ChemComm

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Commun.*, 2020, **56**, 6114

Received 1st March 2020,
Accepted 21st April 2020

DOI: 10.1039/d0cc01604g

rsc.li/chemcomm

Open Access Article. Published on 23 April 2020. Downloaded on 2/9/2026 10:41:30 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

N-Terminal speciation for native chemical ligation†

Oliver R. Maguire, ‡ Jiayun Zhu, William D. G. Brittain, Alexander S. Hudson, Steven L. Cobb and AnnMarie C. O'Donoghue *

Native chemical ligation (NCL) enables the chemical synthesis of peptides *via* reactions between N-terminal thiolates and C-terminal thioesters under mild, aqueous conditions at pH 7–8. Here we demonstrate quantitatively how thiol speciation at N-terminal cysteines and analogues varies significantly depending upon structure at typical pH values used in NCL.

Proteins are versatile biological macromolecules and their total chemical synthesis allows chemists to access important targets that can be difficult to obtain from traditional biological sources.[1] Furthermore, the incorporation of non-natural amino acid residues is more readily achievable allowing the structural and functional properties of proteins to be probed.

Over the past two decades, native chemical ligation (NCL, Fig. 1) has revolutionized peptide science through its ability to couple peptide fragments under mild conditions without additional coupling agents or side chain protection.[2] Protein total syntheses have been achieved chiefly through the combination of solid phase peptide synthesis (SPPS)[3] and native chemical ligation (NCL). SPPS allows for the synthesis of peptide fragments of up to 50 residues in length and NCL allows subsequent bioconjugation of these fragments to give target proteins.[4] Early examples of NCL required the presence of a cysteine residue at the N-terminus of one peptide fragment, however, its scope has been expanded substantially through the use of thiol analogues of natural amino acids[5] and selenocysteine derivatives.[6] These systems can then be transformed into the natural residue *via* desulfurisation[7] or deselenisation[8] reactions. Other major advances have included auxiliary-mediated NCL,[9] kinetically controlled ligation (KCL)[10] and templated NCL.[11]

The mechanism of NCL involves transthioesterification between the cysteine residue and the thioester followed by an intramolecular S-to-N acyl shift to form the native amide bond (Fig. 1).[2c] The addition of aryl thiol additives usually accelerate NCL through an initial thioester exchange step to a thioaryl nucleofuge.[12] Typically, as drawn in Fig. 1, only a single thiolate species is considered as the active nucleophile in the NCL literature based on the higher nucleophilicity of the anionic thiolate relative to neutral thiol.

However, the cysteine residue at the N-terminus of a peptide has two possible ionization sites at the thiol and ammonium groups. The pK_as of an alkyl thiol (~9–11) and amino acid primary ammonium (~9–10) are in sufficiently close proximity such that up to four species may be present in solution depending upon the pH: **i** cationic, **ii** formally neutral zwitterionic, **iii** neutral and **iv** anionic species (Fig. 2). The concentration of each species is controlled by the acid dissociation constants, K_a(A)–K_a(D). Importantly, this means that NCL has the option of two different thiolate species in solution at a given pH (**ii** and **iv**), and it is not strictly correct to quote one pK_a for the cysteine thiol as is commonly done. The use of modified cysteine analogues in NCL will further alter K_a(A)–K_a(D) and the species distribution.

Herein, we report the pK_a values for a series of cysteine and thiolated analogues of amino acid methyl esters and peptides (Fig. 3, **1–10**). We evaluate the speciation (%**i–iv**) over the whole pH range, including at typical NCL pH values. N-terminal acid dissociation constants will be most influenced by substituents in close proximity thus monomeric amino acid derivatives and short peptides are appropriate models to assess N-terminal

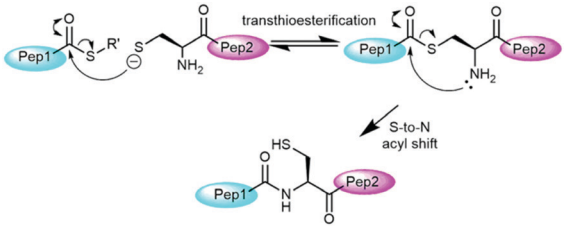

Fig. 1 Mechanism of native chemical ligation.

Department of Chemistry, Durham University, University Science Laboratories, South Road, Durham DH1 3LE, UK. E-mail: annmarie.odonoghue@durham.ac.uk

† Electronic supplementary information (ESI) available: UV-Visible spectrophotometric methods, data and fitting analysis; general synthetic procedures, materials and instrumentation. See DOI: 10.1039/d0cc01604g

‡ Current address: Radboud University Nijmegen, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands.

This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 April 2020. Downloaded on 2/9/2026 10:41:30 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

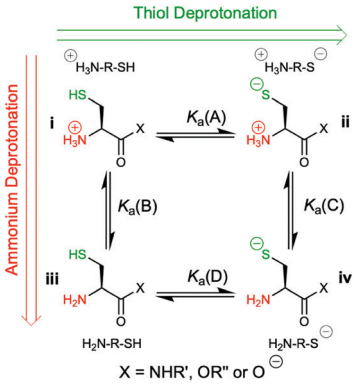

Fig. 2 The four possible cysteine species **i–iv** in solution and the acid dissociation constants that define the interrelationship between each species.

speciation. The dissociation constants K_a(A)–K_a(D) were determined by UV-Vis spectrophotometry using an adapted form of the procedure reported by Benesch[13] for evaluation of the acid dissociation constants of cysteine **1**. The changes in absorbance from thiolate (A_{RS^-}, λ_{max} = 237 nm) were determined across the pH range 1.4–12.5 (Fig. S1–S47, ESI†). A small blue shift of the absorbance to λ_{max} = 235 nm is observed at lower pHs, which is attributed to ^+H_3N-R-S$^-$ **ii** absorbing at shorter wavelengths than H_2N-R-S$^-$ **iv** (ESI,† Section S3). For the range of thiolated substrates employed, it was necessary to conduct measurements in the presence of 2 mM TCEP to prevent interference from thiol(ate) oxidation. The fraction of thiolate species, f_{RS^-}, in solution was calculated from the ratio of A_{RS^-} at a given pH to A_{RS^-}(max) at pH 12.5 where all thiols are in thiolate form (eqn (1)). The dissociation constants K_a(A), K_a(B) and K_a(D) could then be obtained by fitting eqn (2) to the data for f_{RS^-} *versus* pH (*e.g.* Fig. S30 for **6**, ESI†, Section S2) and K_a(C) using eqn (S1) (ESI†). Attempts to fit the data to an alternative model with two non-overlapping pK_as and three species: H_2A, HA^- and A^{2-} did not converge upon a solution (ESI,† Section S5).

$$f_{RS^-} = \frac{A_{RS^-}}{A_{RS^-}(\mathrm{max})} = \frac{\left(A_{obs} - A_{237}^{RSH}\right)}{\left(A_{obs} - A_{237}^{RSH}\right)_{max}} \quad (1)$$

$$f_{RS^-} = \frac{[\mathbf{ii}] + [\mathbf{iv}]}{[\mathbf{i}] + [\mathbf{ii}] + [\mathbf{iii}] + [\mathbf{iv}]} = \frac{\dfrac{K_a(\mathrm{A})}{K_a(\mathrm{B})} + \dfrac{K_a(\mathrm{D})}{[H^+]}}{\dfrac{[H^+]}{K_a(\mathrm{B})} + \dfrac{K_a(\mathrm{A})}{K_a(\mathrm{B})} + \dfrac{K_a(\mathrm{D})}{[H^+]} + 1} \quad (2)$$

The acid dissociation constants pK_a(A)–pK_a(D) determined for **1–10** are shown in Table 1. With the exception of cysteine zwitterion **1**, the pK_as are in the order:

$$\mathrm{p}K_a(\mathrm{B}) < \mathrm{p}K_a(\mathrm{A}) < \mathrm{p}K_a(\mathrm{C}) < \mathrm{p}K_a(\mathrm{D})$$

pK_a(A) and pK_a(D) both refer to deprotonation at the thiol with **i** more acidic than **iii** (pK_a(A) < pK_a(D)). This may be attributed to the greater stability of the zwitterionic conjugate base **ii** relative to **iv** due to internal electrostatic stabilisation of the thiolate anion by the ammonium cation. For acid dissociation

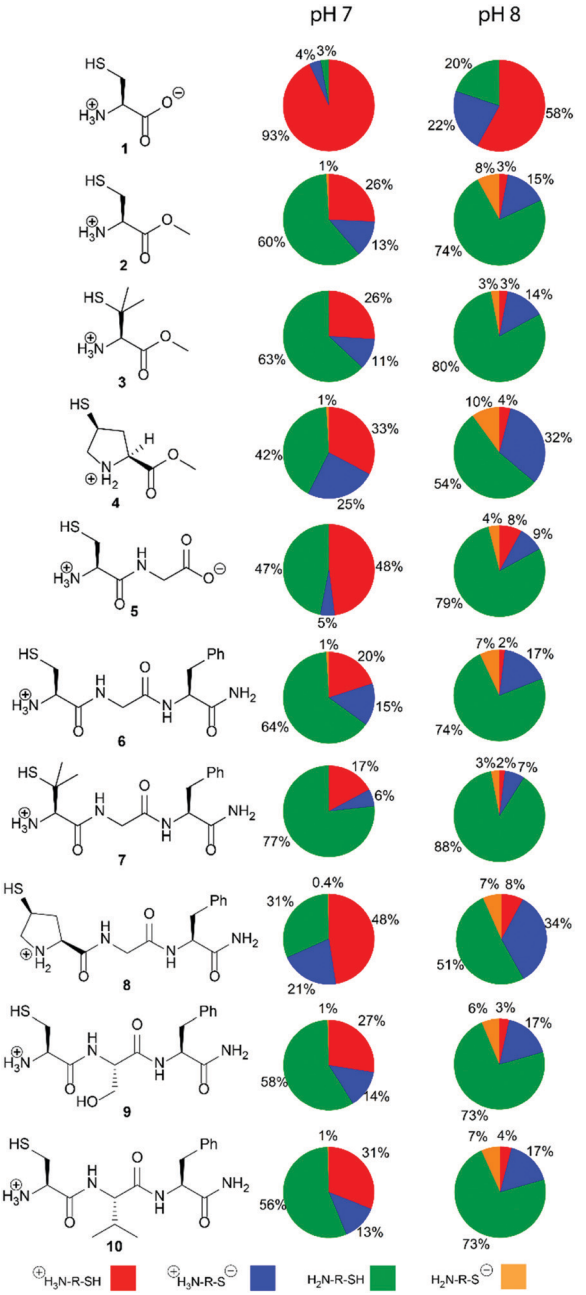

Fig. 3 The cysteine and thiolated analogues of amino acids methyl esters and peptides studied. The percentage abundance of each of the four species **i–iv** in solution at pH 7.0 and pH 8.0, 25 °C and ionic strength I = 0.3 M (NaCl) are calculated using acid dissociation constants, K_a(A)–K_a(D), reported herein. Percentages are accurate to ±6%.

of the two ammonium species, the electrostatic stabilisation of **ii** decreases acidity relative to **i** (pK_a(C) > pK_a(B)). N-terminal pK_a(A)–pK_a(D) values for cysteine zwitterion **1** are all higher than for **2–10** owing to the proximity of the anionic carboxylate,§ which destabilises all conjugate base species **ii–iv**. This effect decreases with distance or upon conversion to ester (**1** > **5** > **2**).

Thiol pK_a(A) and pK_a(D) values vary between cysteine and derivatives **1–10**. The pK_a(A) and pK_a(D) values for penicillamine derivatives **3** and **7** are higher than for cysteine analogues **2** and **6**.

Open Access Article. Published on 23 April 2020. Downloaded on 2/9/2026 10:41:30 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 1 Summary of pK_a(A)–pK_a(D) values for a range of cysteine derivatives at 25 °C, ionic strength I = 0.3 M (NaCl) with 2 mM TCEP.[a] pK_a values are accurate to ±0.07 units

	$^{\oplus}H_3N$-R-SH **i** ⇅ $^{\oplus}H_3N$-R-$S^{\ominus}$ **ii** pK_a(A)	$^{\oplus}H_3N$-R-SH **i** ⇅ H_2N-R-SH **iii** pK_a(B)	$^{\oplus}H_3N$-R-$S^{\ominus}$ **ii** ⇅ H_2N-R-$S^{\ominus}$ **iv** pK_a(C)	H_2N-R-SH **iii** ⇅ H_2N-R-$S^{\ominus}$ **iv** pK_a(D)
1 Cysteine	8.41 (8.53)[b]	8.47 (8.86)[b]	9.88 (10.36)[b]	9.83 (10.03)[b]
2 Cysteine methyl ester	7.31	6.63	8.29	8.98
	7.35[c]	6.99[c]	8.60[c]	8.95[c]
3 Penicillamine methyl ester	7.38	6.61	8.60	9.38
	7.67[c]	7.07[c]	8.71[c]	9.31[c]
4 (4*S*)-Mercaptoproline methyl ester	7.12[d]	6.89[d]	8.52[d]	8.74[d]
5 H-Cys-Gly-OH	7.97	7.01	8.34	9.30
6 H-Cys-Gly-Phe-NH_2	7.13	6.50	8.42	9.04
7 H-Pen-Gly-Phe-NH_2	7.44	6.35	8.39	9.49
8 (4*S*)-Mcp-Gly-Phe-NH_2	7.36	7.18	8.49	8.89
9 H-Cys-Ser-Phe-NH_2	7.31	6.67	8.43	9.06
10 H-Cys-Val-Phe-NH_2	7.39	6.74	8.39	9.03

[a] pK_a(A), pK_a(B) and pK_a(D) values were obtained from a fit of the percentage of thiol in thiolate form (f_{RS^-}, eqn (1)) to eqn (2) and pK_a(C) was determined using eqn (S1) (ESI). [b] Value from Benesch determined in the absence of TCEP.[13] [c] Determined in the absence of 2 mM TCEP. [d] Determined in the absence of 2 mM TCEP only as minimal oxidation was observed on the timescale of the UV-Vis spectrophotometric experiments.

The decreased acidity of the penicillamine thiols can be attributed to the two additional electron donating methyl groups in **3** and **7**, which will inductively destabilise the thiolate and/or reduce solvation of the penicillamine thiolate due to the adjacent methyl groups. Unexpectedly, the (4*S*)-mercaptoproline methyl ester **4** and peptide **8** have lower thiol pK_a(D) values than for the cysteine methyl ester **2** and peptide **6**. It might be predicted that the secondary thiol in **4** and **8** would have a higher pK_a(D) value due to greater inductive destabilisation of thiolate than for the primary thiol in **2** and **6**. To account for the observed decrease, we propose that the conformation of the pyrrolidine coupled with stereoelectronic effects alters speciation in this case. The pyrrolidine ring in proline and proline derivatives has two major C$^{\gamma}$-*endo* pucker and C$^{\gamma}$-*exo* pucker conformations (Fig. 4a). The preferred conformation is dependent upon a combination of stereoelectronic effects, minimisation of unfavourable dipole dipole interactions and whether the substituent C$^{\gamma}$ is the *R* or *S* enantiomer. Moroder has shown that (4*S*)-mercaptoproline favours the C$^{\gamma}$-*exo* ring pucker.[14] The pyrrolidine ring adopts a conformation that places the thiol group in the sterically more favourable equatorial position of the ring pucker. In the favoured conformations the C–N and C–S bonds lie in an *anti*-conformation (Fig. 4b). We postulate that this *anti*-conformation may promote the formation of the thiolate *via* stereoelectronic stabilisation by the more electronegative N counteracting an unfavourable inductive effect (Fig. 4c).

We also examined the effect of the identity of the adjacent amino acid (Gly, Ser or Val) upon the pK_a values of the N-terminal Cys residue in peptides **6**, **9** and **10**. Ser and Val were chosen to represent adjacent residues with hydrogen bonding capabilities and steric bulk, respectively. Compared to Gly, the Ser and Val residues led to increases in both pK_a(A) and pK_a(B), however, no effects upon pK_a(C) and pK_a(D) values were observed. Hydrogen bonding of the Ser of **9** to the terminal ammonium will stabilize cationic species **i** thereby increasing both pK_a(A) and pK_a(B).

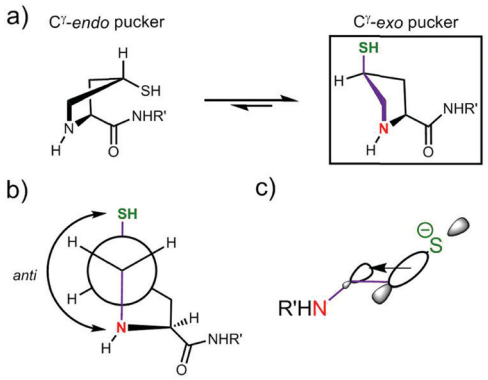

Fig. 4 Proposed stereoelectronic justification for the lower pK_D values for 4*S*-mercaptoproline methyl ester **4** and peptide **8** compared to cysteine analogues **2** and **6**: (a) the pyrrolidine ring in 4*S*-mercaptoproline **4** prefers to adopt the C$^{\gamma}$-*exo* pucker in solution;[14] (b) in the C$^{\gamma}$-*exo* pucker the C–N and C–S bonds are in an *anti*-conformation (purple coloured bonds); (c) the *anti*-conformation allows for stereoelectronic stabilisation of the thiolate by hyperconjugation.

The hydrophobic and inductively donating Val of **10** would also be predicted to favour cationic **i** relative to formally neutral **ii** and **iii** by allowing for increased aqueous solvation.

To assess the impact that our results could have for NCL we evaluated the relative concentrations of the four species **i–iv** across the whole pH range using the pK_a(A)–pK_a(D) values. The variation in the concentration of **i–iv** for **1–10** at pH 0–14 is shown in Fig. S57–S66 (ESI†). As NCL is typically performed at pH 7–8, the populations (%) of species **i–iv** at pH 7 and 8 are given in Fig. 3 and Table S20 (ESI†). Importantly, our results show that the major thiolate species is zwitterion **ii** rather than anion **iv** as commonly represented in NCL literature. For all substrates **1–10**, ≤ 1% of anion species **iv** is present at pH 7. For the cysteine peptide **6** at pH 8, the percentage of the

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 April 2020. Downloaded on 2/9/2026 10:41:30 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

concentration of **ii** (17%) is over twice that of **iv** (7%). A similar 2-fold higher concentration of **ii** *vs.* **iv** exists for the penicillamine peptide **7** at pH 8, with 10% total thiolate. For (4*S*)-Mcp peptide **8** at pH 8 the effect is more pronounced with a 5-fold higher concentration of **ii** *vs.* **iv**, and 41% in thiolate form.

We have established that four thiol(ate) species **i–iv** must be considered in NCL. Our speciation diagrams (ESI,† Fig. S57–S66) illustrate that structural modifications to the cysteine scaffold significantly change the concentrations of **i–iv** as a function of pH. While our study provides details of structure-population properties of cysteine species **i–iv**, their relative reactivities (*i.e.* nucleophilicities), and thus individual contributions to NCL, remain unknown. However, our study delivers an essential step towards delineating the complex, parallel processes that contribute to NCL, where all species **i–iv** and their associated rate constants, which are quantitative measures of nucleophilicity, must be considered (see $k_{\mathrm{NCL\text{-}i}}$–$k_{\mathrm{NCL\text{-}iv}}$, eqn (S8) and Scheme S3, Section S8, ESI†).

These speciation differences between the various cysteine and thiolated analogues could be potentially exploited to perform N-terminal kinetically controlled ligations. C-terminal kinetically controlled ligations have been demonstrated based upon reactivity differences between thioesters.[10] N-terminal kinetically controlled ligations could function *via* adjusting the rate of ligation based upon differences in thiolate concentrations and nucleophilicities using pH control. Used in conjunction with C-terminal kinetically controlled ligations, this could permit even finer control over peptide ligation.

In conclusion, we have demonstrated that four different species are present in solution for N-terminal cysteines and thiolated amino acid analogues, and have determined their pK_a(A)–pK_a(D) values. Our data highlight that two thiolate species, **ii** and **iv**, are present under NCL conditions, and must be considered. The anionic form **iv** will only be the most abundant thiolate species at higher pH values (pH > 8.3). Notably, (4*S*)-mercaptoproline methyl ester **4** and peptide **8** are unusually acidic, with lower thiol pK_a(D) values than cysteine analogues **2** and **6**, and thus higher thiolate **iv** populations at pH 7–8. We propose that additional stereoelectronic stabilisation of the thiolate in the favoured C$^{\gamma}$-*exo* ring puckered conformation of (4*S*)-Mcp favours acid dissociation. The effects of an amino acid adjacent to Cys were confined to cationic species **i**, and hence only the pK_a(A) and pK_a(B) values. These data permit the evaluation of the percentage of active thiolate species at typical NCL pH values for a range of widely used N-terminal cysteine derivatives offering (bio)chemists quantitative insight into the structural factors that influence NCL.

The authors thank Dr Jingyi Zong for help with peptide synthesis, Dr David Hodgson for helpful discussions, EPSRC (ORM, EPSRC DTA 1213247) and Cambridge Research Biochemicals Ltd. (ASH) for studentships.

Conflicts of interest

There are no conflicts to declare.

Notes and references

§ C-terminal COOH groups are deprotonated at pH > 3 (pK_a = 2.2 ± 0.4).[15]

1 S. B. H. Kent, *Annu. Rev. Biochem.*, 1988, **57**, 957.
2 (*a*) P. A. Cistrone, M. J. Bird, D. T. Flood, A. P. Silvestri, J. C. J. Hintzen, D. A. Thompson and P. E. Dawson, *Curr. Protoc. Chem. Biol.*, 2019, **11**, e61; (*b*) P. E. Dawson and S. B. H. Kent, *et al.*, *Science*, 1994, **266**, 776; (*c*) S. B. H. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338; (*d*) P. Thapa, R. Y. Zhang, V. Menon and J. P. Bingham, *Molecules*, 2014, **19**, 14461; (*e*) A. C. Conibear, E. E. Watson, R. J. Payne and C. F. W. Becker, *Chem. Soc. Rev.*, 2018, **47**, 9046.
3 R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149.
4 (*a*) D. Bang, N. Chopra and S. B. H. Kent, *J. Am. Chem. Soc.*, 2004, **126**, 1377; (*b*) D. Bang and S. B. H. Kent, *Angew. Chem., Int. Ed.*, 2004, **43**, 2534; (*c*) D. M. M. Jaradat, *Amino Acids*, 2018, **50**, 39.
5 (*a*) J. Chen and S. J. Danishefsky, *et al.*, *Tetrahedron*, 2010, **66**, 2277; (*b*) D. Crich and A. Banerjee, *J. Am. Chem. Soc.*, 2007, **129**, 10064; (*c*) C. Haase and O. Seitz, *Angew. Chem., Int. Ed.*, 2008, **47**, 1553; (*d*) S. S. Kulkarni, J. Sayers, B. Premdjee and R. J. Payne, *Nat. Rev. Chem.*, 2018, **2**, 0122; (*e*) P. Siman, S. V. Karthikeyan and A. Brik, *Org. Lett.*, 2012, **14**, 1520; (*f*) R. E. Thompson and R. J. Payne, *et al.*, *Angew. Chem., Int. Ed.*, 2013, **52**, 9723; (*g*) S. D. Townsend and S. J. Danishefsky, *et al.*, *J. Am. Chem. Soc.*, 2012, **134**, 3912; (*h*) R. L. Yang and C. F. Liu, *et al.*, *J. Am. Chem. Soc.*, 2009, **131**, 13592.
6 (*a*) R. J. Hondal, B. L. Nilsson and R. T. Raines, *J. Am. Chem. Soc.*, 2001, **123**, 5140; (*b*) R. Quaderer, A. Sewing and D. Hilvert, *Helv. Chim. Acta*, 2001, **84**, 1197; (*c*) M. D. Gieselman and W. A. van der Donk, *et al.*, *Org. Lett.*, 2001, **3**, 1331; (*d*) L. R. Malins and R. J. Payne, *Org. Lett.*, 2012, **14**, 3142; (*e*) R. Quaderer and D. Hilvert, *Chem. Commun.*, 2002, 2620, DOI: 10.1039/b208288h; (*f*) A. L. Braga and D. P. Bottega, *et al.*, *J. Org. Chem.*, 2006, **71**, 4305.
7 (*a*) L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526; (*b*) B. L. Pentelute and S. B. H. Kent, *Org. Lett.*, 2007, **9**, 687; (*c*) Q. Wan and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2007, **46**, 9248; (*d*) K. Jin and X. C. Li, *et al.*, *Angew. Chem., Int. Ed.*, 2017, **56**, 14607; (*e*) T. S. Chisholm, D. Clayton, L. J. Dowman, J. Sayer and R. J. Payne, *J. Am. Chem. Soc.*, 2018, **140**, 9020; (*f*) N. Ollivier, T. Toupy, R. C. Hartkoorn, R. Desmet, J. M. Monbalieu and O. Melnyk, *Nat. Commun.*, 2018, **9**, 2847.
8 (*a*) L. R. Malins and R. J. Payne, *et al.*, *Angew. Chem., Int. Ed.*, 2015, **54**, 12716; (*b*) P. S. Reddy, S. Dery and N. Metanis, *Angew. Chem., Int. Ed.*, 2016, **55**, 992; (*c*) N. Metanis, E. Keinan and P. E. Dawson, *Angew. Chem., Int. Ed.*, 2010, **49**, 7049.
9 (*a*) L. E. Canne and S. B. H. Kent, *et al.*, *J. Am. Chem. Soc.*, 1996, **118**, 5891; (*b*) J. Offer and P. E. Dawson, *et al.*, *J. Am. Chem. Soc.*, 2002, **124**, 4642.
10 (*a*) D. Bang, B. L. Pentelute and S. B. H. Kent, *Angew. Chem., Int. Ed.*, 2006, **45**, 3985; (*b*) J. Lee, Y. Kwon, B. L. Pentelute and D. Bang, *Bioconjugate Chem.*, 2011, **22**, 1645; (*c*) V. Y. Torbeev and S. B. H. Kent, *Angew. Chem., Int. Ed.*, 2007, **46**, 1667; (*d*) E. E. Watson, X. Liu, R. E. Thompson, J. Ripoll-Rozada, M. Wu, I. Alwis, A. Gori, C.-T. Loh, B. L. Parker and G. Otting, *ACS Cent. Sci.*, 2018, **4**, 468.
11 (*a*) S. Ficht, A. Mattes and O. Seitz, *J. Am. Chem. Soc.*, 2004, **126**, 9970; (*b*) J. Sayers, P. M. T. Karpati, N. J. Mitchell, A. M. Goldyss, S. M. Kwong, N. Firth, B. Chan and R. J. Payne, *J. Am. Chem. Soc.*, 2018, **140**, 13327.
12 (*a*) P. E. Dawson, M. J. Churchill, M. R. Ghadiri and S. B. H. Kent, *J. Am. Chem. Soc.*, 1997, **119**, 4325; (*b*) E. C. B. Johnson and S. B. H. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640.
13 R. E. Benesch and R. Benesch, *J. Am. Chem. Soc.*, 1955, **77**, 5877.
14 S. A. Cadamuro and L. Moroder, *et al.*, *Angew. Chem., Int. Ed.*, 2008, **47**, 2143.
15 A. Fersht, *Structure and Mechanism in Protein Science*, Freeman, New York, 2002, ch. 1, pp. 2–3.