

Cite this: *Chem. Sci.*, 2024, 15, 1297

All publication charges for this article have been paid for by the Royal Society of Chemistry

# Rapid building block-economic synthesis of long, multi-O-GalNAcyated MUC5AC tandem repeat peptides†

Arseniy Galashov,<sup>a</sup> Ekaterina Kazakova,<sup>a</sup> Christian E. Stieger,<sup>b</sup> Christian P. R. Hackenberger<sup>ab</sup> and Oliver Seitz<sup>ab\*</sup>

The study of mucin function requires access to highly O-glycosylated peptides with multiple tandem repeats. Solid-phase synthesis would be a suitable method, however, the central problem in the synthesis of mucin glycopeptides is the need to use precious and potentially vulnerable glycoamino acid building blocks in excess. In this article, we report the development of a method based on SPPS and native chemical ligation/desulfurization chemistry that allows the rapid, reliable, and glyco-economical synthesis of long multi-O-GalNAcyated peptides. To facilitate access to the glycosyl donor required for the preparation of Fmoc-Ser/Thr( $\alpha$ Ac<sub>3</sub>GalNAc)-OH we used an easily scalable azidophenylselenenylation of galactal instead of azidonitration. The problem of low yield when coupling glycoamino acids in small excess was solved by carrying out the reactions in 2-MeTHF instead of DMF and using DIC/Oxyma. Remarkably, quantitative coupling was achieved within 10 minutes using only 1.5 equivalents of glycoamino acid. The method does not require (microwave) heating, thus avoiding side reactions such as acetyl transfer to the N-terminal amino acid. This method also improved the difficult coupling of glycoamino acid to the hydrazine-resin and furnished peptides carrying 10 GalNAc units in high purities (>95%) of crude products. Combined with a one-pot method involving native chemical ligation at a glycoamino acid junction and superfast desulfurization, the method yielded highly pure MUC5AC glycopeptides comprising 10 octapeptide tandem repeats with 20  $\alpha$ -O-linked GalNAc residues within a week.

Received 22nd September 2023  
Accepted 18th December 2023

DOI: 10.1039/d3sc05006h

rsc.li/chemical-science

## Introduction

Mucins are the major constituent of the mucus layer that protects the epithelium from direct contact with the external environment.<sup>1–3</sup> A characteristic hallmark of mucins is the presence of highly O-glycosylated multiple tandem repeats (TRs). For example, MUC5AC contains approximately 240 copies of a TTSTTSXP motif, with Thr and Ala being the most frequent amino acids in position X (Fig. 1A).<sup>4</sup> Each Ser or Thr residue can occur in O-glycosylated form. In diseases, glycosylation patterns and the number of TRs can change dramatically.<sup>5</sup> For example, malignancies are typically accompanied by increased occurrence of the Tn ( $\alpha$ GalNAc), T (Gal $\beta$ 1-3GalNAc $\alpha$ 1) and sialyl-Tn antigen structures.<sup>6,7</sup> Therefore, this region is of the highest interest to researchers looking for ways to decipher the functional consequences.

Due to difficulties with the cloning of extended repeated sequences and the microheterogeneity of native glycans, conventional methods based on genetic engineering cannot provide access to mucin TR regions of defined length and glycan composition. Instead, chemical synthesis has been used to provide glyco-defined mucins, most frequently for applications in immunological research.<sup>8–10</sup> Many research groups have contributed to the advancement of the synthesis methodology.<sup>11</sup> While N-linked glycopeptides can be prepared by connecting elaborated carbohydrates with the full-length peptide, no such



Fig. 1 (A) Tandem repeat motif (TTSTTSXP) for MUC5AC. (B) Fmoc-Ser/Thr( $\alpha$ Ac<sub>3</sub>GalNAc)-OH building blocks.

<sup>a</sup>Department of Chemistry, Humboldt-Universität zu Berlin, Brook-Taylor-Straße 2, 12489, Berlin, Germany. E-mail: oliver.seitz@chemie.hu-berlin.de

<sup>b</sup>Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP), Robert-Rössle-Strasse 10, 13125, Berlin, Germany

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3sc05006h>

approach is feasible for the synthesis of long mucin peptides. Instead, the synthesis of *O*-glycopeptides relies on preformed glycoamino acids, with Fmoc-Ser/Thr( $\alpha$ Ac<sub>3</sub>GalNAc) as the most frequently used building blocks (Fig. 1B).

Synthesis of *O*-glycopeptides is challenging. Couplings to and with glycoamino acids are more difficult than with conventional amino acid building blocks. For one thing, reactivity is low due to steric requirements. For another, glycoamino acids run the risk of racemization or  $\beta$ -elimination reactions under forced coupling conditions.<sup>12,13</sup>

Most of the previous work yielded *O*-glycopeptides that rarely contained more than 20 amino acids and only a few glycan units, some of which were of impressive complexity.<sup>14–25</sup> However, the inherent properties of mucins, such as viscoelasticity, gelation, shielding, and multivalency, emerge from multirepeat TRs.<sup>26–28</sup> We, therefore, embarked on a research program focusing on the development of methods for the synthesis of multi TR mucins. In the synthesis of long mucin peptides containing many GalNAc units, the difficulties of *O*-glycopeptide synthesis become even more apparent. A typical approach to get around difficult steps in solid-phase peptide synthesis is to use the building blocks in large excess and employ double couplings. Even with commercially available Fmoc-Ser/Thr( $\alpha$ Ac<sub>3</sub>GalNAc), the easiest glycoamino acids to synthesize, such an approach is extremely costly if a large number of glycosylated building blocks must be incorporated. Notable contributions to the synthesis of highly GalNAcylated MUC1 and MUC2 TRs, with up to 141 amino acids in length, have been reported.<sup>29–32</sup> Middle-sized segments (<40 amino acids) containing up to 7 GalNAc units were synthesized on solid phase either by single amino acid extension or fragment condensation. Purified fragments were subsequently stitched together by a series of silver-assisted thioester ligations<sup>29</sup> or Cu-click reactions.<sup>30</sup> Recently, Wang<sup>31</sup> described three native chemical ligation (NCL) reactions used to convergently assemble a 240 amino acid antifreeze protein from two 30-mers containing none or 10 GalNAc residues. A common theme of the solid-phase synthesis tactics used to access fragments for ligation is that Fmoc-Thr( $\alpha$ Ac<sub>3</sub>GalNAc) building blocks were coupled by using a twofold excess.<sup>30–32</sup> Considering the reported difficulties,<sup>12</sup> this number of excess equivalents is small. Possible truncations, side reactions, and purity of the products prior to purification have not been reported.

Herein, we describe the development of an efficient and economically viable method for the smooth synthesis of highly *O*-GalNAcylated mucin multi tandem repeats. We report an optimized multigram synthesis of both Fmoc-Thr( $\alpha$ Ac<sub>3</sub>GalNAc) and Fmoc-Ser( $\alpha$ Ac<sub>3</sub>GalNAc) based on azidophenylselenylation of galactal.<sup>33</sup> During the optimization of glycoamino acid coupling, we observed acetyl transfer reactions resulting in the undesired capping of the growing peptide chain. Most surprising were the results from a thorough analysis of coupling conditions, which led to a method enabling quantitative couplings within 10 minutes by using only 1.5 equivalents of glycoamino acid. This method also facilitated the notoriously unreliable loading of the hydrazine resin required for the synthesis of peptide thioesters according to the methods of Liu

and Dawson.<sup>34,35</sup> Combined with an extremely efficient ligation–desulfurization sequence the new method enabled the smooth synthesis of 10 TR long MUC5AC peptides containing 20 GalNAc residues on Ser or Thr.

## Results and discussion

### General considerations

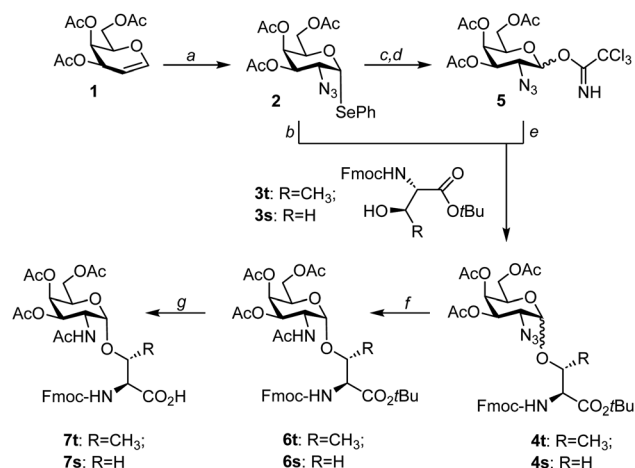
The arguably most robust access to long peptides is provided by native chemical ligation.<sup>36–39</sup> MUC5AC contains multiple copies of a Thr- and Ser-rich octapeptide TR, with TTSTTSXP as a typical representative.<sup>4</sup> Considering the absence of cysteine and the occurrence of Ala as a frequently occurring amino acid at position X, we targeted a Ser-Cys junction, which can be converted to the native Ser-Ala junction by desulfurization. In our retrosynthetic analysis, we planned the solid-phase synthesis of 40-mers comprising GalNAc residues at 25% of the amino acids. Of particular interest for us was to explore NCL with thioesters containing glycoamino acid at the C-terminus. The resulting need for substantial amounts of Fmoc-Thr/Ser( $\alpha$ Ac<sub>3</sub>GalNAc) building blocks prompted us to develop both a reliable and affordable pipeline for their preparation and their economic use in peptide chemistry.

### Multigram-scale preparation of glycoamino acids

A range of different approaches to the synthesis of Tn antigen has been described in the literature.<sup>40–46</sup> Usually, peracetylated galactal serves as a precursor for a 2-azido-2-deoxygalactosyl donor, which is required for achieving  $\alpha$ -selectivity in the glycosylation of serine or threonine. The introduction of the azido group is commonly performed by azidonitration with ceric ammonium nitrate. This reaction yields an anomeric nitrate, which is converted to galactosyl acetate or halide. Azidonitration of galactal is cumbersome, often low yielding, and as Ratcliffe and Lemieux stated: “it must be realized that hazardous conditions may be encountered.”<sup>47</sup> In our labs, azidonitration was a bottleneck, and we therefore opted to explore alternative chemistries. The Nifantiev group reported a safe, reliable, and high-yielding homogeneous azidophenylselenylation of glycals.<sup>33</sup> The formed selenoglycosides have been used for 1,2-*cis*-selective glycosylation reactions.<sup>41,48</sup> Although it is difficult to compare the combined toxicity of ceric ammonium nitrate and sodium azide used in azidonitration with the anticipated toxicity of organic selenides, avoidance of sodium azide in azidophenylselenylation reduces the risk of explosion in large scale synthesis.

For the purpose of an application in the  $\alpha$ -selective glycosylation of Fmoc-threonine *tert*-butyl ester **3t**, 2-azido-2-deoxyphenylselenogalactosyl donor **2** was prepared from triacetylgalactal **1** as previously described (Scheme 1).<sup>33,48</sup> Next, this donor was allowed to react with the threonine aglycon at  $-10\text{ }^{\circ}\text{C}$  under activation with NIS and TMSOTf in a mixture of diethyl ether and dichloromethane. The 2-azido-galactosyl-threonine conjugate **4t** was formed as a 2.5:1 mixture of  $\alpha$ - and  $\beta$ -isomers (Table 1). Chromatographic purification proved challenging. Therefore, the  $\alpha/\beta$ -mixture was subjected to the





**Scheme 1** Synthesis of Fmoc-Thr(αAc<sub>3</sub>GalNAc)-OH and Fmoc-Ser(αAc<sub>3</sub>GalNAc)-OH. Conditions: (a) Se<sub>2</sub>Ph<sub>2</sub>, trimethylsilyl azide, PhI(OAc)<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>, −30 °C (55%); (b) *N*-iodosuccinimide, trimethylsilyl triflate, CH<sub>2</sub>Cl<sub>2</sub>, Et<sub>2</sub>O, 0 °C; (c) *N*-iodosuccinimide, acetone, H<sub>2</sub>O, RT (quant.); (d) CCl<sub>3</sub>CN, Cs<sub>2</sub>CO<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, RT (84%); (e) trimethylsilyl triflate, CH<sub>2</sub>Cl<sub>2</sub>, Et<sub>2</sub>O, 0 °C; (f) Zn, THF, AcOH, Ac<sub>2</sub>O, RT; (g) TFA, CH<sub>2</sub>Cl<sub>2</sub>, RT (70–75%).

zinc-mediated azide reduction in the presence of acetic anhydride furnishing the *N*-acetylgalactosamine-threonine conjugate **6t**, which was easier to purify than **4t**. After chromatography, the α-isomer **6t** was isolated in 24% yield. The same synthetic route was applied to the glycosylation of Fmoc-serine *tert*-butyl ester **3s**, providing the α-linked serine conjugate **6s** in 26% yield.

The relatively moderate yields of the glycosylation, caused by difficult separation from by-products, prompted us to consider an alternative donor. The 2-azido-2-deoxygalactosyl trichloroacetimidate **5** was readily prepared from selenogalactoside **2** in 84% yield in two steps. First, **2** was quantitatively hydrolyzed to hemiacetal **S1** (Scheme S1†), followed by the known reaction with trichloroacetimidate. Glycosylation of amino acid **3t** was induced upon the addition of trimethylsilyl triflate. With a 2.3 : 1 ratio of α/β-anomers, this reaction provided comparatively the same α-selectivity as the glycosylation with the selenophenyl donor **2** but almost twice as high yields (Table 1). Importantly, the outcome was independent of the reaction scale (100 mg to 9.3 g). To facilitate the purification of **6** after reductive acetylation, it proved beneficial to adjust the donor/acceptor ratios. A 1 : 1.1 ratio allowed the isolation of Fmoc-Thr(αAc<sub>3</sub>GalNAc)-O<sup>t</sup>Bu in 47% yield over two steps. For

the synthesis of Fmoc-Ser(αAc<sub>3</sub>GalNAc)-O<sup>t</sup>Bu, the donor/acceptor ratio was adjusted to 1.2 : 1. The α-selectivity and the yields were slightly higher than for glycosylation of the threonine derivative. Cleavage of the *tert*-butyl ester was uneventful and furnished the glycoamino acid building blocks **7t** and **7s** on a multigram scale.

### Loading of (glyco)amino acid to the NH<sub>2</sub>-NH-TRT resin

The synthesis of glycopeptide thioesters used in NCL chemistry is most easily carried out on hydrazine resins.<sup>50,51</sup> The glycopeptide hydrazides obtained after TFA cleavage are converted to glycopeptide thioesters under mildly acidic conditions.<sup>55</sup> The retrosynthetic analysis called for a loading of the H<sub>2</sub>N-NH-trityl resin with Fmoc-Ser(O<sup>t</sup>Bu)-OH or Fmoc-Ser(αAc<sub>3</sub>GalNAc)-OH. In our laboratories, loading this resin led to notoriously inconsistent results. While problematic couplings to N-terminal azapeptides have been described in a recent review,<sup>52</sup> difficulties with the H<sub>2</sub>N-NH-trityl resin have apparently not been reported. Usually, loading is performed with HATU/DiPEA, although similar methods have been reported with HBTU,<sup>53</sup> HCTU,<sup>54</sup> and PyAOP.<sup>55</sup>

To evaluate various coupling conditions, we prepared a Tentagel R Fmoc-NH-NH-TRT resin from Tentagel R Cl-TRT resin.<sup>56</sup> This low loading resin was used because of its known advantages in the synthesis of long and difficult peptides. However, instead of suspending Fmoc-NH-NH<sub>2</sub> in DMF/CH<sub>2</sub>Cl<sub>2</sub>, we used anhydrous DMSO as a solvent. The Fmoc-load was determined before Fmoc cleavage and after subsequent coupling of the starting amino acid. Conventional double coupling of Fmoc-Ser(O<sup>t</sup>Bu)-OH (10 eq.) under activation with HATU in the presence of DiPEA (condition 1, Table 2) afforded only 49% loading

**Table 2** Coupling systems for loading of Fmoc-Ser(O<sup>t</sup>Bu)-OH to the NH<sub>2</sub>-NH-TRT resin

| Coupling conditions <sup>a</sup>              | Yield <sup>b</sup> (%) |
|---|------------------------|
| (1) 2 × HATU/DiPEA, 45 min in DMF, rt         | 49                     |
| (2) 2 × HATU/DiPEA, 30 min each in DMF, 50 °C | 26                     |
| (3) 3 × PyOxym/DiPEA, 5 min in DMF, 75 °C     | 25                     |
| (4) 1 × DIC/Oxyma/DiPEA, 17 h in DMF, rt      | 63                     |
| (5) 1 × DIC/Oxyma, 5 min in 2-MeTHF, rt       | 92                     |
| (6) 1 × DIC/Oxyma, 10 min in 2-MeTHF, rt      | 100                    |

<sup>a</sup> 10 eq. of amino acid and activating reagents were used, 20 eq. of DiPEA was used. In condition **4** 1 eq. of DiPEA was added. <sup>b</sup> Yields calculated according to Fmoc-test. In conditions with 2-MeTHF preactivation (2 min) was used.

**Table 1** Approaches to the glycosylation of Fmoc-Thr(OH)-O<sup>t</sup>Bu and Fmoc-Ser(OH)-O<sup>t</sup>Bu

| Donor    | Acceptor  | α : β ratio after glycosylation <sup>a</sup> | Yield for two steps (glycosylation + azide reduction), <sup>b</sup> % |
|----------|-----------|--|---|
| <b>2</b> | <b>3t</b> | 2.5 : 1                                      | 24  |
| <b>2</b> | <b>3s</b> | 2.7 : 1                                      | 26  |
| <b>5</b> | <b>3t</b> | 2.3 : 1                                      | 47 <sup>c</sup>   |
| <b>5</b> | <b>3</b>  | 2.7 : 1                                      | 51 <sup>c</sup>   |

<sup>a</sup> Ratio was determined by NMR of crude. <sup>b</sup> Only for α-isomer. <sup>c</sup> Gram-scale reactions yields are shown.

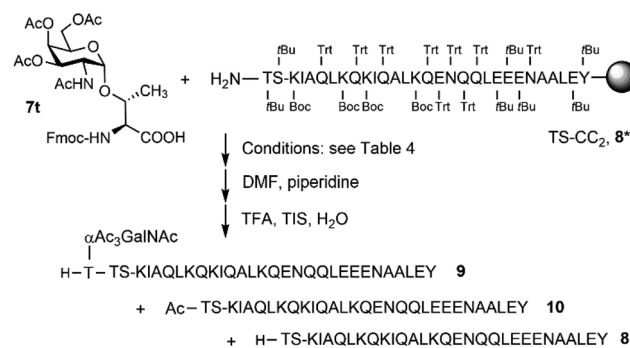


yield. Microwave heating did not remedy the situation (condition 2). To avoid the possible formation of guanidino adducts, PyOxim was used as an activator with little success (condition 3). A higher loading yield was achieved when a coupling with DIC/Oxyma activation was performed overnight (condition 4). In our further attempts, we were inspired by the work of de la Torre/Albericchio<sup>57</sup> and the Schönleber/Pedersen groups,<sup>58</sup> who used 2-MeTHF as a solvent for coupling reactions promoted by DIC/Oxyma. Replacing DMF with 2-MeTHF dramatically improved reaction rates (conditions 5 and 6). Quantitative yields were obtained after 10 min reaction time.

Next, we investigated the non-trivial loading of the H<sub>2</sub>N-NH-TRT resin with Fmoc-Ser( $\alpha$ Ac<sub>3</sub>GalNAc)-OH (Table 3). Given its high cost, coupling with 10 equivalents is not an option. Gratifyingly, with the DIC/Oxyma system in 2-MeTHF, a 75% loading yield was obtained by using only 1 eq. glycoamino acid (condition 1, Table 3). Increasing the coupling time from 2 to 20 hours did not improve the yield (condition 2), but with only 0.5 excess equivalents, >90% yield was achieved (condition 3), which was deemed sufficient in terms of yield and building block economy.

### Optimizing coupling of glycoamino acid to peptide chain

Microwave-assisted SPPS is a powerful tool that can facilitate the synthesis of aggregating peptides and dramatically improve the yield of synthesized peptide while reducing the time required for synthesis.<sup>59</sup> It has been reported that the method has the potential to improve the synthesis of *O*-glycosylated mucin and antifreeze peptides.<sup>60–62</sup> Therefore, we investigated the coupling of Fmoc-Thr( $\alpha$ Ac<sub>3</sub>GalNAc)-OH onto a resin-bound peptide under microwave conditions. Due to our interest in using coiled coil interactions for the assembly of large glycopeptide structures, we explored the extension of a slightly modified coiled coil peptide<sup>63</sup> TS-CC<sub>2</sub>, **8\*** (Scheme 2). This peptide was prepared according to a recently described method<sup>64</sup> involving Fmoc-deprotection for 1 minute at 90 °C with 20% piperidine in DMF and coupling of amino acids with DIC/Oxyma for 2 minutes at 90 °C. HPLC-MS analysis of the crude material obtained upon TFA cleavage (Fig. 2A) confirmed the efficiency of this synthesis protocol, which provided the 30-mer peptide in very high purity. For coupling of **7t**, we extended the coupling time to 5 min (condition 1, Table 4). Spectrometric monitoring of Fmoc loads suggested quantitative coupling yields. Furthermore, the HPLC trace of the crude material from the TFA cleavage indicated smooth coupling, as inferred from



Scheme 2 Optimizing coupling of Fmoc-Thr( $\alpha$ Ac<sub>3</sub>GalNAc)-OH (**7t**).

peak-to-peak conversion (Fig. 2B). However, MS analysis showed that the extension of **8\*** with Fmoc-Thr( $\alpha$ Ac<sub>3</sub>GalNAc)-OH led to the formation of two products. The *m/z*-ratios correspond to the desired extension product **9** and a truncation product **10** obtained by acetylation of **8\***. Since no capping was used, the acetyl group must be derived from the glycoamino acid. Aminolysis of acetic acid esters by the N-terminal amino group is probably facilitated by microwave heating. This and the difficulty in separating the two products prompted us to explore alternative coupling conditions.

In search of (glyco) efficient coupling, we reviewed conditions based on the use of HATU as one of the strongest activation reagents.<sup>65</sup> We used DMF as a solvent and varied the base, reaction time, and temperature (Table 4). Again, couplings were carried out at the resin-bound 30-mer peptide **8\***, and again, rather than simply determining the Fmoc load, we monitored the reactions by HPLC analysis of crudes obtained after TFA cleavage (Fig. S5†). With only 1.5 equivalents of glycoamino acid/HATU/HOAt, the reaction required 90 minutes to afford a 92% yield (conditions 9). Based on a report from the Gildersleeve lab on the epimerization of glycoamino acids with DIPEA<sup>12</sup> and slightly slower couplings (compare conditions 2

Table 3 Optimizing loading of NH<sub>2</sub>-NH-TRT resin with Fmoc-Ser( $\alpha$ Ac<sub>3</sub>GalNAc)-OH (**7s**)<sup>a</sup>

| Coupling conditions  | Yield <sup>b</sup> (%) |
|--|------------------------|
| (1) 1 eq. <b>7s</b> /DIC/Oxyma, 2 h in 2-MeTHF, rt           | 75                     |
| (2) 1 eq. <b>7s</b> /DIC, 10 eq. Oxyma, 20 h in 2-MeTHF, rt  | 71                     |
| (3) 1.5 eq. <b>7s</b> /DIC, 15 eq. Oxyma, 2 h in 2-MeTHF, rt | 92                     |

<sup>a</sup> AA = Fmoc-Ser( $\alpha$ Ac<sub>3</sub>GalNAc)-OH. <sup>b</sup> Yields calculated according to Fmoc-test.

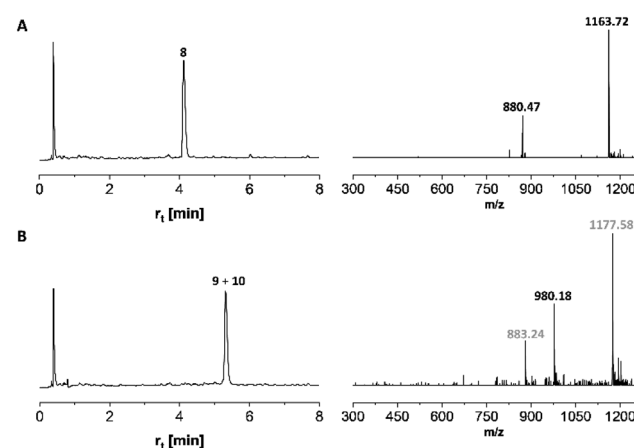


Fig. 2 UPLC and ESI-MS analysis of (A) peptide **8** cleaved from **8\*** before and (B) after coupling of glycoamino acid **7t**. Gray numbers correspond to mass peaks of peptide **10**. Gradient 18–30% ACN in water + 0.1% TFA.  $\lambda$  = 210 nm.



**Table 4** Optimizing the coupling of Fmoc-Thr( $\alpha$ Ac<sub>3</sub>GalNAc)-OH (**7t**) acid to the resin-bound peptide **8\*** in DMF

| Coupling conditions  | Yield <sup>a</sup> (%) |
|--|------------------------|
| (1) 1.5 eq. <b>7t</b> /DIC/Oxyma, 5 min, 75 °C                       | NA                     |
| (2) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. DiPEA, 30 min, rt            | 75                     |
| (3) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. DiPEA, 60 min, rt            | 78                     |
| (4) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. TMP, 30 min, rt              | 70                     |
| (5) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. TMP, 60 min, rt              | 82                     |
| (6) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. TMP, 90 min, rt              | 82                     |
| (7) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. TMP, 30 min, rt              | 82                     |
| (8) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. TMP, 60 min, rt              | 90                     |
| (9) 1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. TMP, 90 min, rt              | 92                     |
| (10) 2.25 eq. <b>7t</b> /HATU/HOAt, 4.5 eq. TMP, 90 min, rt          | 94                     |
| (11) 2 × [0.95 eq. <b>7t</b> /HATU/HOAt, 1.9 eq. TMP, 30 min, 40 °C] | 89                     |
| (12) 2 × [1.15 eq. <b>7t</b> /HATU/HOAt, 2.3 eq. TMP, 30 min, 40 °C] | 93                     |
| (13) 2 × [1.3 eq. <b>7t</b> /HATU/HOAt, 2.6 eq. TMP, 30 min, 40 °C]  | 97                     |
| (14) 2 × [1.5 eq. <b>7t</b> /HATU/HOAt, 3 eq. TMP, 18 min, rt]       | 100                    |

<sup>a</sup> Yields calculated by integrating peaks in UPLC traces. Based on four independent replicates performed for condition 10, a 4% standard deviation is estimated. TMP, 2,4,6-trimethylpiperidine.

and 3 with 4 and 5), we focused on the use of TMP as a base. Increasing the excess from 1.5 to 2.25 equivalents at a 90 minute reaction time provided only a minor improvement in coupling yields (conditions 10). With an aim to reduce coupling time, we investigated mild microwave irradiation (25 W) and increased the temperature to 40 °C (conditions 11–13). At the same time, double coupling (18–30 min each) was performed with a total of 1.9–3 equivalents. At this temperature, we did not observe acetyl transfer. A double coupling for 18 min with 1.5 equivalents each provided a quantitative yield (condition 14). However, the need for 3 equivalents in each glycoamino acid coupling is costly.

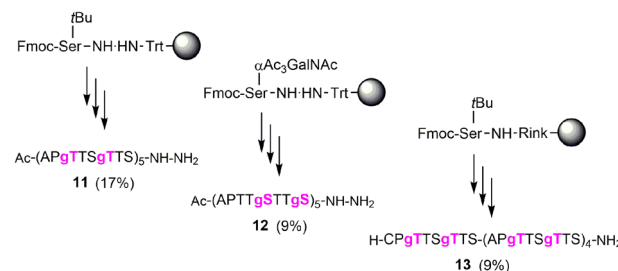
Motivated by success with difficult couplings to the hydrazine resin, we explored reactions in 2-MeTHF with activation by the DIC/Oxyma system (Table 5). In our first attempt, a double coupling with substoichiometric amounts of glycoamino acid was carried out with little success. Remarkably, a single coupling with 1.5 equivalents of glycoamino acid in 2-MeTHF for 10 min proved sufficient to give a quantitative yield. No heating was required, and there was no evidence of acetyl transfer (Fig. S6†).

The highly efficient coupling of Fmoc-Ser/Thr( $\alpha$ Ac<sub>3</sub>GalNAc) in 2-Me-THF is noteworthy. Research by de la Torre and Albericchio has shown that coupling reactions with DIC/Oxyma activation proceed remarkably swiftly in 2-MeTHF, whereas DMF seemed to be a better solvent for activation with HATU or HBTU.<sup>57</sup> There is an increasing amount of evidence to suggest that the rate of coupling reactions increases with decreased polarity of the solvent.<sup>58,66</sup> We assume that the high reactivity in

apolar solvents can counterbalance the reduction of reaction rate caused by low concentration of the building block. Hence, apolar solvents such as 2-MeTHF could prove particularly advantageous when expenses for building blocks are high.

### Solid-phase synthesis of glycopeptides

Next, we embarked on the solid phase synthesis of 5 TRs long MUC5AC peptides that can be used in NCL chemistry (Scheme 3). Couplings of the glycoamino acids were performed in 2-MeTHF with 1.5 eq. of glycoamino acid, 1.5 eq. of DIC, and 1.5 eq. of Oxyma. The other amino acids were introduced in DMF



**Scheme 3** Solid-phase synthesis of MUC5AC glycopeptides and their yields after isolation in parentheses. gS, gT = Ser/Thr( $\alpha$ GalNAc). Conditions: Fmoc-cleavage: 20% piperidine in DMF; coupling: HATU/Oxyma/DiPEA (5 eq., 5 eq., 15 eq.); coupling of **7t** or **7s**: see condition 3, Table 5; capping: Ac<sub>2</sub>O/DiPEA/DMF (70 : 20 : 10, v/v/v); TFA cleavage: for **11** and **12** TFA/TIS/H<sub>2</sub>O (96/2/2). For **13** TFA/TIS/EDT/H<sub>2</sub>O (94/2/2).

**Table 5** Coupling of Fmoc-Thr( $\alpha$ Ac<sub>3</sub>GalNAc)-OH (**7t**) in 2-MeTHF to peptide **1**

| Coupling conditions   | Yield <sup>a</sup> (%) |
|---|------------------------|
| (1) 2 × [0.78 eq. <b>7t</b> /DIC, 10 eq. Oxyma in 2-Me-THF, 18 min, rt] | 56                     |
| (2) 1.09 eq. <b>7t</b> /DIC/Oxyma in 2-MeTHF, 10 min, rt                | 94                     |
| (3) 1.5 eq. <b>7t</b> /DIC/Oxyma in 2-MeTHF, 10 min, rt                 | 100                    |

<sup>a</sup> Yields calculated by integrating peaks in UPLC chromatograms. For condition 1 double coupling was performed.



solutions by using 5 eq. of amino acids and activation with HATU/Oxyma in the presence of DiPEA. In test cleavages of glycopeptides, we observed that *O*-acetyl-protected glycopeptides are unstable in aqueous acidic solution. The glycan residues lose acetyl protecting groups, resulting in complicated HPLC profiles (Fig. S7†). If necessary, this side reaction can be suppressed by dissolving the glycopeptides in buffer solutions at pH 7. Preferably, however, the acetyl protecting groups were removed prior to cleavage by treating the resin with a  $N_2H_4 \cdot H_2O$ /DMF mixture. Under these conditions, the synthesis of the 40-mer peptide hydrazides **11** and **12**, containing 10 glycosylated amino acids, proceeded remarkably smoothly, as evidenced by the high purity of crude material (Fig. 3A and B). The situation was somewhat different for the synthesis of glycopeptide **13**, which contains an N-terminal

cysteine for NCL. Here, the crude material showed a few by-products that were removed by preparative HPLC (Fig. 3C).

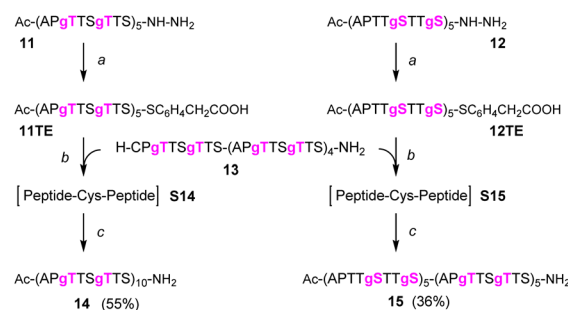
MS analysis of the glycopeptides showed peaks with *m/z* values (−200 Da and −400 Da) corresponding to a loss of one or two GalNAc residues (Fig. 3D–F). This loss did not occur when glycopeptides were detached in *O*-acetylated form (Fig. S8†). Hydrolytic cleavage of sugar residues could occur during detachment with TFA. We therefore prepared glycopeptide **12** with *O*-acetyl groups on the GalNAc residues (which is known to increase their hydrolytic stability) and performed the *O*-deacetylation in solution. The mass spectra did not change (Fig. S9†). Therefore, we attributed the loss of GalNAc-residues to the conditions of the ESI-MS measurement. In the absence of readily ionizable side chains, protonation may occur also at the sugar, followed by possible cleavage of the glycosidic bond. Most importantly, this fragmentation reaction can be prevented by altering the conditions of ESI-MS (*vide infra*).

### One-pot native chemical ligation/desulfurization of glycopeptides

We continued the synthesis with an NCL reaction between glycopeptides **11** and **13** (Scheme 4). First, **11** was converted to the peptide thioester using Dawson's Knorr-pyrazole method.<sup>34</sup> After treatment with acetylacetone in the presence of mercaptophenylacetic acid (MPAA) for 90 minutes, the peptide hydrazide was quantitatively converted to a mixture of peptide-pyrazolate and peptide-MPAA thioester, and complete conversion to the peptide-MPAA thioester was observed after a total of 150 minutes (Fig. S10†). Without purification, glycopeptide **13** was added, and the buffer was adjusted to pH = 7.1. At the unhindered Ser-Cys junction, NCL was fast and afforded 95% ligation yield after only 30 min (Fig. 4A, left). We next examined the NCL between fragments **12** and **13**. This reaction involves the formation of a Ser(GalNAc)-Cys junction, which cannot be avoided if GalNAcylation also extends to serine residues. We expected reactions at the more sterically demanding Ser(GalNAc) site to proceed slower than reactions at the C-terminal Ser. Contrarily, thioester formation at **12** proceeded faster than at **11**, requiring only 60 min instead of 150 minutes for completion



Fig. 3 UPLC analysis of glycopeptides (A) **11**, (B) **12**, and (C) **13** before (left) and after (right) HPLC purification. Gradient: 07–17% ACN in water + 0.1% TFA.  $\lambda$  = 210 nm. ESI-MS spectra of glycopeptides (D) **11**, (E) **12** and (F) **13**. Gray numbers indicate mass peaks due to fragmentation during ESI-MS measurement.



Scheme 4 NCL/desulfurization affords MUC5AC peptides with 10 TRs and 20 GalNAc moieties. (a) 1.5 eq. acac, 200 mM MPAA, 6 M Gd-HCl, pH 3.1; (b) 2.5 mM peptides, 25 mM TCEP-HCl in 6 M Gd-HCl, 0.2 M Na<sub>2</sub>HPO<sub>4</sub>, pH 7.1; (c) EtOAc extraction, then: 0.8 mM peptide, 6 M Gd-HCl, 0.2 M sodium citrate, 200 mM TCEP-HCl, 100 mM NaBEt<sub>4</sub>, pH 4.5.





Fig. 4 HPLC analysis of (A) NCL 13 + 11TE → S14 (left) or NCL 13 + 12TE → S15 (right), (B) desulfurization S14 → 14 (left) or S15 → 15 (right) and (C) purified products. 11OH and 12OH are hydrolysis products formed from corresponding thioesters. (D) ESI-HRMS analysis of purified MUC5AC peptides 14 (left) and 15 (right). *m/z* assignments for most abundant isotope.

(Fig. S11†). Likewise, NCL with glycopeptide 13 proceeded smoothly, providing 95% ligation yield after only 30 minutes (Fig. 4A, right).

The mixtures of the NCL reactions were forwarded for desulfurization. We applied a very recent method from the Li group, which has been reported to enable almost instantaneous desulfurization upon reaction with tetraethylborate.<sup>67</sup> A slight shift of retention to shorter times (compare Fig. 4A with 4B), consistent with the loss of a hydrophobic mercapto group, indicated that the reaction was completed after 1 minute. No side reactions were observed in the process of desulfurization. Of note, the apparent purity of crude products 14 and 15, both

of which carry 20 GalNAc residues on 10 MUC5AC tandem repeats, was higher than 95%. After preparative HPLC and lyophilization, the glycopeptides 14 and 15 were obtained in 55% and 36% yield, respectively. To prevent fragmentation during high-resolution ESI-MS, it proved necessary to reduce the concentration of the collision gas (Fig. 4D).

## Conclusion

We developed a glyco-economic synthesis of long and highly *O*-GalNAcylated mucin peptides. The method relies on Fmoc-SPPS and NCL/desulfurization chemistry. Several bottlenecks had to be removed during method development.

(1) We introduced a facilitated, more reliable, and easily scalable synthesis of Fmoc-Thr/Ser( $\alpha$ Ac<sub>3</sub>GalNAc)-OH, in which the potentially hazardous azidonitration of tri-*O*-acetyl galactal was replaced with azidophenylselenylation.

(2) For the challenging loading of hydrazine resin with a glycoamino acid coupling was performed in 2-MeTHF with DIC/Oxyma activation, which reliably provided higher yields than coupling in DMF.

(3) This approach also eliminated the third and probably most limiting bottleneck, namely the need for large amounts of expensive glycoamino acids in coupling reactions. By using 2-MeTHF as a solvent with DIC/Oxyma activation agents, only 1.5 equivalents of Fmoc-Thr/Ser( $\alpha$ Ac<sub>3</sub>GalNAc)-OH were sufficient to achieve quantitative coupling yields within 10 minutes of reaction time. There was no need in time-consuming double coupling procedures or potentially harmful microwave heating. Without microwave heating users of “traditional” synthesizers can perform glyco-economic yet speedy syntheses of *O*-glycopeptides. In addition, the absence of microwave heating avoids side reactions, such as the acetyl transfer reaction we observed. Remarkably high purities of crude products of 90–95% show that the new method allows very smooth solid-phase synthesis of mucin peptides encompassing up to 5 MUC5AC tandem repeats and carrying 10 GalNAc units.

(4) Native chemical ligation between two highly glycosylated peptides proceeded rapidly and cleanly also at the glycoamino acid junction.

With the presented method, solid-phase synthesis and purification of a peptide 40-mer with  $\alpha$ -*O*-linked GalNAc residues on 25% of all amino acids takes 4 days. Within 2 days, subsequent solution steps, including thioester preparation, ligation, and desulfurization, are accomplished. We believe that these advances will further mucin research by facilitating access to a greater variety of GalNAcylated peptides in a short time.

## Data availability

The datasets supporting this article have been uploaded as part of the ESI.†

## Author contributions

O. S. formulated the research goals and aims. A. G. and E. K. conducted the research. C. S. performed measurements. A. G.,



E. K. and C. S. contributed to visualization of data. O. S., A. G., E. K. contributed to the original draft. All authors contributed to reviewing and editing.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 431232613 – SFB 1449.

## References

- 1 C. E. Wagner, K. M. Wheeler and K. Ribbeck, *Annu. Rev. Cell Dev. Biol.*, 2018, **34**, 189.
- 2 A. P. Corfield, *Biochim. Biophys. Acta, Gen. Subj.*, 2015, **1850**, 236.
- 3 C. Werlang, G. Cárcarmo-Oyarce and K. Ribbeck, *Nat. Rev. Mater.*, 2019, **4**, 134.
- 4 T. Lang, *Sci. Rep.*, 2022, **12**, 20652.
- 5 C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak, *Nat. Rev. Nephrol.*, 2019, **15**, 346.
- 6 C. Fu, H. Zhao, Y. Wang, H. Cai, Y. Xiao, Y. Zeng and H. Chen, *HLA*, 2016, **88**, 275.
- 7 J. M. Burchell, R. Beatson, R. Graham, J. Taylor-Papadimitriou and V. Tajadura-Ortega, *Biochem. Soc. Trans.*, 2018, **46**, 779.
- 8 O. Ouerfelli, J. D. Warren, R. M. Wilson and S. J. Danishefsky, *Expert Rev. Vaccines*, 2005, **4**, 677.
- 9 N. Martínez-Sáez, J. M. Peregrina and F. Corzana, *Chem. Soc. Rev.*, 2017, **46**, 7154.
- 10 M.-M. Wei, Y.-S. Wang and X.-S. Ye, *Med. Res. Rev.*, 2018, **38**, 1003.
- 11 W. Doelman and S. I. van Kasteren, *Org. Biomol. Chem.*, 2022, **20**, 6487.
- 12 Y. Zhang, S. M. Muthana, D. Farnsworth, O. Ludek, K. Adams, J. J. Barchi Jr and J. C. Gildersleeve, *J. Am. Chem. Soc.*, 2012, **134**, 6316.
- 13 Y. Zhang, S. M. Muthana, J. J. Barchi and J. C. Gildersleeve, *Org. Lett.*, 2012, **14**, 3958.
- 14 O. Seitz and C.-H. Wong, *J. Am. Chem. Soc.*, 1997, **119**, 8766.
- 15 W. Ma, Y. Deng, Z. Xu, X. Liu, D. G. Chapla, K. W. Moremen, L. Wen and T. Li, *J. Am. Chem. Soc.*, 2022, **144**, 9057.
- 16 A. Gabba, R. Attariya, S. Behren, C. Pett, J. C. van der Horst, H. Yurugi, J. Yu, M. Urschbach, J. Sabin, G. Birrane, E. Schmitt, S. J. van Vliet, P. Besenius, U. Westerlind and P. V. Murphy, *J. Am. Chem. Soc.*, 2023, **145**, 13027.
- 17 N. J. Mitchell, S. S. Kulkarni, L. R. Malins, S. Wang and R. J. Payne, *Chem.–Eur. J.*, 2017, **23**, 946.
- 18 R. M. Wilson and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2013, **135**, 14462.
- 19 N. Takeda, T. Takei, Y. Asahina and H. Hojo, *Chem.–Eur. J.*, 2018, **24**, 2593.
- 20 N. T. Supekar, V. Lakshminarayanan, C. J. Capicciotti, A. Sirohiwal, C. S. Madsen, M. A. Wolfert, P. A. Cohen, S. J. Gendler and G.-J. Boons, *ChemBioChem*, 2018, **19**, 121.
- 21 C. Bello and C. F. W. Becker, *Biorg. Med. Chem.*, 2017, **25**, 5016.
- 22 L. R. Malins, N. J. Mitchell, S. McGowan and R. J. Payne, *Angew. Chem., Int. Ed.*, 2015, **54**, 12716.
- 23 C. Xu, H. Y. Lam, Y. Zhang and X. Li, *Chem. Commun.*, 2013, **49**, 6200.
- 24 M. Fumoto, H. Hinou, T. Ohta, T. Ito, K. Yamada, A. Takimoto, H. Kondo, H. Shimizu, T. Inazu, Y. Nakahara and S.-I. Nishimura, *J. Am. Chem. Soc.*, 2005, **127**, 11804.
- 25 C. Pett and U. Westerlind, *Chem.–Eur. J.*, 2014, **20**, 7287.
- 26 M. C. Rose and J. A. Voynow, *Physiol. Rev.*, 2006, **86**, 245.
- 27 G. W. Hughes, C. Ridley, R. Collins, A. Roseman, R. Ford and D. J. Thornton, *Sci. Rep.*, 2019, **9**, 17350.
- 28 C. R. Shurer, M. J. Colville, V. K. Gupta, S. E. Head, F. Kai, J. N. Lakins and M. J. Paszek, *ACS Biomater. Sci. Eng.*, 2018, **4**, 388.
- 29 H. Hojo, Y. Matsumoto, Y. Nakahara, E. Ito, Y. Suzuki, M. Suzuki, A. Suzuki and Y. Nakahara, *J. Am. Chem. Soc.*, 2005, **127**, 13720.
- 30 M. Galibert, V. Piller, F. Piller, V. Aucagne and A. F. Delmas, *Chem. Sci.*, 2015, **6**, 3617.
- 31 N. Ohyabu, K. Kakiya, Y. Yokoi, H. Hinou and S.-I. Nishimura, *J. Am. Chem. Soc.*, 2016, **138**, 8392.
- 32 X. Liu, J. Liu, Z. Wu, L. Chen, S. Wang and P. Wang, *Chem. Sci.*, 2019, **10**, 8694.
- 33 Y. V. Mironov, A. A. Sherman and N. E. Nifantiev, *Tetrahedron Lett.*, 2004, **45**, 9107.
- 34 D. T. Flood, J. C. J. Hintzen, M. J. Bird, P. A. Cistrone, J. S. Chen and P. E. Dawson, *Angew. Chem., Int. Ed.*, 2018, **57**, 11634.
- 35 G.-M. Fang, Y.-M. Li, F. Shen, Y.-C. Huang, J.-B. Li, Y. Lin, H.-K. Cui and L. Liu, *Angew. Chem., Int. Ed.*, 2011, **50**, 7645.
- 36 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. H. Kent, *Science*, 1994, **266**, 776.
- 37 J. D. Lin and X.-W. Liu, *Chem.–Asian J.*, 2020, **15**, 2548.
- 38 V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu and O. Melnyk, *Chem. Rev.*, 2019, **119**, 7328.
- 39 M. H. S. Marqvorsen, C. Araman and S. I. van Kasteren, *Bioconjugate Chem.*, 2019, **30**, 2715.
- 40 H. Kunz and S. Birnbach, *Angew. Chem. Int. Ed. Engl.*, 1986, **25**, 360.
- 41 T. S. Kärkkäinen, K. P. Ravindranathan Kartha, D. MacMillan and R. A. Field, *Carbohydr. Res.*, 2008, **343**, 1830.
- 42 C. Plattner, M. Höfener and N. Sewald, *Org. Lett.*, 2011, **13**, 545.
- 43 D. Straßburger, M. Glaffig, N. Stergiou, S. Bialas, P. Besenius, E. Schmitt and H. Kunz, *ChemBioChem*, 2018, **19**, 1142.
- 44 S. M. Piazza, M. R. Reynolds, J. Chiaramonte, P. Xu, F. A. Chapa-Villarreal and J. F. Trant, *New J. Chem.*, 2021, **45**, 19224.
- 45 K. M. Koeller, M. E. B. Smith and C.-H. Wong, *Bioorg. Med. Chem.*, 2000, **8**, 1017.





- 46 H. Paulsen and J.-P. Hölck, *Carbohydr. Res.*, 1982, **109**, 89.
- 47 R. U. Lemieux and R. M. Ratcliffe, *Can. J. Chem.*, 1979, **57**, 1244.
- 48 E. D. Kazakova, D. V. Yashunsky, E. A. Khatuntseva and N. E. Nifantiev, *Pure Appl. Chem.*, 2020, **92**, 1047.
- 49 R. R. Schmidt and J. Michel, *Angew Chem. Int. Ed. Engl.*, 1980, **19**, 731.
- 50 Y.-C. Huang, G.-M. Fang and L. Liu, *Natl. Sci. Rev.*, 2015, **3**, 107.
- 51 J.-S. Zheng, S. Tang, Y.-C. Huang and L. Liu, *Acc. Chem. Res.*, 2013, **46**, 2475.
- 52 R. Chingle, C. Proulx and W. D. Lubell, *Acc. Chem. Res.*, 2017, **50**, 1541.
- 53 Y. Maki, R. Okamoto, M. Izumi and Y. Kajihara, *J. Am. Chem. Soc.*, 2020, **142**, 20671.
- 54 D.-L. Huang, Y. Li, J. Liang, L. Yu, M. Xue, X.-X. Cao, B. Xiao, C.-L. Tian, L. Liu and J.-S. Zheng, *J. Am. Chem. Soc.*, 2020, **142**, 8790.
- 55 N. Hananya, S. K. Daley, J. D. Bagert and T. W. Muir, *J. Am. Chem. Soc.*, 2021, **143**, 10847.
- 56 M. J. Bird and P. E. Dawson, *Pept. Sci.*, 2022, **114**, e24268.
- 57 Y. E. Jad, G. A. Acosta, T. Govender, H. G. Kruger, A. El-Faham, B. G. de la Torre and F. Albericio, *ACS Sustain. Chem. Eng.*, 2016, **4**, 6809.
- 58 V. Martin, S. Jadhav, P. H. G. Egelund, R. Liffert, H. Johansson Castro, T. Krüger, K. F. Haselmann, S. Thordal Le Quement, F. Albericio, F. Dettner, C. Lechner, R. Schönleber and D. S. Pedersen, *Green Chem.*, 2021, **23**, 3295.
- 59 S. L. Pedersen, A. P. Tofteng, L. Malik and K. J. Jensen, *Chem. Soc. Rev.*, 2012, **41**, 1826.
- 60 C. Heggemann, C. Budke, B. Schomburg, Z. Majer, M. Wißbrock, T. Koop and N. Sewald, *Amino Acids*, 2010, **38**, 213.
- 61 T. Matsushita, H. Hinou, M. Kurogochi, H. Shimizu and S.-I. Nishimura, *Org. Lett.*, 2005, **7**, 877.
- 62 D. J. Lee, P. W. R. Harris, R. Kowalczyk, P. R. Dunbar and M. A. Brimble, *Synthesis*, 2010, **2010**, 763.
- 63 T. Lebar, D. Lainšček, E. Merljak, J. Aupič and R. Jerala, *Nat. Chem. Biol.*, 2020, **16**, 513.
- 64 J. M. Collins, K. A. Porter, S. K. Singh and G. S. Vanier, *Org. Lett.*, 2014, **16**, 940.
- 65 L. A. Carpino, *J. Am. Chem. Soc.*, 1993, **115**, 4397.
- 66 J. Lopez, S. Pletscher, A. Aemissegger, C. Bucher and F. Gallou, *Org. Process Res. Dev.*, 2018, **22**, 494.
- 67 Z. Sun, W. Ma, Y. Cao, T. Wei, X. Mo, H. Y. Chow, Y. Tan, C. H. P. Cheung, J. Liu, H. K. Lee, E. C. M. Tse, H. Liu and X. Li, *Chem*, 2022, **8**, 2542.

