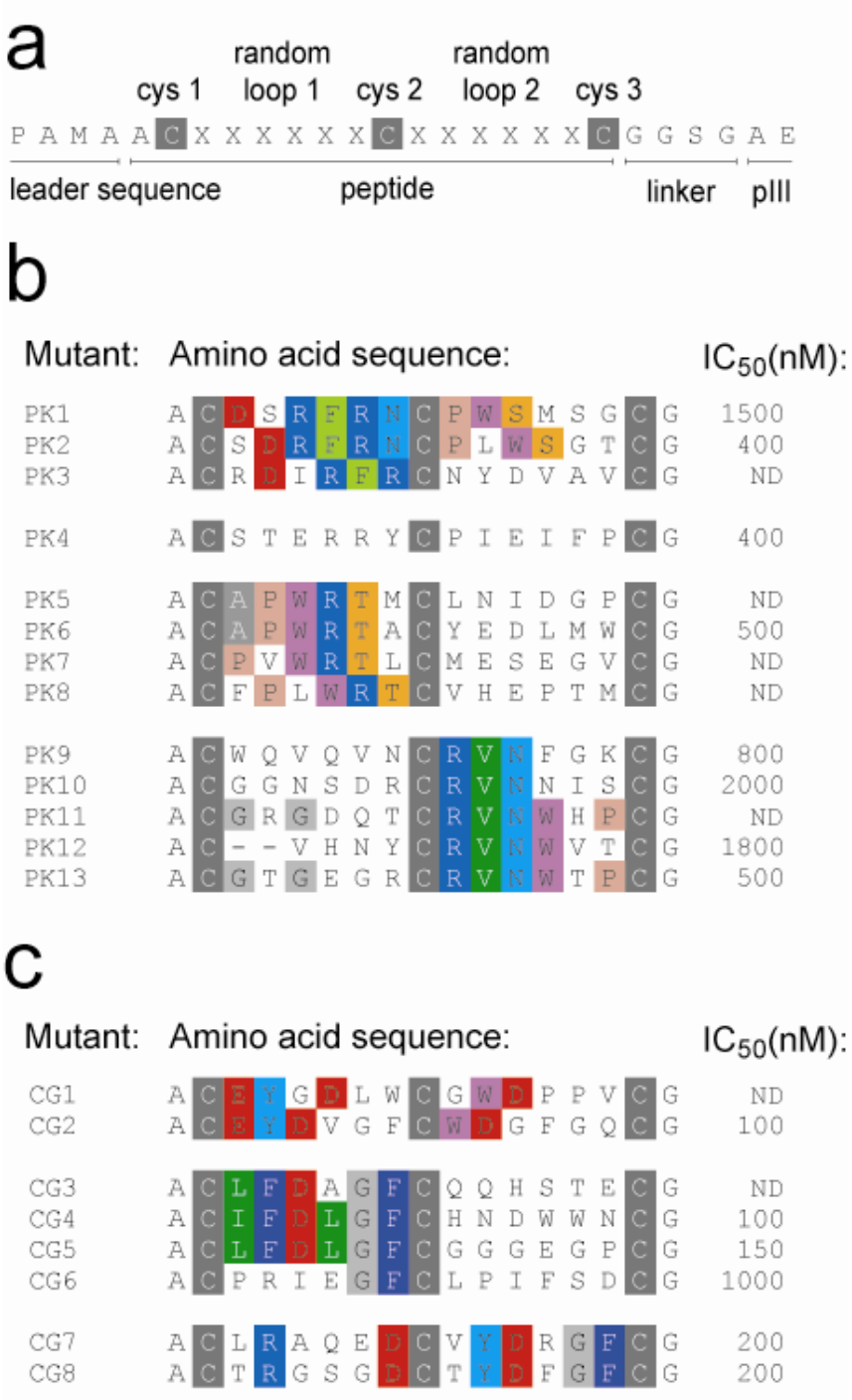

a
random
random
cys 1
loop 1
cys 2
loop 2
cys 3
P A M A A C X X X X X X C X X X X X X C G G S G A E
leader sequence
peptide
linker
pIII
b
Mutant:
Amino acid sequence:
IC50(nM):
PK1 A C D S R F R N C P W S M S G C G 1500
PK2 A C S D R F R N C P L W S G T C G 400
PK3 A C R D I R F R C N Y D V A V C G ND
PK4 A C S T E R R Y C P I E I F P C G 400
PK5 A C A P W R T M C L N I D G P C G ND
PK6 A C A P W R T A C Y E D L M W C G 500
PK7 A C P V W R T L C M E S E G V C G ND
PK8 A C F P L W R T C V H E P T M C G ND
PK9 A C W Q V Q V N C R V N F G K C G 800
PK10 A C G G N S D R C R V N N I S C G 2000
PK11 A C G R G D Q T C R V N W H P C G ND
PK12 A C - - V H N Y C R V N W V T C G 1800
PK13 A C G T G E G R C R V N W T P C G 500
c
Mutant:
Amino acid sequence:
IC50(nM):
CG1 A C E Y G D L W C G W D P P V C G ND
CG2 A C E Y D V G F C W D G F G Q C G 100
CG3 A C L F D A G F C Q Q H S T E C G ND
CG4 A C I F D L G F C H N D W W N C G 100
CG5 A C L F D L G F C G G G E G P C G 150
CG6 A C P R I E G F C L P I F S D C G 1000
CG7 A C L R A Q E D C V Y D R G F C G 200
CG8 A C T R G S G D C T Y D F G F C G 200

Figure 4

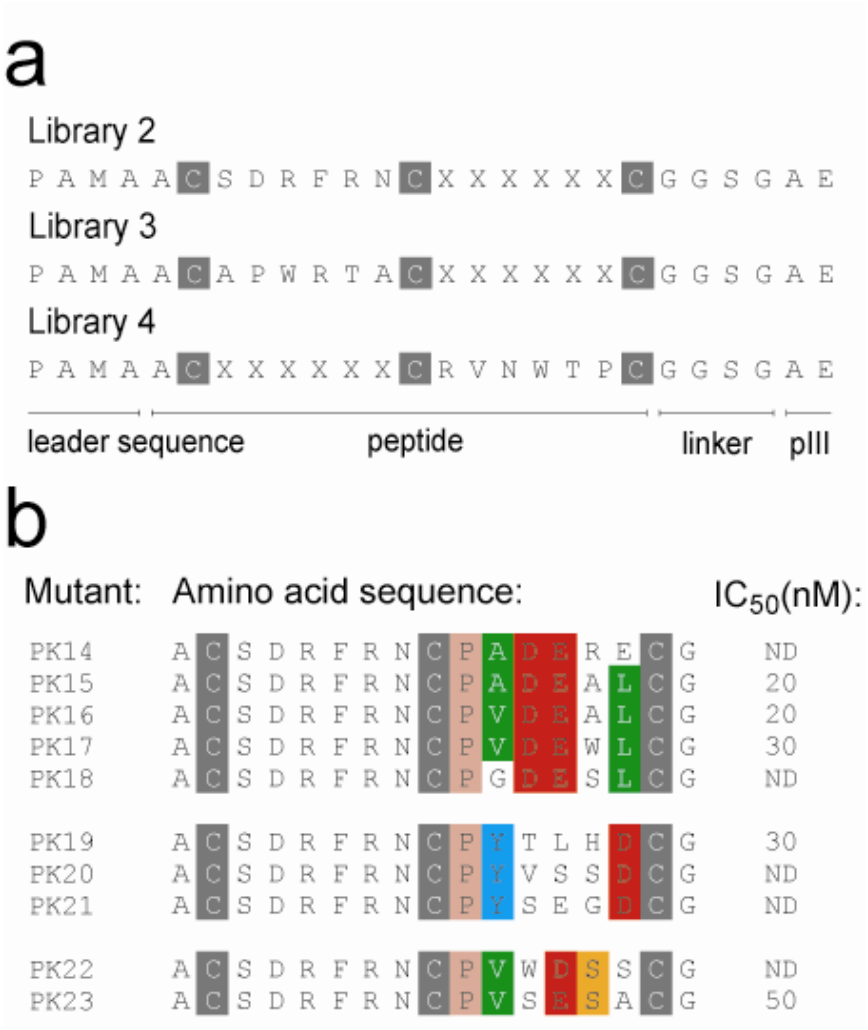

a
Library 2
P A M A A C S D R F R N C X X X X X X C G G S G A E
Library 3
P A M A A C A P W R T A C X X X X X X C G G S G A E
Library 4
P A M A A C X X X X X X C R V N W T P C G G S G A E
leader sequence
peptide
linker
pIII
b
Mutant:
Amino acid sequence:
IC50(nM):
PK14 A C S D R F R N C P A D E R E C G ND
PK15 A C S D R F R N C P A D E A L C G 20
PK16 A C S D R F R N C P V D E A L C G 20
PK17 A C S D R F R N C P V D E W L C G 30
PK18 A C S D R F R N C P G D E S L C G ND
PK19 A C S D R F R N C P Y T L H D C G 30
PK20 A C S D R F R N C P Y V S S D C G ND
PK21 A C S D R F R N C P Y S E G D C G ND
PK22 A C S D R F R N C P V W D S S C G ND
PK23 A C S D R F R N C P V S E S A C G 50

Figure 5

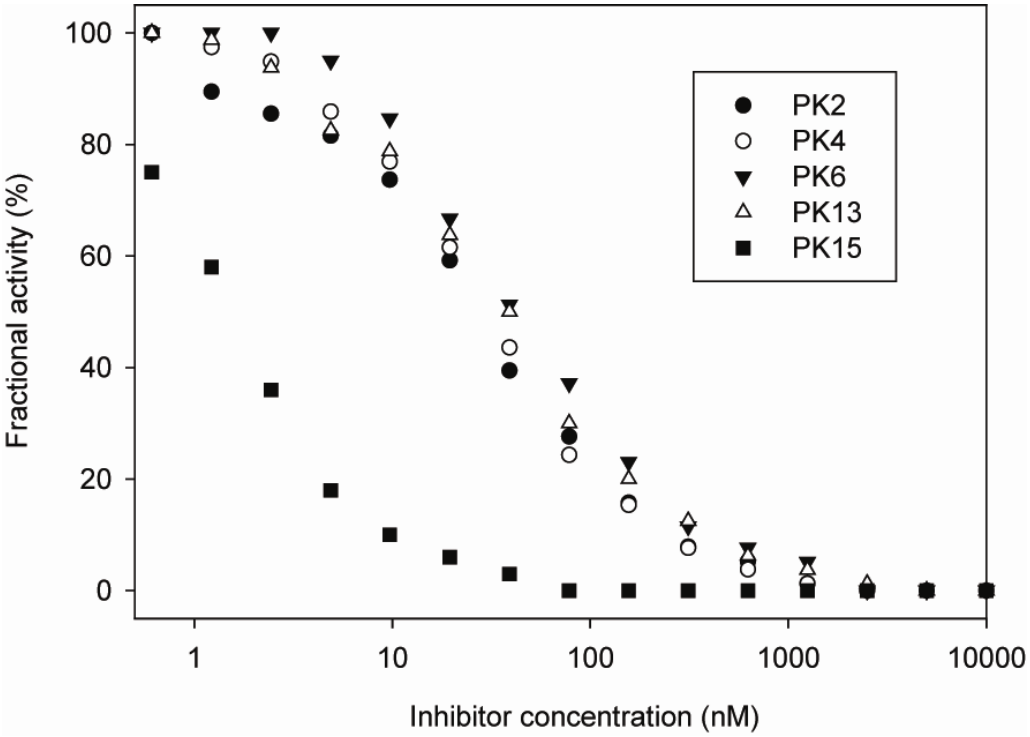
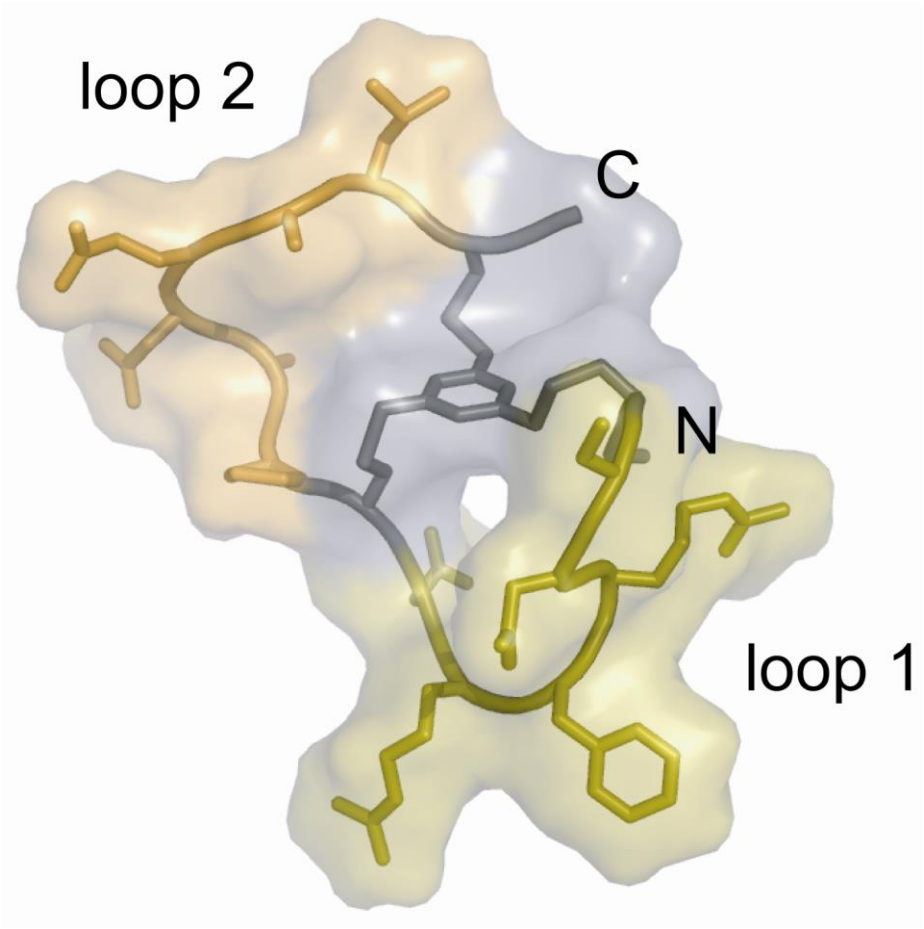
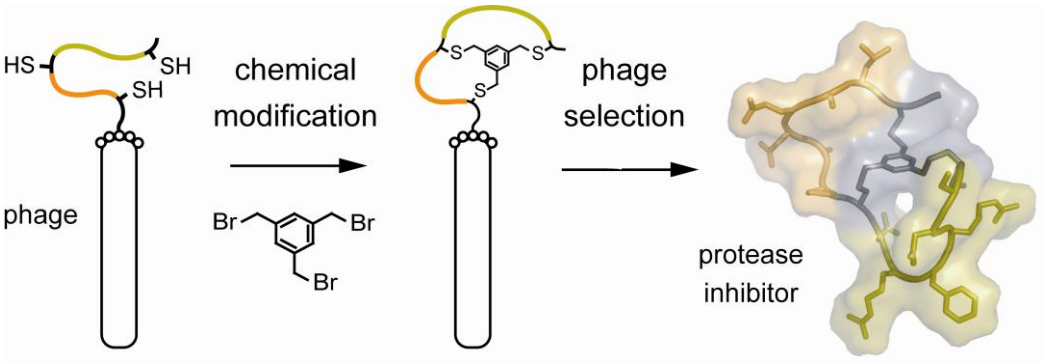

100
80
60
40
20
0
Fractional activity (%)
1
10
100
1000
10000
Inhibitor concentration (nM)
PK2
PK4
PK6
PK13
PK15

Figure 6



Graphical Abstract



Phage-encoded combinatorial chemical libraries based on bicyclic peptides

Christian Heinis, Trevor Rutherford, Stephan Freund and Greg Winter

Supplementary Methods

Cloning and expression of peptide-D12 fusion proteins

The domains D1-D2 of the g3p (comprising amino acid residues 2 to 217 of the mature fd-g3p) with and without the N-terminally fused peptide ^{N}ACGSGCGSGCG^{C} were expressed in *E.coli*. The pUC119 based expression vector with a leader sequence and the D1-D2 gene with a C-terminal hexa-histidine tag (here termed pUC119H6D12) was kindly provided by Phil Holliger from the Laboratory of Molecular Biology (LMB) in Cambridge. A plasmid for expression of D1-D2 with the N-terminal peptide was cloned by PCR amplification of the D1-D2 gene with the primers pepd12ba (encoding the peptide sequence) and d12fo and ligation into the *SfiI*/*NotI* digested pUC119H6D12 vector. The gene for the expression of disulfide-free D1-D2 with a total of 20 amino acids was PCR amplified from the vector fdg3p0ss21 with either the primer pair d120ssba/d120ssfo, pepd120ssba/d120ssfo, P2cd120ssba/d120ssfo or P1cd120ssba/d120ssfo and *SfiI*/*NotI* ligated into pUC119H6D12 for expression of disulfide-free D1-D2 with and without the N-terminal fused peptides ^{N}ACGSGCGSGCG^{C}, ^{N}AGSGCGSGCG^{C} or ^{N}AGSGKGSGCG^{C}. All 6 proteins were expressed in TG1 *E.coli* cells at 30°C for 8 hours and the periplasmic fraction was purified stepwise by Ni-affinity chromatography and gel filtration on a Superdex 75 column in 20 mM NH_4HCO_3 pH 7.4.

Mass spectrometric analysis of peptide-D12 fusion proteins

The molecular masses of the proteins (5-20 μM) before and after modification with TBMB were determined by denaturing the proteins in 4 volumes of 50% MeOH, 1% formic acid and analysis on a time of flight mass spectrometer with electrospray ionization (Micromass, Milford, MA, USA). Molecular masses were obtained by deconvoluting multiply charged protein mass spectra using MassLynx version 4.1.

Creation of the phage peptide library 1

The genes encoding a semi-random peptide with the sequence Ala-Cys-$(Xaa)_6$-Cys-$(Xaa)_6$-Cys, the linker Gly-Gly-Ser-Gly and the two disulfide-free domains D1 and D2 were cloned in the correct orientation into the phage vector fd0D12 to obtain Library 1. The vector fd0D12, lacking the genes of the D1 and D2 domains of gene 3 and having a second *SfiI* restriction site was previously created by whole-plasmid PCR amplification of fdg3p0ss21 using the primer ecoG3pNba and pelbsfiecofo. The genes encoding the peptide repertoire and the two gene 3 domains were step-wise created in two consecutive PCR reactions. First, the genes of D1 and D2 were PCR amplified with the two primer prepcr and sfi2fo using the vector fdg3p0ss21 as a template. Second, the DNA encoding the random peptides was appended in a PCR reaction using the primer sficx6ba and sfi2fo. The ligation of 33 and 9 μg of *SfiI* digested fd0D12 plasmid and PCR product yielded $4.4x10^9$ colonies on 12 20x20 cm chloramphenicol (30 μg/ml) 2YT plates. Colonies were scraped off the plates with 2YT media, supplemented with 15% glycerol and stored at -80°C. Glycerol stocks were diluted to OD_{600}=0.1 in 1 litre 2YT/chloramphenicol (30 μg/ml) cultures and phage were expressed at 30°C over night (12 to 16 hrs).

Biotinylation of antigens

Human plasma kallikrein (activated with factor XIIa) was purchased from Innovative Research (Southfiled, MI, USA) and biotinylated at a concentration of 1.2 μM with a 5-fold molar excess of Sulfo-NHS-LC-biotin (Pierce, Rockford, IL, USA) in PBS, pH 7.4 / 5% DMSO at RT for 1 hr. The biotinylated protein was purified on a PD-10 column using a buffer of 50 mM NaAc, pH 5.5, 200 mM NaCl. Readily biotinylated human cathepsin G was purchased from Lee Biosolutions (St. Louis, MI, USA).

Subcloning and expression screening of phage selected clones

The plasmid DNA of clones selected after the second and third round of biopanning was PCR-amplified in a single tube with the primer 21seqba and flagfo and cloned into the vector pUC119H6D12 at the *SfiI* and *NotI* sites for the periplasmic expression of the peptides fused to the disulfide-free D1 and D2 domains with a C-terminal FLAG tag and a hexa-histidine-tag. The ligated plasmids were electroporated into TG1 cells and plated on 2YT/ampicillin (100 μg/ml) plates. Clones expressing the recombinant protein were identified as follows: Media (2YT with 100 μg/ml ampicillin) in 96-deep well plates (1 ml/well) was inoculated with cells of individual colonies and incubated at 37°C. Protein expression was induced with 1 mM IPTG when the cultures were turbid and the plates were shaken 300 rpm at 30°C o/n. The cells were pelleted by centrifugation at 3500 rpm for 30 minutes, lysed with washing buffer containing 1 mg/ml lysozyme and spun at 3500 rpm to pellet the cell debris. The supernatants were transferred to 96-well Polysorp plates (Nunc, Roskilde, Denmark) for non-specific adsorbtion. The wells were rinsed twice with washing buffer containing 0.1% Tween 20 and blocked with washing buffer containing 1% BSA and 0.1 % Tween 20 for 1 hr. Anti-FLAG M2-peroxidase conjugate (Sigma-

Aldrich, St. Louis, MO, USA) was 1:5000 diluted and blocked in washing buffer containing 1% BSA and 0.1 % Tween 20 and added to the plates for 1 hr. The wells were washed (5 times with washing buffer containing 0.1% Tween 20 and once without detergent) and the bound peroxidase was detected with TMB substrate solution (eBiosciences, San Diego, USA). The plasmid DNA of protein expressing clones was sequenced (60 clones selected with plasma kallikrein, 36 clones selected with cathepsin G, 30 clones selected in the affinity maturation with plasma kallikrein; Geneservice, Cambridge, UK).

Affinity maturation of human plasma kallikrein inhibitors

Three peptide phage libraries were created essentially as the library 1 (see above) but using the degenerate primer sficx6abc (Library 2), sficx6abb (Library 3) and sficx6aba (Library 4) instead of sficx6ba. Electroporation of the ligation reactions into TG1 cells yielded $9.5x10^8$ (library 2), $1.1x10^9$ (library 3) and $1.2x10^9$ (library 4) transformants. Phage of each library were produced in 1 L cultures, purified, pooled and reacted with TBMB. Three rounds of panning were performed essentially as in the selections described in the materials and methods section but using the biotinylated human plasma kallikrein at a lower concentration (1 nM in the 1st and 2nd round, 200 pM in the 3rd round).

Activity and specificity measurement of human plasma kallikrein inhibitors

Inhibitory activities (IC_{50}) were determined by measuring residual activities of the enzyme upon incubation (30 min, RT) with different concentrations of inhibitor (typically ranging from 10 μM to 0.5 nM). The activities of human plasma kallikrein (0.1 nM) and factor XIa (0.8 nM; Innovative Research, Southfiled, MI, USA) were

measured with Z-Phe-Arg-AMC (100 μM) and the activity of human thrombin (2 nM; Innovative Research, Southfiled, MI, USA) with Boc-Phe-Ser-Arg-AMC (100 μM) in 10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1mM $CaCl_2$, 0.1% BSA, 0.01 % triton X-100 and 5% DMSO. Recombinant mouse plasma kallikrein from R&D Systems (Minneapolis, MN, USA) with a signal peptide was proteolytically activated with 0.5 mg/ml thermolysin at 37°C for 1 hr. The activity of mouse plasma kallikrein (3 nM) was measured with Z-Phe-Arg-AMC (100 μM) in 50 mM Tris-Cl pH 7.5, 10 mM $CaCl_2$ and 250 mM NaCl, 0.1% BSA, 0.01% triton X-100 and 5% DMSO. Inhibitor hydrolysed in one binding loop was generated by incubation of TBMB modified peptide PK15 with human plasma kallikrein at a molar ratio of 5:1 for 24 hours at 37°C and subsequent heat inactivation of the enzyme at 60°C (30 min). Apparent K_i values were calculated according to the Cheng and Prusoff equation.

Identification of the human plasma kallikrein cleavage site in the PK15 conjugate

2 μg (1 nmol) of peptide PK15 conjugate dissolved in 1 μL DMSO was added to 6.8 μg (80 pmol) human plasma kallikrein in 20 μL reaction buffer (10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1mM $CaCl_2$) and incubated for 18 hours at 37°C. 1 μL of different 5-fold dilutions (22 μg/μL, 4.4 μg/μL, 880 ng/μL, 176 ng/μL, 35 ng/μL, 7 ng/μL, 1.4 ng/μL, 280 pg/μL; in reaction buffer) of leucyl aminopeptidase from *Aeromonas Proteolytica* (EC 3.4.11.10; Sigma, A8200) was added to 1 μL of the plasma kallikrein-treated PK15 conjugate and incubated for 1 hour at 37°C. 0.5 μL of the reactions was mixed with 0.5 μL of matrix solution (10 mg/mL α-Cyano-4-hydroxycinnamic acid in 1 mL of 50% acetonitrile/50% H_2O/0.1% TFA). 0.5 μL of a 10-fold dilution of the mixture (in matrix solution) was transferred to a MALDI

sample plate and the mass of the inhibitor or inhibitor fragments was measured on a Axima-CFR plus (Shimadsu) MALDI-TOF mass spectrometer.

Measurement of factor XII activation in human plasma

Normal human plasma from single donors was purchased from 3H Biomedical (Uppsala, Sweden). The plasma was centrifuged at 1500xg at 20°C for 15 minutes to obtain platelet-poor plasma (PPP). Aliquots of the PPP were stored in polypropylene tubes at -80°C. Samples of 60 µl PPP containing 5, 50, 500 or 5000 nM of aprotinin (Roche, Mannheim, Germany) or TBMB modified peptide PK15 were prepared. Activation of factor XIIa was measured as follows. 2 µg of kaolin was added to the plasma samples, mixed well and incubated for 20 minutes at 37°C. The samples were diluted 250-fold in 50 mM Tris-Cl pH 7.8, 150 mM NaCl. Plasma kallikrein-like activity was measured with the chromogenic substrate H-D-Pro-Phe-Arg-pNA (100 µM; Bachem, Bubendorf, Switzerland) using an absorption plate reader (absorption at 450 nm; Molecular Devices, Sunnyvale, CA, USA). The same chromogenic substrate is also recognized and modified by plasma kallikrein. However, at the inhibitor concentrations required to reduce the factor XIIa activity by 50% (160 nM for the TBMB modified peptide PK15 and 5 µM for aprotinin), the plasma kallikrein is essentially inhibited and can not be measured with the substrate.

Structure determination of TBMB modified peptide PK15

1 mg of PK15 conjugate was dissolved in 550 µl 10 mM deuterated Tris-Cl pH 6.6, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$ to obtain an inhibitor concentration of 1 mM. TOCSY and NOESY (**Supplementary Fig. 4**) spectra were acquired at 293 K on an 800 MHz Bruker Avance instrument with TCI cryoprobe. The NOE mixing

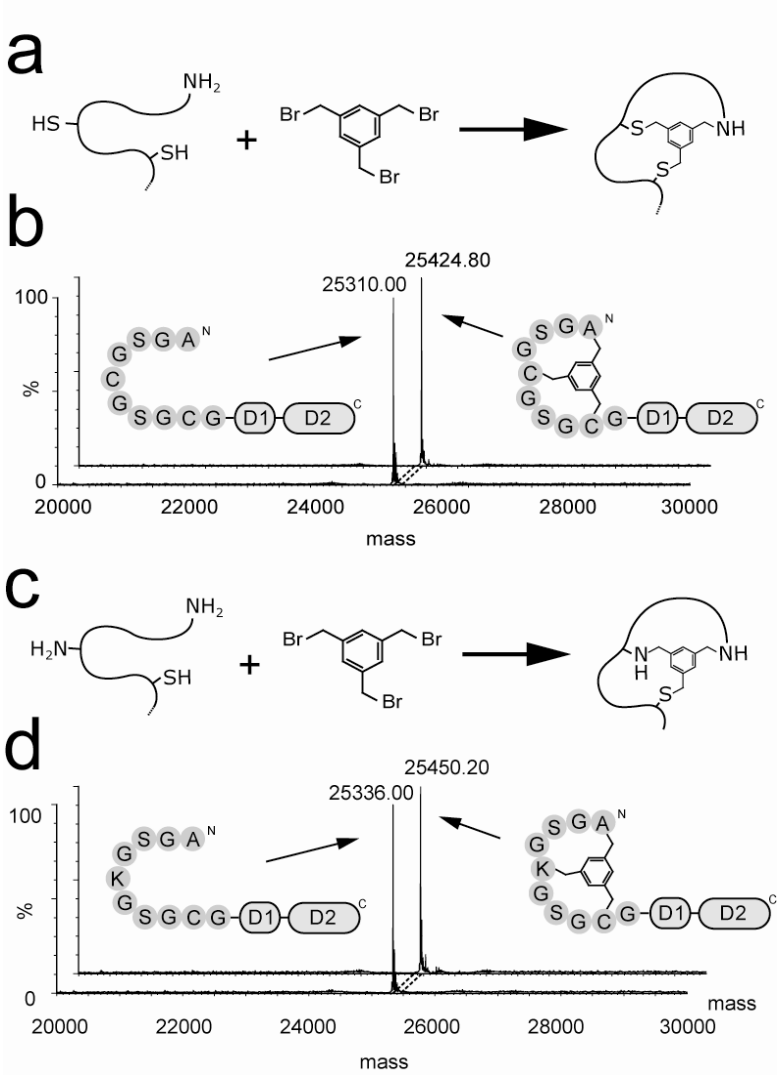
time was 250 ms. The Sparky software package (T. D. Goddard & G. D. Kneller, Sparky 3, University of California, San Francisco) was used for spectral assignments. There were a total of 154 NOE restraints, 71 of which were inter-residue, and 83 intra-residue. Within the inter-residue restraints, 47 were sequential restraints, 24 were medium range restraints and none was a long range restraint. The structure shown in Figure 6 is the average structure of 50 calculated structure conformers. The program PyMOL was used for structure analysis and visualization.

Sequences of DNA primer

The nuclease restriction sites are underlined

pepd12ba	5'-CTCGCGGCCCAGCCGGCCATGGCAGCGTGTGGCTCTGGCTGCGGTTCCGGCTGTGGTGCAGAAACTGTTGAAAGTTG-3'
d12fo	5'-GAGTCATTCTGCGGCCGCATTGACAGGAGGTTGAGGCAGGTC-3'
d120ssba	5'-CATGCCATGACTCGCGGCCCAGCCGGCCATGGCTGAAACTGTTGAAAGTAG-3'
d120ssfo	5'-GAGTCATTCTGCGGCCGCATTGACAGGAGGTTGAGGCAGGTA-3'
pepd120ssba	5'-CTCGCGGCCCAGCCGGCCATGGCAGCGTGTGGCTCTGGCTGCGGTTCCGGCTGTGGTGCTGAAACTGTTGAAAGTAG-3'
p2cd120ssba	5'-CTCGCGGCCCAGCCGGCCATGGCAGCGGGCTCTGGCTGCGGTTCCGGCTGTGGTGCTGAAACTGTTGAAAGTAG-3'
p1cd120ssba	5'-CTCGCGGCCCAGCCGGCCATGGCAGCGGGCTCTGGCAAAGGTTCCGGCTGTGGTGCTGAAACTGTTGAAAGTAG-3'
ecoG3pNba	5'-GCATGAATTCCAGTCAGTACGGCCTCGGGGGCCATGGCTTCTGGTACCCCGGTTAAC-3'
pelbsfiecofo	5'-GCATGAATTCCGATGACTGAGGCCGGCTGGGCCGCATAGAAAGGAACAACTAAAGGAAT -3'
preperba	5'-GGCGGTTCTGGCGCTGAAACTGTTGAAAGTAG-3'
sfi2fo	5'-GAAGCCATGGCCCCCGAGGCCCCGGACGGAGCATTGACAGG-3'
sficx6ba	5'-TATGCGGCCCAGCCGGCCATGGCAGCGTGTNNKNNKNNKNNKNNKNNKTGCNNKNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
flagfo	5'-CATTCTGCGGCCGCCTTATCGTCGTCATCCTTGTAATCTGCAGCATTGACAGGAGGTTGAGGCAGGTA-3'
sficx6aba	5'-TATGCGGCCCAGCCGGCCATGGCAGCGTGTNNKNNKNNKNNKNNKNNKTGCCGCGTTAATTGGACCCCGTGTGGCGGTTCTGGCGCTG-3'
sficx6abb	5'-TATGCGGCCCAGCCGGCCATGGCAGCGTGTGCGCCGTGGCGCACCGCGTGCNNKNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
sficx6abc	5'-TATGCGGCCCAGCCGGCCATGGCAGCGTGTAGCGATCGTTTTCGTAATTGCNNKNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'

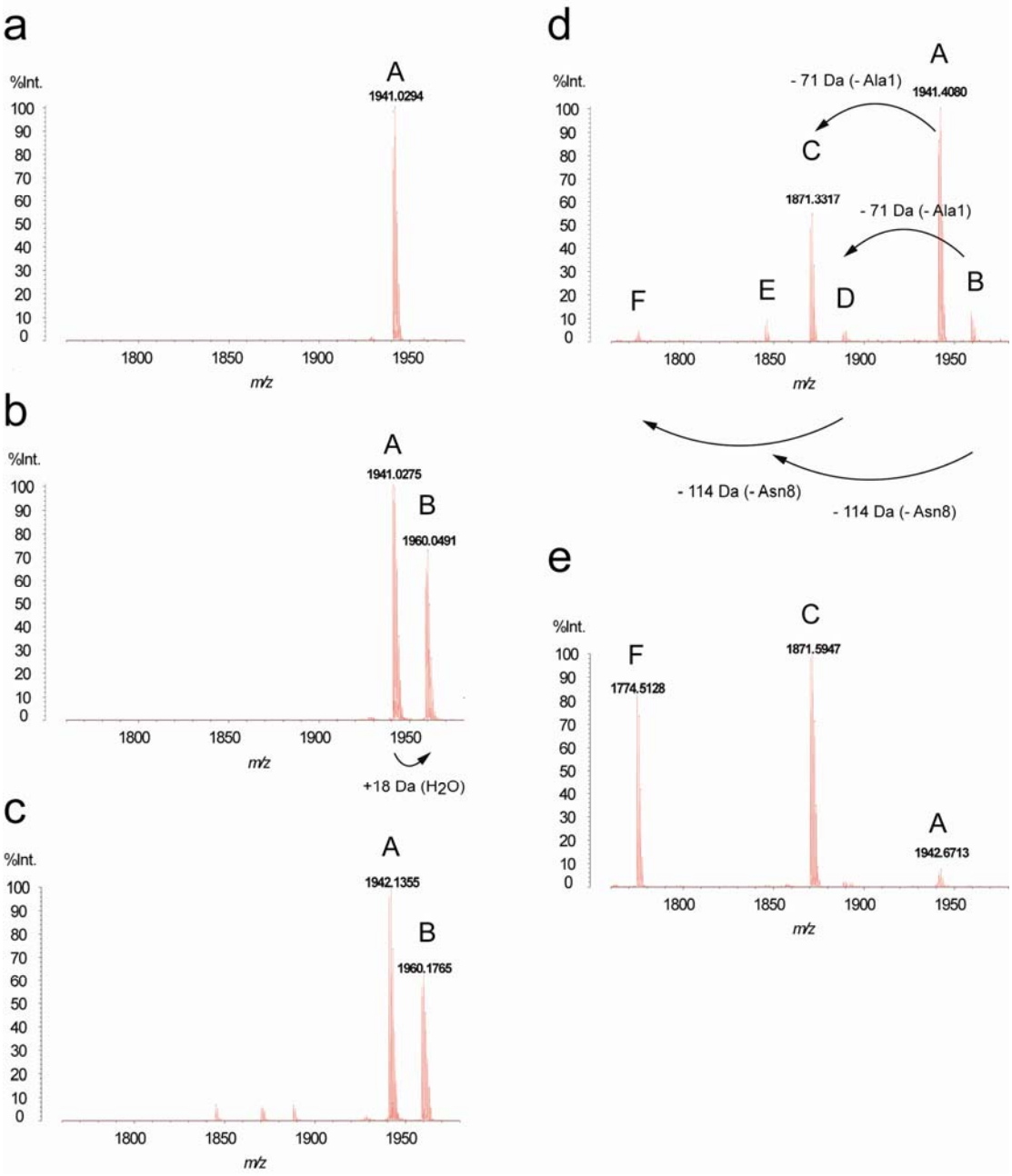
Supplementary Figure 1



Supplementary Figure 1 Chemical reaction of the tri-functional compound TBMB with peptides containing one or two cysteine residues. (**a**) Plausible reaction mechanism of TBMB with a peptide fusion protein containing two cysteine residues. (**b**) Mass spectra of a peptide fusion protein with two cysteines before and after reaction with TBMB. (**c**) Plausible reaction mechanism of TBMB with a peptide fusion protein containing one cysteine and one lysine residue. (**d**) Mass spectra of a

peptide fusion protein with one cysteine and one lysine residue before and after reaction with TBMB.

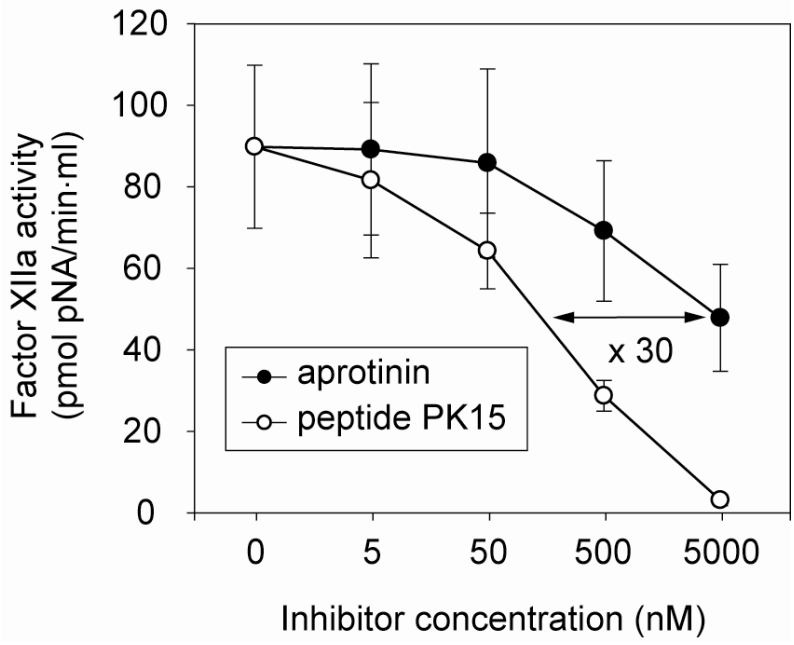
Supplementary Figure 2



Supplementary Figure 2 Mass spectra of PK15 treated with plasma kallikrein and leucyl aminopeptidase from *Aeromonas Proteolytic.* (**a**) PK15, (**b**) PK15 treated with plasma kallikrein, (**c**) PK15 treated with plasma kallikrein and 7 ng leucyl aminopeptidase, (**d**) PK15 treated with plasma kallikrein and 35 ng leucyl

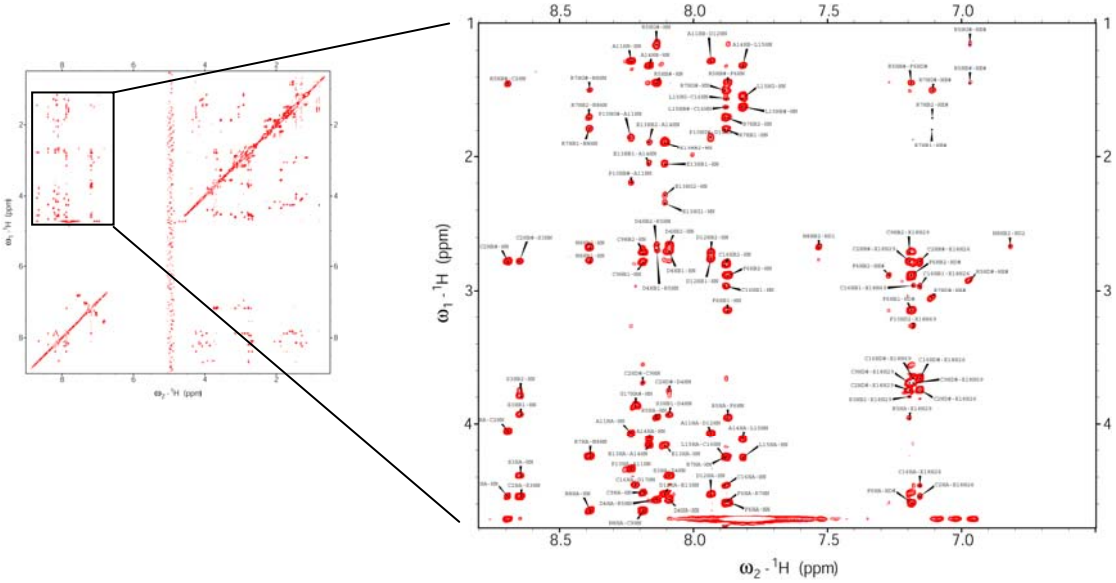
aminopeptidase, (**e**) PK15 treated with plasma kallikrein and 176 ng leucyl aminopeptidase. The mass shifts indicate that the N-terminal Ala1 of PK15 is cleaved off by the aminopeptidase,that PK15 is cleaved by plasma kallikrein between Arg7 and Asn8 and that Asn8 is subsequently cleaved off by the aminopeptidase.

Supplementary Figure 3



Supplementary Figure 3 Suppression of factor XII activation in human plasma through the inhibition of plasma kallikrein with aprotinin or TBMB modified peptide PK15. The intrinsic coagulation pathway in human plasma of three different donors was initiated by addition of kaolin. The negatively charged surface of kaolin activates small amounts of factor XII. Prekallikrein is converted to kallikrein by activated factor XII (XIIa), and kallikrein exerts a positive feedback to activate more factor XII. The activity of factor XIIa was measured with the colorimetric substrate H-D-Pro-Phe-Arg-pNA. Mean values and standard deviations of factor XIIa activity are indicated.

Supplementary Figure 4



Supplementary Figure 4 800 MHz NOESY spectrum of 1 mM PK15 conjugate in 10 mM deuterated Tris-Cl pH 6.6, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, pH 6.6, 293 K. The NOE mixing time was 250 ms. The spectrum was acquired with a standard Bruker pulse program, incorporating W5-WATERGATE solvent signal suppression.

Supplementary Table 1

	PK15
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	199
Intra-residue	122
Inter-residue	77
Sequential ($\|i - j\| = 1$)	52
Medium-range ($\|i - j\| < 4$)	25
Long-range ($\|i - j\| > 5$)	0
Intermolecular	0
Hydrogen bonds	0
Total dihedral angle restraints	0
ϕ	
ψ	
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.01 +/- 0.005
Dihedral angle constraints (º)	
Max. dihedral angle violation (º)	
Max. distance constraint violation (Å)	0.3
Deviations from idealized geometry	
Bond lengths (Å)	0.002 +/- 0.0007
Bond angles (º)	0.44 +/- 0.04
Impropers (º)	0.15 +/- 0.07
Average pairwise r.m.s. deviation* (Å)	
Heavy	(2-16) 2.11 +/- 0.62
Backbone	(2-16) 1.49 +/- 0.56

Supplementary Table 1 NMR and refinement statistics of PK15

* Pairwise r.m.s. deviation was calculated among 50 refined structures.

Supplementary Table 2

Target	Peptide sequence	Affinity	Reference
Prostate specific antigen (PSA)	CVAYCIEHHCWTC	K_D = 2.9 μM	[1]
Human kallikrein 2	SRFKVWWAAF	IC_{50} = 3.4 μM	[2]
Urokinase-type plasminogen activator (uPA)	CSWRGLENHRMC	K_i = 6.7 μM	[3]
Urokinase-type plasminogen activator (uPA)	CPAYSRYLDC	K_i = 0.4 μM	[4]
Chymotrypsin	CCFSWRCRC	K_i = 1 μM	[5]
TF-fVII	EEWEVLCWTWETCER	IC_{50} = 1,5 nM	[6]
Angiotensin converting enzyme 2 (ACE2)	GDYSHCSPLRYYPWWKCTYPDP	K_i = 2.8 nM	[7]
ErbB-2	KCCYSL	K_i = 30 μM	[8]
Urease	YDFYWW	IC_{50} = 30 μM	[9]
Pancreatic lipase	CQPHPGQTC	IC_{50} = 16 μM	[10]
Beta-lactamase	CVHSPNREC	IC_{50} = 9 μM	[11]
DNase II	CLRLLQWFLWAC	K_i = 0.2 μM	[12]

Supplementary Table 2 Phage selected peptide inhibitors. Indicated are the sequences of the peptides, the enzyme targets and the binding affinities. The cysteine residues that form disulfide bridges are underlined.

References

1. Wu, P., Leinonen, J., Koivunen, E., Lankinen, H. & Stenman, U.H. Identification of novel prostate-specific antigen-binding peptides modulating its enzyme activity. *Eur J Biochem* **267**, 6212-20 (2000).
2. Hekim, C. et al. Novel peptide inhibitors of human kallikrein 2. *J Biol Chem* **281**, 12555-60 (2006).
3. Hansen, M. et al. A urokinase-type plasminogen activator-inhibiting cyclic peptide with an unusual P2 residue and an extended protease binding surface demonstrates new modalities for enzyme inhibition. *J Biol Chem* **280**, 38424-37 (2005).
4. Andersen, L.M., Wind, T., Hansen, H.D. & Andreasen, P.A. A cyclic peptidylic inhibitor of murine urokinase-type plasminogen activator: changing species specificity by substitution of a single residue. *Biochem J* **412**, 447-57 (2008).
5. Krook, M., Lindbladh, C., Eriksen, J.A. & Mosbach, K. Selection of a cyclic nonapeptide inhibitor to alpha-chymotrypsin using a phage display peptide library. *Mol Divers* **3**, 149-59 (1997).
6. Dennis, M.S. et al. Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* **404**, 465-70 (2000).
7. Huang, L. et al. Novel peptide inhibitors of angiotensin-converting enzyme 2. *J Biol Chem* **278**, 15532-40 (2003).
8. Karasseva, N.G., Glinsky, V.V., Chen, N.X., Komatireddy, R. & Quinn, T.P. Identification and characterization of peptides that bind human ErbB-2

selected from a bacteriophage display library. *J Protein Chem* **21**, 287-96 (2002).

9. Houimel, M., Mach, J.P., Corthesy-Theulaz, I., Corthesy, B. & Fisch, I. New inhibitors of Helicobacter pylori urease holoenzyme selected from phage-displayed peptide libraries. *Eur J Biochem* **262**, 774-80 (1999).

10. Lunder, M., Bratkovic, T., Kreft, S. & Strukelj, B. Peptide inhibitor of pancreatic lipase selected by phage display using different elution strategies. *J Lipid Res* **46**, 1512-6 (2005).

11. Sanschagrin, F. & Levesque, R.C. A specific peptide inhibitor of the class B metallo-beta-lactamase L-1 from Stenotrophomonas maltophilia identified using phage display. *J Antimicrob Chemother* **55**, 252-5 (2005).

12. Sperinde, J.J., Choi, S.J. & Szoka, F.C., Jr. Phage display selection of a peptide DNase II inhibitor that enhances gene delivery. *J Gene Med* **3**, 101-8 (2001).