Analytical Tools for Dynamic Combinatorial Libraries of Cyclic Peptides

Taleen Peker, Benjamin Zagiel, Lou Rocard, Claudia Bich, Emmanuelle Sachon, Roba Moumné

RESEARCH ARTICLE

Analytical Tools for Dynamic Combinatorial Libraries of Cyclic Peptides

Taleen Peker‡ [a], Benjamin Zagiel‡[a], Lou Rocard[a], Claudia Bich[b], Emmanuelle Sachon[a][c], Roba Moumné [a]*

[a] Dr.Taleen Peker, Dr. Benjamin Zagiel, Dr Lou Rocard, Dr Emmanuelle Sachon, Dr. Roba Moumné
Sorbonne Université, École normale supérieure, PSL University, CNRS
Laboratoire des biomolécules, LBM, 75005 Paris, France.
[b] Dr. Claudia Bich
UMR 5247-CNRS-UM-ENSCM, Institut des Biomolécules Max Mousseron (IBMM), Université de Montpellier, Montpellier, France
[c] Dr. Emmanuelle Sachon
MS^{3}U platform, Fédération de chimie moléculaire de Paris centre – FR2769, Sorbonne Université, 4 place Jussieu, 75005 Paris, France.
‡ These authors contributed equally.
*E-mail: roba.moumne@sorbonne-universite.fr

Supporting information for this article is given via a link at the end of the document.

Abstract: Target-directed dynamic combinatorial chemistry is a very attractive strategy for the discovery of bioactive peptides. However, its application has not yet been demonstrated, presumably due to analytical challenges that arise from the diversity of a peptide library with combinatorial side-chains. We previously reported an efficient method to generate, under biocompatible conditions, large dynamic libraries of cyclic peptides grafted with amino acid's side-chains, by thiol-to-thioester exchanges. In this work, we present analytical tools to easily characterize such libraries by HPLC and mass spectrometry, and in particular to simplify the isomers' distinction requiring sequencing by MS/MS fragmentations. After structural optimization, the cyclic scaffold exhibits a UV-tag, absorbing at 415 nm, and an ornithine residue which favors the regioselective ring-opening and simultaneous MS/MS fragmentation, in the gas-phase, upon CID activation.

Introduction

Peptides are promising candidates to target challenging large interfaces such as those involved in protein-protein or protein-nucleic acid interactions owing to their size and functional richness.[1] In particular, macrocyclic peptides in which either a head-to-tail or a side-chain to side-chain covalent bond is introduced,[2,3] exhibit better drug-like properties and represent a promising class of drug candidates.[4,5] Indeed, cyclic backbones generally improve resistance to proteolytic degradation allowing to enhance the *in vivo* half-life, and most often a better membrane permeability is obtained, which makes them suitable for challenging intracellular protein targets.[6,7] Besides, their constrained 3-dimensional structure provides them with restricted conformational freedom that reduces the entropic penalty upon binding to their target, thus resulting in higher affinities. In general, the design of drug-like bioactive cyclic peptide relies on a structure-guided rational approach which is usually followed by a laborious iterative optimization process, involving cycles of parallel peptide synthesis, purification and individual screening of each compound. Alternatively, a general approach for the discovery of bioactive compounds relies on high throughput screening (HTS) of large chemical libraries.[8,9] The use of dynamic combinatorial chemistry (DCC) to generate such libraries would in principle allow the simultaneous screening of the dynamic combinatorial library (DCL) upon addition of the biological target, a so-called target-directed DCC (*td*DCC) approach.[10] Application of *td*DCC to generate and screen peptide libraries is very attractive and several studies toward this aim have been reported over the last years.[11] We have recently introduced an original approach for the development of large dynamic libraries of cyclic peptides, in which amino acid's side-chains are grafted on the surface of a well-ordered cyclic peptide scaffold by DCC.[12] The strategy is based on thiol-to-thioester exchanges, occurring reversibly under biocompatible conditions (in water, pH 7.0-7.5), between cysteine containing peptides and thioester building blocks (BB) bearing amino acid's side-chains (Figure 1).

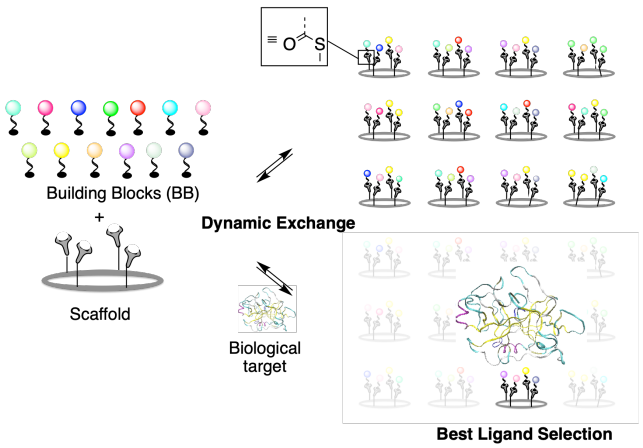


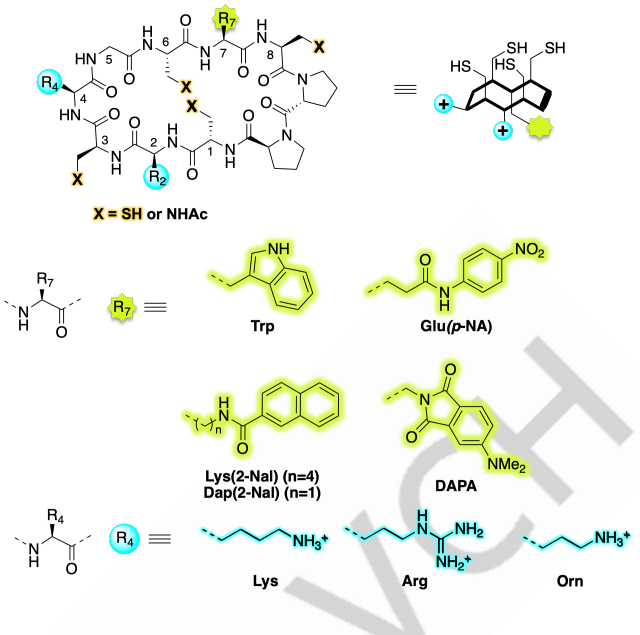
Figure 1. Protein mimetic design by dynamic side-chains grafting of different building blocks (BBs) on a defined 3-D peptide scaffold.

In the presence of the biological target, the latter should in principle select the best binders into the DCL and shift the dynamic library composition, inducing the amplification of the new hit compound(s). Comparison of the mixture's composition at equilibrium, for the target-templated reaction to a blank reaction, allows the identification of the binders.[13,14] The main challenge associated to this approach is the identification of hit sequences after the screening process. The common method consists in comparing the HPLC chromatograms obtained for the libraries in the absence and presence of the target, in order to determine the best binders through their peak amplification, which requires well-resolved chromatograms. The peak identification can be carried out upon coupling HPLC to mass spectrometry (MS). However, MS does not discriminate regioisomeric peptides that bear the same side-chains in different positions. To do so, once regioisomers are separated by HPLC, peptide sequencing by MS/MS (tandem MS) is required. In this report, we have optimized the sequence of cyclic peptides generated by DCC in order to facilitate their analysis, by (i) introducing a UV-Vis active probe with a specific wavelength that does not overlap with the BBs own absorption, (ii) exploring different strategies in order to simplify characterization of the hit sequences by MS/MS fragmentation.

Results and Discussion

UV-Vis probe

We have previously reported 10mer cyclic peptides, as β-hairpin scaffolds containing a ^{D}Pro-^{L}Pro (^{D}P-P) *β*-turn inducing template, and bearing one to four thiol groups.[12] The latter were exposed over one face of the peptides, in positions 1, 3, 6, 8, prone to react in reversible thiol to thioester exchanges (**1-3** Figure 2). A tryptophan residue in position 7 allows the HPLC monitoring of the reaction at a specific wavelength (280 nm), while cationic residues enhance the water solubility required for running the DCC experiments in biocompatible conditions. However, the analysis of the DCLs using tryptophan as a UV probe, (λ_{max}=280 nm) is not ideal as the thioester BBs also absorb at 280 nm, in particular those bearing chromophoric side-chains, such as indole and phenyl groups. Such conditions increase the probability of overlapping BBs with products' signals that can preclude the detection of amplification in the presence of a target. Moreover, the quantification of the grafted peptides is biased by the absorption of the grafted side-chains. In order to simplify the analysis, introduction of a UV-tag with an absorption wavelength that does not overlap with the grafted side-chain's absorption, would allow detecting solely the peptide species. For this purpose, three different UV-active probes were investigated: *p*-nitroanilide (*p*-NA), 2-naphthoyl (2-Nal) and 4-(*N*,*N*-dimethylamino)phtalimidoalanine (4-DAPA) reported by Imperiali *et al*.[15] Apart from the absorption wavelength, the chemical probe must fulfill two crucial criteria: it must be orthogonal to the DCC reaction and stable to the analytical MS method. Commercially available Fmoc-Glu(*p*-NA)-OH was directly used for the solid phase peptide synthesis (SPPS) of peptide **4**, while the 2-Nal and 4-DAPA containing amino acids were synthesized. 2-Nal was easily grafted into Fmoc-Lys-OH/Fmoc-Dap-OH (Fmoc-Dap-OH: *N*-Fmoc-L-2,3-diaminopropionic acid) using the corresponding acyl chloride in respectively 91 and 78% yields (see SI). Fmoc-(4-DAPA)-OH was prepared in a four steps synthesis following reported procedure (49% overall yield).[15] The different probes were introduced in one-to-four cysteine containing peptides **5-9** to study their effect on the DCC reactivity. The peptidic elongations were performed from Gly^5 (*C-ter*) to Cys^6 (*N-ter*) by classical SPPS in Fmoc/*t*Bu strategy using the very acid-labile chlorotrityl resin. After the resin cleavage with hexafluoroisopropanol (HFIPA), the head-to-tail macrocyclization was performed using HATU as coupling agent, followed by TFA/scavengers deprotection, which afforded the cyclic peptides in moderate 10-20% yields after purification by reverse phase HPLC (RP-HPLC) (see SI). Normalized UV-Vis absorption spectra, recorded in water, of labeled peptides **4**, **7** and **9**, bearing respectively the probes *p*-NA, 2-Nal and DAPA, highlight an absorption at λ>300 nm of different intensities (Figure 3). An intermediate intensity was found for peptide-*p*-NA **4** with a broad band centered at 320 nm ($\varepsilon \sim 2.10^4$ $M^{-1}cm^{-1}$), while peptide-(2-Nal) **7** weakly absorbs ($\varepsilon_{320} < 1.10^4$ $M^{-1}cm^{-1}$) and peptide-DAPA **9** remarkably displays two broad and intense bands centered at 325 and 420 nm ($\varepsilon_{325,\,420} \sim 6.10^4$ $M^{-1}cm^{-1}$). MALDI-MS analyses revealed protonated molecules $[M+H]^+$ as well as $[M+Na]^+$ and $[M+K]^+$ cationized ions, but also a fragmentation of the *p*-nitrophenyl moiety for peptide **4** upon laser activation, leading to fragment



Scaffold	AA^1	AA^2	AA^3	AA^4	AA^5	AA^6	AA^7	AA^8
1	Dap(Ac)	Ala	Dap(Ac)	Arg	Gly	Cys	Trp	Dap(Ac)
2	Dap(Ac)	Lys	Cys	Arg	Gly	Cys	Trp	Dap(Ac)
3	Cys	Lys	Cys	Arg	Gly	Cys	Trp	Cys
4	Cys	Lys	Cys	Arg	Gly	Cys	Glu(p-NA)	Cys
5	Dap(Ac)	Ala	Dap(Ac)	Arg	Gly	Cys	Lys(2-Nal)	Dap(Ac)
6	Dap(Ac)	Ala	Dap(Ac)	Arg	Gly	Cys	Dap(2-Nal)	Dap(Ac)
7	Cys	Lys	Cys	Arg	Gly	Cys	Lys(2-Nal)	Cys
8	Dap(Ac)	Ala	Dap(Ac)	Arg	Gly	Cys	DAPA	Dap(Ac)
9	Cys	Lys	Cys	Arg	Gly	Cys	DAPA	Cys
10	Dap(Ac)	Lys	Cys	Orn	Gly	Dap(N-Ac)	Trp	Cys
11	Cys	Lys	Cys	Orn	Gly	Cys	Trp	Cys
12	Cys	Lys	Cys	Orn	Gly	Cys	DAPA	Cys

Figure 2. Structure of ß-hairpin folded peptide scaffolds bearing different UV tags and different cationic residues. The UV-Vis probe is in position 7 (green). Cationic residues are in positions 2 and 4 (cyan).

with $[M+H-16]^+$, $[M+H-14]^+$ and $[M+H-30]^+$ which can be attributed to the reduction of the nitro group. This phenomenon surely occurs in MS/MS experiments thus complexifying the analysis. Peptide **4** has been thus excluded of the DCC studies.

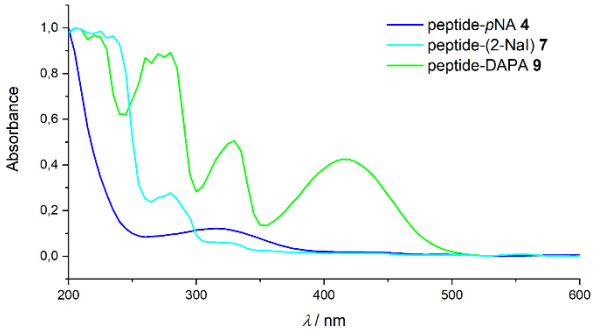


Figure 3. Normalized UV-Vis spectra of peptides **4**, **7**, **9** (0.1 mM range) recorded in water.

The impact of the 2-Nal and DAPA probes on the DCC reaction was then investigated. For the 2-Nal labeling, the DCC experiments were studied on a series of monotopic (**5** and **6**) and tetratopic (**7**) peptides with respectively a single or four cysteine residues on the scaffold surface. The effect of the length of the spacing chain of the probe was evaluated on the monotopic series (**5** *vs.* **6**). Contrary to the monotopic peptides, the tetratopic peptide **7** requires two cationic residues to reach good water solubility at pH 7-7.5 for running the DCC. The thioester exchange reactions were carried out in the previously reported conditions, on peptide **5, 6** and **7,** with a panel of BBs (1 equiv. per thiol) made from an *N*-acetylcysteine (NAC), bearing representative amino acid's side-chains through thioester linkage [**NAC(L)**, **NAC(F)** and **NAC(K),** mimicking respectively leucine, phenylalanine or lysine side-chains]. The reaction was followed by analytical RP-HPLC at 318 nm. Reaction of peptides **5** or **6** containing a single thiol group, with **NAC(L)** or **NAC(K)** led to the corresponding thioester adducts **5(L)/6(L)** and **5(K)/6(K)** as single products with around 50% conversion (Figure 4a and SI figures S27 and S29). As previously reported for the Trp-containing peptide **1-3**, an equilibrium is reached in few hour (2 to 3h), with an apparent equilibrium constant at a fixed pH of 7.5, K', calculated from the concentration values of each component at equilibrium (see SI), lying around 1, reflecting a well-balanced repartition between reactants and products.[12] So, changing the nature of the side-chain at position 7, that does not point toward the same face than the cysteine's side-chains, has no influence on the DCC reaction. Similarly, changing the length of the side-chain, at this position, has no influence on the reaction outcome. With **NAC(F)** the equilibrium is slightly more shifted towards the thioester adduct **5(F)** or **6(F)** with respective K' values of 2.5 and 5 (Figure 4a and SI figures S28). This phenomenon is even more pronounced when four thiol groups are present at the surface of peptide **7**: treating **7** in the DCC reaction with 4 equivalents of thioester **NAC(X)** should in principle lead to $2^4 = 16$ adducts bearing either zero, one (**7(X)**), two (**7(X)$_2$**), three (**7(X)$_3$**) or four (**7(X)$_4$**) thioester side-chains. This is indeed observed for the reaction with **NAC(K)** where the exchanged side chain is polar (Figure 4b).

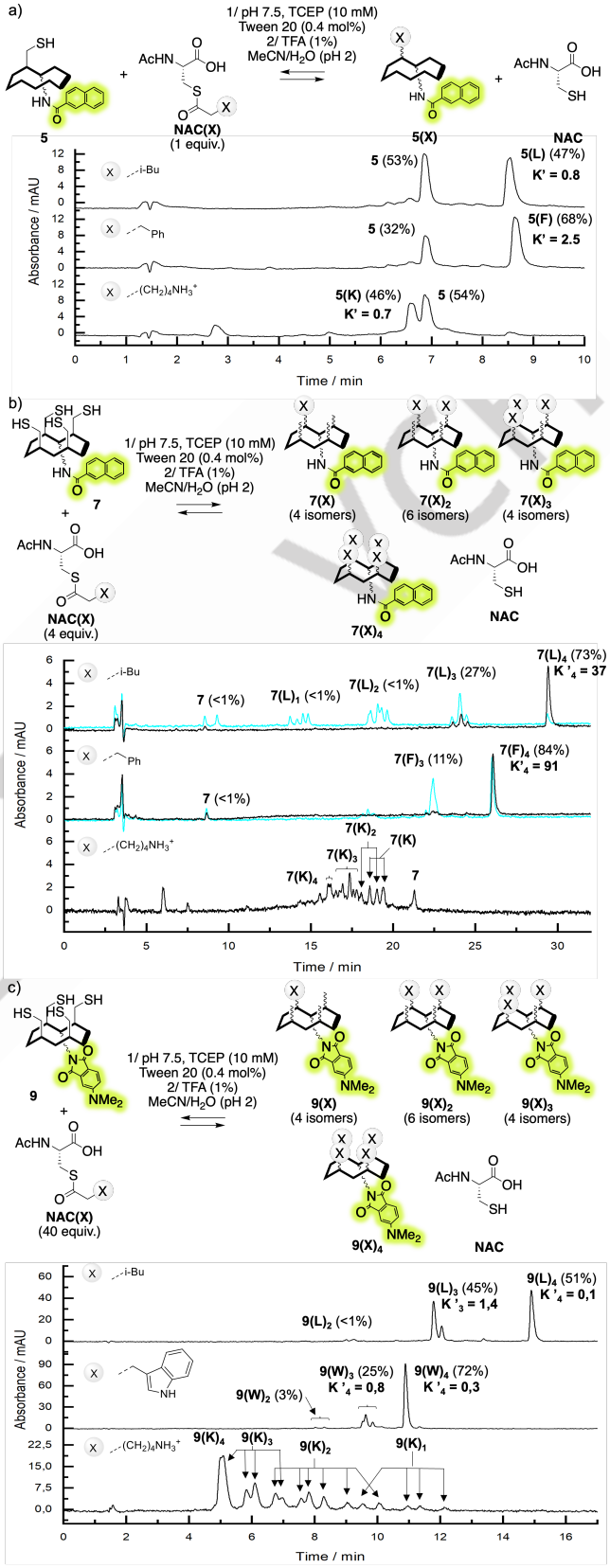


Figure 4. Reaction of: (a) peptide **5**, (b) peptide **7** and (c) peptide **9** with (a) 1, (b) 4, (c) 40 equiv. of respectively **NAC(L)**, **NAC(F)** and **NAC(K)**; and RP-HPLC traces (318 nm) obtained for the reaction at equilibrium (in black) or after 3 h (in cyan).

However, for thioester bearing hydrophobic side-chains such as **NAC(L)** or **NAC(F)**, while such a mixture is obtained in the first 2h, the equilibrium, obtained after 6h, is shifted toward the formation of compounds bearing three or four thioester side-chains. This phenomenon, not observed with the Trp containing scaffold **3**, suggests a stabilization of the corresponding compounds, most probably driven by hydrophobic interactions, favored by the presence of the NaI group. This represents a crucial drawback as it could act as a thermodynamic trap and strongly affect the level of amplification upon addition of a biological target, and prevent the use of this tag for such an application. The DCC experiments were carried out with DAPA labeled mono- and tetratopic peptides **8-9** and were followed by analytical RP-HPLC at 415 nm. When treated with a panel of BBs (1-5 equiv.) in the same DCC conditions at pH 7.5, peptide **8** reacted smoothly to give the corresponding thioester compound, with a roughly statistical mixture of the thiol and thioester peptide obtained at equilibrium after few hours, whatever the nature of the side-chain X (K' 0.3 to 1 with X=L, F, K, G, A W, E see SI figures S33-S39). Sequential addition of different BBs (**NAC(K)** then **NAC(F)**) to a mixture of peptides **8** and **8(A)**, led at each step to a reorganization of the products distribution with the formation of the new adducts **8(K)** and **8(F)** at the expense of **8(A)**, ultimately demonstrating the reversibility of the process as previously demonstrated for the Trp-containing peptide.[12] (see SI figures S40) Treatment of peptide **9** in the same conditions with few representative BBs individually (**NAC(L)**, **NAC(W)** and **NAC(K)** (40 equiv)) led after reaching the equilibrium (8h) to a mixture of mono-, di-, tri- and tetra-acylated peptide (Figure 4c) this time with close K' values. The compatibility of the chemical probe with MS and MS/MS was then verified. While peptides **5-7** as well as their thioester derivatives revealed good stability in the MS experiments, their MS/MS analysis shows that the main fragmentation occurs at the naphthoyl group, which complicates the sequencing of the peptide by MS/MS. On the opposite, peptide **8** and **9** bearing the DAPA tag appeared to be stable both in MS and MS/MS (see SI). Altogether, these results confirm that DAPA is the most appropriate probe for this approach with a strong absorption at 415 nm, orthogonal to that of the BBs, chemically inert, and highly stable in MS and MS/MS. Moreover, this chemical probe has the advantage to possess fluorescent properties, which could be exploited for the analysis of the library and for the study of interaction between library members and a biological partner.

MS/MS peptide sequencing

In this DCC approach, regioisomeric peptides bearing the same thioester side-chains in different positions can be obtained. Having the same *m/z* values, these isomers cannot be distinguished by MS analysis, but as previously shown, characterization of such isomers can be achieved by MS/MS fragmentation. For example, starting from peptide **2**, bearing two cysteine residues, and two different BBs, **NAC(A)** and **NAC(L)**, the generated library contains 9 peptides with 0, 1 or 2 thioester side-chains (Figure 5a). Among them, different pairs of isomers are obtained. Isomers **2(A^3)**/**2(A^6)** and **2(L^3)**/**2(L^6)** bearing respectively a single alanine or leucine side-chain on one of the two thiol groups whereas **2(A^{3}L^6)**/**2(L^{3}A^6)** bear both alanine and leucine side-chains on the two thiols. In all the studies, ESI or MALDI ions sources have been used. While MALDI is known to mainly produce singly charged ion, multiply charged species are usually obtained with ESI. Collision induced dissociations (CID) was used to activate peptide ions in MS/MS experiments. CID is known to lead to the fragmentation of peptide bonds C(O)–NH, resulting in b-, y- and a-ions, the latest being obtained by a consecutive fragmentation of b-ions with a loss of CO.[16,17] Sequencing cyclic peptides usually requires more energy than linear corresponding sequences since two steps are needed: (i) ring-opening upon protonation, randomly, at each amide bond, yielding a series of *n* isomeric linear peptide ions in oxazolone form and (ii) sequential fragmentation of these linear ions, resulting in *n* series of ions, for a total *n* × (*n*-1) b ions. This high energy generally leads to low signal-to-noise ratio. Moreover, additional signals corresponding to internal fragments and/or scrambling sequences are also often observed, rendering the sequence mapping arduous.[18-20] In order to enhance cyclic peptide fragmentation, doubly charged precursor ions were fragmented using ESI-MS/MS. Multiply charged ions are known to undergo fragmentation at lower collision energies than their singly charged counterparts, due to the creation of Coulombic repulsions between the two charges of equal polarity present on the same peptide molecule. In Figure 5b, each fragment ion obtained from the fragmentation of the doubly protonated ion precursor $[M+2H]^{2+}$ at m/z 669.8, is described according to Liu *et al*,[20] with the general formula xmAB, where "x" is the nature of the ion (b, a, etc.) and "m" the number of residues making up this ion. A and B are the two residues connecting the peptide bond, A-B, which has been broken to generate the linear ion corresponding to a b-type ion ending with a positively charged oxazolone, A being the *N*-terminal residue and B the *C*-terminal one. For example, for the peptide **2(A^{3}L^6)**, the ESI-MS/MS spectrum of the doubly protonated ion reveals the presence of fragment ions, arising from the ring-opening of at least five different positions, X^{D}P, ^{D}PP, PX, XK, KC(A), (X= Dap(Ac)) (Figure 5b). Cleavage of the amide bond at the *N*-terminal side of the proline residues is prevailing. This phenomenon, known as the "proline effect", is due to the increased basicity of the tertiary prolyl-amide site and the steric strains generated by the presence of the pyrrolidine ring.[21] However, this effect is not sufficient to fully control the regioselectivity of the ring-opening. Despite complex MS/MS pattern, the sequence of the different isomers could be assigned. Depending on the position of the thioester side-chains, some of the fragment ions have specific *m/z* values which can be used as diagnostic ions to identify the position of the grafted side chain. For example, for peptide isomers **2(A^{3}L^6)** and **2(L^{3}A^6)**, the a6^{D}PP ion of the sequence P^{10}-X^1-K^2-**C(A)**^{3}-R^4-G^5-**C(L)**^{6}-W^7-X^8-^{D}P^9 and the b5PX of X^1-K^2-**C(L)**^{3}-R^4-G^5-**C(A)**^{6}-W^7-X^8-^{D}P^9-P^{10} are displayed respectively at *m/z* 698.4 and 671.4 (Figure 5b and S45, respectively). The identification of these single ions is enough to assign each spectrum. Although the developed method has succeeded in assigning peptide sequences, the complex fragmentation pattern together with the low signal-to-noise ratio obtained by this method requires a tremendous and careful analysis of the results which makes it difficult to consider as a

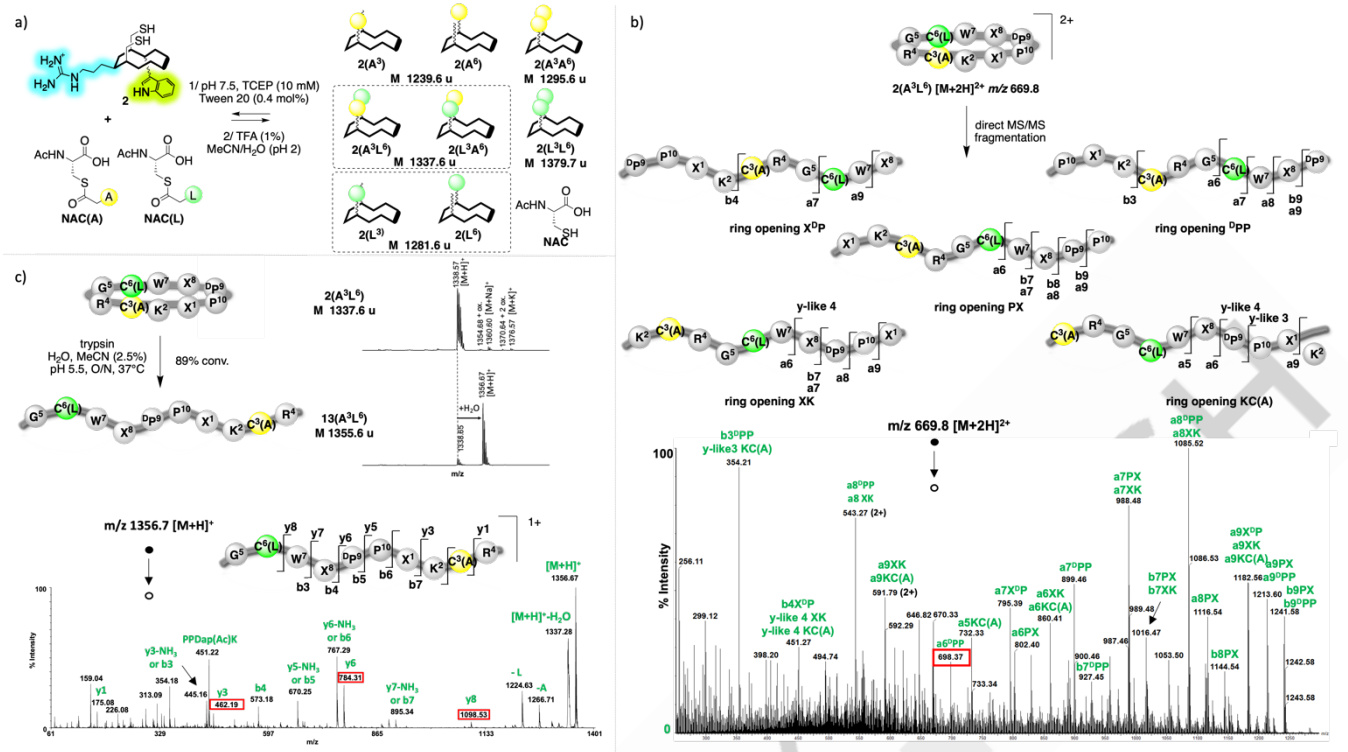


Figure 5. a) Reaction of **2** with **NAC(A)** and **NAC(L)**; b) Schematization of the direct opening of the cyclic peptide in the gas-phase under CID *(on the top)* and fragmentation spectrum of the peptide **2(A^{3}L^6)** obtained by ESI-MS/MS fragmentation (CID, N_2) of the doubly protonated precursor ion $[M+2H]^{2+}$ at *m/z* 669.8 *(at the bottom)*; c) Schematization of the enzymatic ring-opening using trypsin *(on the top)* and fragmentation spectrum by MS/MS of the singly protonated precursor ion $[M+H]^+$ at *m/z* 1356.7 of the linearized peptide **2(A^{3}L^6)** after trypsin opening *(at the bottom)*. Diagnostic ions are squared in red. Annotations of the ions are in green

routine method, especially for libraries obtained from tetratopic peptides, where up to six isomers can be formed and co-elution of their HPLC signal might be observed. A general approach for the characterization of cyclic peptide sequences obtained by combinatorial methods relies on the site-selective ring-opening of the peptide backbone prior to analysis, leading to a single linear sequence. Different strategies were investigated on the first generation of peptides bearing the Trp probe. First, taking advantage of the presence of arginine and lysine residues into the peptide sequence, we sought of using trypsin endoprotease, which hydrolyses amide bonds at the C-terminus of these basic residues.[22] Because cyclic peptides, and particularly those that possess an highly ordered secondary structure, are known to be quite stable to proteases, the challenge here is to find conditions in which the enzyme is able to catalyze the backbone amide hydrolysis while maintaining the thioester side-chains. Notably, the optimal pH for trypsin activity is around 7.4 but this pH could lead to the thioester hydrolysis. The reaction conditions for ring-opening were first established using peptide **2**. The tryptic digestion was studied at different pH, using trypsin at 37°C. In most cases the reaction proceeded smoothly at pH 5.5, in water, to give the corresponding linear peptide, after overnight digestion. Adding 2.5% of acetonitrile to the mixture was shown to improve solubility of some peptide sequences, while maintaining the activity of the enzyme. In these conditions, the ring-opening happened smoothly (~ 90% conversion), as observed on the MALDI-MS spectra with a peak appearing at +18 Da compared to the *m/z* of the cyclic peptide, indicating that the peptide has been hydrolyzed and linearized *C*-terminal to the arginine residue. No thioester hydrolysis was observed in these conditions (Figure 5c for **2(A^{3}L^6)**, see SI figure S48 **2(A^{3}L^6)**). MS/MS fragmentation spectrum of the obtained linear precursor ion, shows a unique hydrolysis site *C*-terminus to the arginine. Noticeably, the amide bond at the *C*-terminal side of the lysine located in the ß-strand is not hydrolyzed by the enzyme, probably because of the high conformational rigidity of the ß-strand fragment. For peptides **2(L^{3}A^6)** and **2(A^{3}L^6)** a single fragmentation pattern corresponding to a unique sequence was obtained. Even though some thioester cleavage is observed, structural information allowed to unambiguously assign the position of the grafted side-chains. However, for peptide **2(L^3)** and **2(L^6)** containing a free thiol group, diagnostic ions corresponding to the two isomers were detected in both fragmentation spectra (SI figure S47), suggesting that a thioester exchange between the functionalized position and the free thiol occurred during digestion, even at pH 5.5. Attempts to run the reaction at lower pH revealed unsuccessful. To improve further the regioselective ring-opening strategy of the cyclic backbone, the “ornithine effect” was investigated. It consists in the spontaneous site-specific cleavage of ornithine-containing peptides upon gas-phase activation (CID or UPVD),

at the *C*-terminal side of this residue.[23-25] It involves the intra-residual nucleophilic attack of the ornithine's δ-amine on the carbonyl group, resulting in a characteristic and preferential cleavage at its adjacent amide bond. The selective cleavage at the ornithine residue can be explained by the highly favorable generation of a six-membered lactam ring. This strategy requires a multi-charged precursor ion (number of charges ≥ 2) leading after the ring-opening either to one doubly charged fragment ion and a neutral fragment or to two singly charged linear peptides. MALDI-MS/MS is not adapted since as stated above, MALDI ion sources mainly produce singly charged ions, consequently, all the following MS/MS studies were performed by ESI-MS/MS. Attempts to obtain ornithine-containing peptides by hydrazinolysis of the arginine residue of peptide **2** or its thioester derivatives according to reported procedure failed,[23] leading to the decomposition of the peptides. So, new peptides **10**-**11**, in which the arginine residues were replaced by ornithine, were prepared by SPPS. ESI-MS/MS experiments of doubly protonated peptide ions **10**-**11** revealed diagnostic ions, which fingerprinted the regioselective ring-opening at the ornithine residue (SI figures S17 and S20). Treatment of peptide **10** with 4 equiv. of **NAC(L)** and 4 equiv. of **NAC(A)**, in the established conditions, led after 3h to an equilibrium in which the eight expected new compounds were separated by RP-HPLC and identified by MALDI-MS. Two pairs of isomers **10(A^3)/10(A^8)** and **10(L^{3}A^8)/10(A^{3}L^8)** were selected as representative peptides for the MS/MS study. The four corresponding peaks were collected and analyzed by ESI-MS and MS/MS. The MS/MS spectra show a single fragmentation pattern corresponding to the linear precursor ion obtained by the regioselective ornithine amide-opened peptide, confirming thus the ornithine site-specific cleavage upon CID activation. For instance, for the assignment of **10(L^{3}A^8)/10(A^{3}L^8)**, despite thioester side-chains neutral losses (observed at m/z 599 for L and m/z 620 for A), characterization of the regioisomers was possible (Figure 6).

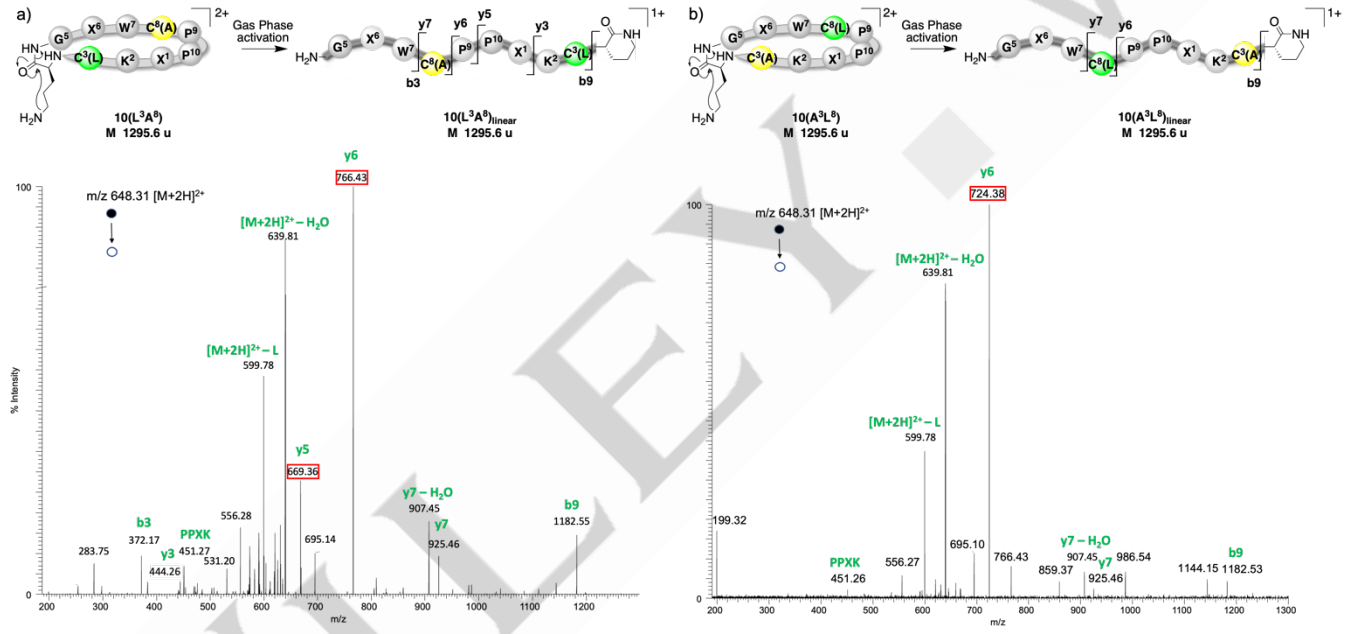


Figure 6. ESI-MS/MS (LTQ/Orbitrap) fragmentation (CID, N_2) of the ornithine-containing peptide ions $[M+2H]^{2+}$ at m/z 648.31 leading to the concomitant ring-opening and fragmentation of the linearized peptides a) **10(L^{3}A^8)$_{linear}$** and b) **10(A^{3}L^6)$_{linear}$**. Diagnostic ions of the position of the L or A side-chain are squared in red. These diagnostic ions are singly charged ions, explaining that their *m/z* is higher than the mass of the precursor ion. Annotations of the ions are in green.

Ions y6, found at *m/z* 724.4 or 766.4, fingerprint **10(L^{3}A^8)** and **10(A^{3}L^8),** respectively. Besides, the fragmentation patterns of **10(A^3)/10(A^8)** are well distinct, which confirm that no thioester exchange occurred in the gas-phase (SI figures S50 and S51). A more complex DCL was generated from tetratopic peptide **11** and the two thioesters **NAC(A)** and **NAC(L),** in large excess (Figure 7). While 3^4=81 compounds could be in principle formed, in such conditions where a large excess of BBs is used, the 16 compounds bearing four thioester side-chains are obtained as major compounds, as observed by HPLC-UV analysis at 280 nm. All the HPLC peaks were collected and submitted to MS analysis for their assignment. As an example of isomer's distinction, the four isomers **11(A^{1,3,6}L^8),** **11(A^{1,3,8}L^6),** **11(A^{1,6,8}L^3),** and **11(A^{3,6,8}L^1)** (*m/z* values of 1357.6), elute between 15.1 and 15.9 min. The MS/MS spectrum of the first eluted isomer exhibits a diagnostic ion y3 at *m/z* 444.3, consistent with the sequence of **11(A^{1,6,8}L^3)**. The second isolated peak was attributed to **11(A^{1,3,8}L^6)**, with diagnostic ions observed at *m/z* 896.4 and 755.4, corresponding to the fragmented ion y7(-H_2O) and y6, respectively. The third peak eluted at 15.9 min contains both **11(A^{1,3,6}L^8)** and **11(A^{3,6,8}L^1),** as proven by the presence of the diagnostic ions y6, at *m/z* 755.4 and 797.4 respectively (SI figure S53). The obtained fragmentation pattern allowed to easily assign each signal and showed that the simplified pattern obtained for the site-specific opening peptide allowed

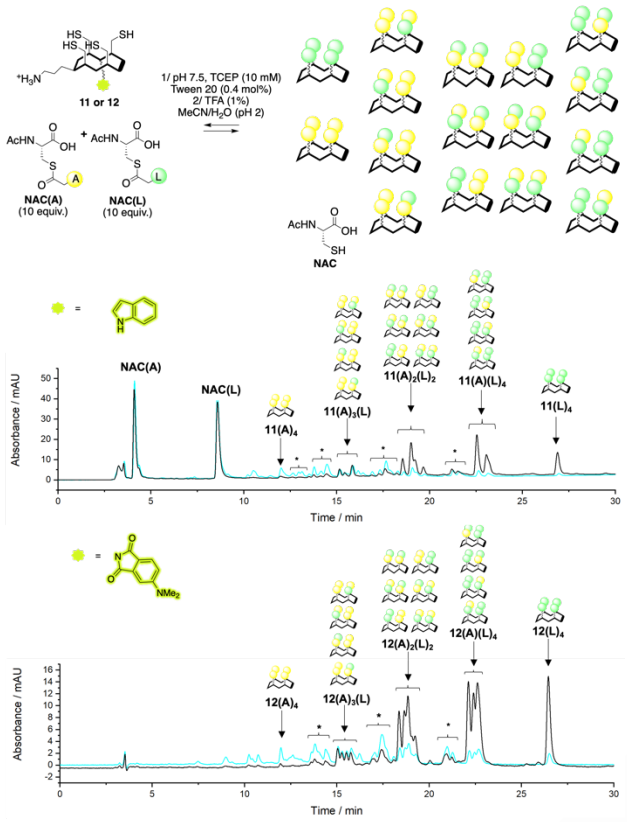


Figure 7. a) Thiol to thioester exchanges between peptides **11** or **12** with **NAC(A)** and **NAC(L)** (10 equiv. of each); RP-HPLC traces at the equilibrium (in black) or after 3 h (in cyan) between b) peptide **11** and **NAC(A)/NAC(L)** (280 nm) and c) peptide **12** and **NAC(A)/NAC(L)** (415 nm). * Only the major peaks have been assigned for the sake of clarity, the other peaks correspond to di- and tri-acylated compounds (see SI for the complete assignment)

to access the sequence even for mixture of isomers obtained when overlap of the HPLC signals occurs, which would have been very difficult to do from the MS/MS fragmentation spectra of the original cyclic compound. Finally, having validated the use of ornithine for the ring-opening and sequence attribution of the scaffold, a new peptide **12** was prepared, combining both the DAPA as a UV tag in position 7 and the ornithine moiety in position 4 for MS-linearization, with 4 cysteine residues at the scaffold surface. Different libraries were generated by treatment of **12** in the above-mentioned conditions with a panel of BBs bearing representative amino acid side-chains [A, L, hR (= homoarginine), W] (see SI, figures S54-57). In all cases, the HPLC traces recorded at 415 nm revealed the presence of different peptides bearing 0, 1, 2, 3 and 4 thioester side-chains, with the amount of each different species mostly depending on the BB to peptide ratio, as expected for a thermodynamically controlled reaction. In all cases, the K' values obtained for each sequential side-chain grafting step (mono-, di-, tri- or tetra-acylation) are very close (same order of magnitude) and roughly around 1, confirming that the different species composing the libraries are isoenergetic, as previously observed for peptide **3**, **9** or **11**. Next a larger library was generated using simultaneously two BBs, **NAC(L)** and **NAC(A)**, in a large excess (10 equiv.). (Figure 7) As expected, a dynamic library containing the 4-thioester side-chains as major compounds in different relative positions is observed, in agreement to the previously obtained results. As illustrated in Figure 7, the different adducts are easier to detect with the DAPA probe, allowing the HPLC monitoring at 415 nm, which excludes the BB absorbance. Moreover, we could verify on few of the obtained compounds that the sequence can be easily assigned by ESI-MS/MS analysis, thus confirming the compatibility of the sequencing method with the DAPA probe (SI figure S59).

Conclusion

Generating large dynamic libraries of cyclic peptides is of great interest for screening inhibitors of complex biological targets. In this respect, our strategy, which consists in the combinatorial grafting of multiple amino acid side-chains on the surface of a cyclic peptide with defined 3-D structure, by thiol-to-thioester exchanges, is very efficient. However, the analysis of such libraries for the identification of the hit compounds, can be very arduous, in particular when the library grows up and involves regioisomers. In this article, we have investigated structural tools to simplify the downstream analysis of the libraries. Specifically, we focused our attention on the implementation of a UV-Vis probe, which allows the HPLC monitoring at a specific wavelength, does not affect the reaction outcomes of the DCC and is stable to MS and MS/MS conditions. The DAPA probe, successfully introduced in the scaffold, fulfills all those criteria and its green fluorescence could be advantageously exploited for studying the interaction between library members and a biological partner. On the other hand, methods for site-selective ring-opening were established in order to facilitate the identification of regioisomers by MS/MS fragmentation. Upon tryptic digestion, peptides were selectively hydrolyzed after the arginine residue, however, thiol-to-thioester exchanges can happen during the process even under relatively acidic conditions. Alternatively, an ornithine residue was introduced to conduct a preferential cleavage upon gas-phase activation, driven by the formation of a six-membered lactam ring. This one-step strategy leading to the simultaneous opening and fragmentation of the linearized peptide in the gas-phase is fast, in comparison to the enzymatic digestion, and easy to implement on an ESI-MS/MS system, which is the most widespread and widely used ionization source. Finally, this strategy appeared to be compatible with our grafted peptides, with no thioester exchanges occurring during analysis. We are currently exploring the potential of this DCC-based strategy as a screening tool of peptide ligands in different relevant biological contexts.

Supporting Information

The Supporting Information includes detailed experimental procedure for chemical synthesis and peptide synthesis, thiol-to-thioester exchanges, MS, MS/MS, NMR spectra and HPLC analyzes (PDF file).

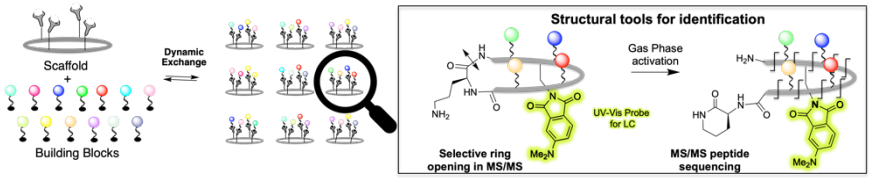
Acknowledgements

We thank the Mass Spectrometry MS^{3}U and the SPPS Platforms (Fédération de chimie moléculaire de Paris centre, FR2769, Sorbonne Université) for services. We acknowledge the French Ministère de l'Enseignement Supérieure et de la Recherche, Sorbonne Université and the doctoral program *Interfaces Pour le Vivant* for PhD fundings.

Keywords: Dynamic Combinatorial Chemistry • thioester exchange • cyclic peptide • mass spectrometry

[1] M. Muttenthaler, G. F. King, D. J. Adams, P. F. Alewood, *Nat. Rev. Drug Discov.* **2021**; *20*, 30.
[2] T. A.; Hill; N. E. Shepherd; F. Diness; D. P. Fairlie *Angew. Chem. Int. Ed.* **2014**, *53*, 13020
[3] E. Lenci; A. Trabocchi *Chem. Soc. Rev.* **2020**, *49*, 3262.
[4] H. Zhang, S. Chen, *RSC Chem. Biol.* **2022**, *3*, 18–31.
[5] A. Zorzi, K. Deyle, C. Heinis, *Curr. Op. Chem. Biol.* **2017**, *38*, 24–29.
[6] D. S. Nielsen, N. E. Shepherd, W. Xu, A. J. Lucke, M. J. Stoermer, D. P. Fairlie, *Chem. Rev.* **2017**, *117*, 8094–8128.
[7] P. G. Dougherty, A. Sahni, D. Pei, *Chem. Rev.* **2019**, *119*, 10241–10287.
[8] C. Sohrabi, A. Foster, A. Tavassoli, *Nat. Rev. Chem.* **2020**, *4*, 90–101.
[9] Y. Zhang, J. Guo, J. Cheng, Z. Zhang, F. Kang, X. Wu, Q. Chu, *J. Med. Chem.* **2023**, *66*, 95.
[10] P. Frei, R. Hevey, B. Ernst, B. *Chem. Eur. J.* **2019**, *25*, 60–73.
[11] A. Rodrigues, L. Rocard, R. Moumné, *ChemSystemsChem* **2023**, e202300011.
[12] B. Zagiel, T. Peker, R. Marquant, G. Cazals, G. Webb, E. Miclet, C. Bich, E. Sachon, R. Moumné, *Chem. Eur. J.* **2022**, e202200454
[13] A. M. Hartman, R. M. Gierse, A. K. H. Hirsch, *Eur. J. Org. Chem.* **2019**, 3581.
[14] A. Canal-Martín, R. Pérez-Fernández, *ACS omega* **2020**, *5*, 26307–26315.
[15] B. Imperiali, M. Sainlos, *Nat. Protoc.* **2007**, *2*, 3201–3209.
[16] B. Paizs, S. Suhai, *Mass Spectrom. Rev.* **2005**, *24*, 508–548.
[17] P. Roepstorff, J. Fohlman, *Biol. Mass Spectrom.* **1984**, *11*, 601–601.
[18] A. P. Jonsson, *Cell. Mol. Life Sci.* **2001**, *58*, 868–884.
[19] J. R. Yates, *J. Mass Spectrom.* **1998**, *33*, 1–19.
[20] W.-T. Liu, J. Ng, D. Meluzzi, N. Bandeira, M. Gutierrez, T. L. Simmons, A. W. Schultz, R. G. Linington, B. S. Moore, W. H. Gerwick, P. A. Pevzner, P. C. Dorrestein, *Anal. Chem.* **2009**, *81*, 4200–4209.
[21] B. L. Schwartz, M. M. Bursey, *Biol. Mass Spectrom.* **1992**, *21*, 92–96.
[22] M. J. Page, E. Di Cera, *Cell. Mol. Life Sci.* **2008**, *65*, 1220–1236.
[23] W. M. McGee, S. A. McLuckey, *J. Mass Spectrom.* **2013**, *48*, 856–861.
[24] C. M. Crittenden, W. R. Parker, Z. B. Jenner, K. A. Bruns, L. D. Akin, W. M. McGee, E. Ciccimaro, J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 856–863.
[25] B. M. Prentice, W. M. McGee, J. R. Stutzman, S. A. McLuckey, *Int. J. Mass Spectrom.* **2013**, 354–355.

Entry for the Table of Contents



Identification of combinatorial cyclic peptides: Complex dynamic combinatorial libraries of cyclic peptides can be generated by thiol-to-thioester exchanges. Structural tools were developed onto the peptidic scaffold to facilitate the identification of a hit by chromatography and mass spectrometry. A chemically inert UV-Vis probe was introduced as well as an ornithine residue for site-selective ring-opening in MS/MS.

Institute and/or researcher Twitter usernames: @LBM_lab, @LouRocard