

Showcasing research from Professor David M. Chenoweth's laboratory, School of Arts & Sciences, Department of Chemistry, University of Pennsylvania, Philadelphia, PA, USA.

**Aza-proline effectively mimics L-proline stereochemistry in triple helical collagen**

Recently, biophysical studies have demonstrated that natural L-amino acids can be replaced with non-natural achiral aza-amino acids in folded protein structures. However, the structural consequences of achiral aza-amino acid incorporation have not been elucidated in any relevant folded biomolecule. Herein, Chenoweth and co-workers present an atomic resolution crystal structure and computational analysis illustrating that aza-proline mimics L-proline stereochemistry in triple helical collagen. In addition to providing fundamental insight into peptide and protein structure, the incorporation of aza-amino acids could have impacts in areas ranging from synthetic materials to drug design.

**As featured in:**



See David M. Chenoweth *et al.*, *Chem. Sci.*, 2019, 10, 6979.

Cite this: *Chem. Sci.*, 2019, 10, 6979

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

## Aza-proline effectively mimics L-proline stereochemistry in triple helical collagen†

Alexander J. Kasznel,<sup>ab</sup> Trevor Harris,<sup>a</sup> Nicholas J. Porter,<sup>a</sup> Yitao Zhang<sup>a</sup> and David M. Chenoweth<sup>\*a</sup>

The prevalence of L-amino acids in biomolecules has been shown to have teleological importance in biomolecular structure and self-assembly. Recently, biophysical studies have demonstrated that natural L-amino acids can be replaced with non-natural achiral aza-amino acids in folded protein structures such as triple helical collagen. However, the structural consequences of achiral aza-amino acid incorporation has not been elucidated in the context of any relevant folded biomolecule. Herein, we use X-ray crystallography to provide the first atomic resolution crystal structure of an achiral aza-amino acid residue embedded within a folded protein structure, definitively illustrating that achiral aza-proline has the capacity to effectively mimic the stereochemistry of natural amino acids within the context of triple helical collagen. We further corroborate this finding with density functional theory computational analysis showing that the natural L-amino acid stereochemistry for aza-proline is energetically favored when arranged in the aza-proline-hydroxyproline-glycine motif. In addition to providing fundamental insight into peptide and protein structure, the incorporation of achiral stereochemical mimics such as aza-amino acids could have far reaching impacts in areas ranging from synthetic materials to drug design.

Received 5th May 2019  
Accepted 19th June 2019

DOI: 10.1039/c9sc02211b

rsc.li/chemical-science

## Introduction

A prevalent feature of structural biology is the “monochirality of life,” best exemplified by the ubiquity of isotactic polymers in biological systems.<sup>1</sup> In nature, both proteins and nucleic acids are predominately homochiral. This property has been shown to facilitate biomolecular interaction and also enable expedient protein folding by lowering the entropy of biomolecular assembly.<sup>1b,2</sup> The chirality of life's amino acid building blocks is thus a crucial factor in the synthesis and stability of essential biomolecules.

Collagen is a critical structural protein in mammals. The principal characteristics of collagen are its triple helical structure and triplet repeat amino acid sequence (XaaYaaGly)<sub>n</sub>; the Xaa and Yaa positions are typically filled by proline (Pro, P) and the post-translationally modified amino acid hydroxyproline (Hyp, O), respectively.<sup>3</sup> Our lab has investigated the structure and function of collagen by synthesizing an array of collagen model peptides with novel biophysical properties, most notably by integrating the non-natural amino acids aza-glycine (azGly, azG)<sup>4</sup> and aza-proline (azPro, azP).<sup>5</sup> These aza-peptides have the

α-CH group from one or more amino acid residues in their peptide chain replaced with a N atom.<sup>6</sup> This seemingly small structural change can have substantial effects on peptide structure and electronics such as increasing intrinsic chemical stability and resistance towards biodegradation<sup>7</sup> and improving collagen stability and self-assembly.<sup>4b,c</sup> These properties have led to the pursuit of aza-peptides as promising candidates for therapeutic applications,<sup>8</sup> which generated a wave of new solid-phase synthetic methods<sup>9</sup> and fundamental studies.<sup>10</sup> Collagen peptides containing non-natural amino acid substitutions that do not disturb the overall collagen topology would be useful biological tools to probe the collagen interactome.<sup>3b</sup>

The added nitrogen atom in azPro is notable for its sp<sup>3</sup>-like pyramidalized conformation (Fig. 1A).<sup>10a,11</sup> Although the alteration of the α-CH stereocenter in Pro to N results in a net isoelectronic change, this modification has a large effect on the conformational properties of the α-stereocenter and gives the azPro ring the ability to mimic either D- or L-amino acid stereochemistry. Unlike electronically neutral carbon, nitrogen has a lone pair that can rehybridize during pyramidal inversion, turning this inversion into a dynamic process.<sup>12</sup>

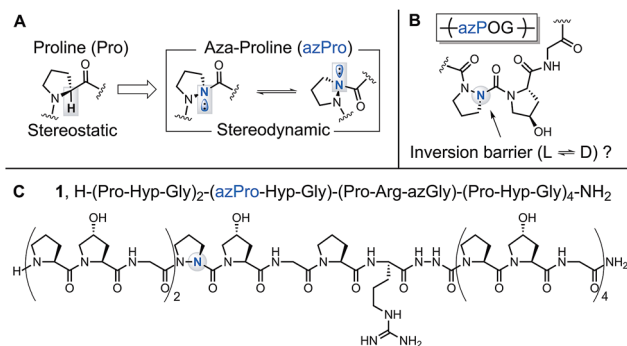
In a previous study of a collagen model peptide (CMP), the added stereodynamic center in azPro slowed the rate of triple helical self-assembly.<sup>5</sup> However, because azPro is typically studied in the context of short β-turn mimics,<sup>10a,f,11,13</sup> many details of azPro's structural behavior in collagen remain uncharacterized: (1) Does the added N atom in azPro pyramidalize in collagen? If so, does the triple helical structure

<sup>a</sup>Department of Chemistry, University of Pennsylvania, 231 S. 34th St., Philadelphia, PA 19104-6323, USA. E-mail: dcheno@sas.upenn.edu

<sup>b</sup>Department of Bioengineering, University of Pennsylvania, 210 S. 33rd St., Philadelphia, PA 19104-6323, USA

† Electronic supplementary information (ESI) available: A detailed explanation of all procedures for synthesis, purification, characterization, crystallography, and computational analysis. See DOI: 10.1039/c9sc02211b





**Fig. 1** Synthetic modification to collagen peptide backbone using azPro. (A) Backbone structures of Pro and azPro illustrating stereodynamic N atom in azPro. (B) Inversion of azPro in the context of an azPOG triplet in collagen. (C) Backbone structure of CMP 1 showing site of azPro substitution.

force azPro's pyramidal conformation, or do the underpinning steric and conformational constraints of the azPro backbone play a greater role in determining this? (2) Does azPro mimic the conformation of the canonical Pro residue in collagen? (3) Finally, what is the barrier to azPro's pyramidal inversion in the context of a collagen amino acid sequence (Fig. 1B)?

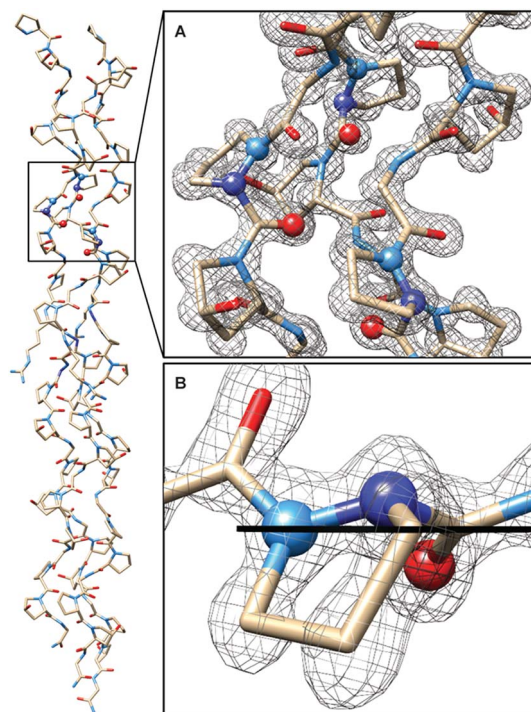
To answer these questions, we synthesized azPro-containing CMP 1, H-(POG)<sub>2</sub>(azPOG)(PRazG)(POG)<sub>4</sub>-NH<sub>2</sub> (Fig. 1C), and determined its structure using X-ray crystallography. We also performed density functional theory (DFT) calculations to elucidate azPro's stereodynamic nature in a collagen peptide. To date, there have been few published examples of azPro in the crystal state.<sup>11a,14</sup> Notably, none of these have shown azPro in the context of any biologically relevant macromolecule or higher-order molecular assembly. In this publication, we present the first atomic resolution crystal structure of collagen containing azPro, and our findings indicate that azPro is a structural mimic of its natural amino acid counterpart in collagen.

## Results and discussion

### Crystallography

In CMP 1, one Pro residue from each strand was replaced with its aza-derivative azPro (Fig. 1C). This amino acid sequence was engineered to produce a homotrimeric peptide containing a total of 3 azPro residues when fully assembled. CMP 1 was synthesized on solid phase and crystallized by sitting-drop vapor diffusion (detailed procedures in ESI†). The crystal structure of 1 (PDB 6M80) reveals that this CMP still forms a triple helix despite the introduction of azPro to its backbone and that azPro does adopt a pyramidal configuration within this helix (Fig. 2). This structural evidence illustrates that azPro is not held in a quasiplanar state and that the superseding energetics of the collagen assembly could compensate for this destabilized azPro configuration.

Since the dominating structural framework and energetics of the triple helix can theoretically hold azPro in either a quasiplanar or pyramidal configuration, we investigated whether the overall triple helical structure or azPro's fundamental backbone



**Fig. 2** Crystal structure of CMP 1 solved to 1.10 Å resolution. (Left) Structure confirming formation of triple helix ( $R_{\text{work}} = 11.7\%$ ,  $R_{\text{free}} = 14.4\%$ ,  $P2_1$ ). (A) Magnified view of triple helix with electron density at site of azPro integration (2mFo-DFc map contoured at  $1.8\sigma$ ). (B) Close-up of single azPro residue with plane bisecting atoms connected to pyramidalized nitrogen. Light blue = nitrogens already present in collagen, dark blue = added nitrogen in azPro.

architecture were responsible for azPro's pyramidal configuration. To determine this, we examined 26 unique structures from the Cambridge Structural Database (CSD) containing *N*-amido-urea moieties of the general form  $\text{RN}_x(\text{O})\text{C}-\text{N}_y\text{R}-\text{N}_z\text{R}-\text{C}(\text{O})\text{R}$  (where  $\text{N}_y$  is the pyramidal atom in azPro) in their respective backbones (see Fig. S1–S3 and Table S1†). We observed an increase in pyramidalization in the  $\text{N}_y$  atom position as the degree of substitution increased at the  $\text{N}_x$ ,  $\text{N}_y$ , and  $\text{N}_z$  atoms. Pyramidalization becomes more pronounced at the  $\text{N}_y$  atom when it is incorporated into a cycle as it is in azPro. Integrating  $\text{N}_y$  into a cycle creates steric and conformational constraints on the conjugated amide framework by restricting the main-chain phi dihedral ( $\text{C}-\text{N}_y-\text{N}_z-\text{C}$ ). This leads the system to balance several structural and electronic factors such as amide resonance, ring strain, and sterics. As a result, amide resonance is weakened in azPro, the effects of which can be seen in its non-planar structure and long amide N–CO bonds. The pyramidal N sites of the three azPro rings in the triple helix of CMP 1 have N–CO bond distances of 1.42, 1.44, and 1.48 Å, longer than a typical tertiary amide bond length of  $\sim 1.34$  Å.<sup>15</sup> This indicates that azPro's pyramidalization is connected to the underlying functionality in its backbone rather than the long-range forces of the triple helix.

The combined structural distortions observed in the azPro residues also resemble the features of classical twisted amides



(Table S2†).<sup>16</sup> However, because of delocalization throughout the conjugated amide framework, azPro residues do not exhibit the same high reactivity as classical twisted amides. The extended conjugated backbone provides azPro with increased resonance stabilization although, due to the aforementioned structural distortions (pyramidalization and twisting), resonance is slightly weakened.

To further investigate the structural implications of azPro substitution in collagen, we compared the crystal structure of CMP **1** to that of previously published triple-helical CMP **2**, H-(POG)<sub>3</sub>(PRazG)(POG)<sub>4</sub>-NH<sub>2</sub>, which contains Pro instead of azPro.<sup>4a</sup> This comparison indicates an overall RMSD of 0.36 Å for all atoms (468 atom pairs; H atoms excluded), and more specifically reveals that the azPro residues in **1** adopt the same conformation as the corresponding Pro residues in **2** (Fig. 3).

The angles around the pyramidal N sites of the azPro residues in CMP **1** are slightly shallower than the angles around the tetrahedral C sites of the corresponding canonical Pro residues in CMP **2** (Tables S3 and S4†). Taken together, these data indicate that CMPs containing azPro can still self-assemble into the biologically essential collagen triple helix and that this modification does not disturb the conformations of the peptide backbone or amino acid side chains.

### Computational analysis

To study the self-assembly of this azPro-containing CMP, we characterized the stereodynamic nature of azPro using DFT. We first compared the crystal structure fragment azPro-Hyp-Gly (**1-a**) and the low-energy conformer of the dipeptide model system Ac-azPro-Hyp-OMe (**3**).<sup>5</sup> The overlay of these two structures reveals that their conformations match closely, with an overall RMSD of 0.12 Å (Fig. 4A). This statistical analysis leads us to conclude that azPro readily adopts the conformation necessary for collagen self-assembly rather than being forced into a pyramidal conformation by the overall triple helical structure of the peptide strands. Since the calculated collagen model system mimics the azPro-Hyp-Gly fragment from the collagen triple helix accurately, we used it to determine the energetic cost for pyramidal inversion of azPro. Pyramidal inversion of amines

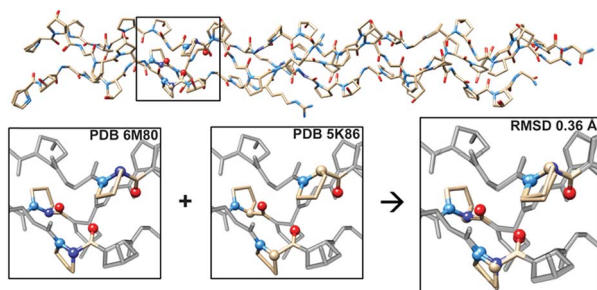


Fig. 3 Aza-proline mimics proline in backbone of CMP. (Top) Superposition of crystal structures of CMPs **1** and **2** showing close matching of backbone and side chain conformations throughout; RMSD = 0.36 Å. (Bottom) Close-up of boxed region in **1** (PDB 6M80), **2** (PDB 5K86), and overlay of the two structures, illustrating correspondence between conformation of azPro7 in **1** and Pro7 in **2**.

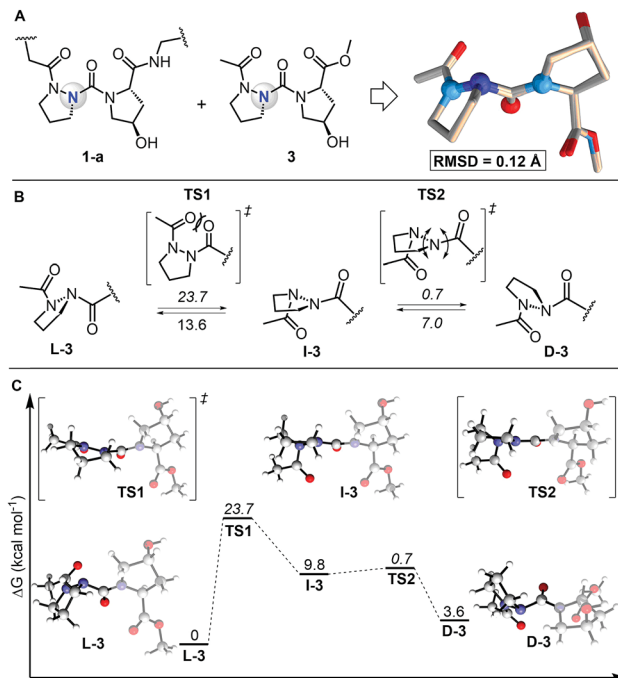


Fig. 4 Comparison of calculated model system with experimental collagen fragment and free energy profile for interconversion of backbone N atom in azPro. (A) Overlay of azPro-Hyp-Gly crystal structure fragment (PDB 6M80) and model dipeptide system Ac-azPro-Hyp-OMe (**3**). **3** was calculated using M06-2X/6-31+G(d,p)/SMD = H<sub>2</sub>O. 20 atom pairs were used for RMSD; H atoms were excluded. (B) Pyramidal inversion of azPro in context of **3**. (C) Free energy profile for inversion of azPro. Energies are relative to L-**3** except for barriers (italicized), which are relative to preceding minima.

typically proceeds through one planar transition state in which nitrogen's bonding orbitals are  $\sim$ sp<sup>2</sup> hybridized and its lone pair occupies a p orbital.<sup>17</sup> Conversely, we found that the overall inversion of azPro (L-**3** to D-**3**, where the prefixes indicate the stereochemical notation for azPro) is composed of two transition states connected by intermediate I-**3** (Fig. 4B). The first transition state TS1 involves steric repulsion between the central carbonyl group, which connects the Hyp ring to the azPro ring, and the acyl carbonyl group as the two cross paths. The distance between the two oxygen atoms of their respective carbonyls in the transition state (2.7 Å) is smaller than the sum of their van der Waals radii (The van der Waals radius of oxygen is 1.52 Å). This barrier, >20 kcal mol<sup>-1</sup>, also reflects the endergonic nature of I-**3**, which is almost 10 kcal mol<sup>-1</sup> higher than L-**3**.

The second transition state TS2, with an activation barrier of only  $\sim$ 1 kcal mol<sup>-1</sup>, returns the quasi-planar azPro to a completely pyramidal geometry in D-**3**. The free energy profile for the inversion pathway reveals a conformational bias as D-**3** is  $\sim$ 4 kcal mol<sup>-1</sup> less stable than the conformer needed to form the triple helix, L-**3** (Fig. 4C). Since D-amino acids have been shown to destabilize the triple helix,<sup>5,18</sup> azPro's energetic preference for the L-configuration when arranged in the azPOG triplet should help facilitate the stable triple helical assembly of **1**. While the system shows a conformational preference for L-**3**, the inversion barrier indicates that this does not preclude D-**3**

from being populated. This contrasts with a stereostatic Pro equivalent in which only both enantiomers can exist after racemization. This distinction could facilitate the use of entropic or enthalpic external factors to control the stereo-dynamics of azPro in peptides.<sup>19</sup>

In this study, crystallographic data verifies that the added N atom in azPro adopts a pyramidalized conformation and that azPro is a near perfect mimic for Pro in this CMP. Furthermore, DFT calculations indicate that the conformation of azPro in this CMP is energetically favorable and that azPro is not simply forced into this conformation by the overall helical structure of the peptide. Elucidating these previously uncharacterized structural properties of azPro in collagen will enable and inform future studies of the biophysical properties of aza-amino acids in biologically relevant molecules. In addition, these findings highlight the teleological importance of stereochemical pre-organization on self-assembly in collagen. Although azPro can structurally mimic L-proline in collagen, because of the slower rate of triple helical self-assembly, configurationally labile amino acids would dramatically alter collagen's evolutionary timescales. Using azPro as a stereodynamic probe can reveal fundamental insight into the interplay of structural forces and folding processes in biomolecular systems.

## Experimental procedures

A detailed explanation of all procedures for synthesis, purification, characterization, crystallography, and computational analysis can be found in the ESI.† Crystallographic data for CMPs 1 and 2 including atomic coordinates and structure factors are deposited in the Protein Data Bank (PDB) with accession codes 6M80 and 5K86, respectively.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by funding from the University of Pennsylvania and from the National Science Foundation (Penn MRSEC Grant No. DMR-1720530). Instruments were supported by the National Science Foundation and the National Institutes of Health including MALDI-TOF MS (NSF MRI-0820996) and HRMS (NIH RR-023444). The authors would like to thank Dr David Christianson for use of his crystallography facilities, Dr Travis Pemberton for his guidance with data collection and structure refinement, Dr Mason Smith for his guidance with structure refinement, and Andrew Bright for his contributions to peptide synthesis, characterization, and crystallization. For synchrotron access, we are grateful to the Northeastern Collaborative Access Team beamline 24-ID-C at the Advanced Photon Source, Argonne National Laboratory. We thank Dr Gabriel dos Passos Gomes for helpful discussions regarding calculations and the University of Pennsylvania SAS GPC for computational resources.

## Notes and references

- (a) J. H. Brewster, *J. Chem. Educ.*, 1986, **63**, 667–670; (b) D. G. Blackmond, *Cold Spring Harbor Perspect. Biol.*, 2010, **2**, a002147.
- V. Nanda, A. Andrianarijaona and C. Narayanan, *Protein Sci.*, 2007, **16**, 1667–1675.
- (a) M. D. Shoulders and R. T. Raines, *Annu. Rev. Biochem.*, 2009, **78**, 929–958; (b) B. An, Y.-S. Lin and B. Brodsky, *Adv. Drug Delivery Rev.*, 2016, **97**, 69–84.
- (a) A. J. Kasznel, Y. Zhang, Y. Hai and D. M. Chenoweth, *J. Am. Chem. Soc.*, 2017, **139**, 9427–9430; (b) Y. Zhang, M. Herling and D. M. Chenoweth, *J. Am. Chem. Soc.*, 2016, **138**, 9751–9754; (c) Y. Zhang, R. M. Malamakal and D. M. Chenoweth, *J. Am. Chem. Soc.*, 2015, **137**, 12422–12425.
- (a) Y. Zhang, R. M. Malamakal and D. M. Chenoweth, *Angew. Chem., Int. Ed.*, 2015, **54**, 10826–10832; *Angew. Chem.*, 2015, **127**, 10976–10982; (b) M. R. Aronoff, J. Egli, M. Menichelli and H. Wennemers, *Angew. Chem., Int. Ed.*, 2019, **58**, 3143–3146; *Angew. Chem.*, 2019, **131**, 3175–3178.
- J. Gante, *Synthesis*, 1989, **21**, 405–413.
- (a) J. M. Gassman and J. Magrath, *Bioorg. Med. Chem. Lett.*, 1996, **6**, 1771–1774; (b) P. Wipf, W. Li, C. M. Adeyeye, J. M. Rusnak and J. S. Lazo, *Bioorg. Med. Chem.*, 1996, **4**, 1585–1596; (c) C. Dugave and L. Demange, *Lett. Pept. Sci.*, 2003, **10**, 1–9; (d) A. Fässler, G. Bold, H.-G. Capraro, R. Cozens, J. Mestan, B. Poncioni, J. Rösel, M. Tintelnot-Blomley and M. Lang, *J. Med. Chem.*, 1996, **39**, 3203–3216; (e) S. V. Andurkar, C. Béguin, J. P. Stables and H. Kohn, *J. Med. Chem.*, 2001, **44**, 1475–1478.
- (a) C. Proulx, D. Sabatino, R. Hopewell, J. Spiegel, Y. G. Ramos and W. D. Lubell, *Future Med. Chem.*, 2011, **3**, 1139–1164; (b) A. Begum, D. Sujatha, K. V. S. R. G. Prasad and K. Bharathi, *Asian J. Chem.*, 2017, **29**, 1879–1887; (c) A. Zega, *Curr. Med. Chem.*, 2005, **12**, 589–597.
- (a) H. Han and K. D. Janda, *J. Am. Chem. Soc.*, 1996, **118**, 2539–2544; (b) F. Bizet, N. Tonali, J.-L. Soulier, A. Oliva, J. Kaffy, B. Croussea and S. Onger, *New J. Chem.*, 2018, **42**, 17062–17072; (c) R. Chingle, C. Proulx and W. D. Lubell, *Acc. Chem. Res.*, 2017, **50**, 1541–1556; (d) M. Arujõe, A. Ploom, A. Mastitski and J. Järv, *Tetrahedron Lett.*, 2018, **59**, 2010–2013; (e) A. S. Dutta and J. S. Morley, *J. Chem. Soc., Perkin Trans. 1*, 1975, **17**, 1712–1720; (f) N. S. Freeman, Y. Tal-Gan, S. Klein, A. Levitzki and C. Gilon, *J. Org. Chem.*, 2011, **76**, 3078–3085; (g) R. E. Melendez and W. D. Lubell, *J. Am. Chem. Soc.*, 2004, **126**, 6759–6764; (h) D. Sabatino, C. Proulx, S. Kloczek, C. B. Bourguet, D. Boeglin, H. Ong and W. D. Lubell, *Org. Lett.*, 2009, **11**, 3650–3653.
- (a) Y. Che and G. R. Marshall, *J. Org. Chem.*, 2004, **69**, 9030–9042; (b) X.-S. Yan, H. Luo, K.-S. Zou, J.-L. Cao, Z. Li and Y.-B. Jiang, *ACS Omega*, 2018, **3**, 4786–4790; (c) H. J. Lee, K. B. Lee, I. A. Ahn, S. Ro, K. H. Choi and Y. S. Choi, *J. Pept. Res.*, 2000, **56**, 35–46; (d) E. Gloaguen, V. R. Brenner, M. Alauddin, B. Tardivel, M. Mons, A. Zehnacker-Rentien, V. R. Declerck and D. J. Aitken, *Angew. Chem., Int. Ed.*,



- 2014, **53**, 13756–13759; *Angew. Chem.*, 2014, **126**, 13976–13979; (e) M. Alauddin, E. Gloaguen, V. Brenner, B. Tardivel, M. Mons, A. Zehnacker-Rentien, V. Declerck and D. J. Aitken, *Chem.–Eur. J.*, 2015, **21**, 16479–16493; (f) M. Zouikri, A. Vicherat, A. Aubry, M. Marraud and G. Boussard, *J. Pept. Res.*, 1998, **52**, 19–26.
- 11 (a) A. Lecoq, G. Boussard, M. Marraud and A. Aubry, *Biopolymers*, 1993, **33**, 1051–1059; (b) Y. Che and G. R. Marshall, *Biopolymers*, 2006, **81**, 392–406.
- 12 I. V. Alabugin, S. Bresch and G. d. P. Gomes, *J. Phys. Org. Chem.*, 2015, **28**, 147–162.
- 13 (a) W.-J. Zhang, A. Berglund, J. L.-F. Kao, J.-P. Couty, M. C. Gershengorn and G. R. Marshall, *J. Am. Chem. Soc.*, 2003, **125**, 1221–1235; (b) Y. K. Kang and B. J. Byun, *J. Phys. Chem. B*, 2007, **111**, 5377–5385.
- 14 C. Didierjean, A. Aubry, M. Zouikri, G. Boussard and M. Marraud, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 1995, **51**, 688–690.
- 15 F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen and R. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1987, S1–S19.
- 16 C. Liu and M. Szostak, *Chem.–Eur. J.*, 2017, **23**, 7157–7173.
- 17 J. M. Lehn, *Top. Curr. Chem.*, 1970, **15**, 311–377.
- 18 (a) J. C. Horng, F. W. Kotch and R. T. Raines, *Protein Sci.*, 2007, **16**, 208–215; (b) N. K. Shah, B. Brodsky, A. Kirkpatrick and J. A. M. Ramshaw, *Biopolymers*, 1999, **49**, 297–302.
- 19 S. Wang, T. Taniguchi, K. Monde, M. Kawahata, K. Yamaguchi, Y. Otani and T. Ohwada, *Chem. Commun.*, 2016, **52**, 4018–4021.

