ChemPubSoc Europe

DOI: 10.1002/chem.201400125

CHEMISTRY
A European Journal
Full Paper

Peptide Synthesis

Long-Range Chemical Ligation from N→N Acyl Migrations in Tryptophan Peptides via Cyclic Transition States of 10- to 18-Members

Suvendu Biswas,[a] Roger Kayaleh,[a] Girinath G. Pillai,[a, b] Christopher Seon,[a] Ian Roberts,[a] Vadim Popov,[a] Khalid A. Alamry,[c] and Alan R. Katritzky*[a, c]

Abstract: Chemical ligations to form native peptides from N→N acyl migrations in Trp-containing peptides via 10- to 18-membered cyclic transition states are described. In this study, a statistical, predictive model that uses an extensive synthetic and computational approach to rationalize the chemical ligation is reported. N→N acyl migrations that form longer native peptides without the use of Cys/Ser/Tyr residues or an auxiliary group at the ligation site were achieved. The feasibility of these traceless chemical ligations is supported by the N–C bond distance in *N*-acyl isopeptides. The intramolecular nature of the chemical ligations is justified by using competitive experiments and theoretical calculations.

Introduction

Native chemical ligation (NCL), first reported by Wieland et al.[1] and later developed by Kent and co-workers,[2] is a chemoselective and regioselective reaction of a peptide thioester and a terminal Cys residue of another peptide that gives a native amide bond at the ligation site through a rapid S→N acyl transfer via a cyclic transition state (TS). This is one of the best techniques for the synthesis of peptides and semisynthesis of proteins, takes place in aqueous solution, and is becoming widely used.[3–5] In biomedical research, NCL has been used in the synthesis of the cancer protein NY-ESO-1,[6] cytochrome b562,[7] and dendrimers, monodisperse macromolecules with highly branched three-dimensional architectures.[8] Although still of great use, NCL is limited by the need for an N-terminal Cys residue at the ligation site to afford a peptide containing an internal Cys and by the low abundance of Cys in globular proteins (1.7% of the residues).[9–11] To overcome the limitation of the low abundance of Cys, great effort has gone into developing thiol auxiliary groups. However, subsequent ligations were found to be hard to complete, because of steric hindrance,[12,13] and problematic, because the extraneous groups in the ligated product can be challenging to remove.[11–15]

There are a number of approaches to circumvent the need for an N-terminal Cys residue at the ligation site in classical NCL. These techniques include 1) the use of an auxiliary group followed by removal of that group after ligation,[16–18] 2) NCL followed by the conversion of penicillamine into Val,[17] 3) sugar-assisted NCL,[19–21] 4) Cys-free "direct aminolysis",[22] and 5) traceless Staudinger ligation.[23] However, the development of new ligation methods is still an important area of research in order to access modified peptides and proteins.

In an attempt to address these limitations, our group has devoted considerable effort toward developing alternative techniques. We have reported ligations of S-acylated Cys-containing peptides to form the corresponding native peptides through long-range chemical ligations via 8- to 19-membered transition states.[24–28] This methodology utilizes the selective S-acylation of Cys-containing peptides by *N*-acyl benzotriazoles followed by microwave-assisted high-yielding chemical ligations of the resulting *S*-acyl isopeptides under mild conditions and with no auxiliary groups. However, the low abundance of Cys in the natural peptide sequence remains an obstacle. To address this challenge, we reported recently the chemoselective O→N acyl migration of *O*-acyl serines via 5-, 8-, and 11-membered transition states,[29] and we successfully demonstrated long-range O→N acyl migrations of *O*-acyl Tyr units via 10- to 18-membered transition states.[30]

To the best of our knowledge, N→N acyl migration for the synthesis of native peptides has not been explored. We discovered recently the first examples of successful chemoselective N→N acyl migration involving Trp-containing isopeptides via 10-, 11-, and 12-membered cyclic transition states.[31] These chemical ligations were achieved by migration of the *N*-pepti-

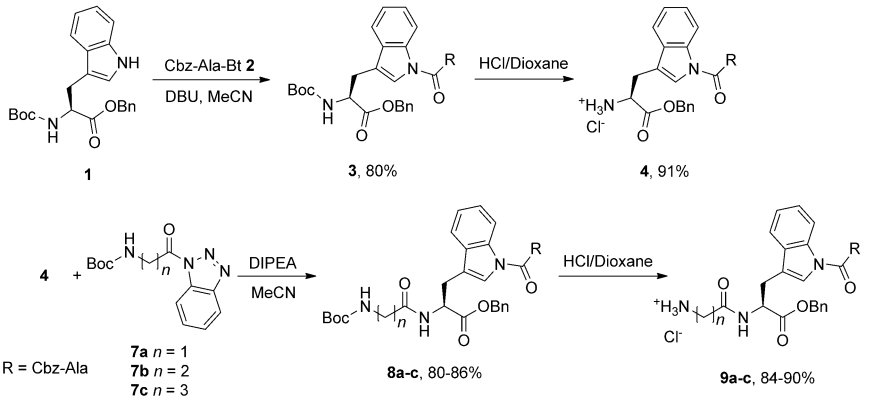
[a] Dr. S. Biswas, R. Kayaleh, G. G. Pillai, C. Seon, I. Roberts, Dr. V. Popov, Prof. A. R. Katritzky
Department of Chemistry, University of Florida
Gainesville, FL 32611 (USA)
E-mail: katritzky@chem.ufl.edu

[b] G. G. Pillai
Department of Chemistry, University of Tartu, Tartu 50411 (Estonia)

[c] Prof. K. A. Alamry, Prof. A. R. Katritzky
Department of Chemistry, King Abdulaziz University
Jeddah 21589 (Saudi Arabia)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201400125.

doyl unit of a Trp isopeptide unit to produce natural peptides. They utilize neither Cys nor Ser residues, nor an auxiliary group at the ligation site. However, this methodology still needs to be fully developed and explored by examining the following factors: 1) the range of cyclic transition states, 2) the best conditions for the ligation step, and 3) the effects of substituents in the amino acid residue and rationalization of the relative abundance of ligated product. The purpose of this work is to identify the structural features controlling the ligation and to correlate the variation in reactivity of *N*-acyl peptides. We have documented the synthetic and computational investigation of *N*-acyl isopeptide ligations to form native peptides from nonterminal Trp residues via 10- to 18-membered cyclic transition states.



Scheme 1. Synthesis of isodipeptide **4** and isotripeptides **9a–c**.[31] Bn: benzyl; Boc: *tert*-butoxycarbonyl; Bt: benzotriazole; Cbz: benzyloxycarbonyl; DBU: 1,8-diazabicyclo[5.4.0]undec-7-ene; DIPEA: *N*,*N*-diisopropylethylamine.

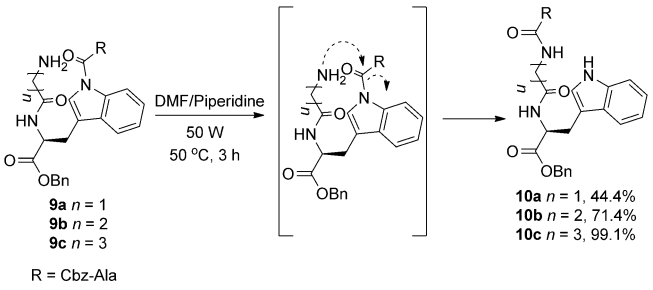
Results and Discussion

Intermediate isodipeptide **4** was synthesized and served as the starting material for the investigation of N→N acyl migration via 10- to 18-membered cyclic transition states. In this study, we aim to investigate a novel chemical ligation strategy for N→N acyl transfer by developing a general methodology and a computational model to predict the feasibility of N→N acyl migration in peptide synthesis. Compound **4** was coupled with the benzotriazolides of dipeptides **12a–o** and tripeptides **17a–c** of α-, β-, or γ-amino acids to afford isotetrapeptides **14a–o** and isopentapeptides **19a–c**, which were required for the ligation studies involving 13- to 18-membered cyclic transition states. To enhance the migration rates, possibly by lowering the steric hindrance at the ligation sites, we placed Gly, β-, and γ-amino acid units within the isopeptides in order to study the feasibility of N→N acyl migrations in 13- to 18-membered cyclic transition states. A statistical model was generated by using conformation analysis and molecular descriptors.

Preliminary results on N→N acyl migrations via 10- to 12-membered cyclic TSs

N-acylation of the indole nitrogen atom in Trp was challenging but was achieved when Boc-protected Trp (**1**) was treated with Cbz-Ala-Bt (**2**) in MeCN in the presence of a strong base, such as DBU, which resulted in the Boc-protected *N*-acyl isodipeptide **3** (80%; Scheme 1). Subsequent Boc deprotection was conducted by using 4 N HCl in dioxane solution to afford the hydrochloride salt of the unprotected isodipeptide **4** (91%).

In our previous studies,[31] we tried chemical ligation experiments on **4** and observed that chemical ligation via a 7-membered cyclic TS was not favored in either aqueous buffer or basic DMF/piperidine conditions. However, in the longer isopeptides **9a–c**, which were prepared by the usual coupling and deprotection protocol (Scheme 1) under DMF/piperidine conditions, N→N acyl migration occurred via a 10- to 12-membered cyclic TS to give Cbz-protected tripeptides **10a–c** (44.4, 71.4, and 99.1%, respectively) as the ligation products (Scheme 2). This N→N acyl chemical ligation occurs in Trp, one of the important natural amino acids present in peptides and proteins. This encouraged us to explore the area of chemical ligation by developing a general, high-yielding, and feasible pathway to the synthesis of Trp-containing peptides via ligation techniques.



Scheme 2. Chemical ligation of *N*-acyl isopeptides **9a–c** in DMF/piperidine.[31]

Feasibility of *N*→*N* acyl migrations via a 13-membered cyclic TS

The starting isotetrapeptides **14a–d** for the N→N acyl transfer via a 13-membered TS were prepared by a straightforward coupling reaction. *N*-Acyl benzotriazoles are advantageous reagents to construct peptides, peptidomimetics, and peptide conjugates.[32–34] Compound **4** and four different Boc-protected dipeptide benzotriazolides **12a–d** were first coupled in MeCN/DIPEA to afford the Boc-protected isotetrapeptides **13a–d**. No chromatography was needed, and compounds **13** were purified by acidic and basic workups. Boc deprotection of the Trp-containing isotetrapeptides in 4 N HCl/dioxane afforded the

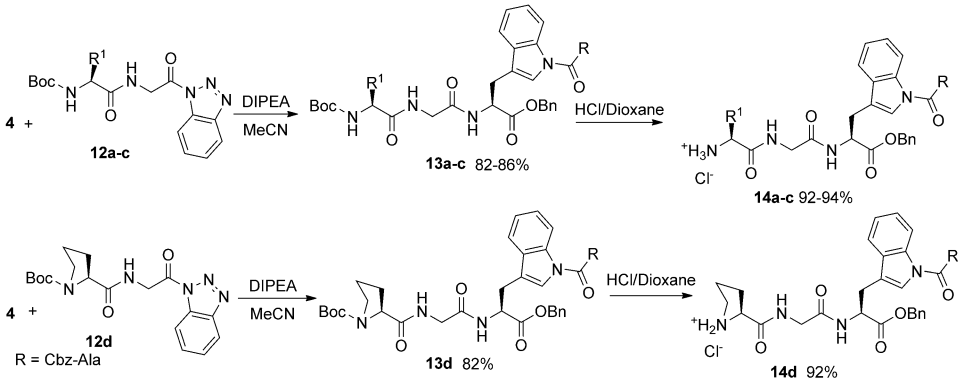
HCl salts of the unprotected isotetrapeptides **14a–d** (Scheme 3). The amino acids Gly, Ala, Phe, and Pro were chosen for the isotetrapeptide sequences to enable a comparative study on the effect of the substituents at the chemical ligation site for the given 13-membered cyclic TS.

Initially, the chemical ligation experiments were carried out for **14a** under aqueous conditions (pH 7.4, 1 M buffer strength, microwave irradiation, 50 °C, 50 W, 3 h) to produce the expected ligation products. However, the expected ligation product was observed in relatively low yield (2%). The ligation experiments were then switched to basic piperidine/DMF conditions (Scheme 4). To our delight, this resulted in the expected ligation products, which, in some cases, were produced in almost quantitative yield (Table 1). In the cases of Phe- and Pro-containing isotetrapeptides **14c** and **14d**, the relative abundances of ligated peptides **15c** and **15d** were 5 and 17%, respectively. This result was expected, because it was anticipated that the chemical ligation would be less feasible, due to steric hindrance by the CH_2Ph group in the case of **14c** or, for **14d**, as a result of the Pro residue inducing a turn in the peptide chain, which would result in too large a distance between the reaction sites. HPLC/MS, with (−)ESI-MS/MS, confirmed the formation of the ligated products **15a–d**, because these produce different MS fragmentation patterns from those of the initial isotetrapeptides **14a–d**. The relative abundances of the crude ligated mixtures were analyzed by HPLC (Table 1).

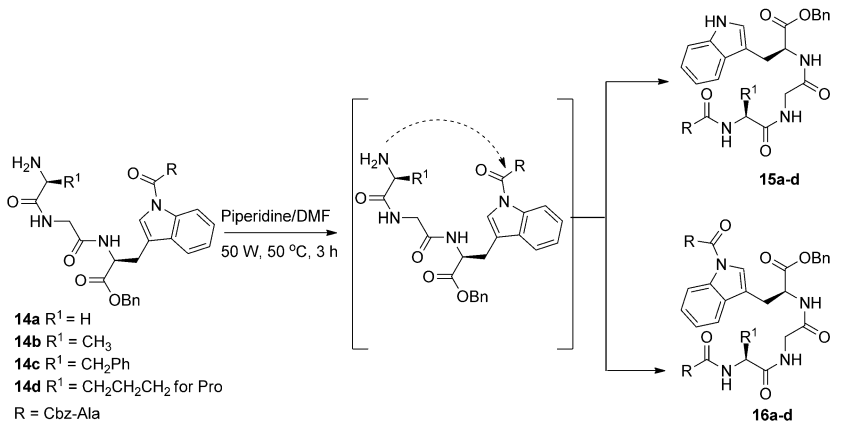
Table 1. Chemical ligation of *N*-acyl isopeptides **14a–d** via 13-membered TSs.

Compd	Cyclic TS size	Yield [%]	Relative area [%][b,c] SM (t_R)	LP (t_R)	BA (t_R)	Product characterization[d] ligated peptide (LP) LP	$[M+H]^+$[e]
14a	13	87	39.11 (41.3)	60.89 (46.3)	0.00	**15a**	614.3
14b	13	88	27.20 (46.3)	72.80 (56.5)	0.00	**15b**	628.3
14c	13	84	94.99 (40.2)	5.01 (46.8)	0.00	**15c**	704.3
14d	13	84	83.05 (45.7)	16.95 (54.6)	0.00	**15d**	654.3

[a] Total crude yield of products isolated. [b] Determined by HPLC-MS (semiquantitative). The area of ion peak resulting from the sum of the intensities of the $[M+H]^+$ and $[M+Na]^+$ ions of each compound was integrated (and corrected for starting material). [c] SM: starting material; LP: ligated peptide; BA: bisacylated product; t_R: retention time. [d] By HPLC-MS. [e] Found.



Scheme 3. Synthesis of isotetrapeptides **14a–d** for ligation studies through 13-membered TSs.

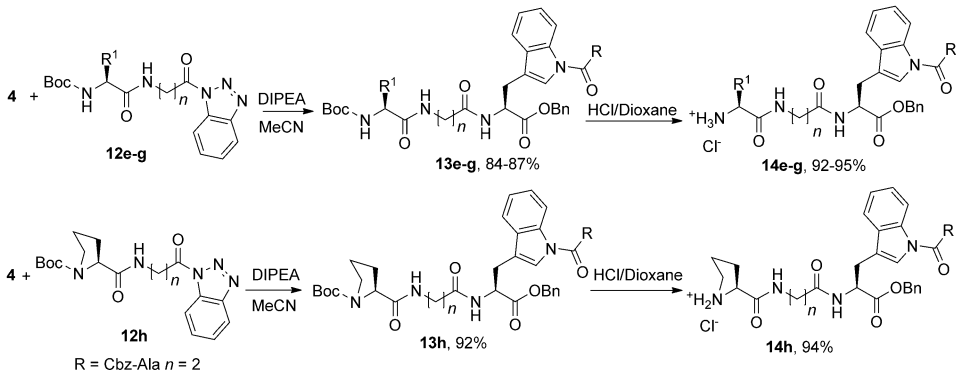


Scheme 4. Chemical ligation of *N*-acyl isopeptides **14a–d** through 13-membered TSs.

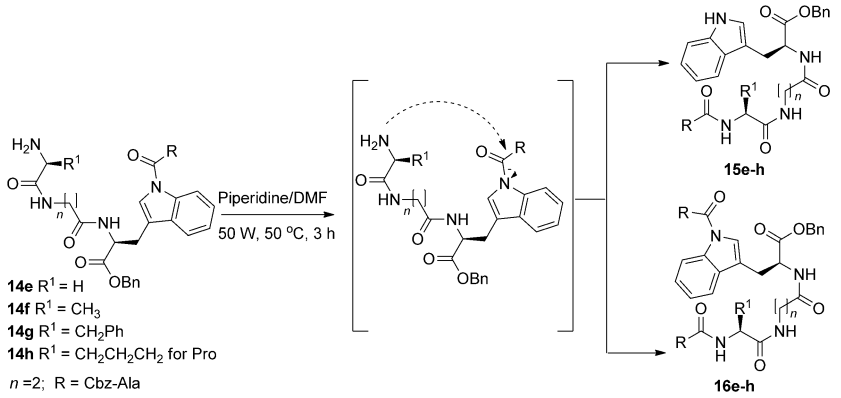
Feasibility of N→N acyl migrations via a 14-membered cyclic TS

Coupling reactions between **4** and four different Boc-protected dipeptide benzotriazolides **12e–h** were performed to study N→N acyl transfer via a 14-membered TS. Boc deprotection of tryptophan tetrapeptides **13e–h** in 4 N HCl/dioxane gave the HCl salts of the unprotected tetrapeptides **14e–h**, which were chosen as potential substrates for the ligation study via a 14-membered TS (Scheme 5). The amino acids Gly, Ala, β-Ala, Phe, and Pro and other amino acids in the isotetrapeptide sequence were chosen to enable a comparative study on the effect of chemical ligation between the given 14-membered cyclic TSs and the 13-membered cyclic TSs.

Chemical ligation via a 14-membered cyclic TS was investigated by subjecting isotetrapeptides **14e–h** to microwave irradiation at 50 °C and 50 W for 3 h under basic piperidine/DMF conditions (Scheme 6). The reaction mixtures were cooled, the sol-



Scheme 5. Synthesis of isotetrapeptides **14e–h** for ligation studies through 14-membered TSs.



Scheme 6. Chemical ligation of *N*-acyl isopeptides **14e–h** through 14-membered TSs.

Table 2. Chemical ligation of *N*-acyl isopeptides **14e–h** via 14-membered TSs.

Compd	Cyclic TS size	Yield [%][a]	Relative area [%][b,c] SM (t_R)	LP (t_R)	BA (t_R)	Product characterization[d] ligated peptide (LP) LP	$[M+H]^+$[e]
14e	14	83	11.06 (39.3)	88.94 (45.7)	0.00	**15e**	628.3
14f	14	88	1.75 (48.0)	98.25 (56.5)	0.00	**15f**	642.3
14g	14	87	69.50 (40.3)	30.50 (47.3)	0.00	**15g**	718.3
14h	14	84	97.37 (44.8)	2.63 (55.0)	0.00	**15h**	668.3

[a] Total crude yield of products isolated. [b] Determined by HPLC-MS (semiquantitative). The area of ion peak resulting from the sum of the intensities of the $[M+H]^+$ and $[M+Na]^+$ ions of each compound was integrated (and corrected for starting material). [c] SM: starting material; LP: ligated peptide; BA: bisacylated product; t_R: retention time. [d] By HPLC-MS. [e] Found.

vent was removed under reduced pressure, and the ligation mixtures (1.0 mg mL^{-1} in methanol) were analyzed by HPLC/MS. The NH_2 site of the unprotected *N*-acyl isotetrapeptides **14e–h** attacked intramolecularly the amide carbonyl carbon atom (C=O) linked to the indole nitrogen atom of Trp via a 14-membered cyclic TS (Scheme 4) to give the ligation peptides **15e–h**. Formation of the expected ligation products in the cases of Gly and Ala (in **15e** and **15f**) was almost quantitative. However, in the case of **14g** (Phe at the N terminus), the ligation product was observed in only 30% yield, and only 3% of the ligated product **15h** was observed in case of **14h** (Pro at the N terminus). This is consistent with our findings for 13-membered TSs for ligation products with similar amino acids. HPLC/MS, with (−)ESI-MS/MS, confirmed that the ligated products **15e–h** each produced different MS fragmentation patterns from those of the initial isotetrapeptides **14e–h**. The relative abundances of the crude ligated mixtures as analyzed by HPLC are shown in Table 2.

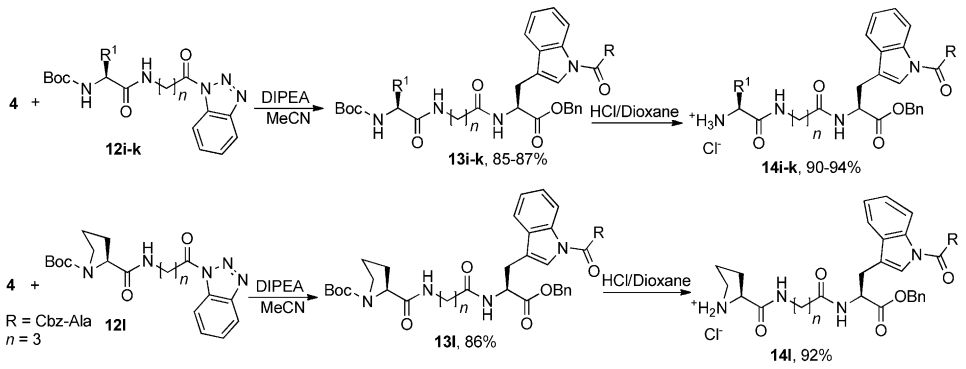
Feasibility of N→N acyl migrations via a 15-membered cyclic TS

The initial isotetrapeptides for the N→N acyl transfer via 15-membered TSs were prepared by following a similar protocol to that in Scheme 5. Four different Boc-protected dipeptide benzotriazolides **12i–l** were first coupled with compound **4** in MeCN/DIPEA to afford the Boc-protected isotetrapeptides **13i–l**. Boc deprotection in 4 N HCl/dioxane gave the HCl salts of the free isotetrapeptides **14i–l** (Scheme 7).

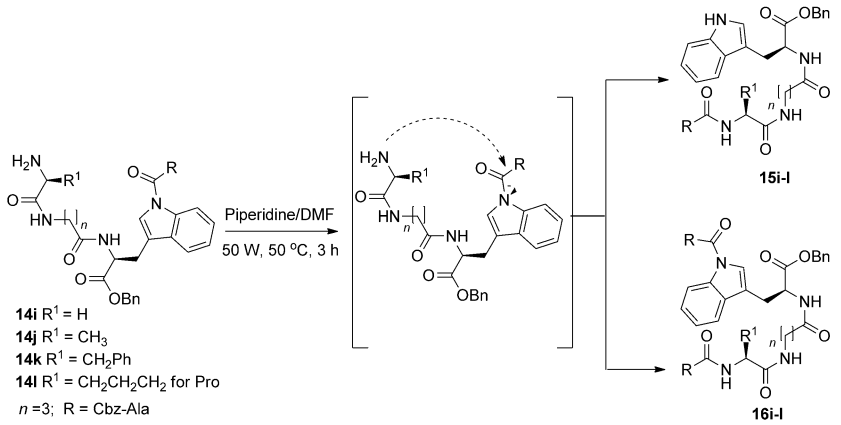
Chemical ligation via a 15-membered cyclic TS was investigated under similar conditions to those in Scheme 4. The abundance of the expected ligation products was low in most cases (Scheme 8). In the case of the proline-containing isotetrapeptide **14l**, only 1% of the ligated product **15l** was observed. This result was probably due to a Pro-induced turn in the peptide chain. HPLC/MS, with (−)ESI-MS/MS, confirmed that the ligated products **15i–l** each produced different MS fragmentation patterns from those of the initial isotetrapeptides **14i–l**. The relative abundances of the crude ligated mixtures, as analyzed by HPLC, are shown in Table 3.

Feasibility of *N*→*N* acyl migrations via a 16- to 18-membered cyclic TS

N→N acyl transfer via a 16- to 18-membered TS would facilitate the synthesis of longer peptides. Coupling reactions between **4** and three different Boc-protected tripeptide benzotriazolides, **17a–c**, in MeCN in the presence of DIPEA (3 equiv)



Scheme 7. Synthesis of isotetrapeptides **14i–l** for ligation studies through 15-membered TSs.



Scheme 8. Chemical ligation of *N*-acyl isopeptides **14i–l** in DMF/piperidine.

Table 3. Chemical ligation of *N*-acyl isopeptides **14i–l** via 14-membered TSs.

Compd	Cyclic TS size	Yield [%][a]	Relative area [%][b,c] SM (t_R)	LP (t_R)	BA (t_R)	Product characterization[d] ligated peptide (LP) LP	$[M+H]^{+}$[e]
14i	15	85	8.29 (37.7)	91.71 (43.8)	0.00	**15i**	642.3
14j	15	90	53.60 (45.4)	46.40 (54.1)	0.00	**15j**	656.3
14k	15	86	95.87 (40.7)	4.13 (47.4)	0.00	**15k**	732.3
14l	15	87	99.14 (45.4)	0.86 (55.9)	0.00	**15l**	682.3

[a] Total crude yield of products isolated. [b] Determined by HPLC-MS (semiquantitative). The area of ion peak resulting from the sum of the intensities of the $[M+H]^{+}$ and $[M+Na]^{+}$ ions of each compound was integrated (and corrected for starting material). [c] SM: starting material; LP: ligated peptide; BA: bisacylated product; t_R: retention time. [d] By HPLC-MS. [e] Found.

at 20 °C gave the Boc-protected *N*-acyl isopentapeptides **18a–c** (Scheme 9). Compounds **18** were purified by acidic and basic workups, and no chromatography was required. The HCl salts of the unprotected isopentapeptides **19a–c** were achieved upon Boc deprotection of *N*-acyl isopentapetides **18a–c** in 4 N HCl/dioxane.

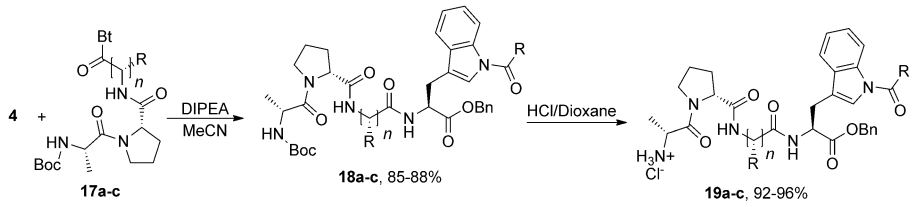
The chemical ligation experiments were performed first for **19a** under basic piperidine/DMF conditions (microwave irradiation, 50 °C, 50 W, 3 h) to produce the expected ligation products. Ligation did not occur for 16-membered cyclic transition states (we recovered mainly the starting material **19a**). Compounds **19b** and **19c** were microwave irradiated in basic piperidine/DMF conditions (50 °C, 50 W, 3 h), and the reaction mixtures were analyzed by HPLC/MS, which showed significant amounts of ligated products (Scheme 10). The abundance of the expected ligation products from **19b** was 31% and from **19c** was 21%. HPLC/MS, with (−)ESI-MS/MS, confirmed that the ligated products **20b** and **20c** each produced different MS fragmentation patterns from those of the initial isotetrapeptides **19b** and **19c**. The abundances of the crude ligated mixtures as analyzed by HPLC are shown in Table 4. On combining all of our experimental results, we observed that the intramolecular N→N acyl migration is highly favored for medium-ring-size cyclic transition states, whereas there is a decrease in the expected ligation products for larger-ring-size cyclic transition states.

Isolation of ligated product

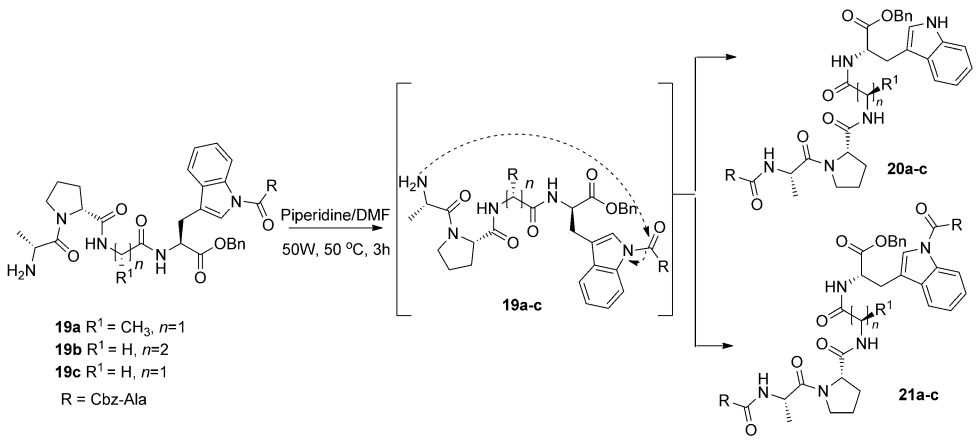
The formation of ligated product **15f** from compound **14f** was further confirmed by isolation with semi-preparative HPLC and characterization by ^{1}H and ^{13}C NMR spectroscopy, elemental analysis, and analytical HPLC. The ^{1}H NMR spectra showed a clear difference in the $\delta = 11.00–7.00$ ppm range (Figure 1). The appearance of the new peak in the case of **15f** at $\delta = 10.85$ ppm in the ^{1}H NMR spectrum (which was absent in **14f**) indicated the formation of the desired ligation product. This peak at $\delta = 10.85$ ppm is a typical NH proton peak from the indole ring in a Trp moiety and clearly confirmed the intramolecular N→N acyl migration of *Z*-alanine to the N terminus to form native peptide **15f** (Scheme 6).

Competitive ligation experiments

To further support the intramolecular nature of the chemical ligation of unprotected isopeptides **14a–o** and **19a–c**, we studied the chemical ligation of isotetrapeptide **14f** in the pres-



Scheme 9. Synthesis of isopentapeptides **19a–c** for ligation studies.



Scheme 10. Chemical ligation of *N*-acyl isopeptides **19a–c** through 16- to 18-membered TSs.

Table 4. Chemical ligation of *N*-acyl isopeptides **19a–c** in DMF/piperidine.

Compd	Cyclic TS size	Yield [%][a]	Relative area [%][b,c] SM (t_R)	LP (t_R)	BA (t_R)	Product characterization[d] ligated peptide (LP) LP	$[M+H]^{+}$[e]
19a	16	83	100 (49.4)	0 (NA)	0.00	**20a**	–
19b	17	86	69.40 (53.9)	30.60 (66.9)	0.00	**20b**	739.3
19c	18	87	79.30 (53.8)	20.70 (67.5)	0.00	**20c**	753.3

[a] Total crude yield of products isolated. [b] Determined by HPLC-MS (semiquantitative). The area of ion peak resulting from the sum of the intensities of the $[M+H]^{+}$ and $[M+Na]^{+}$ ions of each compound was integrated (and corrected for starting material). [c] SM: starting material; LP: ligated peptide; BA: bisacylated product; t_R: retention time; NA: no ligated product. [d] By HPLC-MS. [e] Found.

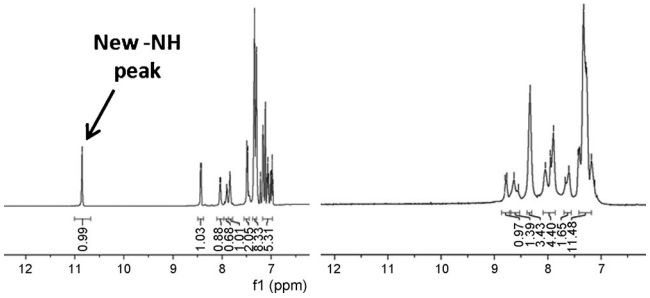
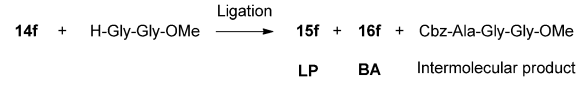


Figure 1. Difference in the ^{1}H NMR spectrum of isolated ligated peptide **15f** (left-hand side) relative to the starting material **14f** (right-hand side).

ence of 20 equivalents of dipeptide **22** (H-Gly-Gly-OMe) under the same reaction conditions as those of Scheme 6 (Scheme 11). HPLC-MS analysis of the isolated crude product (Scheme S72 in the Supporting Information) confirmed the formation of 20% of the desired ligation product **15f** with a retention time at 48.0 min, along with 80% of the starting material **14f** with a retention time at 56.5 min. None of the bisacylated product **16f** was observed. It was also observed that there is no Cbz-protected tripeptide **23** (which is the N-acylated product of dipeptide **22** through an intermolecular ligation pathway) in the HPLC-MS analysis. This competitive experiment supports the hypothesis that the N→N acylation is intramolecular rather than intermolecular.

Computational analysis

Internal chemical ligations can be considered mechanistically as intramolecular nucleophilic reactions between the amide carbonyl carbon atom (C=O) linked to the indole nitrogen atom of Trp and the unprotected N terminus (the distance between these two reactive termini is represented as *b*(N–C)). A cyclic transition state is attained for this transformation, and molecular structure is therefore an important factor. The lowest energy conformer facilitates N→N acyl transfer due to the formation of an amide bond in the ligated peptide. We applied techniques previously employed[27,35,36] for similar ligation reactions including a full conformation search followed by scoring of the conformers based on the spatial distances between the reactive termini, *b*(N–C). The full conformation search considered both rotatable bonds and the phenyl ring of twenty-four related compounds, **9a** to **20c** (including Leu/Val-containing isopeptides, which were used for prediction purposes), and the analysis was performed by using the MMX force field,[37] in PCMODEL v.9.3 software.[38] The best preorganized conformer for each compound, with the *b*(N–C) values, is shown in



Scheme 11. Competitive chemical ligation of *N*-acyl isopeptide **14f** in DMF/piperidine.

Table 5. Computational data for preorganized (conformer) compounds **9a** to **20c**.

Compd	TS size[a]	Sequence	Relative abundance[b]	*b*(N–C) [Å]	Balaban index
9a	10	H-Gly-Trp(Cbz-Ala)-OBn	44.40	4.013	1.437
9b	11	H-β-Ala-Trp(Cbz-Ala)-OBn	71.40	3.171	1.444
9c	12	H-Gaba-Trp(Cbz-Ala)-OBn	99.10	3.010	1.447
14a	13-G	H-Gly-Gly-Trp(Cbz-Ala)-OBn	60.89	3.245	1.454
14b	13-A	H-Ala-Gly-Trp(Cbz-Ala)-OBn	72.80	3.328	1.456
14c	13-F	H-Phe-Gly-Trp(Cbz-Ala)-OBn	5.012	4.845	1.244
14d	13-P	H-Pro-Gly-Trp(Cbz-Ala)-OBn	16.95	3.995	1.253
–	13-V	H-Val-Gly-Trp(Cbz-Ala)-OBn	NA	3.253	1.455
–	13-L	H-Leu-Gly-Trp(Cbz-Ala)-OBn	NA	3.188	1.448
14e	14-G	H-Gly-β-Ala-Trp(Cbz-Ala)-OBn	88.94	2.996	1.444
14f	14-A	H-Ala-β-Ala-Trp(Cbz-Ala)-OBn	98.25	2.978	1.443
14g	14-F	H-Phe-β-Ala-Trp(Cbz-Ala)-OBn	30.50	3.731	1.221
14h	14-P	H-Pro-β-Ala-Trp(Cbz-Ala)-OBn	2.630	4.526	1.236
–	14-V	H-Val-β-Ala-Trp(Cbz-Ala)-OBn	NA	3.233	1.437
–	14-L	H-Leu-β-Ala-Trp(Cbz-Ala)-OBn	NA	3.555	1.426
14i	15-G	H-Gly-Gaba-Trp(Cbz-Ala)-OBn	91.71	3.126	1.430
14j	15-A	H-Ala-Gaba-Trp(Cbz-Ala)-OBn	46.40	3.655	1.428
14k	15-F	H-Phe-Gaba-Trp(Cbz-Ala)-OBn	4.130	4.045	1.198
14l	15-P	H-Pro-Gaba-Trp(Cbz-Ala)-OBn	0.861	4.324	1.218
–	15-V	H-Val-Gaba-Trp(Cbz-Ala)-OBn	NA	3.887	1.417
–	15-L	H-Leu-Gaba-Trp(Cbz-Ala)-OBn	NA	3.562	1.405
20a	16	H-Ala-Pro-Ala-Trp(Cbz-Ala)-OBn	0	4.911	1.281
20b	17	H-Ala-Pro-β-Ala-Trp(Cbz-Ala)-OBn	30.60	3.743	1.231
20c	18	H-Ala-Pro-Gaba-Trp(Cbz-Ala)-OBn	20.70	4.090	1.206

[a] The transition state size and amino acid code for the N-terminal residue. [b] The relative abundance of the ligated product. NA: Test compound for model validation.

Table 5. For understanding of the colinearity between the calculated 127 3D molecular descriptors by using PADEL descriptors and the relative abundance of the ligated peptide, the highly contributing descriptors, that is, the spatial distance, *b*(N–C), and the Balaban index, are shown in Table 5. The Balaban index estimates the isomeric discrimination ability of the peptide.[39] Two of the preorganized 3D conformers with the highest and lowest *b*(N–C) values for the 14-membered-TS compounds **14f** and **14h** are shown in Figure 2a and b, respectively (3D representations for all TSs are given in Table S1 in the Supporting information). These results confirms that the *b*(N–C) value is highly dependent on the amino acid sequence chosen for the peptide chain.

To correlate the spatial distance, *b*(N–C), and the stabilities of selected conformers, quantum chemical calculations were carried out at the DFT/def-SVP level[40–42] of theory by using Turbomole Software.[43,44] To determine whether these results are due to the congestion offered by the steric energy, we calculated those energy values by using molecular mechanics. Steric energy can also influence the relative abundance of the ligated product; thus, we determined the steric energy to prioritize the conformers, in addition to the spatial distance, *b*(N–C). The steric energy of the preorganized conformers (ΔE_S) is derived as the difference between the global minimum energy (E_{global}) and the energy of the best preorganized conformer ($E_{conformer}$), as shown in Equation (1).

$$\Delta E_S = E_{global} - E_{conformer} \quad (1)$$

In general, the feasibility of ligation depends on the spatial distance between the reactive sites, steric hindrance, and hydrogen-bond distances.[27] The calculated data with steric energy and heat of formation for all twenty-four compounds **9a** to **20c** (including Leu/Val-containing isopeptides, which were used for prediction purposes) are given in Table S2 in the Supporting Information. We extended our studies by using the data in Table 5 to design a predictive model with statistical techniques to correlate the relative abundance (ligated product %) with the quantitative structure–activity/property relationship (QSAR/QSPR).[45] The genetic algorithm linear regression method was performed by using QSARINS software,[46] which establishes a correlation between the dependent variable (property/response is the relative abundance of the ligated peptide) and the independent variables (molecular descriptors or factors). The quality of the regression is reflected in the numerical values of statistical parameters, including the coefficient of determination (R^2), the standard error (*s*), the Fisher criterion (*F*), the Student's test (*t*), and the cross-validation coefficient of the determination (R^2_{CV}). These statistical parameters for the model are shown in Table 6, the structure–activity relationship is given by Equation (2) (in which g_1 and g_2 are *b*(N–C) and the Balaban index, respectively), and the regression plot is shown in Figure 3. With reference to the generated model, the spatial distance *b*(N–C) has 85% correlation with the relative abundance, and the Balaban index has 71% correlation with the relative abundance, results that are shown as

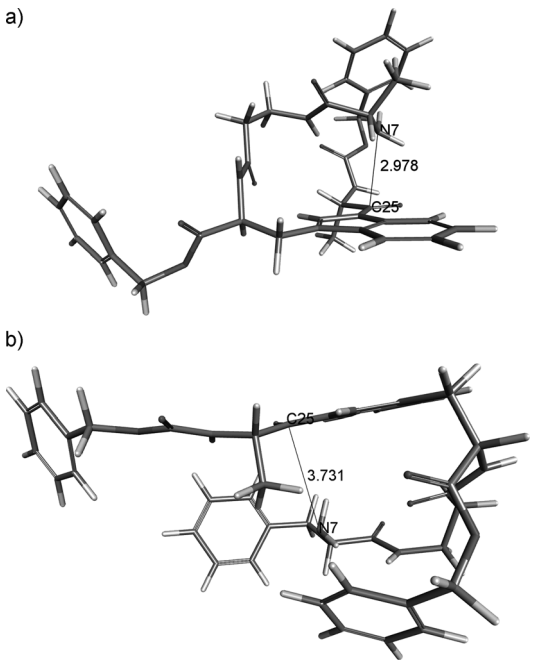


Figure 2. a) Preorganized 3D representation with *b*(N–C) value for compound **14f**. b) Preorganized 3D representation with *b*(N–C) value for compound **14h**.

Table 6. Statistical model for the relative abundance of the ligated peptide.

	Factor[a]	Coefficient[b]	s[c]	t[d]
$R^2 = 0.93$, $F = 106.1$, $s = 9.92$, $n = 2$, $N = 18$, $R^2_{cvloo} = 0.90$, $R^2_{adj} = 0.88$				
I	intercept	0.1439		
g_1	distance b(N–C)	−36.143	−0.626	42.50
g_2	Balaban index	134.304	0.406	37.50

[a] The molecular descriptor based on the BMLR stepwise model. [b] Coefficients of the respective factors including the intercept. [c] Standard error. [d] Student's t test (t criterion).

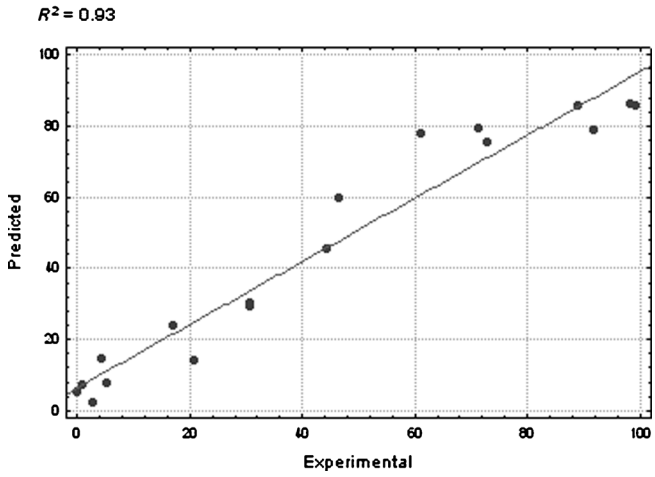


Figure 3. Correlation plot for the relative abundance model of ligated product in Table 7.

regression plots in Figures S1 and S2 in the Supporting Information.

$$\text{Relative abundance} = 0.1439 + ((g_1 \times -36.143 \pm 0.62) + (g_2 \times 134.304 \pm 0.4)) \quad (2)$$

In this present study, the percentage of ligated peptide increases with a) shorter spatial b(N–C) distance and b) higher Balaban index.

By using the Equation (2) model, the relative abundance of ligated product for the twenty-four compounds, **9a** to **20c** (including Leu/Val-containing isopeptides, which were used for prediction purposes), was calculated. The cross-validation coefficient (R^2_{cv}) was very close to the regression coefficient (R^2), which resembles the quality of the model. To our delight, the experimental relative abundances of the ligated peptides for compounds **9a**, **14b**, **14c**, **14e**, **14g**, **14h**, and **20b** were found to be very close to the predicted values. The experimental relative abundances of ligated product for the remaining compounds were close to the predicted values, and the results are shown in Table 7.

Experimental validation of the model

To test the statistical model, three further compounds, **14m–o** (with Val at the N terminus), were synthesized by using a similar protocol (Scheme 12). The amino acids Gly, β-Ala, and Gaba were chosen for the isotetrapeptide sequence to afford a systematic investigation into the feasibility of chemical ligation for the 13- to 15-membered cyclic TSs.

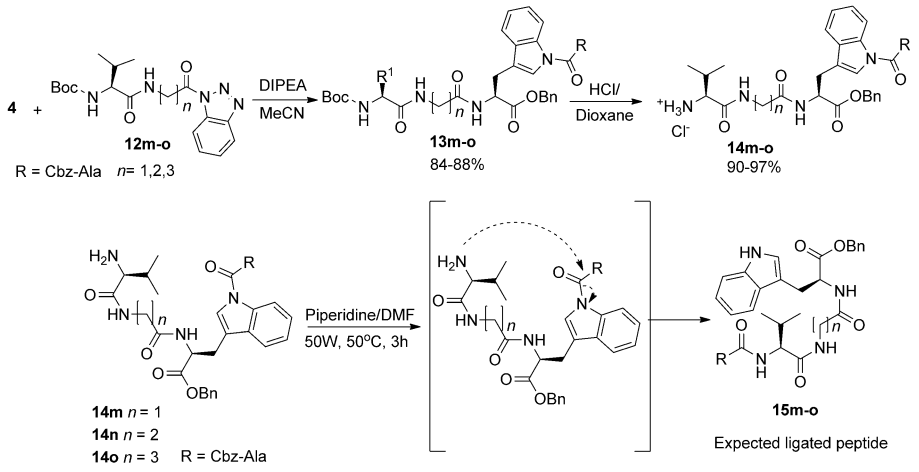
After the ligation experiment was carried out under similar conditions to those used previously (Scheme 12), the crude reaction mixture was analyzed by HPLC-MS. The expected ligated products **15m–o** each produced different MS fragmentation patterns from those of the starting isotetrapeptides **14m–o**, and the relative abundances of the ligated peptides are shown in Table 8. The experimental and predicted abundances of the peptides were not a quantitative match, but a qualitative correlation was observed. These results indicate that the 13- and 14-membered cyclic TSs are more feasible than the 15-membered cyclic TS. This was consistent with our findings with all other compounds for the 13- to 15-membered cyclic TSs (Table 7).

The chemical ligation through an intramolecular N→N acyl transfer to form the native peptides was highly favorable. Intramolecular chemical ligation occurs smoothly in smaller (10- to 12-membered cyclic TS) isotripeptides. It is also highly favored in medium-sized (13- to 15-membered cyclic TS) isotetrapeptides, but it is highly dependent on the peptide sequence and spatial distance, b(N–C). If the terminal amino acids are Phe and Pro, the large steric bulk of the substituents on those amino acids force the two reactive sites apart. If the

Table 7. Experimental and predicted relative abundances of ligated products.

Compd	Cyclic TS size[a]	Ligated peptide sequence	Exptl[b]	Calcd[c]	Diff.
9a	10	Cbz-Ala-Gly-Trp-OBn	44.40	45.67	1.27
9b	11	Cbz-Ala-β-Ala-Trp-OBn	71.40	79.50	8.1
9c	12	Cbz-Ala-Gaba-Trp-OBn	99.10	85.98	13.12
14a	13-G	Cbz-Ala-Gly-Gly-Trp-OBn	60.89	78.14	17.25
14b	13-A	Cbz-Ala-Ala-Gly-Trp-OBn	72.80	75.41	2.61
14c	13-F	Cbz-Ala-Phe-Gly-Trp-OBn	5.012	7.90	2.88
14d	13-P	Cbz-Ala-Pro-Gly-Trp-OBn	16.95	24.03	7.08
–	13-V	Cbz-Ala-Val-Gly-Trp-OBn	NA	77.29	NA
–	13-L	Cbz-Ala-Leu-Gly-Trp-OBn	NA	77.57	NA
14e	14-G	Cbz-Ala-Gly-β-Ala-Trp-OBn	88.94	85.79	3.15
14f	14-A	Cbz-Ala-Ala-β-Ala-Trp-OBn	98.25	86.31	11.94
14g	14-F	Cbz-Ala-Phe-β-Ala-Trp-OBn	30.50	29.28	1.22
14h	14-P	Cbz-Ala-Pro-β-Ala-Trp-OBn	2.630	2.56	0.07
–	14-V	Cbz-Ala-Val-β-Ala-Trp-OBn	NA	78.21	NA
–	14-L	Cbz-Ala-Leu-β-Ala-Trp-OBn	NA	62.17	NA
14i	15-G	Cbz-Ala-Gly-Gaba-Trp-OBn	91.71	79.21	12.5
14j	15-A	Cbz-Ala-Ala-Gaba-Trp-OBn	46.40	59.83	13.43
14k	15-F	Cbz-Ala-Phe-Gaba-Trp-OBn	4.130	14.84	10.71
14l	15-P	Cbz-Ala-Pro-Gaba-Trp-OBn	0.861	7.44	6.579
–	15-V	Cbz-Ala-Val-Gaba-Trp-OBn	NA	49.96	NA
–	15-L	Cbz-Ala-Leu-Gaba-Trp-OBn	NA	60.10	NA
20a	16	Cbz-Ala-Ala-Pro-Ala-Trp-OBn	0	5.31	5.31
20b	17	Cbz-Ala-Ala-Pro-β-Ala-Trp-OBn	30.60	30.19	0.41
20c	18	Cbz-Ala-Ala-Pro-Gaba-Trp-OBn	20.70	14.29	6.41

[a] The transition state size and amino acid code for the N-terminal residue. [b] Experimental relative abundance of ligated peptide. NA: Test compound for model validation. [c] Predicted relative abundance of ligated peptide based on the model.



Scheme 12. Synthesis of **14 m–o** and chemical ligation to validate the statistical model.

Table 8. Chemical ligation of *N*-acyl isopeptides **14 m–o** in DMF/piperidine.

Compd	Cyclic TS size	Yield [%][a]	Relative area [%][b,c] SM (t_R)	LP (t_R)	BA (t_R)	Product characterization[d] ligated peptide (LP) LP	$[M+H]^+$[e]
14 m	13	83	2.50 (48.1)	97.5 (57.1)	0.00	**15 m**	656.3
14 n	14	88	1.80 (46.8)	98.2 (57.2)	0.00	**15 n**	670.3
14 o	15	87	91.65 (47.2)	8.35 (57.6)	0.00	**15 o**	684.3

[a] Total crude yield of products isolated. [b] Determined by HPLC-MS (semiquantitative). The area of ion peak resulting from the sum of the intensities of the $[M+H]^+$ and $[M+Na]^+$ ions of each compound was integrated (and corrected for starting material). [c] SM: starting material; LP: ligated peptide; BA: bisacylated product; t_R: retention time. [d] By HPLC-MS. [e] Found.

cyclic TS gets larger (16- to 18-membered cyclic TSs), the chemical ligation becomes more difficult.

Conclusions

We have demonstrated an efficient and convenient synthesis of novel *N*-acyl isopeptides containing Trp residues. The subsequent intramolecular chemical ligation through an N→N acyl migration was favored through a long-range 10- to 18-membered cyclic TS to form the native peptides. This novel methodology was achieved without using Cys/Ser/Tyr residues or an auxiliary group at the ligation site. A statistical model was generated to predict and rationalize the feasibility of ligation. The model was further supported with the synthesis and experimental ligation data. Given that there are an increasing number of studies involving the synthesis of longer peptides and isopeptides to evaluate their biological activities, we believe this new ligation approach represents a significant development in the field.

Acknowledgements

We thank the University of Florida and the Kenan foundation, and we also acknowledge King Abdulaziz University for technical and financial support of this work. We thank Dr. C. D. Hall, Mr. Wang, and G. Valuvenko for helpful discussions. G.G.P. is grateful to the European Social Fund under project 1.2.0401.09-0079 supported by graduate school "Functional materials and technologies".

Keywords: acyl migration · chemical ligation · peptides · transition states · tryptophan

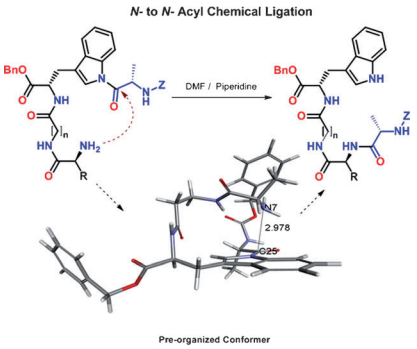
[1] T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, H. Lau, *Liebigs Ann. Chem.* **1953**, *583*, 129–149.
[2] a) J. Wilken, S. B. H. Kent, *Curr. Opin. Biotechnol.* **1998**, *9*, 412–426; b) G. G. Kochendoerfer, S. B. H. Kent, *Curr. Opin. Chem. Biol.* **1999**, *3*, 665–671.
[3] G. A. Lemieux, C. R. Bertozzi, *Trends Biotechnol.* **1998**, *16*, 506–513.
[4] T. Nakamura, A. Shigenaga, K. Sato, Y. Tsuda, K. Sakamoto, A. Otaka, *Chem. Commun.* **2014**, *50*, 58–60.
[5] B. Dang, T. Kubota, K. Mandal, F. Bezanilla, S. B. H. Kent, *J. Am. Chem. Soc.* **2013**, *135*, 11911–11919.
[6] P. W. R. Harris, M. A. Brimble, *Biopolymers* **2009**, *94*, 542–550.
[7] D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554–6559.
[8] A. W. Bosman, H. M. Jansen, E. W. Meijer, *Chem. Rev.* **1999**, *99*, 1665–1688.
[9] B. L. Nilsson, L. L. Kiessling, R. T. Raines, *Org. Lett.* **2000**, *2*, 1939–1941.
[10] B. L. Nilsson, L. L. Kiessling, R. T. Raines, *Org. Lett.* **2001**, *3*, 9–12.
[11] J. A. Restituyo, L. R. Comstock, S. G. Petersen, T. Stringfellow, S. R. Rajski, *Org. Lett.* **2003**, *5*, 4357–4360.
[12] H. Hojo, C. Ozawa, H. Katayama, A. Ueki, Y. Nakahara, Y. Nakahara, *Angew. Chem.* **2010**, *122*, 5446–5449; *Angew. Chem. Int. Ed.* **2010**, *49*, 5318–5321.
[13] D. Macmillan, D. W. Anderson, *Org. Lett.* **2004**, *6*, 4659–4662.
[14] T. Kawakami, S. Aimoto, *Tetrahedron Lett.* **2003**, *44*, 6059–6061.
[15] J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, *124*, 4642–4646.
[16] D. Crich, A. Banerjee, *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.
[17] C. Haase, H. Rohde, O. Seitz, *Angew. Chem.* **2008**, *120*, 6912–6915; *Angew. Chem. Int. Ed.* **2008**, *47*, 6807–6810.
[18] J. Chen, Q. Wan, Y. Yuan, J. Zhu, S. J. Danishefsky, *Angew. Chem.* **2008**, *120*, 8649–8652; *Angew. Chem. Int. Ed.* **2008**, *47*, 8521–8524.
[19] A. Brik, Y.-Y. Yang, S. Ficht, C.-H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 5626–5627.
[20] S. Ficht, R. J. Payne, A. Brik, C.-H. Wong, *Angew. Chem.* **2007**, *119*, 6079–6083; *Angew. Chem. Int. Ed.* **2007**, *46*, 5975–5979.
[21] R. J. Payne, S. Ficht, S. Tang, A. Brik, Y.-Y. Yang, D. A. Case, C.-H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 13527–13536.
[22] R. J. Payne, S. Ficht, W. A. Greenberg, C.-H. Wong, *Angew. Chem.* **2008**, *120*, 4483–4487; *Angew. Chem. Int. Ed.* **2008**, *47*, 4411–4415.
[23] E. Saxon, J. I. Armstrong, C. R. Bertozzi, *Org. Lett.* **2000**, *2*, 2141–2143.
[24] A. R. Katritzky, N. E. Abo-Dya, S. R. Tala, Z. K. Abdel-Samii, *Org. Biomol. Chem.* **2010**, *8*, 2316–2319.
[25] A. R. Katritzky, S. R. Tala, N. E. Abo-Dya, T. S. Ibrahim, S. A. El-Feky, K. Gyanda, K. M. Pandya, *J. Org. Chem.* **2011**, *76*, 85–96.

[26] J.-C. M. Monbaliuab, A. R. Katritzky, *Chem. Commun.* **2012**, *48*, 11601 – 11622.
[27] A. A. Oliferenko, A. R. Katritzky, *Org. Biomol. Chem.* **2011**, *9*, 4756 – 4759.
[28] K. Ha, M. Chahar, J.-C. M. Monbaliu, E. Todadze, F. K. Hansen, A. A. Oliferenko, C. E. Ocampo, D. Leino, A. Lillicotch, C. V. Stevens, A. R. Katritzky, *J. Org. Chem.* **2012**, *77*, 2637 – 2648.
[29] M. El Khatib, M. Elagawany, F. Jabeen, E. Todadze, O. Bol'shakov, A. Oliferenko, L. Khelashvili, S. A. El-Feky, A. Asiri, A. R. Katritzky, *Org. Biomol. Chem.* **2012**, *10*, 4836 – 4838.
[30] V. Popov, S. S. Panda, A. R. Katritzky, *J. Org. Chem.* **2013**, *78*, 7455 – 7461.
[31] V. Popov, S. S. Panda, A. R. Katritzky, *Org. Biomol. Chem.* **2013**, *11*, 1594 – 1597.
[32] S. Biswas, N. E. Abo-Dya, A. Oliferenko, A. Khiabani, P. J. Steel, K. A. Alamry, A. R. Katritzky, *J. Org. Chem.* **2013**, *78*, 8502 – 8509.
[33] N. E. Abo-Dya, S. Biswas, A. Basak, I. Avan, K. A. Alamry, A. R. Katritzky, *J. Org. Chem.* **2013**, *78*, 3541 – 3552.
[34] S. Biswas, I. Avan, A. K. Basak, N. E. Abo-Dya, A. Asiri, A. R. Katritzky, *Amino Acids* **2013**, *45*, 159 – 170.
[35] B. E.-D. M. El-Gendy, E. H. Ghazvini Zadeh, A. C. Sotuyo, G. G. Pillai, A. R. Katritzky, *Chem. Biol. Drug Des.* **2013**, *81*, 577 – 582.
[36] K. Ha, I. Lebedyeva, Z. Li, K. Martin, B. Williams, E. Faby, A. Nasajpour, G. G. Pillai, A. O. Al-Youbi, A. R. Katritzky, *J. Org. Chem.* **2013**, *78*, 8510 – 8523.
[37] J. J. Gajewski, K. E. Gilbert, J. McKelvey, *Advances in Molecular Modelling*, JAI, Greenwich, CT, **1990**.
[38] PCModel v9.3, Serena Software, **2011**, K. E. Gilbert, **2011**.
[39] a) A. T. Balaban, *Chem. Phys. Lett.* **1982**, *89*, 399; b) E. Estrada, Y. Gutierrez, *Commun. Math. Comput. Sci.* **2001**, *44*, 455 – 467.
[40] M. Levy, *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 6062 – 6065.
[41] E. Cartier, P. Pfluger, *Phys. Rev. B* **1986**, *34*, 8822 – 8827.
[42] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571 – 2577.
[43] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165 – 169.
[44] R. A. Bachorz, F. A. Bischoff, S. Höfener, W. Klopper, P. Ottiger, R. Leist, J. A. Frey, S. Leutwyler, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2758 – 2766.
[45] W. M. Berhanu, G. G. Pillai, A. A. Oliferenko, A. R. Katritzky, *ChemPlusChem* **2012**, *77*, 507 – 517.
[46] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, *J. Comput. Chem.* **2013**, *34*, 2121 – 2132.

Received: January 12, 2014
Published online on ■■ ■■, 0000

FULL PAPER

Going the distance: Chemical ligations to form native peptides from N→N acyl migrations in Trp-containing peptides via 10- to 18-membered cyclic transition states are described (see scheme; Bn: benzyl; Z: benzyloxycarbonyl). The ligations are achieved without the use of Cys/Ser/Tyr residues or an auxiliary group. This new approach represents a significant development in the field.



Peptide Synthesis

*S. Biswas, R. Kayaleh, G. G. Pillai, C. Seon, I. Roberts, V. Popov, K. A. Alamry, A. R. Katritzky**

■■–■■

Long-Range Chemical Ligation from N→N Acyl Migrations in Tryptophan Peptides via Cyclic Transition States of 10- to 18-Members