Organic & Biomolecular Chemistry

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2018, **16**, 5097

Received 10th June 2018,
Accepted 25th June 2018

DOI: 10.1039/c8ob01365a

rsc.li/obc

The discovery of a freezing-induced peptide ligation during the total chemical synthesis of human interferon-ε†

Yin-He Yang, Bin Di * and Da-Song Yang *

A counterintuitive freezing-induced peptide ligation was discovered during the total synthesis of human interferon-ε (hIFN-ε) which blocks HIV infection through unique mechanisms. The successful synthesis of hIFN-ε (187 amino acids) in this research laid the foundation for related anti-AIDS drug development. Moreover, alanine mutation based on sequence alignment to solve the maldistribution of the ligation site and freezing-induced dominant conformation that facilitates peptide ligation are expected to be helpful for the synthesis of macrobiomolecules.

As a fatal disease caused by HIV infection, AIDS has led to the global death of at least 35 million people in the past three decades and more than 1.8 million new cases are diagnosed annually. The quick spread of HIV has become a serious threat to human health and placed a huge burden on the economy of both individuals and countries, which has gained increasing attention in medicine and politics.[1] Although antiretroviral therapy can slow the progression of the symptoms, the absence of a cure or an effective HIV vaccine is an indisputable fact.[2] Sexual contact is the major route of HIV transmission (80% of adults),[3] and new targeted strategies for prevention are urgently needed.

As a part of the innate immune system, interferons (IFNs) play important roles in the defense against viral infection,[4] while their low selectivity results in numerous side effects. IFN-ε is a special type I interferon; it shares low amino acid sequence homology with others (only about 30%), it is constitutively expressed in the epithelial cells of female and male reproductive organs, it is regulated by estrogen rather than by pattern-recognition receptor pathways,[5] and it cannot be secreted by producing cells.[6] Some interesting research indicate that the expression of IFN-ε can be induced by seminal plasma,[7] and its content in HIV negative female sex workers is increased.[8] IFN-ε upregulates the expression of various host cell restriction factors and reduces the production of HIV-required host proteins, which resulting in the impairment of HIV infection at multiple stages of its life cycle.[9] A recent study has revealed that IFN-ε blocks the invasion of HIV through enhanced phagocytosis and produces reactive oxygen species which is different from other IFNs.[10] The aforementioned results suggest that IFN-ε has evolved as a specific guardian against sexually transmitted infection, so the exploration of its action mechanism and structural modification may result in new anti-AIDS drugs with unique mechanisms and high selectivity. The acquisition of these special proteins for drug investigation is the premise.

Since it is difficult to produce unnatural proteins by using recombinant methods, such as labeling with spectroscopic probes for functional analysis, post-translational modification as well as their intermediates for structure–activity studies, D-protein for racemic protein crystallography[11] and mirror-image genetic replication,[12] as well as the introduction of small molecules or polymers to improve the pharmacokinetic properties.[13] Therefore, a total chemical synthesis strategy of hIFN-ε based on solid phase peptide synthesis (SPPS)[14] and native chemical ligation (NCL)[15] was established in this research. Moreover, the first example of a freezing-induced peptide ligation was discovered upon the process of overcoming the unreasonable condensation failure between peptide segments. Described herein is the validated synthetic approach for hIFN-ε, the discovery of a counterintuitive freezing-induced ligation and its plausible mechanism.

Currently, a peptide containing less than 50 amino acids can be easily synthesized by SPPS,[14] and two peptide segments each owning a C-terminal thioester and N-terminal cysteine (Cys) can be selectively assembled by NCL.[15] However, Cys is rare in proteins, which leads to the development of various thio-derived amino acids (AA) as surrogates,[16] and they can be facilely converted into proteinogenic AA by post-ligation metal-free desulfurization (MFD).[17,18] Furthermore, a convergent ligation strategy has been proved as an effective approach for the complete condensation among four or more peptide frag-

Jiang Su Key Laboratory of Drug Design and Optimization, Key Laboratory on Protein Chemistry and Structural Biology, China Pharmaceutical University, Nanjing 210009, China. E-mail: dibin@cpu.edu.cn, yangds@cpu.edu.cn

†Electronic supplementary information (ESI) available: Experimental procedures, HPLC profiles and MS spectra of the products. See DOI: 10.1039/c8ob01365a

ments.[13] These revolutionary techniques form the cornerstone of modern chemical protein synthesis and also govern which proteins are suitable for chemical synthesis as well as how to construct them.

For hIFN-ε (Uni-Prot Q86WN2), the mature protein consists of 187 AA after the cleavage of the signal peptide (21 AA) at the N-terminus of the original sequence. As Cys is the best ligation site for sequence splitting, a check on hIFN-ε found 3 Cys at positions 32, 142 and 154 (Fig. 1). Obviously, these Cys are maldistributed, especially between positions 32 and 142, resulting in a too long fragment that is beyond the ability of SPPS, and making it an undesirable strategy. Considering that there are 6 alanines (Ala) almost uniformly distributed in the sequence of hINF-ε (at positions 58, 73, 101, 107, 143 and 145), and that Cys can readily convert into Ala through post-ligation desulfurization,[17] the Ala based splitting strategy seems like a good candidate. The main drawback of this route is that the first peptide segment (57 AA) at the N-terminus of the protein is too long for synthesis. To solve this problem, an amino acid sequence alignment of IFN-ε is performed among human and other mammals including mouse (Uni-Prot Q80ZF2), rat (Uni-Prot F1LVL8) and dog (Uni-Prot J9P321) using BioEdit.[19] The results (Fig. 1) show that the proline in hIFN-ε at position 34 is an Ala in other mammals. The homology of the peptide sequence and the ideal position of proline motivated us to mutate it to Ala for a more feasible synthesis scheme.[20] Based on the above considerations, a synthetic strategy of the hIFN-ε protein is designed, which includes splitting it into five peptide segments with approximately equal lengths (33–39–34–38–43) and condensing them through a kinetically controlled convergent ligation. As shown in Fig. 2, all Ala located at the ligation site is temporarily replaced with Cys.

The peptide synthesis was performed on a CS Bio-CS336X automated peptide synthesizer by using an HBTU/DIEA activated Boc-chemistry protocol on a PAM resin.[14] Some important points on the peptide fragment synthesis are listed as follows. To make sure that the original Cys in the sequence and the introduced ones at the ligation site could selectively engage in its specific reaction, different protecting groups of thiol were used. Specifically, acetamidomethyl (Acm) was chosen to protect the original Cys (positions 32, 142 and 154) as it is stable under the conditions of cleavage, ligation and

```
Human|Q86WN2 LDLKLIIFQQ RQVNQESLKL LNKLQTLSIQ QCLPHRKNFL LPQKSLSPQQ
Mouse|Q80ZF2 LEPKRIPFQL W-MNRESLQL LKPLPSSSVQ QCLAHRKNFL LPQQPVSPHQ
Rat|F1LVL8   LEPKLILFQS R-RNRESLQL LKALPSSSVQ LCLAHRKNFL LPQQSVSPHQ
Dog|J9P321   LELKLAFFQQ R-VNRESLKL LNRFQKSSIQ QCLAHRKNFL LPQQSVNRHQ
Human|Q86WN2 YQKGHTLAIL HEMLQQIFSL FRANISLDGW EENHTEKFLI QLHQQLEYLE
Mouse|Q80ZF2 YQEGQVLAVV HEILQQIFTL LQTHGTMGIW EENHIEKVLA ALHRQLEYVE
Rat|F1LVL8   HQEGHVLAVL HEILQQIFVL FQTHSSRGVW EENHIEEVLA ALHRQLGYLE
Dog|J9P321   YQKGQALAIL HEMLQQIFNL FRANISLEGW EERHMENFLT ELHQQLEYLE
Human|Q86WN2 ALMGLEAEKL SGTLGSDNLR LQVKMYFRRI HDYLENQDYS TCAWAIVQVE
Mouse|Q80ZF2 SLGGLNAAQK SGGSSAQNLR LQIKAYFRRI HDYLENQRYS SCAWIIVQTE
Rat|F1LVL8   SLGGLKAERK SGGLSVKDLR LQIKAYFRRI HDYLENHRYS TCAWIIVQIE
Dog|J9P321   VLMSLEAEPN SGILRSDNPR LQIKR----- ---LE----- ----------
Human|Q86WN2 ISRCLFFVFS LTEKLSKQGR PLNDMKQELT TEFRSPR
Mouse|Q80ZF2 IHRCMFFVFR FTTWLSRQDP DP-------- -------
Rat|F1LVL8   VNRCMFFVSR LARRLNSQDI D--------- -------
Dog|J9P321   ---------- ---------- ---------- -------
```

Fig. 1 Amino acid sequence alignment of IFN-ε among different species (human, mouse, rat and dog).

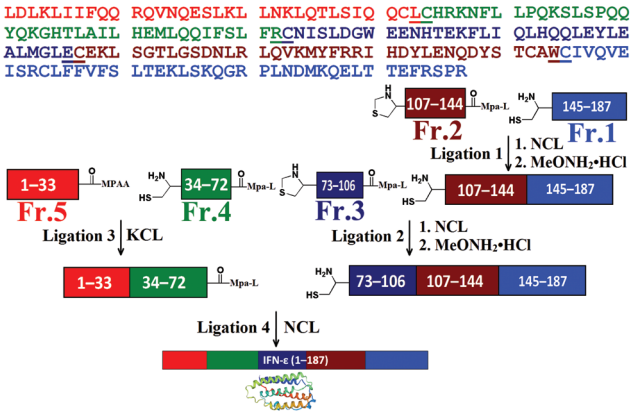

Fig. 2 The strategy for the convergent synthesis of the hIFN-ε protein (187 AA) based on native chemical ligation.

desulfurization.[18] For the introduced Cys at the N-terminus of Fr.2 and Fr.3, 1,3-thiazolidine-4-carbonyl (Thz) could serve as the masking groups to prevent intramolecular cyclization and intermolecular ligation during NCL. As the Cys at the N-terminus of peptide Fr.1 would directly participate in the ligation after synthesis, HF labile *p*-methylbenzyl was a good choice for protection. The same could be said for Fr.4, while the protecting group of imidazole, benzyloxymethyl, was concomitantly introduced with the appearance of histidines in the sequence and it would lead to formylation, methylation, or the formation of thiazolidines on Cys during the HF cleavage.[21] To reduce these by-products, Thz was temporarily introduced to protect the Cys at the N-terminus of Fr.4 and it would be removed through methoxyamine treatment after HF cleavage. The C-terminal thioester of peptides Fr.2 to Fr.5 was established using a general procedure based on a mercaptopropionic acid (Mpa)-leucine coupled PAM resin which would straightforwardly yield the alkyl thioester after cleavage.[22–24] This peptide thioalkylester could be exchanged by 4-mercaptophenylacetic acid (MPAA) and converted into the reactive thiophenylester which could quickly react with Cys-peptide to complete the ligation. The lack of effective protecting tactics for the thioester moiety made NCL unsuitable for the sequential assembly of peptide fragments from the N- to the C-terminus.[25,26] To solve this problem, kinetically controlled ligation (KCL) was developed based on the ligation speed of the thiophenylester was much faster than that of the thioalkylester without the addition of an exogenous thiophenol.[27] Therefore, the thioalkylester of peptide Fr.5 was converted into the thiophenylester in advance using MPAA after the synthesis, to make it compatible with KCL. The synthesized peptide segments Fr.1 to Fr.5 were purified by preparative RP-HPLC and their molecular weights were checked by ESI-MS. The measured masses: 5146.95, 4839.66, 4258.27, 4881.92 and 4131.77 Da (Fig. S2, S4, S6, S10 and S14†) were consistent with the theoretical values: 5146.83, 4839.34, 4258.85, 4881.79 and 4131.58 Da, respectively.

The complete condensation of the five peptide fragments needs at least four ligation steps, considering that the NCL is

only suitable for the sequential assembly of peptide fragments from the C- to the N-terminus and its ligation efficiency is higher than that of KCL; a ligation strategy consists of 3 steps of NCL and 1 step of the KCL is designed (Fig. 2). For ligation 1 between Fr.1 and Fr.2 (C-terminus is tryptophan), in general, it should be completed within 9 h under typical NCL conditions (50 mM MPAA, 20 mM TCEP, pH 7.1, rt),[22] let alone the amount of Fr.1 was added in excess to facilitate the conversion of Fr.2, as it could be recycled after ligation. However, no ligation product could be detected by HPLC and MS, even when the reaction time had been extended to 72 h and the concentration of MPAA had been increased to 100 mM. To solve this problem, some efforts have been devoted to the optimization of the reaction conditions. Considering that the pH of the ligation buffer is a key factor to balance the solubility of MPAA and the stability of the thioester, which further affect the ligation efficiency, a series of pH values including 7.1, 8.5, 9.5, 10.5 and 11.0 have been tested, and only a MS detectable amount of the ligation product appears under pH 8.5 (Fig. S19†). For further improvement, elevated temperatures (20, 30, 40 and 50 °C) have been tried as chemical reactions will generally be speeded up at higher temperatures. Frustratingly, there was no significant difference in the results of these tests. As before, the conventional optimization of the reaction conditions almost has no effect on the yield which reminds us to verify whether the specific peptide sequence near the junction makes it an unusable ligation site. Therefore, two peptide segments located at both sides of the ligation site with a length of about 10 AA were synthesized: one was the N-terminal cysteine-containing segment of Fr.1 [CIVQVEISRC(Acm)L, ESI-MS 1333.59, Fig. S15,† calcd 1333.45] and the other was the C-terminal thioester of Fr.2 [ENQDYSTC(Acm)AW-Mpa-L, ESI-MS 1487.64, Fig. S16,† calcd 1488.44]. Interestingly enough, these two short fragments could easily attain almost complete ligation under typical NCL conditions within 8 h, which denied the above hypothesis. Until now, it was obvious that the length of the two peptide fragments was the main reason for the failure of ligation 1. For the peptide, the length of the sequence results in a secondary structure and steric hindrance. To check this hypothesis, a preliminary 3D structure of hIFN-ε (Fig. 3) was built through homology modeling based on IFN-β (1au1) using SWISS-MODEL.[28] From the figure we could see the junction between Fr.1 and Fr.2 (red circle in Fig. 3) just located near the turning point between helix D and E. As we know, conformation is a critical factor that affects the molecular recognition and interactions between biomolecules.[29] Just like lock and key, the secondary structure of a peptide decides the type of surface accessible amino acid between two sequences; thus the spatial orientation of the ligation site and the chemical reactivity of the given peptides are controlled in turn. The above analysis implies that the pursuit of the dominant conformation may be helpful for the ligation between peptide Fr.1 and Fr.2. After a literature review, we found that lower temperature can reduce the skeleton flexibility of a peptide and promote the formation of the dominant conformation.[30,31] Hence, a ligation buffer

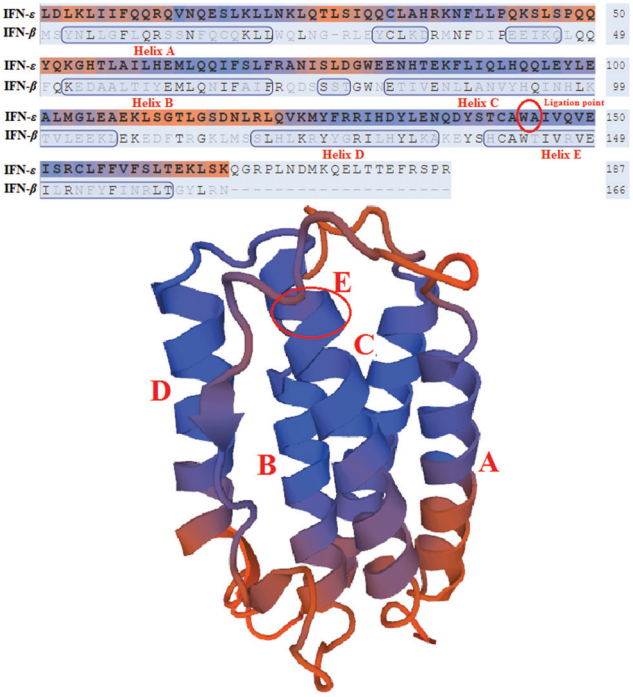

Fig. 3 Predicted 3D structure of hIFN-ε through homology modeling based on IFN-β (1au1) using SWISS-MODEL (red circle: ligation site between Fr.1 and Fr.2; lettering A–E: alpha-helices of hIFN-ε; the background color of the amino acid sequence is consistent with the 3D structure to demonstrate its location).

(50 mM MPAA, 20 mM TCEP, pH 8.5, 6 M Gu·HCl, 0.2 M Na_2HPO_4) containing 20 mg Fr.1 and 10 mg Fr.2 was placed in a fridge (−20 °C). It is amazing that the ligation products were detected by ESI-MS after reaction overnight (measured: 9765.55 Da, Fig. S23,† theoretical: 9766.87 Da), and the HPLC-MS analysis of the ligation buffer indicated that Fr.2 was continuously converted into its thiophenylester which further ligated with Fr.1 to form the products under low temperature (Fig. S20†). The results from the reaction monitoring showed that the ligation products increased with time and reached the maximum within 48 h (Fig. S21†). After purification with HPLC and unmasking of Thz at the N-terminus of the sequence, the products were further ligated with Fr.3 under standard NCL conditions. Ligation 3 was a KCL to assemble peptides from the N-terminus (Fr.5, C-terminus was thiophenylester) to the C-terminus (Fr.4, N-terminus was Cys) without adding MPAA. Ultimately, a final standard NCL between fragment 5–4 thioalkylester and Cys-fragment 3–2–1 resulted in the complete sequence of hIFN-ε (measured: 22 409.23 Da, Fig. 4, calcd: 22 408.41 Da).

In conclusion, a total chemical synthesis route of the hIFN-ε protein (187 AA) has been successfully built in this research based on a 5 fragments native chemical ligation, which laid the foundation for the related anti-AIDS drug development. The difficulties that hinder the synthesis, including insufficient Ala for sequence splitting, the ligation strategy of 5 segments, as well as the low solubility of peptide Fr.5 in acetonitrile–water

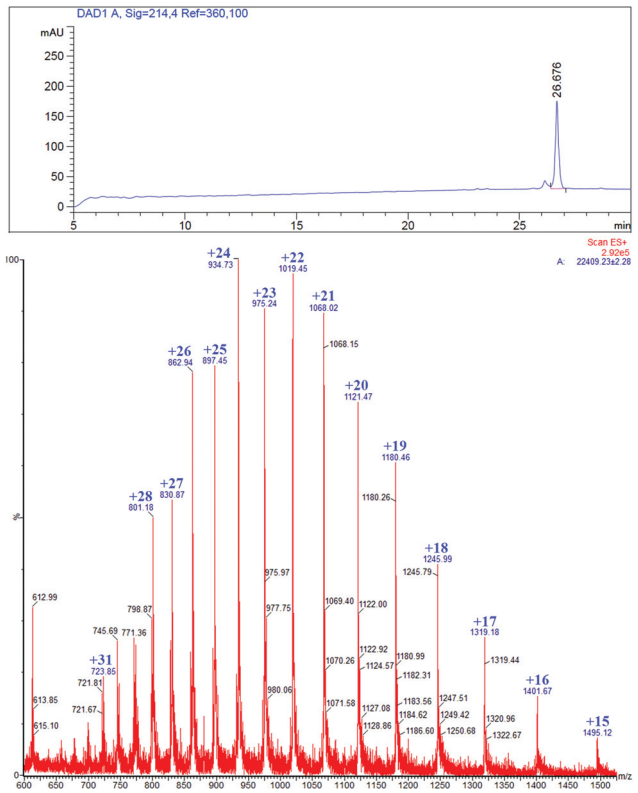

Fig. 4 The liquid chromatogram and ESI-MS spectra of the synthesized hIFN-ε protein.

solution for purification, have been overcome step by step through rational analysis, design and the optimization of the ligation conditions. Moreover, freezing promotes the formation of the dominant conformation which further induces the ligation of peptide segments is discovered upon the process of overcoming the unreasonable condensation failure between peptides Fr.1 and Fr.2. Despite some freezing-catalyzed reactions have been previously reported, and the mechanisms responsible for counterintuitive accelerations under low temperature include increasing concentration, freezing potential, catalytic action of the ice surface, convection effects and temperature differences.[32,33] This is the first example of the freezing-induced reaction reported in protein synthesis and the dominant conformation formation is a supplement to the catalytic mechanism in cryochemistry, which is expected to be helpful for the total synthesis of macrobiomolecules. Furthermore, for this rare and intriguing phenomenon, a systematic mechanistic study including NMR and CD structural studies of the peptide fragments at lower temperature, identification of the structural features, test of the generality of the observation for the ligation of other peptides, *etc*. are important to promote its usage in cryochemistry.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (81573386 and 81773693), the Program for Outstanding Scientific and Technological Innovation Team of Jiangsu Higher Education (2017) and the Fundamental Research Funds for the Central Universities (2016ZPT005).

Notes and references

1 P. M. Sharp and B. H. Hahn, *Cold Spring Harbor Perspect. Med.*, 2011, **1**, a006841.
2 A. S. Fauci, G. K. Folkers and C. W. Dieffenbach, *Nat. Immunol.*, 2013, **14**, 1104–1107.
3 M. S. Cohen, G. M. Shaw, A. J. McMichael and B. F. Haynes, *N. Engl. J. Med.*, 2011, **364**, 1943–1954.
4 G. C. Sen, *Annu. Rev. Microbiol.*, 2001, **55**, 255–281.
5 K. Y. Fung, N. E. Mangan, H. Cumming, J. C. Horvat, J. R. Mayall, S. A. Stifter, N. D. Weerd, L. C. Roisman, J. Rossjohn, S. A. Robertson, J. E. Schjenken, B. Parker, C. E. Gargett, H. P. T. Nguyen, D. J. Carr, P. M. Hansbro and P. J. Hertzog, *Science*, 2013, **339**, 1088–1092.
6 P. Hermant, C. Francius, F. Clotman and T. Michiels, *PLoS One*, 2013, **8**, e71320.
7 D. J. Sharkey, A. M. Macpherson, K. P. Tremellen and S. A. Robertson, *Mol. Hum. Reprod.*, 2007, **13**, 491–501.
8 S. A. Abdulhaqq, C. Zorrilla, G. Kang, X. Yin, V. Tamayo, K. E. Seaton, J. Joseph, S. Garced, G. D. Tomaras, K. A. Linn, A. S. Foulkes, L. Azzoni, M. VerMilyea, C. Coutifaris, A. V. Kossenkov, L. Showe, E. N. Kraiselburd, Q. Li and L. J. Montaner, *Mucosal Immunol.*, 2016, **9**, 1027–1038.
9 A. Garcia-Minambres, S. G. Eid, N. E. Mangan, C. Pade, S. S. Lim, A. Y. Matthews, N. A. de Weerd, P. J. Hertzog and J. Mak, *Immunol. Cell Biol.*, 2017, **95**, 478–483.
10 C. Tasker, S. Subbian, P. Gao, J. Couret, C. Levine, S. Ghanny, P. Soteropoulos, X. Zhao, N. Landau, W. Lu and T. L. Chang, *JCI Insight*, 2016, **1**, e88255.
11 M. Pan, S. Gao, Y. Zheng, X. D. Tan, H. Lan, X. L. Tan, D. M. Sun, L. N. Lu, T. Wang, Q. Y. Zheng, Y. C. Huang, J. W. Wang and L. Liu, *J. Am. Chem. Soc.*, 2016, **138**, 7429–7435.
12 Z. M. Wang, W. L. Xu, L. Liu and T. F. Zhu, *Nat. Chem.*, 2016, **8**, 698–704.
13 F. Liu and J. P. Mayer, in *Protein Ligation and Total Synthesis I. Topics in Current Chemistry*, ed. L. Liu, Springer, Cham, 2015, vol. 362, pp. 183–228.
14 M. Schnölzer, P. Alewood, A. Jones, D. Alewood and S. B. H. Kent, *Int. J. Pept. Res. Ther.*, 2007, **13**, 31–44.
15 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. H. Kent, *Science*, 1994, **266**, 776–779.
16 Y. Gui, L. Q. Qiu, Y. H. Li, H. X. Li and S. W. Dong, *J. Am. Chem. Soc.*, 2016, **138**, 4890–4899.
17 L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526–533.

18 Q. Wan and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2007, **46**, 9248–9252.
19 T. A. Hall, *Nucleic Acids Symp. Ser.*, 1999, **41**, 95–98.
20 M. Izumi, A. Otsuki, M. Nishihara, R. Okamoto and Y. Kajihara, *ChemBioChem*, 2014, **15**, 2503–2507.
21 A. Isidro-Llobet, M. Álvarez and F. Albericio, *Chem. Rev.*, 2009, **109**, 2455–2504.
22 T. M. Hackeng, J. H. Griffin and P. E. Dawson, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 10068–10073.
23 C. Rao and C. F. Liu, *Org. Biomol. Chem.*, 2017, **15**, 2491–2496.
24 J. B. Blanco-Canosa, B. Nardone, F. Albericio and P. E. Dawson, *J. Am. Chem. Soc.*, 2015, **137**, 7197–7209.
25 S. Tang, Y. Y. Si, Z. P. Wang, K. R. Mei, X. Chen, J. Y. Cheng, J. S. Zheng and L. Liu, *Angew. Chem., Int. Ed.*, 2015, **54**, 5713–5717.
26 T. Takei, T. Andoh, T. Takao and H. Hojo, *Angew. Chem.*, 2017, **129**, 15914–15917.
27 D. Bang, B. L. Pentelute and S. B. H. Kent, *Angew. Chem.*, 2006, **118**, 4089–4092.
28 M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede, *Nucleic Acids Res.*, 2014, **42**, W252–W258.
29 J. R. Bock and D. A. Gough, *Bioinformatics*, 2001, **17**, 455–460.
30 C. Mikelis, M. Lamprou, M. Koutsioumpa, A. G. Koutsioubas, Z. Spyranti, A. A. Zompra, N. Spiliopoulos, A. A. Vradis, P. Katsoris, G. A. Spyroulias, P. Cordopatis, J. Courty and E. Papadimitriou, *J. Cell. Biochem.*, 2011, **112**, 1532–1543.
31 G. Taborsky, *Adv. Chem.*, 1979, **180**, 1–26.
32 S. M. Agten, D. P. L. Suylen and T. M. Hackeng, *Bioconjugate Chem.*, 2016, **27**, 42–46.
33 R. O'Concubhair and J. R. Sodeau, *Acc. Chem. Res.*, 2013, **46**, 2716–2724.