ChemComm

Variation in the Yaa Position of Collagen Peptides Containing AzaGlycine

Journal:	*ChemComm*
Manuscript ID	CC-COM-08-2018-006372.R2
Article Type:	Communication

Please do not adjust margins

ChemComm

COMMUNICATION

Variation in the Yaa Position of Collagen Peptides Containing AzaGlycine

Samuel D. Melton and David M. Chenoweth*

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Herein, we report the systematic investigation of amino acid variation in the Yaa position of collagen peptides containing an adjacent azaGlycine residue. We demonstrate the reliability of azaGlycine as a glycine replacement and provide a sequence independent strategy for stabilizing the triple helical assembly of collagen peptides.

Protein folding is carefully regulated through intra- and intermolecular interactions between specific side chains and backbone functional groups of amino acids. Hydrogen bonding (H-bonding) along with a host of other weak interactions serve to stabilize protein structure.[1–4] Unnatural modifications to the protein backbone and side chains have been used in a variety of contexts to modulate the properties of proteins.[5–13] An important class of unnatural peptide modifications are aza-peptides, many of which display biological activity while conveying a host of beneficial properties including resistance to enzymatic degradation and improved therapeutic efficacy.[14–16] Aza-amino acids are characterized by substituting the α-carbon, or some other adjacent position, with nitrogen (Fig. 1a). Some of these amino acid analogs exhibit a higher degree of planarity and restricted dihedral angles compared to natural amide linkages, making them valuable tools in peptide mimicry and structural studies.[7,14,17,18] The aza substitution results in the NH of the backbone amide being attached to an α-nitrogen rather than an α-carbon, which can strengthen NH hydrogen bonds due to increased acidity.[19]

Hydrogen bonds are thought to play a critical role in stabilizing the quaternary structure of the collagen triple helix. This stabilization involves the assembly of three left-handed polyproline II helices into a right-handed triple helix.[20,21] Collagen's primary structure consists of triplet repeats (i.e., Xaa-Yaa-Glycine). When assembled into a triple helix, this repeat

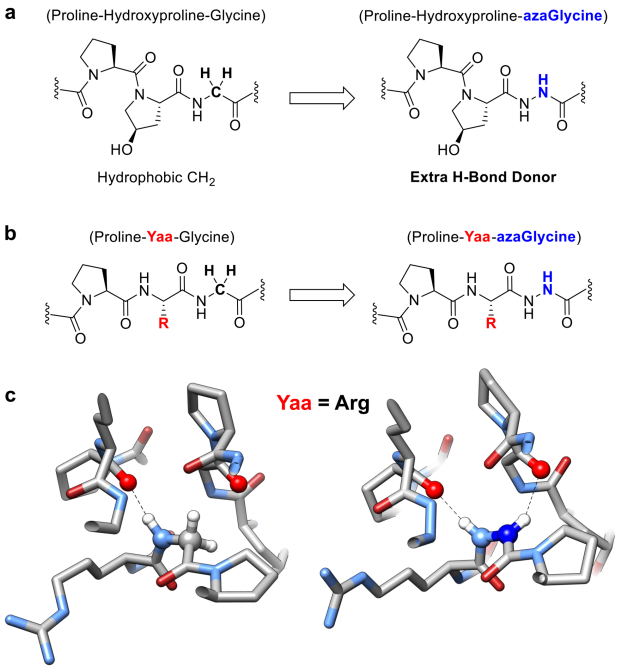


Fig 1. (a) Structural representation of azGly substitution. (b) Structural representation of Yaa and azGly substitutions. (c) Comparison of crystal structure cross sections of an Arg-containing CMP without (left) and with (right) an adjacent azGly substitution, highlighting the hydrogen bonds of the Gly/azGly residue (PDB IDs: 3WN8 (left) and 5K86 (right)).[24,25] The additional nitrogen of azGly is displayed in dark blue.

structure results in interstrand H-bonds between glycine (Gly, G) and the Xaa position residues, producing a densely packed helical core studded with desolvated interstrand H-bonds.[22,23] The Xaa and Yaa positions exhibit a high level of variation with the most common residues being proline (Pro, P) and 4-hydroxyproline (Hyp, O), respectively.[23] Pro-Hyp-Gly (POG) is the most stabilizing repeating unit due to the restricted conformation of the cyclic imino residues, which enforce the preferred geometry for helix formation.[26,27] Variability in the Xaa and Yaa positions is essential for collagen protein-protein interactions, although deviation from stable POG triplets can have a significant destabilizing effect on the triple helix.

Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104, United States. E-mail: dcheno@sas.upenn.edu
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

Recently, our group reported a strategy for stabilizing the collagen triple helix by substituting the conserved Gly residues with azaGlycine (azGly, azG).[28,29]

With the use of azGly in collagen peptides, the substitution of nitrogen for carbon results in an additional H-bond donor within the helical bundle, when present in Proline-Hydroxyproline-azaGlycine (POazG) and Proline-Arginine-azaGlycine (PRazG) triplets (Fig. 1c).[24,28,29] Despite these initial studies, the context dependence of azGly has not been thoroughly investigated for alternate adjacent residues. Given the breadth of residues required for collagen protein recognition and the destabilizing effect of these substitutions with regard to triple helix stability, the synthesis of short, helical collagen mimetic peptides (CMPs) containing protein binding motifs requires improved methods with increased stability.

Herein, we systematically investigate the effect of azGly incorporation when located adjacent to a variable Yaa substitution within a collagen peptide. Compared to the Xaa position, substitutions at the Yaa are often more destabilizing and result in a larger range of thermal stabilities.[30] We synthesized a library of azGly-containing CMPs and evaluated the triple helical stability compared to control peptides using circular dichroism (CD).[26,27,31–33] Our results demonstrate that the degree of azGly-induced stability is dependent on the adjacent Yaa residue. Furthermore, we developed a sequence independent strategy for stabilizing triple helical collagen using azGly residues as H-bond clamps. These results will help inform the design of short, stable, azGly-containing CMPs bearing a host of sequences for collagen protein recognition.

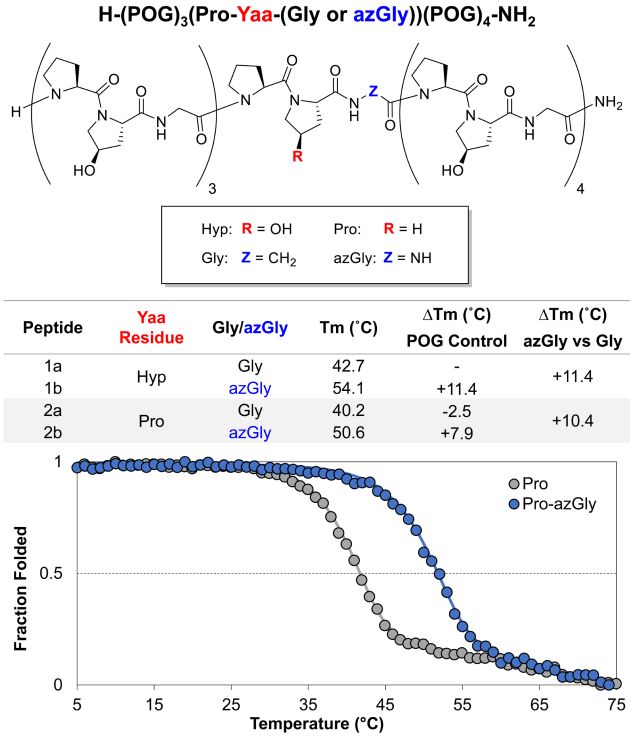


Fig 2. (Top) Sequence and structural representation of Yaa position substitutions. (Middle) T_m values showing the degree of azGly thermal stabilization within each pair of CMPs, e.g. **1a** and **1b**. (Bottom) CD thermal melting data for CMPs **2a** and **2b**.

A control series of peptides were synthesized with the H-$(POG)_3$(Pro-Yaa-Gly)$(POG)_4$-NH_2 general form to determine the effect of a central Yaa substitution on the melting temperature (T_m) of the collagen triple helix compared to the most stable natural CMP in the library (i.e., H-$(POG)_8$-NH_2) (Fig. 2, CMP **1a**). These peptides provided a set of baseline controls for evaluating the effects of azGly substitutions (Chart 1, CMPs **3-22a**). The observed T_m values (obtained in triplicate) were largely consistent with previous reports for a similar host-guest system.[33]

A second series of peptides was synthesized with the H-$(POG)_3$(Pro-Yaa-azGly)$(POG)_4$-NH_2 general form to evaluate the effect of azGly incorporation adjacent to the Yaa substitution (Chart 1, CMPs **3-22b**). The T_m values for these azGly CMPs were directly compared to their corresponding controls with the same Yaa substitution. The azGly stabilization trends can be broadly divided into categories based on the properties of the Yaa side chain as follows: cyclic, charged, polar, hydrophobic, and aromatic.

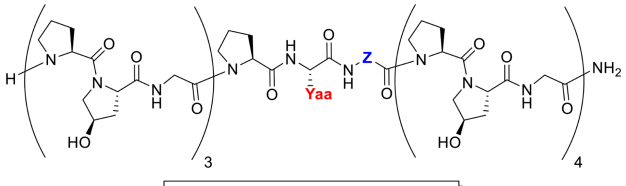
The CMPs containing cyclic side chains and azGly residues (**1b** (Hyp) and **2b** (Pro)) exhibited significant increases in stability (Fig. 2). The 10-11 ˚C increase in thermal stability is consistent with our prior work. As stated in our previous study, the restricted conformation of the imino residues likely assist in positioning the azGly residue for optimal H-bond formation.

The CMPs containing charged side chain and azGly residues (**3b** (aspartate, Asp), **4b** (glutamate, Glu), **5b** (lysine, Lys) and **6b** (arginine, Arg)) exhibited a similar degree of increased stability (Chart 1). The most stabilized CMP was **6b** (Arg), with an 8.6 ˚C increase. In addition, **3b** (Asp), **4b** (Glu), and **5b** (Lys) resulted in increases in T_m of 6-7 ˚C. The values for the non-charged polar side chain azGly CMPs (**7b** (serine, Ser), **9b** (cystine, Cys), **10b** (asparagine, Asn), and **11b** (glutamine, Gln)) showed similar increases in T_m of 7-9 ˚C (Chart 1, middle right). In general, azGly moderately stabilized the triple helix when positioned adjacent to a polar residue. However, **8b** (threonine, Thr) was an exception with enhanced stability, but to a substantially lower degree than other polar residues.

A distinct trend was observed in the CMPs containing hydrophobic side chains and azGly residues (Chart 1). Incorporation of azGly residues adjacent to ß-branched side chains were non-perturbing, as evidenced by CMPs **14b** (valine, Val) and **16b** (isoleucine, Ile) T_m values that were nearly identical to their control counterparts (**14a** and **16a**, respectively). The absence of azGly-induced stability in the hydrophobic ß-branched CMPs is reflected in the reduction in the increased stability of the polar ß-branched CMP **8b** (Thr) compared to the other polar residues. The non-ß-branched side chain CMPs (**13b** (alanine, Ala) and **15b** (leucine, Leu)) exhibited moderate to large increases in thermal stability of 4-10 ˚C. The methionine (Met) azGly CMP **18b** exhibited a small increase in stability of 2.1 ˚C. However, the CMP containing a highly flexible Yaa position Gly residue (CMP **12b**) displayed a moderate increase in thermal stability of 4 ˚C. In comparison to CMPs **1b** (Hyp) and **2b** (Pro), the disparity in stabilization of **12b** is likely due to the flexibility of the Yaa position Gly residue, weakening the adjacent azGly residue's ability to form strong H-bonds. In

This journal is © The Royal Society of Chemistry 20xx

H-(POG)$_3$(Pro-Yaa-(Gly or azGly))(POG)$_4$-NH$_2$



Gly: Z = CH_2 azGly: Z = NH

	Peptide	Yaa Residue	Gly/azGly	Tm (˚C)	ΔTm (˚C) POG Control	ΔTm (˚C) azGly vs Gly
	1a	Hyp	Gly	42.7		
Charged	3a	Asp	Gly	25.1	-17.6	+6.3
	3b		azGly	31.3	-11.4	
	4a	Glu	Gly	31.1	-11.6	+7.4
	4b		azGly	38.6	-4.1	
	5a	Lys	Gly	31.3	-11.4	+6.0
	5b		azGly	37.3	-5.4	
	6a*	Arg	Gly	42.2	-0.5	+8.6
	6b*		azGly	50.8	+8.1	
Polar	7a	Ser	Gly	29.5	-13.2	+7.3
	7b		azGly	36.8	-5.9	
	8a	Thr	Gly	32.6	-10.1	+3.3
	8b		azGly	35.9	-6.8	
	9a	Cys	Gly	29.0	-13.7	+9.1
	9b		azGly	38.1	-4.6	
	10a	Asn	Gly	23.3	-19.4	+9.6
	10b		azGly	32.9	-9.8	
	11a	Gln	Gly	37.0	-5.7	+7.4
	11b		azGly	44.5	+1.8	
Hydrophobic	12a	Gly	Gly	23.8	-18.9	+4.0
	12b		azGly	27.9	-14.8	
	13a	Ala	Gly	34.6	-8.1	+10.2
	13b		azGly	44.8	2.1	
	14a	Val	Gly	34.7	-8.0	+0.0
	14b		azGly	34.7	-8.0	
	15a	Leu	Gly	24.1	-18.6	+4.3
	15b		azGly	28.4	-14.3	
	16a	Ile	Gly	34.4	-8.3	+0.4
	16b		azGly	34.8	-7.9	
	17a	t-Leu	Gly	34.7	-8.0	-0.4
	17b		azGly	34.3	-8.4	
	18a	Met	Gly	36.9	-5.8	+2.1
	18b		azGly	39.0	-3.7	
Aromatic	19a	Phe	Gly	19.5	-23.2	+8.7
	19b		azGly	28.2	-14.5	
	20a	Tyr	Gly	18.3	-24.4	+11.0
	20b		azGly	29.3	-13.4	
	21a	Trp	Gly	15.0	-27.7	+11.1
	21b		azGly	26.1	-16.6	
	22a	His	Gly	27.5	-15.2	+11.1
	22b		azGly	38.6	-4.1	

Chart 1. (Top) Sequence and structural representation of CMPs with a central Yaa position substitution. (Bottom) T_m values of each CMP compared to **1a**, the POG control sequence, and the degree of azGly thermal stabilization within each pair of CMPs with the same Yaa substitution, e.g. **3a** and **3b**. CMPs are grouped based on physical properties, e.g. charged residues. *Data for CMPs **6a** and **6b** are from ref. 28.

general, azGly minimally stabilized the collagen triple helix when positioned adjacent to hydrophobic residues.

To further investigate the behavior observed with ß-branched amino acids, we incorporated a bulky unnatural amino acid (i.e., *tert*-Leucine (*t*-Leu)) in CMPs **17a** and **17b** (Chart 1). This residue was chosen to determine if increasing the steric bulk around the ß-carbon would affect the azGly-induced stability. The experimental T_m values of the *t*-Leu-containing CMPs were nearly identical to those of the valine-containing CMPs. To gain insight, we performed molecular dynamic (MD) simulations based on **17b** (refer to SI for specific details regarding MD calculations). The results suggest that a steric clash exists between a *t*-Leu side chain hydrogen and a backbone oxygen (Fig. S1). Within this MD model, the measured distance between the hydrogen and oxygen atoms is 2.4 Å. This

H-(POG)(POazG)(POG)(Pro-Yaa-Gly)(POG)(POazG)(POG)-NH$_2$

Peptide	Yaa Residue	Tm (˚C)	ΔTm (˚C) POG Control	ΔTm (˚C) Gly Control	ΔTm (˚C) Adjacent azGly
1a	Hyp	42.7			
14c	Val	55.2	+12.5	+20.5	+20.5
17c	t-Leu	55.8	+13.1	+21.1	+21.5
20c	Tyr	41.9	-0.8	+23.6	+12.6
21c	Trp	39.0	-3.7	+24.0	+13.0

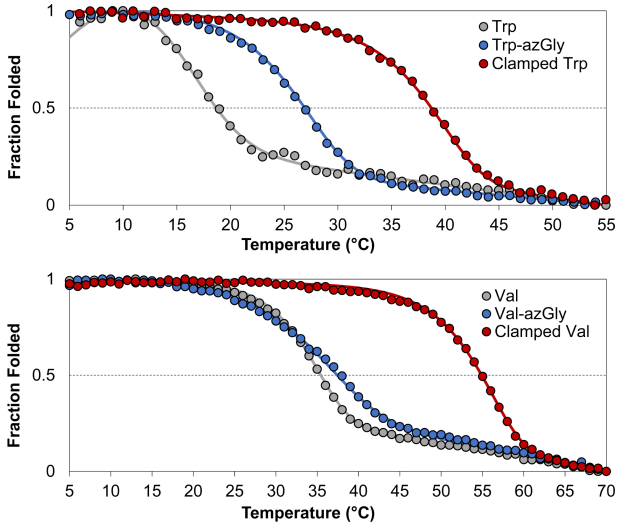


Figure 3. (Top) Sequence representation showing end-position azGly residues and Yaa position substitution for CMPs **14c**, **17c**, **20c**, and **21c**. (Middle) T_m values of clamped CMPs compared to **1a**, the POG control sequence, the corresponding Gly controls (CMPs **14a**, **17a**, **20a**, and **21a**), and the adjacent azGly sequence (CMPs **14b**, **17b**, **20b**, and **21b**). The sequences and data for these comparison CMPs are available in Chart 1 and Fig. S2. (Bottom) CD thermal melting data for Trp-containing CMPs **21a**, **21b**, and **21c** and for Val-containing CMPs **14a**, **14b**, and **14c**.

distance is too short to accommodate the minimum required distance of 2.7 Å based on the van der Waal radii of the two atoms (Fig. S1).[34] These observations are consistent with steric crowding and restricted rotation near the site of azGly substitution, preventing optimal H-bond formation.

The CMPs containing aromatic side chains and azGly residues (**19b** (histidine, His), **20b** (phenylalanine, Phe), **21b** (tryptophan, Trp), and **22b** (tyrosine, Tyr)) exhibited increased thermal stability of 8-11 ˚C (Chart 1). Interestingly, the control CMPs containing aromatic substitutions also exhibited the greatest decrease in thermal stability compared to the all POG control CMP **1a**. In general, when azGly was located adjacent to an aromatic residue, the collagen triple helix was significantly stabilized.

Next, we investigated a more general method for stabilizing CMPs with a variable Yaa residue. Instead of positioning the azGly residue adjacent to the Yaa substitution, we incorporated two azGly residues near both ends of the peptide to act as H-bond clamps, flanking the variable region of interest. We investigated the applicability of this strategy to circumvent the variation in azGly-induced stability that is encountered with adjacent azGly substitution by synthesizing CMPs with the general form H-(POG)(POazG)(POG)(Pro-Yaa-Gly)(POG)$_2$(POazG)(POG)-NH$_2$. We chose Yaa substitutions either unaffected by adjacent azGly incorporation or highly destabilizing to the triple helix.

In all cases, the azGly-induced stability of the clamped CMPs dramatically surpassed both the control and adjacent azGly CMPs (Fig. 3, refer to Chart 1 for sequences and T_m values of counterpart CMPs). CMPs **14b** and **17b** exhibited T_m values similar to their control sequences while the clamped CMPs **14c** and **17c** exhibited a 20 ˚C increase. The CMPs containing Tyr (**20a**, **20b**, **20c**) and Trp (**21a**, **21b**, **21c**) also exhibited similar behaviors. The clamped CMPs **20c** and **21c** exhibited a 13 ˚C increase over the adjacent azGly CMPs **20b** and **21b** and a 24 ˚C increase over their controls (**20a** and **21a**, respectively). These results demonstrate the wide applicability of using multiple azGly residues to surround central substitutions, providing a facile method for stabilizing short CMPs containing any residue.

Although azGly can be used to stabilize collagen triple helices, the identity of the adjacent Yaa residue must be considered (Fig. S2). The results from our CMP library show that incorporation of azGly in place of Gly, adjacent to a central substitution site, results in enhanced stability in most cases, especially with aromatic or positively charged amino acids. When paired adjacent to hydrophobic ß-branched residues, azGly proves a reliable replacement for glycine but results in little to no increase in triple helix stability. Utilizing a new strategy, we have demonstrated a general solution for increasing helical stability by positioning azGly residues outside the substitution site. This strategy was effective for residues showing little benefit from adjacent azGly incorporation as well as residues highly destabilizing to the triple helix. This investigation into azGly-induced stability and the wide applicability of azGly H-bond clamping to stabilize variable sequence CMPs will inform the design of novel azGly-containing peptides as well as future work to elucidate variability at the Xaa position adjacent to azaGlycine containing CMPs.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

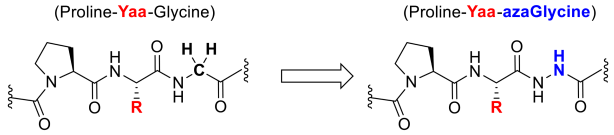
This work and instruments used were supported by funding from the University of Pennsylvania and the National Science Foundation including MALDI-TOF MS (NSF MRI-0820996).

Notes and references

1 A. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci.*, 1936, **22**, 439–447.
2 R. E. Hubbard and M. Kamran Haider, in *Encyclopedia of Life Sciences*, ed. John Wiley & Sons, Ltd, John Wiley & Sons, Ltd, Chichester, UK, 2010.
3 C. N. Pace, H. Fu, K. Lee Fryar, J. Landua, S. R. Trevino, D. Schell, R. L. Thurlkill, S. Imura, J. M. Scholtz, K. Gajiwala, J. Sevcik, L. Urbanikova, J. K. Myers, K. Takano, E. J. Hebert, B. A. Shirley and G. R. Grimsley, *Protein Sci.*, 2014, **23**, 652–661.
4 S.-Y. Sheu, D.-Y. Yang, H. L. Selzle and E. W. Schlag, *Proc. Natl. Acad. Sci.*, 2003, **100**, 12683–12687.
5 D. Dougherty, *Curr. Opin. Chem. Biol.*, 2000, **4**, 645–652.
6 B. Bilgiçer and K. Kumar, *J. Chem. Educ.*, 2003, **80**, 1275.
7 A. Altmayer-Henzien, V. Declerck, J. Farjon, D. Merlet, R. Guillot and D. J. Aitken, *Angew. Chem. Int. Ed.*, 2015, **54**, 10807–10810.
8 E. Beausoleil and W. D. Lubell, *Biopolymers*, 2000, **53**, 249–256.
9 G. B. Liang, J. M. Desper and S. H. Gellman, *J. Am. Chem. Soc.*, 1993, **115**, 925–938.
10 O. Khakshoor and J. S. Nowick, *Curr. Opin. Chem. Biol.*, 2008, **12**, 722–729.
11 C. W. Kang, M. P. Sarnowski, S. Ranatunga, L. Wojtas, R. S. Metcalf, W. C. Guida and J. R. Del Valle, *Chem. Commun.*, 2015, **51**, 16259–16262.
12 Y. Zhang, R. M. Malamakal and D. M. Chenoweth, *Angew. Chemi. Int. Ed.*, 2015, **54**, 10826–10832.
13 C. R. Walters, D. M. Szantai-Kis, Y. Zhang, Z. E. Reinert, W. S. Horne, D. M. Chenoweth and E. J. Petersson, *Chem. Sci.*, 2017, **8**, 2868–2877.
14 C. Proulx, D. Sabatino, R. Hopewell, J. Spiegel, Y. García Ramos and W. D. Lubell, *Future Med. Chem.*, 2011, **3**, 1139–1164.
15 A. S. Dutta, B. J. A. Furr, M. B. Giles and J. S. Morley, *Clin. Endocrinol. (Oxf.)*, 1976, **5**, s291–s298.
16 T. L. Ho, J. J. Nestor, G. I. McCRAE and B. H. Vickery, *Int. J. Pept. Protein Res.*, 1983, **24**, 79–84.
17 R. Chingle, C. Proulx and W. D. Lubell, *Acc. Chem. Res.*, 2017, **50**, 1541–1556.
18 L. Halab, F. Gosselin and W. D. Lubell, *Biopolymers*, 2000, **55**, 101–122.
19 J. Gante, *Synthesis*, 1989, **21**, 405–413.
20 A. A. Adzhubei, M. J. E. Sternberg and A. A. Makarov, *J. Mol. Biol.*, 2013, **425**, 2100–2132.
21 A. Rath, A. R. Davidson and C. M. Deber, *Biopolymers*, 2005, **80**, 179–185.
22 A. Rich and F. H. C. Crick, *J. Mol. Biol.*, 1961, **3**, 483–IN4.
23 K. Gelse, E. Pöschl and T. Aigner, *Adv. Drug Deliv. Rev.*, 2003, **55**, 1531–1546.
24 A. J. Kasznel, Y. Zhang, Y. Hai and D. M. Chenoweth, *J. Am. Chem. Soc.*, 2017, **139**, 9427–9430.
25 K. Okuyama, M. Haga, K. Noguchi and T. Tanaka, *Biopolymers*, 2014, **101**, 1000–1009.
26 M. Kuemin, Y. A. Nagel, S. Schweizer, F. W. Monnard, C. Ochsenfeld and H. Wennemers, *Angew. Chem. Int. Ed.*, 2010, **49**, 6324–6327.
27 R. S. Erdmann and H. Wennemers, *Angew. Chem. Int. Ed.*, 2011, **50**, 6835–6838.
28 Y. Zhang, R. M. Malamakal and D. M. Chenoweth, *J. Am. Chem. Soc.*, 2015, **137**, 12422–12425.
29 Y. Zhang, M. Herling and D. M. Chenoweth, *J. Am. Chem. Soc.*, 2016, **138**, 9751–9754.
30 A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick and B. Brodsky, *Biochemistry*, 2000, **39**, 14960–14967.
31 D. G. Barrett, C. M. Minder, M. U. Mian, S. J. Whittington, W. J. Cooper, K. M. Fuchs, A. Tripathy, M. L. Waters, T. P. Creamer and G. J. Pielak, *Proteins Struct. Funct. Bioinforma.*, 2005, **63**, 322–326.
32 D. D. Jenness, C. Sprecher and W. C. Johnson, *Biopolymers*, 1976, **15**, 513–521.
33 Y. Feng, G. Melacini, J. P. Taulane and M. Goodman, *J. Am. Chem. Soc.*, 1996, **118**, 10351–10358.
34 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.

 This journal is © The Royal Society of Chemistry 20xx

TOC Entry:



The effect of Yaa-position amino acid variation in collagen peptides paired with adjacent azaGlycine residues is investigated.