Late-stage diversification of peptides

Edited by Norbert Sewald
A photochemical C=C cleavage process: toward access to backbone N-formyl peptides

Haopei Wang and Zachary T. Ball*

Abstract
Photo-responsive modifications and photo-uncaging concepts are useful for spatiotemporal control of peptides structure and function. While side chain photo-responsive modifications are relatively common, access to photo-responsive modifications of backbone N–H bonds is quite limited. This letter describes a new photocleavage pathway, affording N-formyl amides from vinylogous nitroaryl precursors under physiologically relevant conditions via a formal oxidative C=C cleavage. The N-formyl amide products have unique properties and reactivity, but are difficult or impossible to access by traditional synthetic approaches.

Findings
The photochemistry of nitroaromatic functional groups has a rich history that dates back decades [1-5]. Photochemical pathways allow access to diverse and interesting target structures [6-10], though photocleavage of C–X bonds for use as photoremoveable protecting groups [11,12] has been the major thrust of the development of 2-nitroaryl compounds. Various 2-nitrobenzyl derivatives are used to photogate heteroatom functional groups, including alcohols, amines, carboxylic acids, and phosphates [11]. Typical photochemical pathways result in cleavage of a benzylic C–X bond following initial benzylic H-atom abstraction [11,13]. In contrast, photorelease systems based on C–C or C=O bond photocleavage are quite rare [14,15]. We recently reported a vinylogous analogue of this photo-deprotection process, which allowed photocleavage of alkenyl sp² C–X bonds, rather than benzylic sp³ C–X cleavage [16,17]. We now report that further studies into this reaction demonstrate two mechanistically distinct photocleavage pathways, with selectivity dependent on pH. In addition to an anticipated alkenyl sp² C–X bond cleavage pathway, we identified a new photochemical reaction pathway, prevalent under neutral and acidic reaction conditions, which leads to formyl products from formal oxidative cleavage of a C≡C bond.
Our interest in vinylogous analogues of 2-nitroaryl photoreactive groups stems from studies into alkenylboronic acid reagents for Chan–Lam-type modification of peptide backbone N–H bonds, directed by a proximal histidine residue (Figure 1C, step, \( i + iv \rightarrow ii \)) [18-20]. Subsequent investigations validated the use of photoreactive boronic acids as an approach to reversible backbone N–H modification via photocleavage of an alkenyl C–N bond [16,17]. Traditional 2-nitroaryl groups allow cleavage of benzylic C–X bonds (e.g. C–O cleavage, Figure 1A) through H-atom abstraction from a photoexcited intermediate, which produces an oxonium-type intermediate (in brackets). Hydrolysis of this intermediate then affords an alcohol product. Recently [16,17], we demonstrated that vinylogous analogues of this mechanism (Figure 1B) provide entry into similar photocaging chemistries for amide release.

Figure 1C shows an example of this concept applied to a peptide substrate. Reaction of the peptide i with an alkenyl-
boronic acid reagent \( iv \) in the presence of a copper(II) salt in water provides access to the backbone N–H alkenylation product \( ii \), directed by a neighboring histidine residue. Upon exposure to 365-nm light, photocleavage of the caging group (red) was observed, producing the free peptide \( i \). Irradiation longer than 10 minutes were sometimes necessary for maximal yield of photo-deprotected product \( i \) [16]. Furthermore, byproducts or transiently stable intermediates were sometimes indicated by HPLC and/or NMR of these photocleavage reactions [16,17]. These observations prompted a more detailed study of the components present during photocleavage reactions of small-molecule models, leading to the identification of the \( N \)-formyl product \( iii \), a possible intermediate on the path to product \( i \) via imide hydrolysis.

To better understand the mechanism of photocleavage and the appearance of the formyl product \( iii \), we first identified the 2-nitroaryl-derived byproducts produced in this reaction. Model compound \( 1 \) was subjected to aqueous photocleavage in the presence of triethylamine, and the resulting reaction mixture was purified by reversed-phase HPLC (Figure 2). We isolated a nitroso product \( 3 \), in addition to two other major identifiable components of the crude reaction: quinoline \( N \)-oxide (4) and quinolinone (5). The compounds 4 and 5 are \( C_9 \) compounds possibly derived from thermal or photochemical rearrangement of compound 3 or another intermediate. The yield of each product was calculated by NMR and verified by isolation (Figure 2).

To test the generality of this process with other functional groups, we prepared and tested alkenyl ether \( 6 \) as a model of C–O-bond cleavage. Photoirradiation of the ether \( 6 \) similarly provided a mixture of \( C_9 \)-containing products 3, 4, and 5. Under these reaction conditions, the C–X cleavage products (MeOH or 2) were observed, but no formyl products were formed. The \( C_9 \) byproducts – the nitroso 3, and related compounds 4 and 5 are all consistent with the classical C–X cleavage mechanism and with hydrolysis of the presumed oxonium intermediate \( 6' \), but are inconsistent with the production of formyl products.

In contrast, photoirradiation of the same alkenyl ether \( 6 \) under acidic conditions at pH 4.0 provided a mixture of methanol (53%) and methyl formate (38%, 7) as determined by NMR, the latter product is the result of formal oxidative C=C cleavage (Figure 3a). Alkenyl amide \( 1 \) at pH 4.0 similarly gave mixtures of the C–N cleavage product 2 and C=C cleavage product 8. We examined product selectivity in the irradiation of alkenyl amide \( 1 \) across a range of pH and found a significant correlation (Figure 3b–d). The formyl product 8 predominated at acidic and neutral pH. The amount of 8 decreased with increasing pH, and above pH 10 the C–X cleavage product 2 became the major product. Unfortunately, no products other than the formyl compound were isolated after photocleavage of compound 1 or 6 in acidic conditions. Instead, when irradiation of alkenyl amide \( 1 \) was conducted in acetone, crude NMR analysis indicated the appearance of product 8 as well as new peaks in the aromatic region.

Following acetylation of the reaction mixture, we were able to isolate small quantities of \( O \)-acetyl \( N \)-hydroxyindole (9,

![Figure 2](image-url)  
Figure 2: Photocleavage of compounds 1 and 6 under basic conditions. Yield of products was calculated from crude \(^1\)H NMR using residual CD\(_2\)OD peaks as internal standard.
Figure 3: (a) Photocleavage of compound 6 under acidic conditions. Yields determined by $^1$H NMR using residual CD$_3$OD as an internal standard. (b–d) Selectivity of photocleavage of alkenyl amide 1 as a function of pH. Product percentage of N-formyl 8 was assessed by crude NMR (c) and graphed (d). Formation of N-formyl-N-methyl acetamide 8 during photocleavage of compound 1. Conditions: 1 (1.8 μmol) was dissolved in MeOD-d$_4$ (200 μL) and deuterated buffer (400 μL). The solution was irradiated at 365 nm for 2 min. (e) Photocleavage reaction of 1 in acetone.
Although the initial byproduct \(\text{N}-\text{hydroxyindole}\) itself proved too unstable to be isolated. It is noteworthy within this context that hydroxyindole is a \(\text{C}_8\) compound, consistent with transfer of the \(\text{C}_1\) formyl group to compound \(\text{8}\). The formation of amide \(\text{2}\) at elevated pH could, in theory, derive from hydrolysis of the initially formed formyl product \(\text{8}\) (i.e. Figure 1C). However, the appearance of primarily \(\text{C}_9\) byproducts in the formation of amide \(\text{2}\) at elevated pH precludes pathways involving the intermediacy of \(\text{8}\). To provide additional support for this analysis, and to assess the stability of \(\text{N}\)-formyl amides formed in this reaction, we irradiated alkenyl amide \(\text{10}\), which contains a 2-phenylethyl substituent that allowed easier isolation of \(\text{N}\)-formyl \(\text{11}\) (Figure 4). After irradiation, the product \(\text{11}\) was isolated in 28% yield, the modest yield reflecting the instability in water and on silica of this compound. The purified \(\text{N}\)-formyl \(\text{11}\) was then dissolved in buffer (pH 8), and its hydrolysis to amide \(\text{12}\) was assessed (Figure 4, inset). We observed clean first-order kinetics to give amide \(\text{12}\) with a half-life \((t_{1/2})\) of 6.4 h.

The observation of \(\text{N}\)-formyl products can be rationalized with a bifurcating mechanism (Figure 5). Following photoactivation, \(\text{H}\)-atom abstraction and nucleophilic addition of water would produce the key intermediate \(\text{B}\). Such hemi-aminol compounds would be unstable under basic conditions, readily forming aldehyde products \(\text{3}\). However, related hemi-aminol compounds are quite stable under non-basic conditions, and the motif is even contained in some natural products, such as zampanolide [21] and spergualin [22]. We propose a competing electrocyclization pathway, affording the heterocycle \(\text{D}\), a pathway which should not be base-catalyzed, and thus may be reasonably predominant under appropriate conditions. From heterocycle \(\text{D}\), a C–C cleavage would produce the \(\text{N}\)-formyl product \(\text{8}\) and a re-aromatized \(\text{C}_9\) heterocyclic byproduct \(\text{E}\). Rearrangement to hydroxyindole \(\text{F}\) would then account for the isolation of the acetylated analogue \(\text{9}\).

The photochemical pathway described here represents a formal oxidative olefin cleavage of vinylogous nitroaryl-modified amides and ethers. The pathway adds to the diversity of photochemical pathways known for 2-nitrophenyl systems, and the concept described here might be useful for the synthetic unmasking of relatively sensitive imido structures. For chemical biology applications, the results point to a far more diverse photochemistry than previously assumed for vinylogous photocleavage systems. Although formyl hydrolysis to the “expected” amide products can and does occur under physiological conditions, the rates of this hydrolysis are slow for the simple models in this study. Within more complex peptides or proteins, selectivity in photocleavage pathways may differ significantly, depending on local chemical environment. It is also worth noting that \(\text{N}\)-formyl products are themselves acylating reagents, and thus could find use in photochemical generation of selective acyl donors.
Supporting Information
Supporting Information File 1
Experimental section and additional information.
[https://www.beilstein-journals.org/bjoc/content/ supplementary/1860-5397-17-202-S1.pdf]

Funding
We acknowledge support from the Robert A. Welch Foundation Research Grant C-1680 (Z.T.B.), and the National Science Foundation under grant number CHE-1904865.

References

ORCID® iDs
Haopei Wang - https://orcid.org/0000-0001-9756-8783
Zachary T. Ball - https://orcid.org/0000-0002-8681-0789

Figure 5: Proposed mechanism for the formation of aldehyde 3 and N-formyl product 8.
Peptide stapling by late-stage Suzuki–Miyaura cross-coupling

Hendrik Gruß1, Rebecca C. Feiner‡2, Ridhiwan Mseya‡3, David C. Schröder1, Michal Jewgiński4, Kristian M. Müller2, Rafal Latajka4, Antoine Marion*3 and Norbert Sewald*1

Abstract
The development of peptide stapling techniques to stabilise α-helical secondary structure motifs of peptides led to the design of modulators of protein–protein interactions, which had been considered undruggable for a long time. We disclose a novel approach towards peptide stapling utilising macrocyclisation by late-stage Suzuki–Miyaura cross-coupling of bromotryptophan-containing peptides of the catenin-binding domain of axin. Optimisation of the linker length in order to find a compromise between both sufficient linker rigidity and flexibility resulted in a peptide with an increased α-helicity and enhanced binding affinity to its native binding partner β-catenin. An increased proteolytic stability against proteinase K has been demonstrated.

Introduction
Peptide cyclisation emerged as a popular approach to limit conformational mobility in order to enhance the binding affinity towards a biological target. Moreover, cyclic peptides are more stable against proteolytic digestion and can provide an improved membrane permeability [1-3]. Hence, peptide-based drugs became of high interest because of their high selectivity combined with low toxicity. Cross-linking of side chain residues results in constrained conformations and can be used to stabilise α-helical secondary structures. This technique is called peptide stapling and the most prominent methodology was de-
cross-couplings are very versatile tools for selective and bio-
phan halogenases opens a broad area of Pd-catalysed late-stage
halogenation at C5, C6, or C7 using FAD-dependent trypto-
Besides addressing the indole C
3-termi-
3-bearing a side chain connection of tryptophan and phenylala-
further improved showing structurally constrained peptides
cyclisation technique by tryptophan C
2–H activation has been
ated stapling reactions were performed only on an analytical
residues of the natural amino acids. However, these Pd-medi-
peptide macrocyclisation of the two above mentioned side chain
formation Pd(II)-aryl complexes. They can further be applied for
peptides and proteins using stochiometric amounts of pre-
labelled cysteine- [38-40] and lysine-containing [41,42]
formation chemistry of biomolecules [35-37]. The approaches by
Buchwald and Pentelute are suitable for selective, bioorthogo-
ligand sulfonated SPhos (sSPhos) and potassium fluoride as a
Pd(dba)
2 was employed [78]. Pd
3, 
-alkenylglycines in a distance of i, i + 3i + 4 for one helix turn
or i, i + 7 for two helix turns, respectively, followed by Ru-cata-
lysed cross-linking [7]. By this robust and reliable approach, a
library of stapled peptides was generated influencing diverse
α-helical dominated protein–protein interactions (PPI) spanning
pathways involved in cancer, infectious diseases, metaba-
olic diseases and neurological disorders [8], which had been
considered undruggable for a long time due to their large con-
tact area and shallow surface [9]. Since then, many other reac-
tions have been investigated for macrocyclisation with the
objective of peptide stapling [10] including lactam- [11,12],
disulfide- [13], thioether- [14-20], triazole- [21,22], oxime- [23]
and hydrazone formation [24] as well as multicomponent reac-
tions such as the Ugi- or Petasis reaction [25-32]. The content
of helicity can moreover be changed by the introduction of a
photo-switchable azobenzene staple [33,34]. Moreover, it has
been shown that Pd-mediated cross-couplings can be success-
fully employed in the generation of cyclic and conformation-
ally stabilised peptides. The groups of Buchwald, Pentelute, and
Ackermann pioneered the development of Pd-mediated aryli-
ation chemistry of biomolecules [35-37]. The approaches by
Buchwald and Pentelute are suitable for selective, bioortho-
nal labelling of cysteine- [38-40] and lysine-containing [41,42]
peptides and proteins using stochiometric amounts of pre-
formed Pd(II)-aryl complexes. They can further be applied for
peptide macrocyclisation of the two above mentioned side chain
residues of the natural amino acids. However, these Pd-medi-
ated stapling reactions were performed only on an analytical
scale and the secondary structures of the cyclic peptides were
not studied. Since tryptophan has only an incidence of about 1%
in proteins, but is highly conserved in binding sites on protein
surfaces mediating PPI [43], it is an attractive target for the de-
v elopment of selective diversifications. C–H activation of the
indole C
2 position by Pd-catalysis allows both selective aryli-
ation [44-48] and formation of macrocycles [49]. The macro-
cyclisation technique by tryptophan C
2–H activation has been
further improved showing structurally constrained peptides
bearing a side chain connection of tryptophan and phenylala-
nine or tyrosine [50]. Moreover, a similar Pd-mediated ap-
proach for C(sp
3)–H activation of phthaloyl-protected N-termi-
nal alanine was also used for macrocyclisation in peptide
stapling [51,52].

Besides addressing the indole C
2, the regioselective enzymatic
halogenation at C5, C6, or C7 using FAD-dependent trypto-
phan halogenases opens a broad area of Pd-catalysed late-stage
diversifications [53-55]. It has been proven that Pd-catalysed
cross-couplings are very versatile tools for selective and bio-

Results and Discussion
Design and synthesis of SMC stapled
peptides
The intramolecular SMC was envisaged as a novel approach
towards one-component peptide stapling by side chain cross-
linking of bromotryptophan and an organoboron moiety. Bromotryptophans are accessible by enzymatic bromination
utilising cross-linked enzyme aggregates (combiCLEAs) con-
taining an FAD-dependent tryptophan halogenase, a flavin
reductase and an alcohol dehydrogenase [73,74]. For this
purpose, tryptophan halogenases RebH and Thal were applied
for the generation of l-7-bromo- and l-6-bromotryptophan, re-
spectively. As a peptide sequence, we chose the β-catenin-
binding domain (CBD) of axin as a benchmark system (PDB
ID 1QZ7) [75]. Axin is a scaffold protein playing an essential
part in the destruction complex for β-catenin labelling in the
canonical Wnt signalling. Loss-of-function mutations in this
pathway lead to a dysregulated signal transduction causing
cancer [75,76]. All-hydrocarbon stapled peptides comprising
amino acids 467 to 481 of the axin CBD had been studied in
the group of Verdine and evaluation of optimised staple posi-
tions at amino acids 471 (i) and 475 (i + 4) resulted in en-
hanced helicity and binding affinity to β-catenin, e.g., for
peptide StAx-3 [77]. Following the StAx-3 peptide, we de-
dsigned peptides including bromotryptophan in i-position and
an organoboron containing side chain in the i + 4-position. The
peptides were synthesised on Rink amide resin by solid-phase
peptide synthesis (SPPS) with Fmoc/tert-Bu strategy followed by
on-resin SMC. For the cross-coupling, a modified protocol by
Planas and co-workers was used [78]. Pd
2(dbh) was employed
as the Pd source together with the water-soluble Buchwald
ligand sulfonated SPhos (sSPhos) and potassium fluoride as a
base. The reaction was performed in a solvent mixture of
dimethoxyethane, ethanol and water (DME/EtOH/H2O 9:9:2) at
120 °C under microwave irradiation for 30 min (Scheme 1)
[78]. The studies were initiated with a macrocyclisation between a
7-bromotryptophan and a 4-pinacoloborono phenylalanine
Scheme 1: Synthesis of SMC stapled axin CBD peptides. Reaction conditions: (a) Pd$_2$(dba)$_3$, sSPhos, KF, DME/EtOH/H$_2$O 9:9:2, 120 °C, µwave, 30 min; (b) TFA/TIS/H$_2$O 95:2.5:2.5, DTT, phenol, 2 × 2 h. B(OR)$_2$ = B(OH)$_2$, B(pin), pin = pinacolato, t-Bu = tert-butyl, Trt = trityl, Pbf = pentamethyl-diatheneboran-5-sulfonfyl.

As a conclusion from those experiments, it was hypothesised that the linker might be too rigid resulting in a distorted structure, which has also been previously reported for thioether cross-linked cysteines bearing a biphenyl template within the staple [20]. Hence, a linker with a higher degree of flexibility was designed. This goal was achieved by a modification of amine-containing amino acids in the $i+4$ position through coupling to 4-carboxyphenylboronic acid, followed by intramolecular SMC. Different linker lengths were achieved by introducing ʟ-2,4-diaminobutyric acid (Dab), ornithine (Orn), or lysine (Lys). Utilising the Alloc protecting group allowed the coupling of 4-carboxyphenylboronic acid once the linear sequence had been synthesised (Scheme 2C). The intramolecular SMC between 6- or 7-bromotryptophan and the boronic acid afforded the stapled peptides P2 to P4 with complete conversion (Scheme 2B). LC–MS analyses revealed broadened or two signals for peptides P1 to P4, which were inseparable by preparative RP-HPLC purification (see Supporting Information File 1). The presence of more than one isomer may be due to the co-existence of diastereomers, i.e. cis/trans isomers or conformers with a high interconversion barrier. For complestatin-based macrocyclic peptides, the existence of biaryl atropisomers caused by the indole of tryptophan has been reported [82]. Recently, the occurrence of isomers was also observed in our group for SMC cyclised RGD peptides. It could be proven that an isomerisation is not caused by the cross-
Scheme 2: Overview of the different cross-linkages obtained by intramolecular SMC. A) General structure of SMC stapled axin CBD peptides. B) Overview of different cross-links: For peptides P1 to P4, 6- and 7-aryl-Trp staples were synthesised, peptide P5 was only generated with 7-aryl-Trp staple. C) On-resin modification of amines (n = 1: Dab, n = 2: Orn, n = 3: Lys) in i+4 position to obtain organoboron side chains with increased linker flexibility. Reaction conditions: (a) Pd(PPh3)4, morpholine in DCM, 2 × 20 min; (b) carboxylic acid, HATU, DIEA in DMF, 1 h. In panel C), amino acids in R1 and R2 are with protecting groups and R3 is resin-bound (Rink amide resin); R3 = 4-phenylboronic acid or (4-ethylphenyl)boronic acid (P2–P5).

coupling but by the presence of stable isomers/conformers. Molecular dynamics (MD) simulations verified the appearance of stable, distinct conformers or atropisomers, which were in accordance with the experimental data [72]. Moreover, the epimerization by the conditions of SMC is unlikely as it has been excluded by total hydrolysis of a late-stage SMC modified RGD peptide [67].

Analysis of the secondary structures of the cyclic peptides P2–P4 by CD spectroscopy also did not show a significantly increased α-helicity in water. Investigation of several derivatives by means of density functional theory (DFT) geometry optimisations indicated that substitution at indole C6 tends to induce more significant deformation of the peptide chain compared to substitution at indole C7. Moreover, introduction of an addition-
al ethylene unit in the linker suggested a conformation with the highest similarity to the linear reference peptide P6 (see Supporting Information File 1), thus representing a good compromise between rigidity and preservation of the target secondary structure. Serine in i-position and glutamic acid in i + 4-position of the linear axin CBD sequence aAxWt were substituted by tryptophan and lysine, respectively, to have a higher analogy to the lysine modified SMC stapled peptides. Following the indications of DFT calculations, stapled peptide P5 was synthesised by modification of lysine in the i + 4-position with 4-(2-carboxyethyl)phenylboronic acid followed by on-resin SMC (Scheme 2B).

LC–MS analysis revealed two isobaric peaks indicating two isomers, which were largely separable by preparative HPLC with the less polar isomer P5.2 being the major one. The secondary structures of both isomers were investigated by CD spectroscopy and an increased helicity was observed for both peptides. In particular, P5.2 shows the characteristic signature of an α-helix with the tendency of minima at λ = 208 and 222 nm and a maximum at λ = 190 nm (Figure 1). The CD spectra provided calculated helicities of 9% for aAxWt, 15% for P5.1, and 21% for P5.2 (see Supporting Information File 1).

Finally, the stability of both peptides P5 and P6 was tested against proteinase K digestion. Whilst the linear analogue P6 is cleaved within a period of 120 min to give three fragments, the stapled peptide P5 allows access to only one of the three cleavage sites: i.e., proteolysis of the Leu–Asp bond within the macrocycle and of the Lys–His bond, which is part of the cross-link, is prevented by the stapling (Figure 3 and Supporting Information File 1, Figures S8 to S10).

Conformational analysis
The identification of two isomers of P5 by LC–MS led us investigate the possibility of diastereomers and conformers in the macrocycle. The amide bond in the staple of P5 is connected to two flexible aliphatic chains and may exist in cis and trans configuration. The energy difference in the analogue N-methylacetamide (NMA) favours the trans isomer by about 2.3 kcal mol⁻¹, which corresponds to an expected cis/trans ratio of about 1:44 at 300 K, with an interconversion barrier of 18.7 kcal mol⁻¹ [83]. The experimental P5.1/P5.2 ratio is nearly 1:3, suggesting an energy difference of only 0.9 kcal mol⁻¹ in favour of P5.2. Strain in the macrocycle might be responsible for such slight decrease in the relative energy between the cis...
and trans isomers compared to isolated NMA. The presence of
diastereomers in the peptide bonds is less likely since the
cis/trans ratio in polypeptides is lower than 1:820 (i.e., an
energy difference greater than 4.0 kcal mol\(^{-1}\) in favour of the
trans isomer) [84]. It is, however, currently hopeless to expect
catching energy differences as small as that between P5.1/P5.2
by molecular modelling. Instead, we discuss qualitatively the
conformational properties of the macrocycle in the two dia-
stereomers of P5 to determine if conformers may exist with
high interconversion barriers, and suggest an assignment based
on other experimental observables, namely, the flexibility of the
overall peptide and its propensity to form a helical secondary
structure. We then proceed to analyse the effect of the staple
and sequence variations on the secondary structure of the
peptidic backbone of aAxWt, P5, and P6.

The conformational preferences of the stapled peptide P5 and of
the linear peptides P6 and aAxW1 were investigated via exten-
sive accelerated molecular dynamics simulations (aMD) as
implemented within the Amber18 program package [85]. The
aMD methodology developed by McCammon and co-workers
[86] has shown to be a highly effective tool to sample the con-
formational space of polypeptides made of sequences of 10 to
30 amino acids [87,88] and of macrocycles [89]. Our simula-
tion strategy, mainly adapted from the latter references, made
use of 15 independent 700 ns-long aMD simulation runs for
each peptide (i.e., a cumulative total of 10.5 \(\mu\)s per peptide) per-
fomed with the ff14SB/GAFF [90,91] and TIP4Pew [91] force
field parameters for the peptides and water, respectively, as well
as specifically derived parameters for non-standard residues of
the linker in stapled peptide P5. The conformation of the
macrocycles was analysed via principal component analysis (PCA)
of the non-hydrogen atoms forming the cycle, and the
structure of the peptide backbone was investigated via secon-
dary structure analysis and backbone root mean square deviation
(RMSD) clustering including amino acids Pro\(^3\) to Met\(^{15}\). Time-
averaged analysis was performed on the ensemble of structures
obtained from the last 500 ns of each simulation run (i.e., a
cumulative total of 7.5 \(\mu\)s per peptide; see Supporting Informa-
tion File 1 for further methodological details and extended anal-
sis).

Figure 4 summarises the conformational analysis on the macro-
cycle in the cis and trans isomer form of P5. PCA reveals that
the first three principal components (PCs) respectively capture
39.0, 21.3, and 13.4\% of the total variance in P5 cis, and 31.2,
25.4, and 9.9\% in P5 trans. PCA-based clustering with a
minimum distance of 4.0 Å in the three-dimensional space
of PC1-3 led to 32 and 38 structural clusters for the
 cis and trans isomers, respectively. The first three representative structures
are depicted in Figure 4, the first six clusters are projected in the
three-dimensional space of PC1-3, and the corresponding repre-
sentative structures are indicated in the two-dimensional projec-
tion in the space of PC1 and PC2, where colouring is made by
relative free energy as obtained from Boltzmann reweighting
using 10th order Maclaurin series expansion [92,93]. In both
isomers, the first three clusters represent about 53\% of the total
conformational space of the macrocycle. Cis and trans isomers
share a fairly similar main conformation (c1, blue in the figure)
with high population (31.8 and 27.9\%, respectively). This con-
formation is stabilised by a hydrogen bond between the CO group of Ile and the NH group of Lys, and triggers the formation of a helical structure in the peptidic portion of the macrocycle. Conformation c2 of the cis isomer (10.9%) presents the Leu side chain pointing towards the centre of the macrocycle and leads to a disruption of helicity. The third conformation for
this isomer (c3 with 10.7%) is also helical in the peptidic portion of the macrocycle and is similar to c1, with the Trp’s indole group pointing in the other direction. In the trans isomer, both conformations are also found, however, with different populations and order. Conformation c2 of P5 trans is helical and resembles c3 of P5 cis, yet with a significantly higher probability of occurrence (17.8%). The non-helical conformation c3 of P5 trans (7.5%) is slightly less populated than the corresponding c2 of P5 cis. The macrocycle of both isomers is found to be rather flexible, forming well-separated conformational clusters in the three-dimensional space of PC1-3. The projection in the two-dimensional space of PC1 and PC2 indicates that the conformers are interconnected with barriers lower than 6 kcal mol$^{-1}$. It is worth noting, that the barriers are likely to be overestimated due to a poor sampling of transition structures compared to local minima. More accurate values would be obtained with methods better adapted for kinetics, such as Markov models (for a general overview see reference [94]). Yet, such low barriers are not sufficient to trap the SMC peptide in conformations that can be separated experimentally at ambient conditions [95] and the analysis therefore rules out the possibility of conformational isomers, within the limits of exhaustivity of our sampling. The cis/trans conversion barrier is likely to be close to that of isolated NMA, and leads us to conclude that the two isomers isolated experimentally are diastereomers of the amide bond in the staple of P5.

The secondary structure analysis of the three peptides aAxWt, P5 (cis and trans) and P6 is summarized in Figure 5. Figure 5A reports the percentage of amino acids adopting a given secondary structure over the simulation time. All peptides show a significant fraction of amino acids in an α-helical conformation, with a smaller yet substantial propensity to participate in turns and bends. Overall, P5 trans is the most helical peptide, followed by aAxWt, and P5 cis, and P6 is significantly less helical than the others, in terms of individual amino acid’s contributions. P6 stands out with a fairly high fraction of amino acids present in an anti-parallel β-sheet backbone conformation. aAxWt also shows a small fraction of anti-parallel β-sheet, while P5 trans only shows a marginal percentage of parallel β-structure. The backbone RMSD-based clustering further breaks down the conformational preferences of the three peptides and is summarized in Figure 5B–E showing the representative structures of the first four structural clusters. The main conformation of aAxWt, P5 cis, and P5 trans is highly populated (23.3, 19.4, and 19.7%, respectively), and shows a full α-helix that closely resembles the active conformation of axin’s binding domain (superimposed in transparent grey). P6 also forms a similar α-helix with a fairly high probability (14.7%). However, the main conformation of P6 (26.2%) is found to be formed by two β-sheets linked by a turn at the middle of the sequence. The second and third conformations of aAxWt, P5 cis, and P5 trans are also significantly helical, with helices formed by at least six consecutive amino acids. The fourth conformation of aAxWt is a β-structure resembling the main geometry of P6. Overall, aAxWt, P5 cis, and P5 trans form helices made of at least six consecutive amino acids with a cumulative probability of 33.1, 37.5, and 45.2%, respectively. Noteworthy, these percentages are not re-weighted and are, therefore, somewhat biased by the aMD protocol. Yet, trends should be qualitatively captured by the analysis, which correlates fairly well with the experimental results in Table S1 (Supporting Information File 1).

The mutations from aAxWt to P6 result in a significant change in conformational preferences and the probability of stable β-structures in the latter. This observation is consistent with the CD spectrum of P6 that presents β-sheets features. The staple in P5 successfully quenches such biologically unfavourable conformation and significantly increases the probability of forming helical structures that closely resemble the active conformation of axin’s binding domain. Both, cis and trans isomers form α-helices with a high probability, yet the trans isomer tends to be more helical than the cis variant. The CD spectra of P5.1 and P5.2 indicate that the latter has a more helical character, which leads us to speculate that P5.1 corresponds to the cis diastereomer, while P5.2 presents the amide bond in a trans configuration. Furthermore, analysis of the structural diversity of the two isomers of P5 indicates that P5 cis (P5.1) is more disordered (see Supporting Information File 1, Table S3), which also correlates with a blue-shifted absorption minimum compared to P5 trans (P5.2).

In the partially helical structures of aAxWt (i.e., second and third conformations in Figure 5b), the helix is formed in the second half of the sequence. In P5 cis and P5 trans, however, the second conformation presents the beginning of the sequence with a helical structure, including the amino acids that participate in the macrocycle. While possible, it would be speculative to link this property of the P5 to its enhanced biological activity. Instead, we find a more likely reason for the greater activity of P5.2 over that of aAxWt in analysing the conformational diversity of the two peptides (see Supporting Information File 1, Table S2). In P5 trans (identified as P5.2), 60% of the conformational space is represented by the first 7 structural clusters against 18 for aAxWt. The latter is, therefore, significantly more flexible and may be found more often in a non-active conformation, including β-structures, compared to P5.2. This last observation tends to correlate with a higher binding affinity of P5.2 over its linear counterparts. Although all peptides can form an α-helix that resembles the active form of axin’s binding domain, P6 and aAxWt occur often in other con-
Figure 5: Molecular modelling of the conformational preferences of the SMC stapled peptides P5 (with cis or trans amide bond in the staple), and the linear references P6 and aAxWt by means of accelerated molecular dynamics (aMD). A. Secondary structure analysis showing the fraction of amino acids found in a specific backbone conformation as normalised over the full sequence and the cumulative last 500 ns of 15 aMD runs for each peptide. B. Results of structural clustering analysis for peptide aAxWt. The representative structure (black) of the four main structural clusters are depicted superimposed to the active conformation of axin’s binding domain (in transparent grey; PDB ID 1QZ7) [75]. Cluster populations are given next to the respective structure and backbone-atoms root mean square deviation with respect to the active conformation are indicated in parenthesis. RMSD reported and used in clustering were calculated from Pro to Met only. A schematic representation of the secondary structure of each amino acid is given below the representative structures as obtained from an average over the whole cluster (beta structures in blue, helices in green, and turn/bend in pink). C–E. Same as panel B for P5.1 (green), P5.2 (red), and P6 (blue), respectively. P5.1 and P5.2 differ in the conformation of the amide bond in the staple (cis and trans, respectively).
formations that are rather far from the active one. Overall, when presenting the peptides to the target domain of β-catenin, P6 and dAXWt are substantially less likely to be in an active or near-active conformation compared to P5.2. In the ensemble of conformations, the fraction of active ones is therefore greater for P5.2, which translates into a greater binding affinity measured experimentally.

Conclusion
In conclusion, suitable reaction conditions were found for the synthesis of stapled peptides by an intramolecular late-stage SMC on resin. The peptide sequences are based on the CBD of axin. Optimisation of the cross-link, guided by DFT geometry optimisation, finally resulted in SMC stapled peptide P5 showing an increased α-helicity. Compared to the linear analogue P6, P5 revealed a five times higher binding affinity to its native binding partner β-catenin. A proteinase K stability assay demonstrated higher stability of the stapled peptide P5 against proteolytic digestion because two of the three cleavage sites are blocked by the macrocycle. Accelerated molecular dynamics simulations verified a significantly higher degree of helicity for SMC stapled peptide P5 compared to the linear analogues P6 and dAXWt, which is in accordance with the experimental data obtained from CD and moreover, explains the increased binding affinity to β-catenin as P5 is more likely to be found in an active conformation.

Supporting Information
Supporting Information File 1
Details on the amino acid and peptide synthesis, analytical data of the peptides, CD spectroscopy, β-catenin expression and purification, fluorescence polarization assay, proteinase K stability assay, and theoretical methods.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-1-S1.pdf]

Acknowledgements
The authors thank A. Nieß, M. Wißbrock, and P. Borchert for technical assistance. A. Merschel and K. D. Stuke are acknowledged for support in the synthetic work.

Funding
The project was supported in part by a Bielefeld University Ph.D. fellowship to H.G. and by Deutsche Forschungsgemeinschaft (SE609/16-1). The DFT calculations were carried out using hardware and software resources of the Supercomputing and Networking Center in Wroclaw (grant no. 197). R.M. and A.M. gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan Xp GPU used for this research as well as TUBITAK ULAKBIM High Performance and Grid Computing Center (TRUBA resources) for additional calculations.

References
Synthesis and late stage modifications of Cyl derivatives

Phil Servatius and Uli Kazmaier*

Abstract
A peptide Claisen rearrangement is used as key step to generate a tetrapeptide with a C-terminal double unsaturated side chain. Activation and cyclization give direct access to cyclopeptides related to naturally occurring histone deacetylase (HDAC) inhibitors Cyl-1 and Cyl-2. Late stage modifications on the unsaturated amino acid side chain allow the introduction of functionalities which might coordinate to metal ions in the active center of metalloproteins, such as histone deacetylases.

Introduction
Among natural products, peptidic structures have entered the limelight due to their extraordinary biological activities [1]. Often found as secondary metabolites for self-defense in different microorganisms, peptidic natural products are assembled either by ribosomal synthesis or by non-ribosomal peptide synthetases (NRPS) [2]. Macro cyclic peptides are pervasive throughout this class of natural products and often show improved stability against proteolytic digest and metabolic processes [3]. Furthermore, cyclization generally helps to fix the active conformation of a peptide needed to interact with the respective cellular target. Incorporation of non-proteinogenic and unusual amino acids often is related to their biological function. For example, trapoxin B (Figure 1) is a cyclic tetrapeptide with a rather unusual epoxaketone side chain and was found to be a strong inhibitor of histone deacetylases (HDACs) [4,5]. HDACs are nuclear isozymes that regulate gene transcription via a dynamic process of acetylation and deacetylation of lysine residues of histones [6-10]. Blockade of the deacylating process causes hyperacetylation of histones and unregulated gene activity, that results in untimely cell death. Eighteen different HDAC enzymes are known so far and they are divided into four classes based on structural homology with yeast proteins [11]. Three of these enzyme classes (I, II, and IV) contain Zn^{2+} within the active site, and therefore these enzymes can be affected by typical Zn^{2+}-binding HDAC inhibitors. In cellular systems, an acetylated lysine of a histone is entering the cavity of the active site and gets coordinated to Zn^{2+}. Subsequent attack of water forms a tetrahedral intermediate which results in a cleavage of the acetylated lysine. Most HDAC inhibitors act as substrate mimics and contain a zinc-binding motif. They competitively...
Figure 1: Naturally occurring HDAC inhibitors.

interact with the HDACs to form stable intermediates and thereby block the active site.

Many naturally occurring HDAC inhibitors are known to date [12]. Acyclic molecules like, e.g., trichostatin A (TSA, Figure 1) were among the first isolated HDAC inhibitors. Isolated in 1976 from *Streptomyces hygroscopicus* by Tsuji et al. [13], TSA played an important role in rationalizing the mode of action of HDACs [14]. Trichostatin contains a hydroxamic acid as zinc-binding motif, inspiring the design of a wide range of synthetic HDAC inhibitors. The essential Zn$^{2+}$-binding group is attached to a non-polar linker, delivering it inside the cavity through a narrow channel. The cap region is responsible for interactions with residues on the rim of the active site [15]. The cap region of acyclic HDAC inhibitors is generally small, resulting in non-specific interactions with the different HDAC isoforms. More diverse cap regions are found in macrocyclic HDAC inhibitors such as trapoxin which contains the unusual, non-proteinogenic amino acid (25S,9S)-2-amino-9,10-epoxy-8-oxodecanoic acid (Aoe) as a zinc-binding group.

Interestingly, Aoe with its α-epoxyketone motif is widespread among this compound class as it is present in other natural products such as Cyl-1 and Cyl-2 [16,17], chlamydocin [18], and many others [12]. The α-epoxyketone is isosteric to an acetylated lysine residue, which makes it a mimic of HDAC’s natural substrate [10]. Although α-epoxyketones and hydroxamic acids show high affinities towards Zn$^{2+}$, other chelating groups are also found in natural products (Figure 2) including ketones (apicidin [19], microsporin A [20]), carboxylic acids (azumamides [21]), α-hydroxy ketones (FR235222 [22]) or thioesters (largazole [23]). These cyclopeptides mainly differ in the amino acid sequence of the peptide backbone, which causes selectivity towards the different HDAC isoforms. In fact, many naturally occurring HDAC inhibitors contain sulfur moieties like, e.g., disulfides or thioesters. They seem to lack a zinc-chelating group at first sight, but the disulfide or thioester acts as a prodrug and are reduced/cleaved in vivo to liberate the free thiol, a strong Zn-binding group [24,25].

Figure 2: Naturally occurring HDAC inhibitors with different zinc-binding motifs.

Results and Discussion

Since a couple of years the focus of our research group is on the synthesis of unnatural amino acids and their incorporation into complex natural products. Many of these show interesting anticancer activities [26-28]. Some linear peptides, such as pretubulysin bind to tubulin [29-31], while others cyclodepsipeptides are strong actin binders [32,33]. Recently, we also became interested in the synthesis of the cyclic HDAC inhibitors. We developed syntheses for chlamydocin [34], Cyl-1 [35], and trapoxin [36] using either Claisen rearrangements [37,38] or Pd-catalyzed allylic alkylation [39,40] as key steps for the synthesis of the unusual Aoe, which was then incorporated into the different tetrapeptides.
Our aim now was to develop a rather flexible protocol that allows us to introduce several types of functionalities onto a given peptide to create libraries of structurally related compounds for SAR studies. Of course, this approach is not limited to the development of HDAC inhibitors, but should be suitable for all kinds of natural product modifications. However, the structural motif of the natural products shown in Figure 2 is suitable to illustrate the concept.

In general, typical syntheses of such natural products start with the synthesis of the unusual building blocks, their incorporation into a linear peptide, which is finally subjected to cyclization at a suitable position. No question, this protocol is well suited to get access to a certain compound, also in large scale, but it is not applicable for the generation of small libraries of related compounds for SAR studies. Therefore, it is more convenient to undertake modifications at a late stage of the synthesis using a suitably modified precursor allowing variations in a straightforward manner. As a model compound, we decided to use the Cyl-1 amino acid backbone and introduce a double unsaturated side chain (Scheme 1). In principle, selective modifications at the two different double bonds (internal and terminal) should be possible. Ring-closing metathesis (RCM) should generate an allylglycine unit, which should undergo a wide range of addition reactions. Ozonolysis, on the other hand, should generate a carbonyl functionality. Radical additions towards the double unsaturated side chain of the Cyl-1 derivative might also allow cyclizations.

Chiral allylic alcohols are easily accessible, either via kinetic resolution of racemic alcohols [46,47], asymmetric catalysis [48], or from chiral pool materials, such as threitol 1 [49]. Using the last approach, 1 was mono-O-allylated to 2 under similar conditions reported previously for monobenzylation (Scheme 3) [50]. Iodination (3) and subsequent elimination of the iodide with zinc dust gave allylic alcohol 4 as a single enantiomer, which was esterified with Boc-protected glycine to allyl ester 5.

Before we incorporated this allylic ester into the desired tetrapeptide, we wanted to make sure that the chelate Claisen rearrangement does not cause any problems. And indeed, Claisen rearrangement of 5 proceeded cleanly, providing the protected amino acid 6 in almost quantitative yield and with perfect chirality transfer.

With this positive results in hand, we incorporated 5 into the desired tetrapeptide 8. So far, we carried out peptide Claisen rearrangements only with small dipeptides, but never used longer peptide chains, such as tetrapeptides. We knew from previous work that the protecting groups on the peptide can have a significant effect on the Claisen rearrangement and therefore we synthesized the Cbz- as well as the Boc-protected peptides 8a and 8b. The tripeptide building blocks were previously also used in the Cyl-1 synthesis. Glycine allyl ester 5 was Boc-deprotected to give amine 7 as hydrochloride salt, using a protocol developed by Nudelman et al. [51]. Coupling with the protected triptides using TBTU occurred without epimerization [52].
The two tetrapeptide allyl esters were subjected to the peptide Claisen rearrangement, the key step of the synthesis. Subjecting allyl ester $8a$ to the usual conditions of an ester enolate Claisen rearrangement with zinc chloride as chelating metal gave the rearranged product in only 33% yield and a diastereomeric ratio of 93:7 (Table 1, entry 1). The reaction was kept at $-45 ^\circ C$ overnight to suppress potential epimerization of the peptide. Generally, epimerization is prevented through deprotonation of amide NH bonds, as argued by Seebach for Li enolates [53,54]. Nevertheless, isoleucine was prone to epimerize under the reaction conditions due to its vicinity to proline and therewith lack of the “protecting” NH group. Since no full conversion was observed in this first attempt, LDA was replaced with LHMDS and the reaction was allowed to warm to room temperature.
overnight (Table 1, entry 2). LHMDS is a weaker base than LDA and should not deprotonate α-substituted amino acid amides [53,54]. No full conversion was observed either and both yield and diastereomeric ratio were similar to the reaction with LDA. If the Boc-protected ester 8b was then treated with LHMDS under the same reaction conditions (Table 1, entry 3) surprisingly no conversion was observed at all. Switching the base back to LDA (Table 1, entry 4) gave similar results than before (Table 1, entry 1). Since the reaction seemed to stop after 30–40% conversion, it was speculated that the ester enolate chelate complex formation was incomplete due to consumption of the base. For instance, deprotonation of tyrosine residues in benzyl position has been observed previously in the derivatization of miuraenamides and would call for an additional equivalent of base. Therefore, the reaction was repeated with 5.5 equiv LDA (Table 1, entry 5) and indeed, the desired tetrapeptide acid was obtained in quantitative yield as crude product, without obvious formation of byproducts. Strikingly, the diastereomeric ratio was also very high. Only prolonged reaction times (>18 h) led to epimerization of the rearranged peptide.

With rearranged tetrapeptide 9b in hand, esterification with pentafluorophenol (Pfp) gave Pfp ester 10, which should readily cyclize after Boc deprotection (Scheme 4). Treatment with HCl in dioxane gave the crude ammonium salt, which was subjected to biphasic ring closure; the hydrochloride salt was added dropwise to a stirred emulsion of saturated NaHCO₃ solution in chloroform [55]. Macrocycle 11 was obtained in acceptable yield as a diastereomeric mixture (dr 87:13), but the diastereomers of 11 could be separated by reversed-phase flash chromatography. It was obviously the C-terminal unusual amino acid which underwent epimerization under the reaction conditions, since already the Pfp ester 10 was partially epimerized, as deter-
mined later on. To access the allylglycine motif for further derivatization, pure cycle 11 was subjected to Grubbs I catalyst in dichloromethane at 45 °C to get the desired product 12. While the reaction proceeded well even with the cyclic tetrapeptide, compound 12 proved to be highly insoluble, which complicated its purification.

An acknowledged method for the removal of metathesis catalysts is the formation of Ru-DMSO complexes, which do not elute from a silica column [56]. This allowed us to remove at least the Ru contamination, but we were unable to subject 12 to further modifications such as cross metathesis or thiol-ene click reactions due to poor solubility. Additionally, providing sophisticated NMR spectra of 12 turned out to be a non-trivial issue. All commercially available deuterated solvents were tested as solvents and finally, recording the spectra in tetrachloroethane-d2 at elevated temperatures (100 °C) led to a clear solution and hence clean NMR spectra.

The solubility issues forced us to investigate also other modification protocols. Thus, macrocycle 11 was subjected to an ozonolysis with subsequent Wittig reaction in a one-pot manner (Scheme 4). Performing the ozonolysis in presence of pyridine led to immediate reduction of the primary ozonide formed during the reaction [57]. Consequently, no PPh3 or Me2S was required to obtain the crude aldehyde. Subsequent addition of a Wittig ylide gave access to a cyclopeptide with an α,β-unsaturated ester side chain as a (E/Z) mixture. Unfortunately, this compound contained triphenylphosphine oxide as impurity, which could not be separated from the product. Subsequent hydrogenation proceeded readily and afforded the saturated cyclopeptide 13. However, the impurity could also not be removed on this stage. Apparently, the Cyl derivatives with a short side chain are not good candidates for further modifications, mainly for solubility reasons.

Therefore, we decided to have a closer look into modifications of the longer side chain present in 11 and subject it to thiol-ene click reactions. Since masked thioles are often found as zinc-coordinating functionalities in HDAC inhibitors, e.g., in the largazoles, we treated 11 with thioacetic acid and BEt3/air in THF to give 82% of the thiol-ene click product (Scheme 4). Performing the ozonolysis in presence of pyridine led to immediate reduction of the primary ozonide formed during the reaction [57]. Consequently, no PPh3 or Me2S was required to obtain the crude aldehyde. Subsequent addition of a Wittig ylide gave access to a cyclopeptide with an α,β-unsaturated ester side chain as a (E/Z) mixture. Unfortunately, this compound contained triphenylphosphine oxide as impurity, which could not be separated from the product. Subsequent hydrogenation proceeded readily and afforded the saturated cyclopeptide 13. However, the impurity could also not be removed on this stage. Apparently, the Cyl derivatives with a short side chain are not good candidates for further modifications, mainly for solubility reasons.

Conclusion
In conclusion, we could show that chelate Claisen rearrangements can be carried out in longer peptides, such as tetrapeptides, as long as acidic positions can be identified, and the amount of base can be adjusted accordingly. The formation of thiol-ene click products 14 substantiated the hypothesis that the insolubility of Cyl derivatives with short side chains limit their synthetic applicability. The fact that 11 underwent rapid intramolecular cyclization after thiol addition renders further investigations into thiol-ene click-initiated cyclization reactions. The chain length in 14 should generally be suitable for effective HDAC inhibition and the thioester moiety might act as a prodrug as described for the natural HDAC inhibitor largazole. Further investigations are currently in progress.

Supporting Information
Supporting Information File 1
Detailed synthetic procedures, characterization of all molecules and copies of NMR spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-18-19-S1.pdf]

Acknowledgements
This article is part of the PhD thesis of Phil Servatius “Total synthesis of natural HDAC inhibitors and derivatives thereof”, Saarland University, 2018.

Funding
Financial support from Saarland University is gratefully acknowledged.

ORCID® iDs
Uli Kazmaier - https://orcid.org/0000-0001-9756-0589

Preprint
A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2021.86.v1

References

License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.18.19