Organic & Biomolecular Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Development and application of serine/threonine ligation for synthetic protein chemistry

Han Liu[a] and Xuechen Li*[a,b]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

Chemical synthesis of proteins, especially those with post-translational modifications, has offered new opportunities to study protein structure-function relationship. During the past four years, we have developed the serine/threonine ligation (STL), which involves the chemoselective reaction between peptide salicylaldehyde esters and peptides with N-terminal serine or threonine. The method has been successfully applied to the synthesis of both linear and cyclic peptides/proteins.

Introduction

The chemistry and biology of proteins is essential for scientist to understand how life works. The knowledge about protein structure and function provides the solid basis for molecular pathology and drug discovery. Armed with modern separation-analysis and gene technology, current biologists can identify trace amounts of proteins from cell lysates, and prepare pure samples for further structure and function investigations. However, the word 'pure' here only means well-defined amino acid sequence, without any information about post-translational modifications. Indeed, most of the eukaryote proteins are post-translationally modified by glycosylation, phosphorylation, lipidation, ubiquitylation, etc. The level of designated PTMs on specific sites tunes the high hierarchical structure and activity of the proteins finely, which works harmoniously under delicate control. Slight variation of PTMs is often associated with pathological and physiological changes, such as tumorigenesis and aging. Since intensive studies on protein PTMs require highly homogenous protein samples with the defined PTM structure at specific sites, which cannot be obtained from biologically controlled methods, the chemical protein synthesis is expected to embrace this challenge.[1]

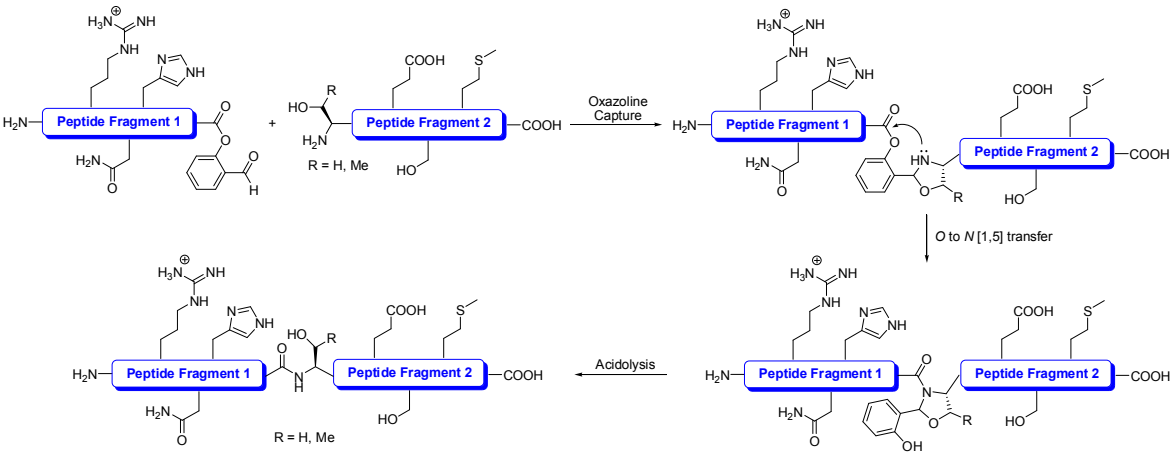
The past half century has witnessed the growth of chemical protein synthesis. In the 1960's, the first chemically synthesized protein, bovine insulin (51 residues in total), was prepared via sequential coupling and deprotection of amino acid building blocks in solution phase.[2] In 1966, the synthesis of the same target was greatly improved by Merrifield[3] based on the solid phase peptide synthesis strategy.[4] Armed with the SPPS technology, scientists are now able to achieve proteins with much larger sizes, such as ribonuclease A (124 residues)[5] and interleukin-3 (140 residues).[6] However, peptide chemists have suffered a lot from the incomplete coupling in each round (especially for long peptides) during the linear SPPS, which in turn resulted in difficult separation of the product from truncated peptides. In this regard, SPPS suites best for the synthesis of peptides with less than 50 amino acids. In the 1990's, the paradigm was changed by the birth of chemical ligation which allowed for the convergent synthesis of large polypeptide from side-chain unprotected peptide segments of small sizes. In 1994, Kent et al. reported the synthesis of interleukin-8 (72 residues) via native chemical ligation (NCL).[7] The thioester of a side chain unprotected C-terminal peptide reacts smoothly with the Cys residue of the side-chain unprotected N-terminal peptide to afford a native peptide linkage via a thioester exchange followed by an *S* to *N* acyl transfer. Based on this strategy, large proteins can be convergently assembled from side chain unprotected peptide fragments (with moderate size and good solubility) in a highly efficient and selective manner. Considering the low existence of Cys in proteins (~ 1.4%), much research attention to extend the peptide ligation at other amino acid sites was paid to the synthesis of amino acids with β-thiol structure,[8] *N*-attached thiol-containing auxiliary,[9] and mild desulfurization conditions.[10] The state of the art of native chemical ligation was recently illustrated incisively in Danishefsky's 12 years' synthesis of a complex glycoprotein-erythropoietin (166 residues, with three *N*-linked dodecasaccharides and one *O*-linked tetrasaccharide).[11]

The chemoselectivity of NCL originates from the chemoselective and reversible thioester exchange, which relies on the thioester modified at the C-terminus and Cys at the N-terminus of the second peptide segment. To explore thiol-independent ligation reactions, scientists have designed new chemoselective amide formation reactions using other mechanisms.[12] In 1994, Tam et al. developed the pseudoproline ligation, based on the reaction of C-terminal glycoaldehyde esters and N-terminal Cys/Ser/Thr residues.[13] The 5-hydroxymethyl pseudoproline unit at the ligation site resembles a proline structure and thus maintains the secondary structure and bioactivity of the product. In 2000, Raines et al. reported the traceless Staudinger ligation, which involves the reaction of a C-terminal thioester containing a phosphine group with an N-terminal azide.[14] Through extensive modification of the phosphine structure, the scope of the ligation can be extended to

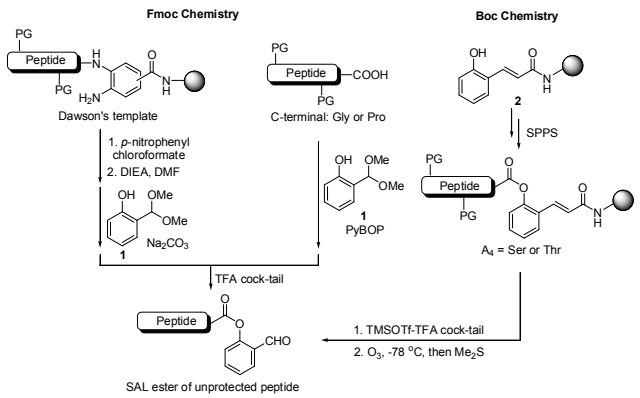
Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE



Scheme 1 Design of serine/threonine ligation (STL).



Scheme 2 Preparation of peptide SAL esters.

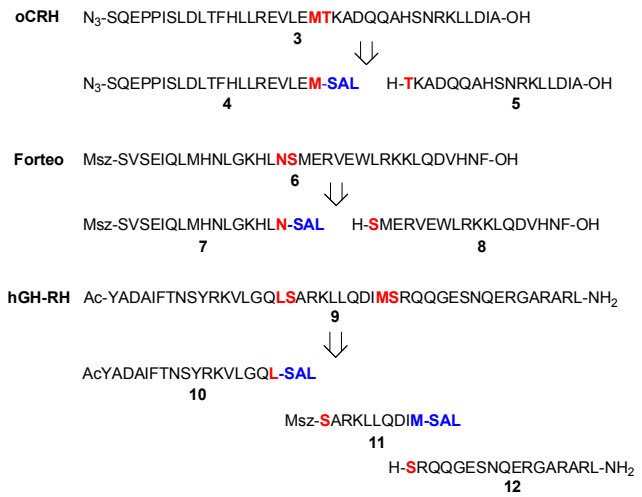
hindered reactants and pure water can be use as solvent. In 2006, Bode et al. reported the α–ketoacid-hydroxylamine (KAHA) ligation, from the decarboxylative coupling between a C-terminal α–ketoacid and an N-terminal hydroxylamine.[15] The usage of 5-oxaproline in place of hydroxylamine has allowed them to synthesize proteins with the generation of a homoserine residue at the ligation site. Among all peptide ligations methods, only NCL and KAHA-mediated homoserine ligation have been successfully applied to deliver synthetic proteins so far.[15b] Nevertheless, the operational simplicity renders NCL the best method of choice, because this method leads to the natural peptides/proteins and only requires one ligating peptide partner to be modified (i.e., the peptide thioester); thus, it has been widely adopted to synthesize proteins and proteins with post-translational modifications.

Construction of linear systems via STL

Our own journey in the world of the chemical protein synthesis began in 2009. Along the line of the development of simple and thiol-independent peptide ligations, we hoped to develop the ligation directly using natural amino acids at the ligating N-terminal peptides. We were interested in using serine (Ser) and threonine (Thr), which are very abundant in natural proteins (~12.7% in total), especially in proteins with post-translational modifications. The key challenge in the development of serine/threonine ligation lies in how to achieve the chemoselectivity, as neither the hydroxyl nor amino group of N-terminal serine/threonine can distinguish them from the internal side-chain functional groups. At the time, we conceived of a two-step strategy, as illustrated in Scheme 1. We hypothesized that the peptide with an aromatic aldehyde at the C-terminus could chemoselectively capture the N-terminal amino alcohol of Ser or Thr to generate an imine, which would be amenable to the further transformation to an oxazolidine via ring-chain tautomerization. Next, the secondary amine in the oxazolidine ring would attack the phenolic ester as a nucleophile to trigger an *O* to *N* [1,5]-acyl transfer. We hoped the peptide phenolic (i.e., salicylaldehyde) ester to be stable for the chemical operations while reactive enough to undergo an acyl transfer. If the proposed reaction worked, the ligated product would be obtained with an *N,O*-benzylidene acetal at the conjunction site. The key issue was whether the natural peptidic linkage could be generated from the benzylidene acetal linked product via simple and peptide-compatible chemical transformations (e.g., acidolysis).

We first tested our idea with a model system using Fmoc protected single amino acid-derived salicylaldehyde (SAL) esters.[16] To our delight, the reaction worked smoothly in pyridine acetate buffer solvent to afford the desired *N,O*-benzylidene acetal intermediate in high yields. More importantly, we found out that the native peptide linkage was readily generated upon the TFA/H_2O/TIPS treatment. On the contrary to Tam's stable pseudoproline unit[13], the acid labile benzylic C-O and C-N bonds ensured the traceless feature of our ligation.

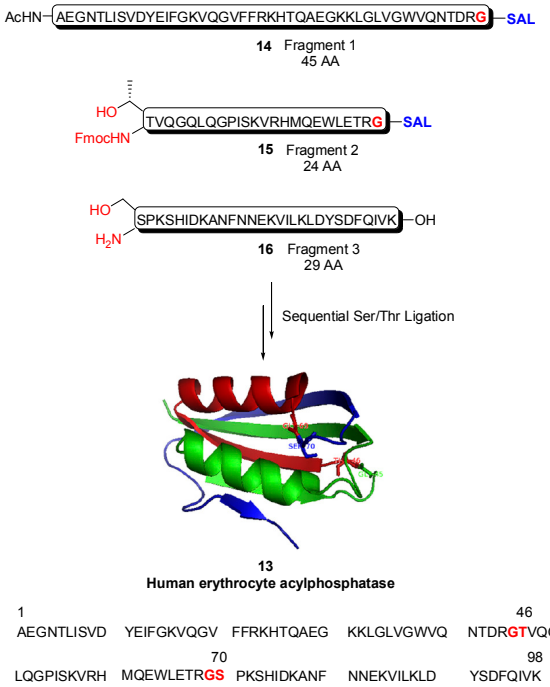
Encouraged by the preliminary results, we tried to apply this serine/threonine ligation (STL) to the side-chain unprotected



Scheme 3 Retrosynthetic analysis of peptide drugs base on STL.

oligopeptide and even protein system. As the prerequisite of this aim, a practical synthesis of side-chain unprotected SAL ester is thus needed. Different methods were developed to accommodate the demands from both Fmoc-SPPS and Boc-SPPS (Scheme 2). For Fmoc chemistry, according to the different amino acid residues at the C-terminus, we have developed two methods. For C-terminal Gly and Pro with no/low propensity of epimerization, a direct coupling reaction with salicylaldehyde dimethyl acetal **1** was efficient for the peptide SAL ester formation.[17] For other types of C-terminal residues, we have developed a on-resin phenolysis strategy using a modified version of Dawson's safety-catch technology[18] After 4-nitrophenyl chloroformate mediated formation of the *N*-acyl benzimidazolinone, the activated resin-bound species was treated with an excess amount of salicylaldehyde dimethyl acetal in the presence of Na_2CO_3. After TFA mediated global deprotection and HPLC purification, the desired peptide SAL ester with different peptide length could be obtained in overall 20–30% yields based on resin loading. The integrity of the α-carbon center at the C-terminus was confirmed by the co-injection of the peptide SAL esters with C-terminal L and D-Ala residues. For the synthesis of peptide SAL esters via Boc-SPPS, we[19] and Liu et al.[20] have developed the 2'-hydroxyl cinnamate modified amino methyl resin **2** independently. The first amino acid was loaded onto the resin via an ester bond, which is stable under TFA mediated deprotection of Boc group. After the Boc-SPPS assembly of the peptide and TMSOTf/TFA mediated global deprotection, the resin bound peptide was treated with ozone to cleave the alkene double bond, and then the desired peptide SAL ester was released into the solution phase in high yields and good purity.

With the desired peptide SAL esters in hand, the ligation with side chain unprotected peptides with *N*-terminal Ser or Thr was tested. Two FDA approved peptide drugs, ovine corticoliberin **3** (oCRH, 41 residues) and Forteo **6** (34 residues), were chosen as targets for their important bioactivities and lack of Cys residue.[17] To avoid the self cyclization of the C-terminal fragment, the N-terminal Ser residues of **4** and **7** were protected with azido and Msz groups, respectively. In both cases, the ligation worked quite well. The >90% conversion was achieved for N-terminal Thr (oCRH) within 12 h, while >95% conversion was observed for N-terminal Ser (Forteo) within 1 h. For the synthesis of human

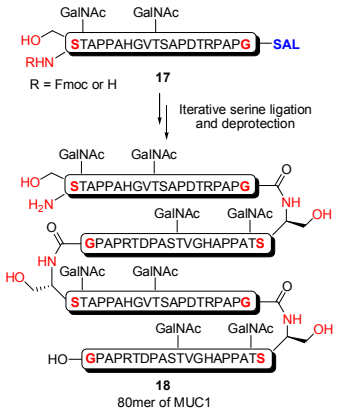


Scheme 4 Synthesis of erythrocyte enzyme acyl phosphatase via STL.

growth hormone-releasing hormone **9** (hGH-RH, 44 residues, no Cys), a C to N three-segment sequential serine/threonine ligation was developed.[21] The drug molecule was divided based on the presence of serine/threonine residues into three fragments **10**-**12**.To allow for the C to N sequential ligation, the N-terminal Ser residue of the middle fragment **11** was protected with *p*-(methylsulfinyl)benzyloxycarbonyl (Msz) group, which is stable to TFA conditions, less hydrophobic than Fmoc and readily removable under the reductive acidolysis treatment (TMSOTf/TFA/thioanisole).

With the goal of assembling complex proteins of biological interest via STL, we undertook the synthesis of human erythrocyte enzyme acyl phosphatase **13** (98 residues, with no Cys residue).[17] As shown in Scheme 4, the target protein was assembled from three fragments **14**-**16**, containing 45, 23, and 29 residues, respectively. To better solubilize the peptide segment and improve the efficiency of the ligation, pyridine/HOAc ratio could be tuned. For instance, the second ligation was conducted in pyridine acetate buffer (pyr/AcOH, 1:10). Gratifyingly, the obtained synthetic protein has exhibited the enzymatic activity in the hydrolysis of benzoylphosphate substrate. This success opens our own door to the world of synthetic protein chemistry.

Chemical synthesis of post-translationally modified proteins offers a good platform to showcase the power of synthetic protein chemistry. We chose MUC1 glycopeptides as our target, for its high importance in the anticancer vaccine development and the interesting structure feature.[22] MUC1 glycopeptides are composed of 20-amino acid long repeating unit, with varying degrees of *O*-glycosylation . Based on our serine ligation method, we prepared the repeating unit derived SAL ester **17** as the key intermediate. The chain-elongated product was obtained in good purity via iterative serine ligation and membrane filtration. After three ligations, the 80-mer product **18** (9.1 kDa) was obtained in good yield, without tedious HPLC purifications.



Scheme 5 Synthesis of MUC1 glycopeptide via serine ligation.

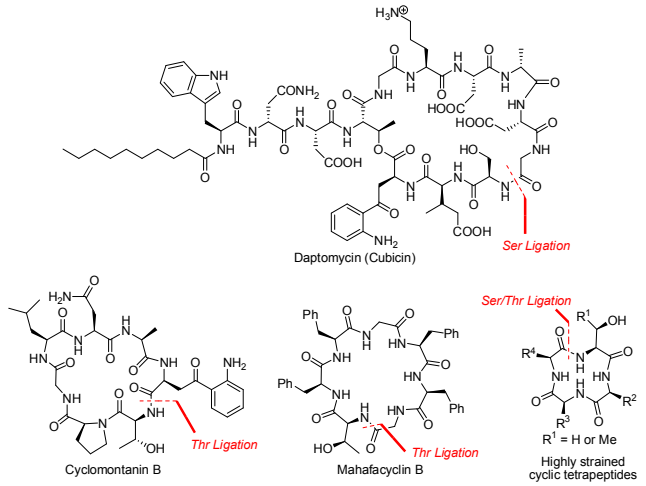


Fig. 1 Selected examples of cyclopeptides synthesized via STL.

Since our preliminary publications, STL has attracted the interest from other research groups. New applications of STL were recently realized in non-peptide systems. Kirshenbaum et al. have investigated the application of STL in the synthesis of peptoid-peptide/protein hybrids.[23] The peptoid SAL esters (with *N*-alkyl glycine at C-terminus) were synthesized via DIC coupling, which reacted efficiently with both chemically synthesized peptides and expressed protein (RNase S)/protein segment. This contribution has demonstrated the combination of STL and EPL in protein synthesis/modification.

Construction of cyclic systems via STL

Having established the efficiency of STL in the convergent synthesis of peptides and proteins, we turned to explore the construction of the cyclic peptides via intramolecular STL. As discovered by medicinal chemists, cyclopeptides have attractive performance in drug discovery, owing to their restricted conformation, high occurrence of non-proteingenic amino acids, strong interactions with targets, and high resistance to degradation. However, the synthesis of cyclopeptides is far from mature. The successful cyclization is often achieved by the trial-and-error selection of the cyclization site and comprehensive screening of reaction conditions (coupling reagents, additives and high dilution condition). Encouraged by the success of the intermolecular STL, we were eager to attempt the synthesis of cyclic peptides/depsipeptides via intramolecular STL.

To our delight, the STL-mediated peptide cyclization worked quite well in the construction of cyclic peptides. As shown in Figure 1, we have achieved the first total synthesis of daptomycin **19**, a FDA approved antibacterial drug containing 13 amino acids and a 31-membered macrocyclic core.. The key cyclization was realized by way of serine ligation at the Gly-Ser site.[24] To compare, the conventional lactamization conditions using HATU and DEPBT failed to cyclize at the same disconnection site. We have also finished the total synthesis of cyclic heptapeptide cyclomontanin B **20**, via Boc SPPS and intramolecular threonine ligation.[25] It is noteworthy that the formation of the peptide SAL ester and the generation of the rare Kyn residue were fulfilled simultaneously via ozonolysis. Based on the Boc SPPS/ozonolysis strategy, Liu' group synthesized a series of cyclic peptides with lengths ranged from 4 to 9 amino acids, including the natural cyclic heptapeptide mahafacyclin B **21**.[20] To test the efficiency of STL-mediated peptide cyclization in the construction of the strained cyclic peptide structure, we have investigated the synthesis of cyclic tetrapeptides without the assistance of internal turn-inducing residues.[19] Both experimental and computational results have indicated that the formation of a 16-membered ring via an imine capture and the subsequent ring contraction via an intramolecular acyl transfer can avoid the skeleton constrain and significantly reduce the activation energy, thereby enabling the efficient tetrapeptide cyclization. The desired cyclic tetrapeptides **22** can be obtained in high monomer/dimer ratio (up to 99: 1), which is very difficult to achieve via other methods.

Conclusions

In summary, we have developed a chemoselective serine/threonine ligation (STL) in which a side chain unprotected peptide with C-terminal SAL esters reacts with a side chain unprotected peptide with N-terminal serine or threonine to afford a ligated product with the generation of the natural peptidic peptide bond (Xaa-Ser/Thr) at the ligation site. As a complementary strategy to native chemical ligation and its variants, our method provides new choice of ligation sites at abundant amino acid residues. We have demonstrated the effectiveness of STL in the convergent synthesis of peptides and proteins. In addition, we have also realized the unique potentials of STL in the construction of both macrocyclic peptides and highly constrained cyclic tetrapeptides. Furthermore, STL has been used to incorporate a peptoid molecule into an expressed protein. These successes have opened the new avenue for the chemical synthesis of biomolecules using STL.

For the future development in view of this strategy, new methods for the synthesis of peptide SAL esters are highly demanded. Since the development of NCL, the synthesis of peptide thioesters via SPPS, requisite for NCL, has drawn a lot of attention. Analogously, the future advancement of STL will require more efficient methods for the peptide SAL ester synthesis. In theory, the methods involving post-SPPS-modification to synthesize peptide thioesters will be likely amenable to the synthesis of peptide SAL esters. In view of the

application, the synthesis of the large proteins including those carrying post-translational modifications (*N*/*O*-glycosylation, phosphorylation, etc.) will be of great importance to demonstrate the efficiency of STL. Recently, there has been an overwhelming demand for studies on protein post-translational modifications to understand the increasing complexity from the level of the genome to the proteome. However, the heterogeneity of post-translational modifications resulted from the non-template biosynthesis has complicated the studies on the structure-activity relationship. The access to the homogeneous post-translationally modified protein with a define modification structure using biologically controlled methods is very difficult, if not impossible. Advances in chemical methods have made it possible to synthesize homogeneous proteins, which has opened up new and exciting opportunities to investigate protein post-translational modifications. To achieve this goal, the development of novel protein synthesis methods and strategies should continue. On the other hand, one should take the available tools, including NCL, EPL, STL and others, and bring the chemical synthesis into full play.

Acknowledgements

The work is supported by the National Basic Research Program of China (973 Program, 2013CB836900), the Research Grants Council-General Research Fund of Hong Kong (HKU703811P, 707412P), the Peacock Program Project Development Fund (KQC201109050074A) and The University of Hong Kong.

Notes and references

[a] *Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, P. R. China. Fax: 852 2857 1586; Tel: 852 2219 4992; E-mail: xuechenl@hku.hk;* [b] *Shenzhen Institute of Research and Innovation of the University of Hong Kong, Shenzhen, P. R. China.*

1 For recent reviews on chemical protein synthesis, see: (a) S. B. H. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338. (b) V. Muralidharan and T. W. Muir, *Nat. Methods*, 2006, **3**, 429. (c) B. L. Nilsson, M. B. Soellner and R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 91.A.

2 (a) J. Meienhofer, E. Schnabel, H. Bremer, O. Brinkhoff, R. Zabel, W. Sroka, H. Klostermeyer, D. Brandenburg, T. Okuda and H. Zahn, *Z. Naturforsch.*, 1963, **18b**, 1130. (b) P. G. Katsoyannis, K. Fukuda, A. Tometsko, K. Suzuki and M. Tilak, *J. Am. Chem. Soc.*, 1964, **86**, 930. (c) Y. T. Kung, Y. C. Du, W. T. Huang, C. C. Chen, L. T. Ke, S. C. Hu, R. Q. Jiang, S. Q. Chu, C. I. Niu, J. Z. Hsu, W. C. Chang, L. L. Chen, H. S. Li, Y. Wang, T. P. Loh, A. H. Chi, C. H. Li, P. T. Shi, Y. H. Yieh, K. L. Tang and C. Y. Hsing, *Sci. Sinica (Peking)*, 1964, **14**, 1710.

3 (a) R. B. Merrifield and J. M. Stewart, *Nature*, 1965, **207**, 1522. (b) A. Marglin and R. B. Merrifield, *J. Am. Chem. Soc.*, 1966, **88**, 5051.

4 (a) R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149. (b) R. B. Merrifield, *Science*, 1965, **150**, 178.

5 B. Gutte and R. B. Merrifield, *J. Biol. Chem.*, 1971, **246**, 1922.

6 I. Clark-Lewis, R. Aebersold, H. Ziltener, J. W. Schrader, L. E. Hood and S. B. H. Kent, *Science*, 1986, **231**, 134.

7 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. H. Kent, *Science*, 1994, **266**, 776.

8 For review, see: C. T. T. Wong, C. L. Tung and X. Li, *Mol. BioSyst.*, 2013, **9**, 826.

9 B. Wu, J. Chen, J. D. Warran, G. Chen, Z. Hua and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, 2006, **45**, 4116.

10 (a) L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526. (b) Q. Wan and S. J. Danishefsky, *Angew. Chem. Int. Ed.,* 2007, **46**, 9248.

11 P. Wang, S. Dong, J.-H. Shieh, E. Peguero, R. Hendrickson, M. A. S. Moore and S. J. Danishefsky, *Science*, 2013, **342**, 1357.

12 For a recent review, see: V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471.

13 (a) C.-F. Liu and J. P. Tam, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 6584. (b) C.-F. Liu and J. P. Tam, *J. Am. Chem. Soc.*, 1994, **116**, 4149. (c) J. P. Tam and Z. Miao, *J. Am. Chem. Soc.*, 1999, **121**, 9013.

14 (a) B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org. Lett.*, 2000, **2**, 1939. (b) B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org. Lett.*, 2001, **3**, 9. (c) N. A. McGrath and R. T. Raines, *Acc. Chem. Res.*, 2011, **44**, 752.

15 (a) J. W. Bode, R. M. Fox and K. D. Baucom, *Angew. Chem. Int. Ed.*, 2006, **45**, 1248. (b) V. R. Pattabiraman, A. O. Ogunkoya and J. W. Bode, *Angew. Chem. Int. Ed.*, 2012, **51**, 5114. (c) A. O. Ogunkoya, V. R. Pattabiraman and J. W. Bode, *Angew. Chem. Int. Ed.*, 2012, **51**, 9693.

16 X. Li, H. Y. Lam, Y. Zhang and C. K. Chan, *Org. Lett.*, 2010, **12**, 1724.

17 Y. Zhang, C. Xu, H. Y. Lam, C. L. Lee and X. Li, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 6657.

18 J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem. Int. Ed.*, 2008, **47**, 6851.

19 C. T. T. Wong, H. Y. Lam, T. Song, G. Chen and X. Li, *Angew. Chem. Int. Ed.*, 2013, **52**, 10212.

20 J.-F. Zhao, X.-H. Zhang, Y.-J. Ding, Y.-S. Yang, X.-B. Bi and C.-F. Liu, *Org. Lett.*, 2012, **15**, 5182.

21 Y. Zhang, T. Li and X. Li, *Org. Biomol. Chem.*, 2013, **11**, 5584.

22 C. Xu, H. Y. Lam., Y. Zhang and X. Li, *Chem. Commun.*, 2013, **49**, 6200.

23 (a) P. M. Levine, T. W. Craven, R. Bonneau and K. Kirshenbaum, *Org. Lett.*, 2014, **16**, 512. (b) P. M. Levine, T. W. Craven, R. Bonneau and K. Kirshenbaum, *Org. Biomol. Chem.*, 2013, **11**, 4142.

24 H. Y. Lam, Y. Zhang, H. Liu, J. Xu, C. T. T. Wong, C. Xu and X. Li, *J. Am. Chem. Soc.*, 2013, **135**, 6272.

25 C. T. T. Wong, H. Y. Lam and X. Li, *Org. Biomol. Chem.*, 2013, **11**, 7616.