Structural and Positional Effects of Peptoid Residues on Stability of the Collagen Triple Helix

Pahweenvaj Ratnatilaka Na Bhuket[1], James R. Eastwood[2], Zhao Qin[3], Julian L. Kessler[4], Yang Li[5], Frank G. Whitby[6], Christopher P. Hill[6], Kent Kirshenbaum[2], S. Michael Yu[1,4,*]

[1] Department of Molecular Pharmaceutics, College of Pharmacy, University of Utah, Salt Lake City, UT 84112, USA

[2] Department of Chemistry, New York University, New York, NY 10003, USA

[3] Department of Civil & Environmental Engineering, College of Engineering & Computer Science, Syracuse University, Syracuse, NY 13244, USA

[4] Department of Biomedical Engineering, College of Engineering, University of Utah, Salt Lake City, UT 84112, USA

[5] Guangdong Provincial Engineering Research Center of Molecular Imaging, the Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong 519000, China

[6] Department of Biochemistry, School of Medicine, University of Utah, Salt Lake City, UT 84132, USA

*Corresponding author: S. Michael Yu (michael.yu@utah.edu)

ABSTRACT

The signature structure of collagen, the triple helix, comprises three collagen strands with GlyXaaYaa repeats intertwining into a right-handed superhelix. Pro and Hyp at the respective Xaa and Yaa positions provides maximum triple helical stability. Previously, we reported that peptoid residues also known as N-substituted Glycines (N-Glys) in the Xaa position generate hyperstable collagen triple helices. Here, we demonstrate that N-Glys at the Yaa position also stabilize the triple helix, with the position (Xaa or Yaa) and side chain structure of N-Gly profoundly affecting the triple helix stability. CD spectroscopy and X-ray crystallography indicated that N-Glys at the Yaa position were more conducive to triple helical folding than amino acids, although they were unable to access the most favorable ϕ-ψ angles for the triple helix represented by Hyp. Metadynamics simulations showed a more diffused conformational space for N-Glys at the Yaa position than at the Xaa position due to minimal steric crowding from neighboring Gly. Only the *S*-isomers of chiral N-C_α branched N-Glys were compatible with triple helical folding, with differences in backbone conformation and accessible rotamers between Xaa and Yaa positions. At the Yaa position, chiral Nspe can stabilize the triple helix better than any other natural or unnatural residues (except Hyp) *via* the intrachain CH•••π interactions. This work not only deepens our understanding of triple helical folding but also demonstrates a new design strategy for stable collagen mimetic peptides with unprecedented side chain diversity, opening new opportunities for applications in biomedicine and biomaterials.

INTRODUCTION

Collagen is the major constituent of the extracellular matrix (ECM) in animals and provides structural and functional support for tissues [1, 2]. There are 28 types of human collagens comprised of 45 different polypeptide chains [3, 4]. All types of collagens contain a triple helix, which is the hallmark of the collagen structure. Three left-handed polypeptides with a polyproline-II (PPII) helical conformation intertwine to form a right-handed triple-helical structure. To fold into the triple helix, each polypeptide chain must have a repetitive GlyXaaYaa sequence, where Gly is present at every third residue. Xaa and Yaa can be any amino acids, but Pro and Hyp [Hyp: (2*S*,4*R*)-hydroxyproline] frequently populate the Xaa and Yaa position, respectively [5]. Glys buried in the interior of the triple helix are solvent-inaccessible; however, their amide N–H hydrogens form interchain hydrogen bonds with the C=O groups of the Xaa residues in another chain, which is the major driving force for triple helical folding [3]. Gly mutations can lead to diseases such as osteogenesis imperfecta [6]. Among canonical amino acids, Pro and Hyp, at the Xaa and Yaa, respectively, provide the highest triple helical stability [7]. Here, the pyrrolidine ring helps preorganize the PPII helix, and the (4*R*)-hydroxyl group on Hyp in the Yaa position exerts a stereoelectronic effect that favors PPII [3] as well as the formation of hydration networks [8] that stabilize the triple helix.

Since natural collagens are large and complex, simple collagen mimetic peptides (CMPs) with GlyXaaYaa triplets have been utilized to study collagen folding and stability [9, 10]. Using CMPs, researchers have investigated triple helical propensities of not only canonical amino acids [5, 7, 11] but also artificial amino acids, which demonstrated that electronegative substituents, such as fluorine, chlorine, and methoxy, with (*R*)-configuration on the Cγ of Pro in the Yaa position, stabilize CMP triple helices via the stereoelectronic effect [12-18] and that substitution of Gly with aza-Gly results in hyperstable CMPs due to extra inter-strand hydrogen bond [19-23]. Egli et al. showed that the attachment of a hydrophobic group to (4*R*)-amidoproline at the Xaa position accelerated folding and enhanced the triple helical stability of CMPs by promoting *cis-trans* isomerization and the formation of a molten globule-like intermediate [24, 25]. Charged amino acids can also influence CMP folding by inter-strand charge-pairing interactions, which can be used to produce heterotrimeric CMPs with predetermined chain compositions and registers [26-31].

Considering that numerous bioactive domains of collagen molecules are triple helical, the folding priniciples of triple helices as applied to non-canonical amino acids are not only scientifically interesting but could also be useful in developing bioactive collagen mimetics with properties superior to natural collagen sequence. Unfortunately, our understanding of various factors responsible for triple helical stabilization applies mostly to α-amino acids and a small number of non-canonical amino acids, which are difficult to synthesize.

Peptoids are oligomers of N-substituted Glys (N-Glys, **Figure 1A**) that have side chains of α-amino acids moved from the C_α position to the N_α position. The submonomer method offers facile synthesis of a large library of peptoids. Early promising studies of peptoid oligomers suggested the possibility of the discovery of a potent new class of therapeutics, owing to their chemical diversity and improved biostability in comparison to conventional peptides. Concerted efforts in the peptoid community to find peptoid sequences with strong folding propensities have produced a growing understanding of side chain – backbone interactions and other stabilizing influences, but progress in developing a robust set of functional secondary structures has been slow. Although N-Glys are structurally similar to Pro and Pro residues are the most critical constituent of collagen protein sequences, to date, there has not been a systematic investigation of incorporating N-Glys into the GlyXaaYaa triple helical collagen sequence. This is a missed opportunity in the study of peptidomimetics, as the collagen triple helix is one of the most stable secondary protein structures and a superb scaffold for displaying N-substituted side chains. In addition, triple helices displaying natural and unnatural side chains offer potential for the discovery of new therapeutics targeting collagen-cellular receptor interfaces as well as collagen-ECM interactions. Previously, it was hypothesized that hydrophobic and steric interactions of N-Glys, especially those at the Xaa position, could stabilize collagen triple helices [32-34]. However, by investigating CMPs substituted with a series of N-Glys at the Xaa position (N^xaaa-CMPs), we showed that the bulkiness of peptoid residues at the Xaa position correlated with triple helical stability rather than hydrophobic interactions. We determined that bulky side chains restrict the rotation of the C_α–N_α bond and preorganized each CMP strand into the PPII structure,

thus increasing triple helical stability [35]. Although this investigation demonstrated for the first time that N-Gly is conducive to forming a stable triple helix when incorporated into the Xaa position of the GlyXaaYaa sequence, there is no knowledge of such for Yaa position, which has different local spatio-chemical environment compared to the Xaa position.

The positional effect of residues within the GlyXaaYaa motif (Xaa vs. Yaa) has been documented, particularly for Pro analogs. In making a stable triple helix, Pro analogs in the Xaa position prefer the C_γ-endo ring pucker, whereas those in the Yaa position prefer the C_γ-exo conformation [3, 10]. For example, (2*S*, 4*S*)-fluoroproline, which has a C_γ-endo ring pucker, stabilizes the triple helix only at the Xaa position and not at the Yaa position, whereas (2*S*, 4*R*)-fluoroproline, which has a C_γ-exo ring pucker, is stabilizing at the Yaa position and not at the Xaa position [15]. This is because the side chains at the Xaa and Yaa positions experience different spatial environments within the triple helix [36, 37].

Although the distinct difference in Xaa and Yaa positions in triple helix stabilization is well known for cyclic Pro derivatives, there is little understanding of such for noncyclic Pro analogs such as N-Glys. In this study, we systematically investigated the triple helical stabilization effect of N-Glys at the Yaa position, results of which were compared to N-Glys at the Xaa position, revealing N-Glys' positional effects on triple helix stabilization. The middle Hyp (Yaa position) of Ac-(GlyProHyp)$_7$-NH_2 was replaced with N-Glys featuring side chains of canonical (natural) and unnatural amino acids (designated as N^yaaa-CMP). CD spectroscopy, X-ray crystallography, and computational modeling including MD simulation, were used to study the molecular structure and triple helical stability of the N^yaaa-CMPs, providing insights into the positional influence of N-Glys on triple helical stabilization, including why N-Glys are more stabilizing at the Xaa position than the Yaa position and the indifference of chemical structure of Yaa substitutents on stability. Our work also revealed stereochemical stabilization mediated by N-C_α substituent of the N-Gly side chain, leading to the discovery of peptoid monomers that specifically produce large stabilization of the triple helix at the Yaa position.

MATERIALS AND METHODS

All chemicals were used as received from commercial suppliers without further purification. The Fmoc-protected amino acids, primary amines, solvents, resin, and reagents for synthesizing and purifying collagen mimetic peptides and peptoids are listed in Supporting Information (SI) **Table S1**.

Host-Guest Peptide synthesis

An intermediate peptide, Fmoc-[GlyProHyp(tBu)]$_3$, was synthesized at a 0.15 mmol scale on TentaGel® R RAM Resin (90 μm) Rink-type (loading density: 0.2 meq/g) via the standard Fmoc-mediated solid-phase peptide synthesis using a Focus XC Solid Phase Peptide Synthesizer (AAPPTec, Louisville, KY, USA). Amino acid coupling was achieved using the HBTU/HOAt chemistry. The resulting intermediate peptide was split to synthesize various host-guest peptides. Each host-guest peptide was manually synthesized. The manual synthesis, cleavage, HPLC purification, and purity determination of the host-guest collagen mimetic peptides are described in the **SI (Tables S2 to S6)**. The purified products were verified using a Waters MALDI Micro MX MALDI-TOF mass spectrometer (Waters Corporation, Milford, MA, USA) at the Mass Spectrometry Facility of the Department of Chemistry at the University of Utah.

Circular dichroism (CD) spectroscopy experiments

After lyophilization, the purified peptides were dissolved in DI water to obtain stock solutions. The concentrations of the host-guest peptide stock solutions were determined by UV-Vis spectroscopy using a SpectraMax M2e microplate reader (Molecular Devices, San Jose, CA, USA). The absorbance of the peptide/peptoid solutions was measured at 214 nm (extinction coefficient: 2200 M^{-1} cm^{-1} per peptide bond [11]) or 280 nm for sequences containing a tyrosine side chain (extinction coefficient: 1490 M^{-1} cm^{-1} per residue) using a quartz cuvette with a path length of 1 cm. Before the CD experiments, peptide stock solutions were heated at 80 °C for 10 min and stored at 4 °C for at least 48 h. Then, the peptide solutions were diluted to 150 μM (total strands) in 1× PBS (pH 7.4). Peptide sequences containing acidic or basic side chains were prepared at 150 μM (total strands) in HCl (0.1 M) or NaOH (0.1 M) solutions. Dilution was performed using cold diluents. For PPII host-guest sequences, peptide solutions were prepared at 100

μM in 5 mM phosphate buffer (pH 7.0). All CD measurements of the peptide solutions were conducted on a JASCO J-1500 CD spectrometer (JASCO, Easton, MD, USA) using a 1-mm quartz cuvette.

The CD spectra of each peptide were recorded from 200 to 250 nm at 4 °C. The CD spectral scanning used the following parameters: data pitch, 0.1 nm; digital integration time, 8 sec; bandwidth, 1 nm; scanning speed, 20 nm/min; scanning mode, continuous. The raw CD spectra were smoothed using a mean movement method with a convolution number of 25. The molar ellipticity ([θ], 10^3 deg cm^2 $dmol^{-1}$) values were converted from the recorded ellipticity using JASCO Spectra Manager Version 2 software (Version 2.15.01). All the spectra are the averages of two independent measurements.

Thermal denaturation of each sequence was performed by monitoring the change in ellipticity at 225 nm from 4−80 °C at a heating rate of 0.5 °C/min. Data processing was done with the JASCO Spectra Manager Version 2 software (Version 2.15.01). The denaturation curves and first-derivative plots were smoothed using a means-movement method with a convolution number of 25. The first derivative was calculated using a subtraction method. The reported T_m values are the averages of two independent thermal denaturation scans. Melting profiles were constructed by plotting the fractions folded against temperature.

X-ray crystallography

Purified and lyophilized N^yphe-CMP and N^ynbz-CMP were dissolved in water at concentrations of 5 and 10 mg/mL, respectively. A Crystal Gryphon LCP robot (Art Robbins Instrument, Sunnyvale, CA, USA) was used for sitting-drop vapor-diffusion crystallization using commercially available crystal screen reagents. Equal amounts (0.25 μL each) of peptides and screen conditions were mixed and sealed in chambers where the reservoirs contained 0.5 mL of screen solutions. N^yphe-CMP crystals were grown at 21 °C with Wizard Cryo 1 & 2 HT-96 (Rigaku, Bainbridge Island, WA, USA) condition D12, corresponding to 40% (v/v) PEG600 containing 100 mM imidazole/hydrochloric acid (pH 8.0) and 200 mM zinc acetate. N^ynbz-CMP crystals were grown at 4 °C in condition H4 of the JBScreen Classic HTS I screen (Jena Bioscience, Thuringia, Germany), corresponding to 30% (w/v) PEG8000 containing 200 mM Ammonium sulfate. In preparation for data collection, crystals were briefly immersed (10-20 sec) in a

cryoprotection solution consisting of a crystallization buffer supplemented with 25% glycerol. The cryoprotected crystals were then harvested using a small rayon loop attached to a mounting pin and quickly plunged into a liquid nitrogen reservoir. Diffraction data from the N^{y}nbz-CMP crystal were collected with a Pilatus-6M detector on beamline 9-2 at the Stanford Synchrotron Radiation Lightsource (SSRL, National Accelerator Laboratory, Menlo Park, CA). The SSRL resources, including Remote Access [38], Blu-Ice [39], Automated Sample Mounting System [40, 41], and AUTOXDS script [42], were used for data collection and analysis. Diffraction data from N^{y}phe-CMP were collected with an Eiger-16M detector on the NECAT beamline 24-ID-E at the Advanced Photon Source (APS, Argonne National Laboratory, Lemont, IL). The diffraction data were processed, integrated, and scaled with XDS [43] and AIMLESS[44]. The structure of N^{y}nbz-CMP was determined by molecular replacement using Phaser in the PHENIX software package [45], where model 1G9W was used as the search model. The initial phases provided electron density maps that allowed us to build a complete model of N^{y}nbz-CMP using COOT [46]. The model was refined with phenix.refine [47]. The model of N^{y}nbz-CMP was used as a search model to determine the structure of N^{y}phe-CMP by molecular replacement. The structure of N^{y}phe-CMP was determined and refined similarly to N^{y}nbz-CMP. The structural figures of N^{y}phe-CMP and N^{y}nbz-CMP were rendered using UCSF Chimera [48]. The crystallographic data and refinement statistics are summarized in **Table S7**.

Metadynamics calculations and simulations

Well-tempered metadynamics was applied to study the conformational stability of Gly, Hyp, Nnbz, Nala, Nleu, Nchx, Nlys, Nphe, and Nval residues within the Yaa-CMP host. We used NAMD and the PLUMED package [49-51] to perform metadynamics simulations for acquiring the free energy landscape as a function of ϕ and ψ angles of the guest residues. Events of energy deposition took place every 200 fs in a Gaussian form with an energy height of 0.01 kcal/mol and a dihedral angle increment of 0.35 rad. We summed 200,000 deposited events to generate the energy landscapes for each selected Yaa-CMP. This Gaussian number yielded the convergence of the energy landscapes. The molecular models and atomic interactions of the selected Yaa-CMP were built from a combination of DFT calculation and CHARMM

general force field (CHARMM GenFF) [52]. We used HyperChem software (Hypercube, Inc., Gainesville, FL, USA) for the DFT calculation. We also updated the standard CHARMM27 force field [53] to enable the modeling and simulation of our peptide-peptoid hybrids. Model optimization was performed to identify the energy favorite coordinates of all atoms using geometric optimization with the conjugate gradient method. Molecular parameters, including charge distribution, van der Waals parameters, bond angles, and dihedral angles, were defined from the optimized molecular models that were made to an automatic analogy with CHARMM GenFF. The penalty scores for all parameters were less than 10, indicating that the analogy was fair [52]. The intrinsic coordinates of all the tested hybrids were also defined according to the optimized structure to ensure that the initial geometry was at the optimum energy level. All structures were simulated under the same conditions as those used for the Xaa-CMP system reported previously [35].

Molecular dynamics (MD) simulations

Molecular models of capped peptoid residues ("dipeptoid") were prepared using the ETDG method [54] to generate approximate starting conformations. These conformations were refined by quantum mechanical geometry optimization using the semi-empirical AM1 level of theory. Partial charges were assigned using the AM1-BCC method [55] to ensure compatibility with the Generalized Amber Force Field [56]. CMP models were prepared from the X-ray crystal structure of N^{γ}nbz-CMP by truncating the *N*-substituted side chains at the C_{β} position and rebuilding the side chain with coordinates from the dipeptoid model. A hybrid parameter set was constructed for the molecular mechanics force field, with Amber ff19SB parameters [57] used for peptide residues and GAFF2 parameters used for peptoid substitution. The resulting models were minimized using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno method [58]. All simulations were performed using the Generalized Born/Surface Area implicit solvation model as implemented by Onufriev, Bashford, and Case [59]. The minimized models were heated to 310 K over 200 ps, allowed to equilibrate for 800 ps, and simulated for 20 ns. The analysis was performed only on this 20 ns simulation.

RESULTS AND DISCUSSION

N-Glys at the Yaa position stabilize the triple helix.

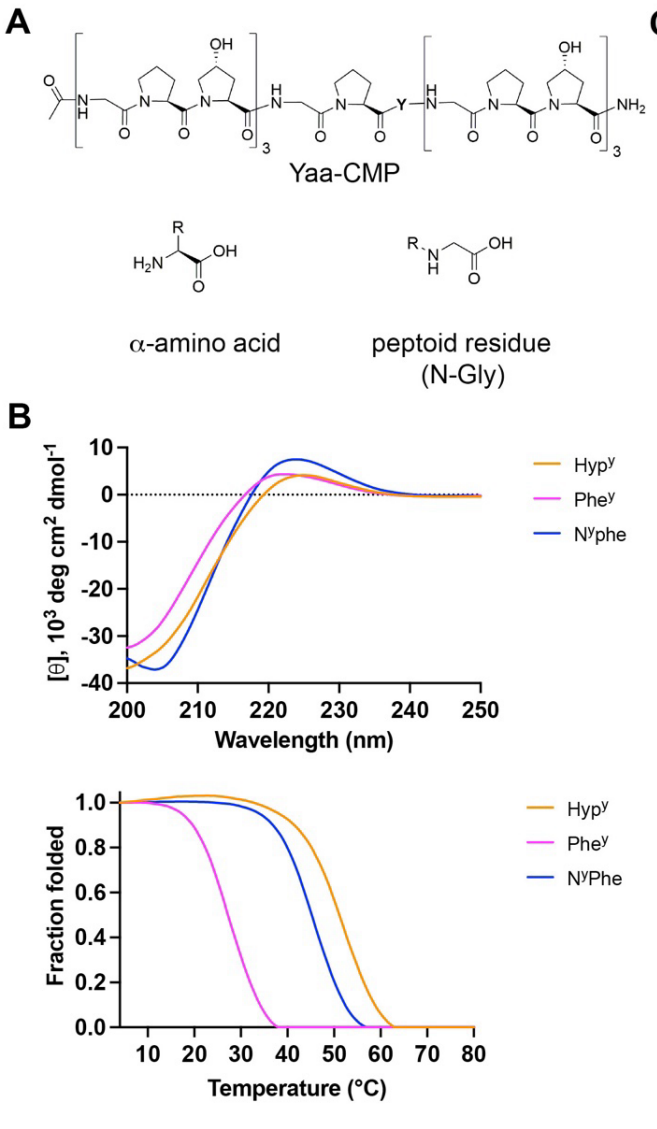
We examined the triple helical propensity of N-Glys (Naaa) at the Yaa position using the host-guest peptide Ac-(GlyProHyp)$_3$GlyProYaa(GlyProHyp)$_3$-NH_2 (Yaa-CMP or N^{y}aaa-CMP), where the central Hyp of the (GlyProHyp)$_7$ host sequence was replaced with peptoid guest residues (**Figure 1**). We first synthesized N^{y}aaa-CMPs with side chains covering the natural side chains and conducted CD spectroscopy at room temperature as well as CD melting experiments, the results of which were compared to those of the corresponding peptides with natural amino acids in **Figure 1** (**Figures S1** to **S3)**. The results clearly showed that N^{y}aaa-CMPs folded into a stable triple helix, as evidenced by the characteristic CD spectrum and sigmoidal shape of the melting transition [32]. Similar to N-Glys at the Xaa position, all N-Glys with natural side chains, except three, improved the triple helix stability compared to their α-amino acid versions, although none were more stable than the host peptide with Hyp at the guest position. Interestingly, N^{y}phe-CMP and N^{y}tyr-CMP showed the greatest stabilization with a T_m elevation of 18 °C (**Figure 1C**). These results demonstrate that moving side chains from the C_α position to the N_α position generally improves the triple helical stability for amino acids not only for the Xaa but also for the Yaa position.

As mentioned above, all N^{y}aaa-CMPs had T_ms lower than that of the host peptide (Hyp^y-CMP), indicating that substituting Hyp with N-Glys destabilized the triple helix. The T_m reduction ranged from 5 – 18 °C, which is in stark contrast to the substitution at the Xaa position, where some N-Glys improved triple helical stability (**Table 1)**. Even when compared to Pro^y-CMP (T_m = 48 °C), the T_m values were generally lower, except for N^{y}val-CMP (**Figure 1C**). These results suggest that, at the Yaa position, N-Glys are less stabilizing than Pro to varying degrees and that hydroxylation of Pro at the Yaa position has exceptional stabilization for the triple helix beyond what could be achieved by the N-Glys.

242 **Table 1** Comparison of triple helical stability of N^xaaa-CMPs and N^yaaa-CMPs.

Residue	ΔT_m for GlyXaaHyp (°C)*, **	ΔT_m for GlyProYaa (°C)*
Nrpe	No triple helix	Perturbed triple helix
Nval	-19	-6
Nspe	-7	-5
Nasp	-5 (PBS)	-18 (PBS)
	+1 (HCl)	-13 (HCl)
Nala	-5	-16
NEt	-3	-11
Nhcys	-1	-11
Nmet	0	-11
Nasn	0	-15
Nlys	0 (PBS)	-10 (PBS)
	-1 (NaOH)	-10 (NaOH)
Nleu	+3	-12
Ndxn	+4	-11
Ntyr	+6 (PBS)	-9 (PBS)
	+4 (NaOH)	-8 (NaOH)
Nphe	+7	-9
Nnbz	+7	-17
Nchx	+8	-11

243 * Compared with Ac-(GlyProHyp)$_7$-NH_2, which has a T_m of 55 °C; **From Ref [35].



C

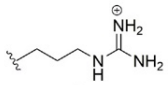
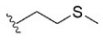
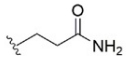
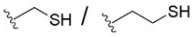
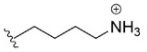
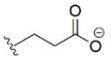
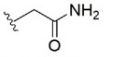
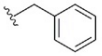
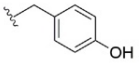
Side chain	T_m (°C)*	
	AA	N-Gly
	55**	-
	48	-
	48	43 (Nʸarg)
	47	44 (Nʸmet)
	46	44 (Nʸgln)
	44	39 (Nʸala)
	43	49 (Nʸval)
	40	44 (Nʸhcys)***
	40	45 (Nʸlys)
	40	45 (Nʸglu)
	33	43 (Nʸleu)
	33	-
	32	37 (Nʸasp)
	31	40 (Nʸasn)
	28	46 (Nʸphe)
	28	46 (Nʸtyr)

Figure 1 Substituting the middle Hyp residue of $(GlyProHyp)_7$ with N-Glys produces stable triple helices. (A) Chemical structure of host CMP with the sequence of Ac-$(GlyProHyp)_3$-GlyPro<u>Yaa</u>-$(GlyProHyp)_3$-NH_2, where the central Yaa position was substituted with various α-amino acids and N-Glys. (B) Representative CD spectra (upper plot) and melting profiles (lower plot) of Yaa-CMPs. Nʸphe-CMP produced CD spectral characteristics similar to those of Hypʸ-CMP and Pheʸ-CMP. A two-state transition is also observed in the thermal melting experiment. These results suggest that Nʸphe-CMP folded into a triple helix. Although Nʸphe-CMP was less stable than Hypʸ-CMP, moving the Phe side chain from the C_α to the N_α (Pheʸ-CMP vs. Nʸphe-CMP) increased the thermal stability of the host peptide. (C) Comparison of triple helical stability of Yaa-CMPs containing α-amino acids and N-Glys with natural side chains. Most N-Glys had higher T_m values than their α-amino acid counterparts. However, none of the N-Glys produced more stable triple helices than Hyp. *The thermal denaturation experiments were performed in 1× PBS. ** The value is from Ref [35]. ***Nʸhcys-CMP is not exactly equivalent to Cysʸ-CMP but structurally similar.

Effects of side chain structures on the triple helical stability

Next, we examined the effect of the side chain structures on the stability of the triple helix. We first investigated the effects of hydrophobic side chains, as Goodman and coworkers suggested that hydrophobic interactions could be the primary driver for the triple helical stabilization of N-Glys with alkyl (e.g. N^xleu) and aryl side chains in the Xaa position [32]. Using the same host-guest peptide system, we determined the CD melting temperatures of N^yaaa-CMPs incorporating a series of N-Glys with natural (**Figure 2A**) and unnatural side chains with varying levels of hydrophobicity (**Figure 2B**). There was no correlation between the side chain hydrophobicity and the T_m of the N-Glys. Similar to our previous study at the Xaa position [35], hydrophobic interactions are not a major factor contributing to the triple helical stability [60]. This is most evident in the similar T_m values between N^ychx-CMP and N^ydxn-CMP which is comparatively more hydrophilic.

N^yphe-CMP and N^ytyr-CMP have similar T_m values, although their aromatic rings have different electron densities. Surprisingly, N^ynbz, an electron-deficient N^yphe derivative, produced a much less stable triple helix than N^yphe and N^ytyr (**Figure 2B**), although it produced CMP with a high T_m when incorporated at the Xaa position [35]. We hypothesized that at the Yaa position, the nitro group in the *ortho* position might cause steric clashes that destabilize the triple helix. To test this hypothesis, we tested N^ytol, which has a structure similar to that of N^ynbz but is more electron-rich. The T_m of N^ytol-CMP was similar to that of N^ynbz-CMP (**Figure 2B**) and markedly lower than those of N^yphe- and N^ytyr-CMPs. In addition, deprotonation of N^ytyr (with 0.01 M NaOH) which further increases the electron density of the aromatic ring had little effect on the T_m of N^ytyr-CMP (47 °C vs. 46 °C, **Figures S4** and **S5**). Therefore, the position of the ring substituent, not the electronic properties of the aromatic ring, seems to have an effect on the triple helical stability of N^yaaa-CMP containing aromatic side chains.

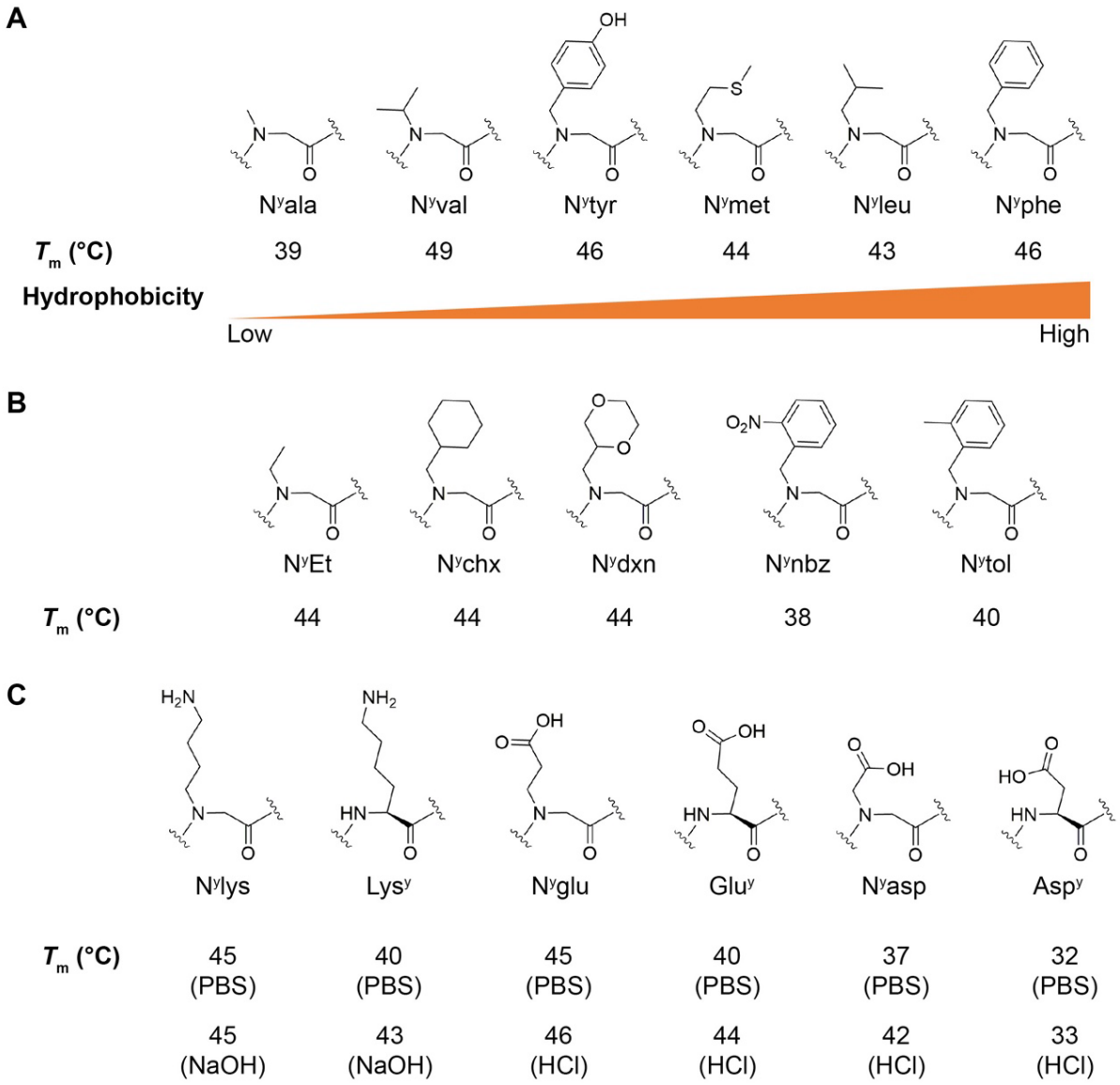


Figure 2 Side chain chemistry of N-Gly does not have a strong correlation with triple helical stability. (A) T_m values for a series of hydrophobic side chains are shown in the order of low to high hydrophobicity [60]. (B) The T_m values of the noncanonical side chains are similar except Nʸnbz- and Nʸtol-CMPs that have substituents at the *ortho* position. Comparing the side chains in (A) and (B), the triple helical stability has little dependency on the hydrophobicity or bulkiness of the side chain. (C) In contrast to amino acids, the ionization state of Nʸ-Glys had little effect on triple helical stability except for Nʸasp, which bears a charge closest to the peptide backbone.

We also examined the effects of side-chain ionization. First, although Arg is one of the most stabilizing amino acids at the Yaa position, Nʸarg was only moderately stabilizing because the T_m of Nʸarg-CMP belonged to the middle of *the* T_m range for natural side chains (**Figure 1C**). Unlike their α-amino acid

versions, the side-chain ionization of N^{y}lys and N^{y}glu had little effect on their T_m (**Figure 2C**). Only N^{y}asp-CMP exhibited T_m differences between the protonated and deprotonated forms. The results suggest that, in the case of N-Glys at the Yaa position, the effect of side chain charges on the triple helical stability might be more pronounced when charged groups are closer to the CMP backbone, which is similar to the observation found previously at the Xaa position [35].

Finally, the CD melting data indicated that the thermal stability of the N^{y}aaa-CMPs was independent of the bulkiness of the side chains, which is surprising because it contrasts the trend observed in the N^{x}aaa-CMPs, where an increase in side chain bulkiness clearly elevated the T_m of CMP. T_m values of N^{y}Et-CMP, N^{y}leu-CMP, and N^{y}chx-CMP were essentially the same (*Tm* 43 – 44 °C), despite their drastic difference in side-chain bulkiness. Based on this observation, we conclude that the bulkiness of the side chains is not a major driver of triple helical stability in the case of N-Glys at the Yaa position. As explained below, when compared to Xaa position, Yaa position provides more space to accommodate the bulky N-Gly side chains, which diminishes their effect on the triple helical stability as represented by the T_m.

Crystal structures of N^{y}phe-CMP and N^{y}nbz-CMP

We chose N^{y}phe-CMP and N^{y}nbz-CMP for X-ray crystallographic analysis because they have a large difference in thermal stability despite their similarity in side chain structures. Both CMPs crystallized with a single triple helical molecule in the asymmetric unit. The triple helical structure of N^{y}phe-CMP and N^{y}nbz-CMP were determined to a resolution of 1.30 Å (**Table S7**). When models for both structures were refined (**Table S7**), they contained all peptide atoms, except the N-terminal acetyl group and Gly$_1$ residue of the leading strand in the N^{y}phe-CMP structure, which have high flexibility. The refined structures of N^{y}phe-CMP and N^{y}nbz-CMP are shown in **Figure 3A**.The average (ϕ, ψ) angles of Gly, Pro, and Hyp from both structures were (-71.3° ± 3.8°, 174.6° ± 3.0°), (-71.1° ± 5.3°, 163.9° ± 5.8°), and (-58.6° ± 4.6°, 148.6° ± 4.3°), respectively, which are consistent with those of N^{x}aaa-CMPs reported previously [35]. The average (ϕ, ψ) angles of the N^{y}phe and N^{y}nbz residues were (-66.7° ± 2.9°, 159.1° ± 5.0°) and (-74.3° ± 1.2°, 162.2°

± 1.7°) respectively. Angles shifted from Hyp were more pronounced in N^{y}nbz than N^{y}phe suggesting that the triple helix becomes more unstable as the (ϕ, ψ) angles move away (in the direction of wider angles) from those of Hyp. On average, the angles shifted from those of Hyp by Δϕ = -14.3°, Δψ = +14.3°. Despite these substantial shifts from the typical collagen peptide geometry at the substituted position, no obvious distortions were observed in either crystal structure. The Pro-peptoid peptide bonds were in the *trans* configuration, similar to other peptide bonds in the triple helix, suggesting that N-Glys did not alter the backbone structure.

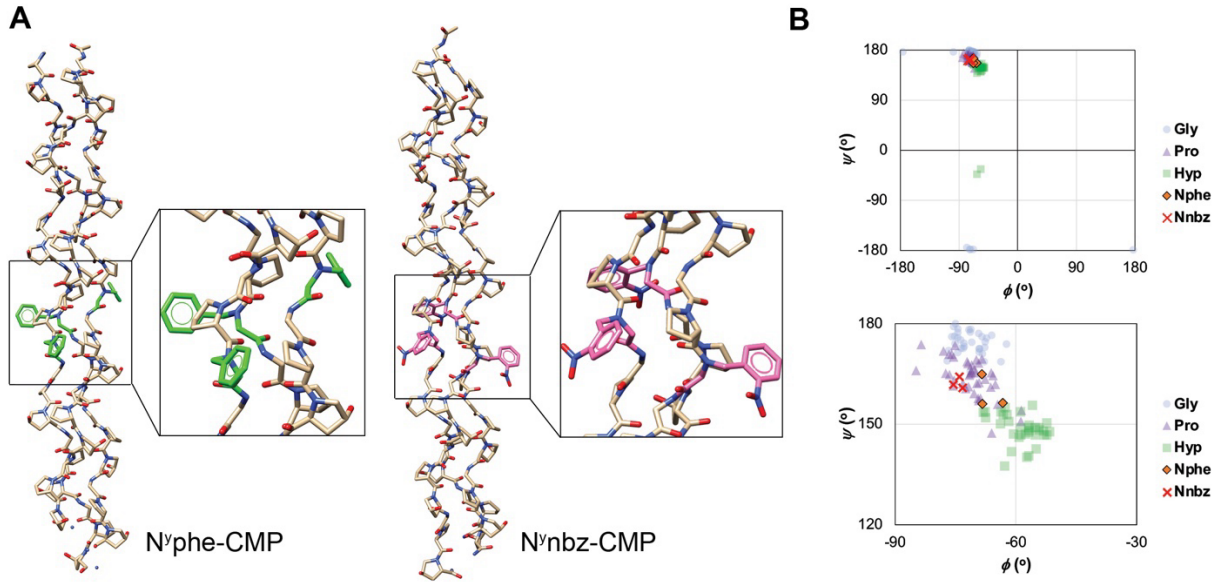


Figure 3 N-Glys do not disrupt the supersecondary structure of the host CMPs. (A) Crystal structures of N^{y}phe-CMP and N^{y}nbz-CMP show a characteristic conformation of the collagen triple helix without noticeable structural distortion. The N^{y}phe and N^{y}nbz residues are labeled in green and magenta, respectively. Both residues adopt the *trans*-peptide bond. (B) The top panel is the Ramachandran plot showing (ϕ, ψ) angles of all residues in the crystal structures of N^{y}phe-CMP and N^{y}nbz-CMP. The (ϕ, ψ) angles of most residues are found in the PPII region of the Ramachandran plot. The lower panel shows the magnified Ramachandran plot of N^{y}phe-CMP and N^{y}nbz-CMP. The (ϕ, ψ) angle coordinates of the N^{y}phe and N^{y}nbz residues are near the Pro region.

Although the deviations of the N-Glys' average (ϕ, ψ) angles did not distort the triple helical structure, these shifts have implications for destabilization of the triple helix conformation. The average ϕ and ψ values of up-puckered Pro^{y} from a quantum mechanical/molecular mechanical study of triple helical (ProProGly)$_{10}$ were -51.3° ± 1.1° and 149.6° ± 6.2°, respectively [37]. Furthermore, the preferred ψ angle at the Yaa position of the collagen triple helix for efficient n → π* interactions is 151° ± 4° [8]. The N-

Glys at the Yaa position from our crystal structures adopted (ϕ, ψ) angles between the Pro and Hyp regions, which are less populated regions in comparison to the residues at the Yaa position from CMP structures available in the PDB database (**Figures 3B** and **S6; Table S8**). These observations imply that these deviations might negatively impact the efficiency of the n → π* interactions exerted by N-Glys, resulting in destabilized triple helices compared to $(GlyProHyp)_7$. This is in contrast to N-Glys at the Xaa position, where the backbone dihedral angles are within the range of down-puckered Pro^x in triple helices [35-37]. Furthermore, when Hyp is in the Yaa position, its hydroxyl group participates in hydration networks that contribute to triple helical stability [61]. Such hydration networks were absent in the crystal structures of N^yphe- and N^ynbz-CMPs (**Figures S7 and S8**). Because the above data were extracted from static crystal structures and do not reflect all possible structural conformations in solution, we performed computational simulations to further dissect the factors affecting the triple helical stability contributed by N-Glys.

Metadynamics simulations of N-Glys in the Yaa position

The metadynamics calculations were performed for the triple helical structures of N^yaaa-CMPs. We selected residues covering different side chains, including small aliphatic, branched aliphatic, alicyclic, aromatic, and hydrophilic side chains. We also conducted simulations of the host CMP, $(GlyProHyp)_7$, and Gly^y-CMP for comparison. **Figures 4A** and **S9** show the free-energy landscapes of the selected residues. We assigned regions to the energy landscapes following the Ramachandran nomenclature proposed by Hollingsworth and Karplus [62]. Although both Hyp^y and Gly^y had the lowest-energy valleys distributed within the PPII (P_{II}) region and slightly higher-energy valleys in the inverse γ-turn (γ′) and bridge (δ) regions, due to additional flexibility, Gly^y also accessed an additional (ϕ, ψ) angle space encompassing the extended chains (ε) as well as mirror images of the PPII helix (P_{II}') and bridge (δ′) regions. N-Glys skewed the energy landscapes of Gly^y towards that of Hyp^y. We believe that N-Glys restrict conformational flexibility via steric hindrance between their side chains and carbonyl groups (**Figure 4B**), resulting in limited access to all allowed (ϕ, ψ) dihedral angles observed for Gly^y. The energy landscape of N^yala

resembled that of Gly^{y} because its small side chain exerts less steric hindrance than the other peptoid residues. None of the N-Glys completely diminished the δ′- and P_{II}′-regions down to Hyp^{y}, suggesting that the N-Glys could sample conformers that did not have the lowest energy levels, thus suppressing global triple helical stability. These metadynamics simulation results are in agreement with our CD melting data that substituting Hyp with the peptoid residues lowers T_m for all peptoid residues.

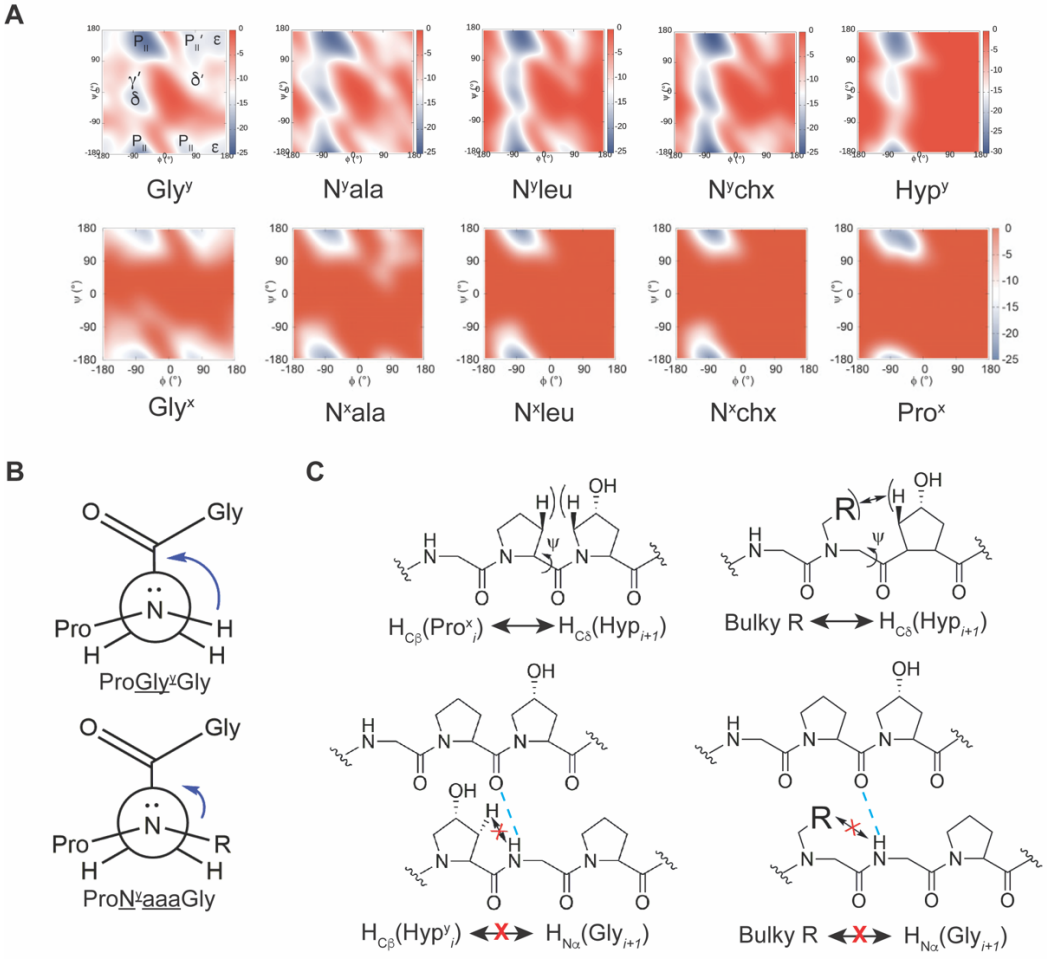


Figure 4 (A) Free energy landscapes (kcal/mol) of the guest residues Gly, Nala, Nleu, and Nchx, in the CMPs are compared with that of the native residue (Pro for Xaa position and Hyp for Yaa position). All residues had the lowest-energy valleys localized within the PPII basin. However, those at the Yaa position also had additional energy wells in other regions. These results suggest that residues in the Yaa position could access more (ϕ, ψ) space than those in the Xaa position. The free energy landscapes of the residues in the Xaa position were adapted with permission from ref. [35], Copyright © 2021 American Chemical Society. (B) Comparing Newman projections of an N-Glys and Gly in the Yaa position, the presence of a side chain on the N_{α} atom restricts the rotation around the N_{α}–C_{α} bond via the steric interaction with the C=O of residue *i+1* (Gly). This results in reduced conformational flexibility compared to that of Gly^{y}. (C) Residues in the Xaa and Yaa positions experience different surroundings. The Xaa residue can have steric interactions with the $H_{C\delta}$ of the succeeding Hyp, resulting in less accessible ψ angles. In contrast, the Yaa residue precedes Gly, which has no side chain. The amide H of Gly is buried in the helical core and hydrogen bonded to the

C=O of Pro in an adjacent strand, which avoids potential steric interaction with the Yaa residue's side chain. These factors may allow the Yaa residue to access more ψ angles than the Xaa residue.

In our previous work, metadynamics calculations showed that the energy distributions of N-Glys in the Xaa position and Pro^x were highly confined within the P_{II} region and had no energy wells in other regions of the Ramachandran plot, except for N^xala-CMP, as shown in **Figure 4A** [35]. Interestingly, the energy landscapes of the same N-Glys at the Yaa position were not limited to the P_{II} region. Notably, the Hyp^y residue also had some conformers accessing other regions, although it has a pyrrolidine ring that restricts backbone conformations, similar to Pro^x. A plausible explanation for this discrepancy is the difference in steric interactions with the residues following the N-Gly. The residue at the Xaa position precedes Hyp. As shown in **Figure 4C**, the ψ angle of the Pro^x_i residue could be restricted by steric interactions between its $H_{C\beta}$ and the $H_{C\delta}$ of the Hyp^y_{i+1}. A bulky N-Gly side chain at the Xaa position may also exhibit similar steric interactions, which limits the accessible ψ angles. On the other hand, Gly succeeds the residue at the Yaa position. Its $H_{N\alpha}$ forms a hydrogen bond with a Pro^x residue in an adjacent CMP strand and is buried within the triple helix. Therefore, potential steric interactions between Gly_{i+1} residue and $H_{C\beta}$ of Hyp^y_i or an N-Gly side chain at the Yaa position are unlikely to occur. This allows Hyp^y_i and N-Glys in the Yaa position to access more conformations than those in the Xaa position.

Effect of N-C_α chiral residues on the triple helical stability of CMPs

N-C_α chiral side chains introduce chirality into N-Glys and can limit the conformational space accessible by the N-Gly backbone. Previously, we found that two model N-C_α chiral peptoid residues, Nspe and Nrpe (**Figure 5A**), in the Xaa position destabilized the triple helix. N^xspe-CMP (T_m: 48 °C) had a T_m value noticeably lower than the unbranched N^xphe-CMP (T_m: 62 °C), and N^xrpe-CMP completely nullified the triple helical folding [35]. When the two stereoisomers were tested at the Yaa position, N^yspe-CMP (T_m: 50 °C) surprisingly formed a more stable triple helix than the N^yphe-CMP (T_m: 46 °C), with its T_m highest among all N-Glys tested at the Yaa position, whereas N^yrpe-CMP showed an unclear transition in its thermal melting profile (**Figure 5A**), suggesting the loss of the triple helical structure. These findings show that the N-C_α chirality of N-Glys influences triple helical folding differently at the Xaa and Yaa positions.

Therefore, we conducted a set of experimental and computational investigations to understand this intriguing stereo-positional effect.

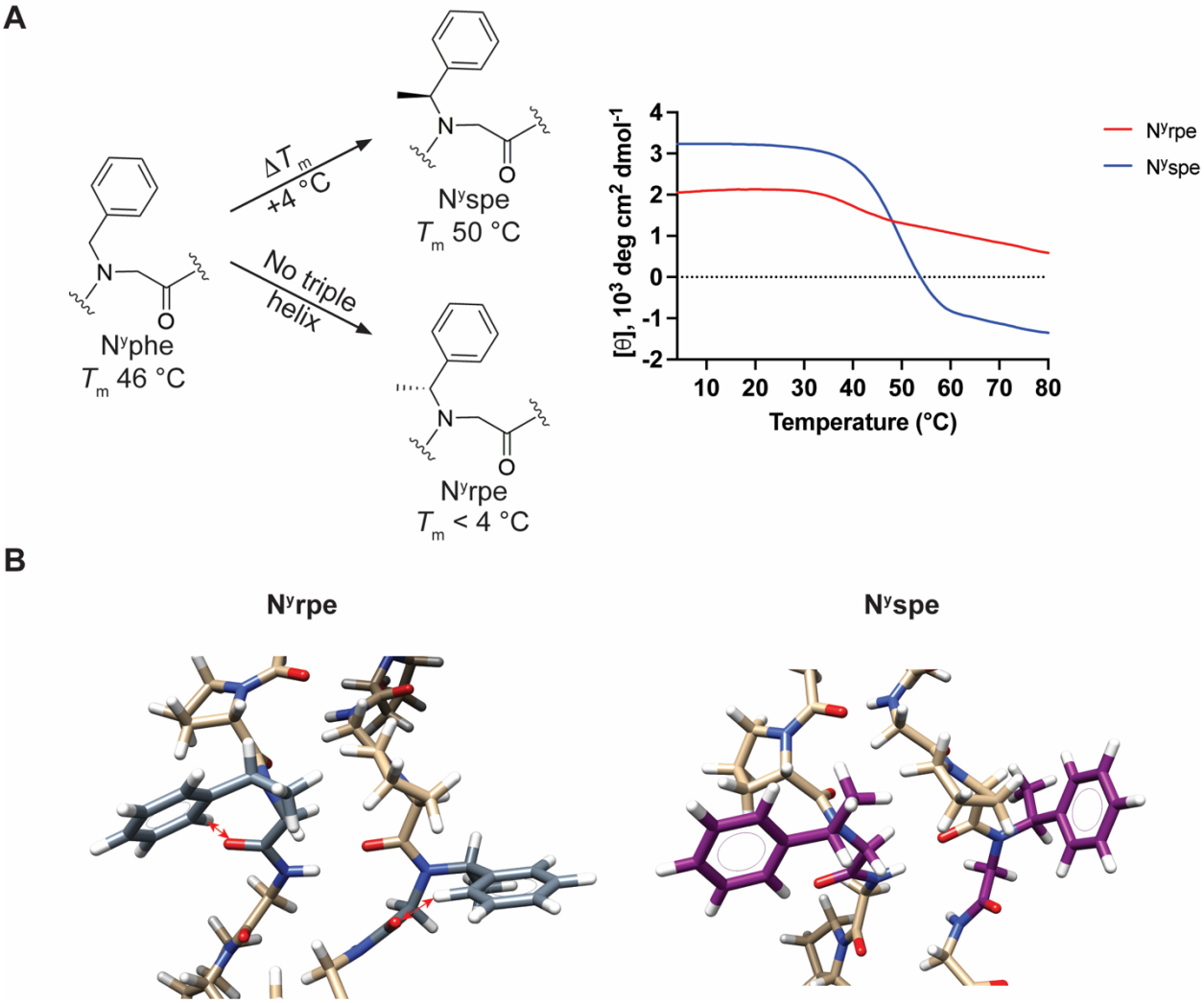


Figure 5 Stereochemistry of N-Cα branched N-Glys affects the stability of collagen triple helices. (A) Nrpe and Nspe-CMPs have a methyl branch on their N-Cα atoms. The melting curve of Nʸrpe-CMP does not show a clear two-state transition. In contrast, Nʸspe-CMP folded into a more stable triple helix with a T_m 4 ºC higher than that of Nʸphe-CMP. It appears that the Yaa position prefers (*S*)-N-Cα branching. (B) Molecular models of CMP triple helices containing Nʸrpe (gray) and Nʸspe (purple) residues. Potential clash of intra-residue atoms for Nʸrpe are indicated by red arrows. The N-Cα methyl group of Nʸspe is positioned between the two peptide strands. The third CMP strands in both models were removed for clarity

We first examined whether the inherent PPII propensity of Nrpe and Nspe affect triple helical folding, because the collagen triple helix consists of three PPII strands. We incorporated Nspe and Nrpe into the Zaa guest position of another host peptide with the sequence Ac-GlyProProZaaProProGly-NH_2 (Nᶻaaa-PP5) to compare their PPII propensities. This host-guest system was previously used to assess the

PPII propensities of canonical amino acids and N-Glys [35, 63]. Surprisingly, N^zspe-PP5 showed a CD trace similar to that of the random coil, whereas N^zrpe-PP5 exhibited a CD trace similar to a combination of PPI and PPII helices with negative and positive peaks appearing around 204 nm and 223 nm, respectively (**Figure S10**). Nrpe and Nspe normally prefer *cis* amide configuration; however a recent study showed that N-Glys favoring *cis* configuration can form stable triple helices [64]. These results indicated that factors other than PPII propensity of the N-Gly itself were responsible for the stereochemical effects of the Nrpe and Nspe residues on the triple helical folding.

Three dimensional computational models of N^xr/spe-CMPs and N^yr/spe-CMPs were generated from the crystal structures of N^xphe-CMP (PDB entry 7JX5) and N^yphe-CMP, respectively, using UCSF Chimera [48]. The models revealed that the van der Waals radii of the α-methyl groups of N^xspe and N^xrpe residues overlapped with the OH group of Hyp_{i+1} and the C=O group of Pro_i, respectively, in an adjacent CMP strand (**Figure S11**). Rotations of the N-C_α bond to avoid these clashes in N^xspe-CMP and N^xrpe-CMP led to new inter-strand steric clashes between the phenyl ring and Hyp residues, which explains their lower T_m compared to N^xphe-CMP. In the case of N^yrpe-CMP, the phenyl branch of the N^yrpe residue oriented towards the backbone clashing with its C=O group (**Figures 5B** and **S12A**). In contrast, the (*S*)-phenylethyl side chain of the Nspe residue at the Yaa position did not have inter- and intra-strand steric clashes. The spatial arrangement of the N^yspe side chain appeared to restrict the N-C_α bond rotation and possibly favor CH···π interactions between the aromatic ring and the intra-strand Pro_{i-1} (**Figures 5B** and **S12B**) as observed in other peptides and proteins [65]. This may have led to the stabilization of the triple helix and an increase in the T_m by 4 °C.

To further understand the stereo-positional effects of N^xr/spe and N^yr/spe on triple helical folding (**Figure 6**), we conducted MD simulations of the $N^{(x/y)}$s/rpe-CMPs with focus on not only the backbone structure but also the side chain rotamers. The MD simulations showed that a number of N^xrpe and N^xspe rotamers adopted ϕ and ψ angles outside the P_{II} region (**Figure 6A**), indicating the disruption of the PPII and triple helices. N^xrpe caused more disruption than N^xspe, which is consistent with our previous

experimental data [35]. In contrast, the rotamers of N^{y}rpe and N^{y}spe were more localized in the P_{II} region. These results suggest that substitutions of Nrpe and Nspe at the Yaa position cause less triple helical perturbation than substitutions at the Xaa position. This is consistent with the CD thermal melting results of N^{y}spe being the most stabilizing among the N^{x}r/spe and N^{y}r/spe series. Additionally, the (ϕ, ψ) maps of Nrpe in the Xaa and Yaa poisitons were more spread out than those of Nspe, emphasizing that Nrpe is less compatible with the collagen triple helix than Nspe.

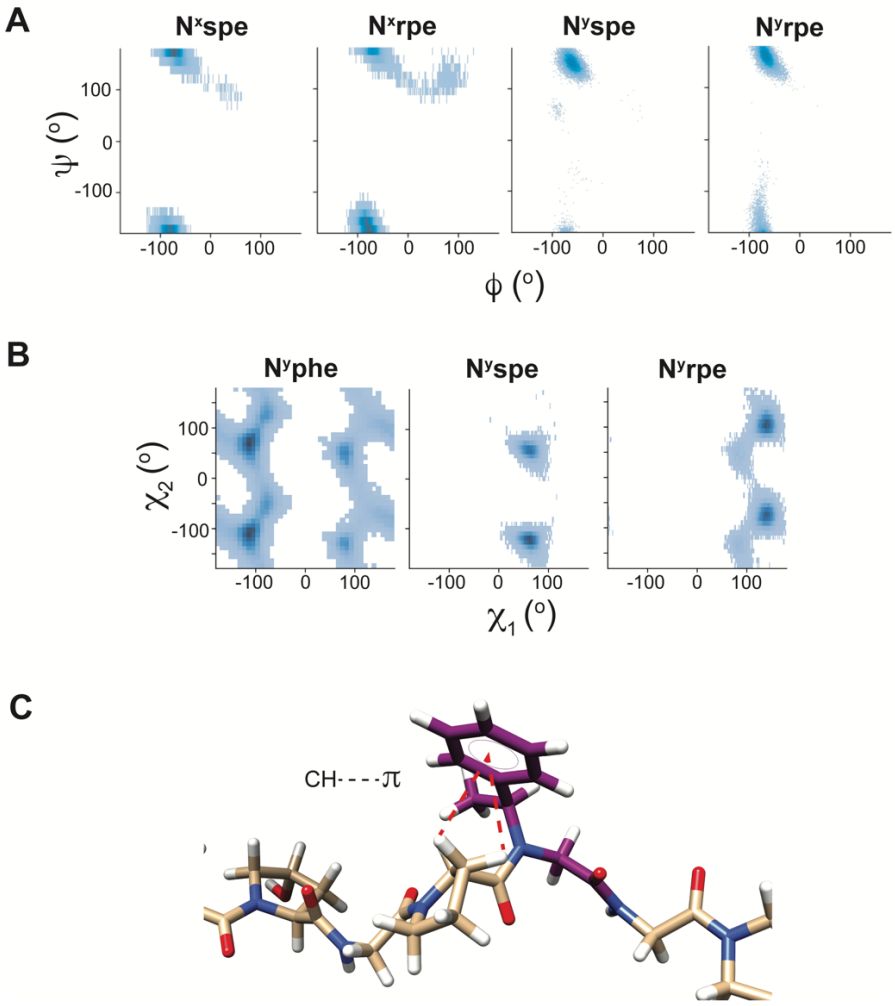


Figure 6 Molecular dynamics (MD) simulations of Nrpe and Nspe showing the potential origins of stereo-positional effects of N^{x}r/spe and N^{y}r/spe in the triple helical stabilization. (A) The ϕ–ψ angle plots of N^{x}r/spe and N^{y}r/spe show the stereochemical and positional differences in backbone rotamer distributions. (B) The χ_1–χ_2 plots of the N^{y}phe, N^{y}spe, and N^{y}rpe side chains from the MD simulations demonstrate that the α-methyl branch restricts the rotation of the N-Cα bond, resulting in fewer rotamers adoptable by the phenyl ethyl side chains. Each blue dot in the dihedral angle plots in (A) and (B) represents one rotamer. The dark blue area indicates the highly populated region. (C) The

potential CH•••π interactions (as indicated by red dashed lines) between the phenyl ring and intrastrand Pro_{i-1} as a result of the restricted N-C_α bond rotation within N^yspe-CMP.

In addition to the backbone dihedral angles, we speculated that side chain conformations of N^yrpe and N^yspe may also contribute to the difference in the triple helical stabilization because their (ϕ, ψ) maps are generally similar. Many experimental results [66] have established that the Nspe and Nrpe residues favor a *cis* amide bond, and that peptoid side chains generally orient away from the backbone (χ_1 has the same sign as ϕ). In the context of the collagen triple helix, the amide is in the *trans* configuration and ϕ is necessarily around −75°, so the capacity of the side chain to accommodate these features may determine the stability of the construct. The three key contacts that these side chains can make are with the carbonyl oxygen, the C_β of the preceding Pro_{i-1}, and the C_δ of a Pro on an adjacent chain of the triple helix.

The MD simulation results for side chain conformations are in agreement with the aforementioned molecular modeling. As shown in **Figures 6B** and **S13-S15**, the N^yphe side chain is free to adopt more rotamers than N^yrpe and N^yspe but favors rotamers with χ_1 around -120°, which pack the phenyl ring against C_γ and C_δ of the adjacent chain. Both N^yrpe and N^yspe adopted rotamers with χ_1 values between 60° and 180°, which orient the phenyl ring away from the adjacent chain. These positive χ_1 values are generally preferred by Nrpe in other chemical structures [66], but they lead to clashes between the *ortho* carbon atoms of the aromatic ring and the carbonyl oxygen (**Figures 5B**, **S12A**, and **S14**) in the context of the collagen triple helix. In N^yspe, however, one set of rotamers (χ_1 ~90°, χ_2 ~60°, ~-120°) avoids all of these clashes and comprises only rotamers populated in the simulations (**Figures 6B** and **S15**). Although both of these side chains show an effect of distorting their backbone ψ value away from its ideal value (~170°) to reduce the clashes between the aromatic C and C=O (**Figures S14 and S15**), ψ distortion is more pronounced in N^yrpe, which may have resulted in the perturbation of the overall helical conformation of CMP, as evidenced by the ill-defined melting transition.

Since computational experiments suggested that the N^yspe residue is conducive to triple helical folding and that a highly populated rotamer show potential CH•••π interactions between the aromatic ring

and intra-strand Pro$_{i-1}$, we synthesized N^{y}sch-CMP which has phenyl ring of N^{y}spe-CMP replaced with cyclohexyl ring. The phenyl to cyclohexyl substitution reduced the T_m of this CMP by a 10 °C (T_m of N^{y}sch-CMP: 40 °C, **Figures S3** and **S16**), suggesting that the aromaticity of the N^{y}spe residue plays a major role in stabilizing the triple helix. In contrast, the T_m remained the same for N^{y}rch-CMP and N^{y}rpe-CMP (**Figures S3** and **S16**). These results support the implications from molecular modeling that the highly populated rotamers of N^{y}spe-CMP are stabilized by CH•••π interactions between the aromatic ring and intra-strand Pro$_{i-1}$ (**Figure 6C**). N^{y}phe-CMP and N^{y}tyr-CMP, which can have similar CH•••π interactions, are also the ones that showed the greatest stabilization from C$_\alpha$ to N$_\alpha$ side chain substitution among all canonical amino acids. Thus, our work shows that Nspe stands out as the most stabilizing residue for the triple helix at the Yaa position, surpassing all other canonical and non-canonical amino acids except Hyp.

CONCLUSION

We demonstrated that at both Xaa and Yaa positions of the GlyXaaYaa repeat, moving the amino acid side chain from C$_\alpha$ to the amino group to produce N-Glys results in stabilization of the triple helix. At the Xaa position, some N-Glys were able to stabilize the triple helix more than Pro; however N-Glys at the Yaa position were less stabilizing than Pro and hydroxylation of Pro at the Yaa position created exceptional stabilization for the triple helix beyond what could be achieved by natural amino acids or N-Glys. X-ray crystallography confirmed that the N-Glys at both positions are conducive to triple helical folding; however, it also showed clear difference in preferred backbone ϕ and ψ angles, which explains the different level of triple helix stabilization at the Xaa and Yaa position compared to Pro and Hyp, respectively. Metadynamic simulation demonstrated a more dispersed free energy landscape map [wider (ϕ, ψ) space] for N-Glys at the Yaa position than at the Xaa position, likely due to steric influence from nearby residues. Chirality introduced at the N-C$_\alpha$ position had distinctively contrasting effects on triple helical folding at the Xaa and Yaa position, evidenced by extensive experimental and computational data which included dihedral angle plots of side chain rotamers. Finally, our work led to the discovery of CH•••π interactions of chiral Nspe at

the Yaa position, which are able to stabilize the triple helix much more than any other canonical and non-canonical residues, except Hyp. This work expands our understanding of the N-Glys's influence on triple helical folding at both Xaa and Yaa position and also provides new strategies for designing N-Glys that can not only produce large triple helix stabilization but also selective stabilization at the Xaa or Yaa positions. The knowledge gained from our work enables incorporation of a variety peptoid monomers into the triple helical peptides at any non-Gly positions to mimic or enhance the structure and functions of native collagens, offering a new platform strategy for the discovery of collagen-like therapeutics and biomaterials.

AUTHOR INFORMATION

Corresponding Author

S. Michael Yu – Departments of Biomedical Engineering and Molecular Pharmaceutics, University of Utah, Salt Lake City, UT 84112, USA.

ORCID: https://orcid.org/0000-0002-5171-2644

Email: michael.yu@utah.edu

Authors

Pahweenvaj Ratnatilaka Na Bhuket – Department of Molecular Pharmaceutics, University of Utah, Salt Lake City, UT 84112, USA.

ORCID: https://orcid.org/0000-0001-9506-4477

James R. B. Eastwood – Department of Chemistry, New York University, New York, New York 10003, USA.

ORCID: https://orcid.org/0000-0003-3895-5227

Zhao Qin – Department of Civil & Environmental Engineering, College of Engineering & Computer Science, Syracuse University, Syracuse, NY 13244, USA

ORCID: https://orcid.org/0000-0001-5696-8553

Julian L. Kessler – Department of Biomedical Engineering, College of Engineering, University of Utah, Salt Lake City, UT 84112, USA

ORCID: https://orcid.org/0000-0002-4497-235X

Yang Li – Guangdong Provincial Engineering Research Center of Molecular Imaging, the Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong 519000, China.

ORCID: https://orcid.org/0000-0003-4729-7860

Frank G. Whitby – Department of Biochemistry, School of Medicine, University of Utah, Salt Lake City, UT 84132, USA

Christopher P. Hill – Department of Biochemistry, School of Medicine, University of Utah, Salt Lake City, UT 84132, USA

ORCID: https://orcid.org/0000-0001-6796-7740

Kent Kirshenbaum – Department of Chemistry, New York University, New York, New York 10003, USA.

ORCID: https://orcid.org/0000-0003-0426-2772

ACKNOWLEDGEMENTS

This research received financial supports from the National Institute of Health (R01AR071358, R21EY029430, and R21OD026618) awarded to S. M. Yu. The X-ray crystallography work reported in this publication was supported by the Office of the Director, the National Institutes of Health, under Award Number S10OD030326 to C. P. Hill. The collection of the diffraction data for N^{y}phe-CMP is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beamline is funded by an NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, for collecting the data for N^{y}nbz-CMP is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of

561 Biological and Environmental Research and by the National Institutes of Health, National Institute of
562 General Medical Sciences (P30GM133894). Molecular graphics and analyses were conducted using UCSF
563 Chimera, a software developed by the Resource for Biocomputing, Visualization, and Informatics at the
564 University of California, San Francisco, with support from NIH grant P41-GM103311. The contents of this
565 publication are solely the responsibility of the authors and do not necessarily represent the official views of
566 NIGMS or NIH.

567

568 **CONFLICT OF INTEREST**

569 The authors declare no competing financial interest to this work.

570

References

1. Frantz, C., K.M. Stewart, and V.M. Weaver, *The extracellular matrix at a glance.* Journal of Cell Science, 2010. **123**(24): p. 4195-4200.
2. Kular, J.K., S. Basu, and R.I. Sharma, *The extracellular matrix: Structure, composition, age-related differences, tools for analysis and applications for tissue engineering.* Journal of Tissue Engineering, 2014. **5**: p. 2041731414557112.
3. Shoulders, M.D. and R.T. Raines, *Collagen Structure and Stability.* Annual Review of Biochemistry, 2009. **78**(1): p. 929-958.
4. Ricard-Blum, S., *The Collagen Family.* Cold Spring Harbor Perspectives in Biology, 2011. **3**(1).
5. Ramshaw, J.A.M., N.K. Shah, and B. Brodsky, *Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host–Guest Triple-Helical Peptides.* Journal of Structural Biology, 1998. **122**(1): p. 86-91.
6. Beck, K., et al., *Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine.* Proceedings of the National Academy of Sciences, 2000. **97**(8): p. 4273.
7. Persikov, A.V., et al., *Amino Acid Propensities for the Collagen Triple-Helix.* Biochemistry, 2000. **39**(48): p. 14960-14967.
8. Bella, J., *Collagen structure: new tricks from a very old dog.* Biochemical Journal, 2016. **473**(8): p. 1001-1025.
9. Sakakibara, S., et al., *Synthesis of poly-(L-prolyl-L-prolylglycyl) of defined molecular weights.* Bulletin of the Chemical Society of Japan, 1968. **41**(5): p. 1273-1273.
10. Kubyshkin, V., *Stabilization of the triple helix in collagen mimicking peptides.* Organic & Biomolecular Chemistry, 2019. **17**(35): p. 8031-8047.
11. Kar, K., Y.-H. Wang, and B. Brodsky, *Sequence dependence of kinetics and morphology of collagen model peptide self-assembly into higher order structures.* Protein Science, 2008. **17**(6): p. 1086-1095.

12. Holmgren, S.K., et al., *Code for collagen's stability deciphered.* Nature, 1998. **392**(6677): p. 666-667.
13. Holmgren, S.K., et al., *A hyperstable collagen mimic.* Chemistry & Biology, 1999. **6**(2): p. 63-70.
14. DeRider, M.L., et al., *Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations.* Journal of the American Chemical Society, 2002. **124**(11): p. 2497-2505.
15. Hodges, J.A. and R.T. Raines, *Stereoelectronic Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers.* Journal of the American Chemical Society, 2003. **125**(31): p. 9262-9263.
16. Hodges, J.A. and R.T. Raines, *Stereoelectronic and Steric Effects in the Collagen Triple Helix: Toward a Code for Strand Association.* Journal of the American Chemical Society, 2005. **127**(45): p. 15923-15932.
17. Kotch, F.W., I.A. Guzei, and R.T. Raines, *Stabilization of the Collagen Triple Helix by O-Methylation of Hydroxyproline Residues.* Journal of the American Chemical Society, 2008. **130**(10): p. 2952-2953.
18. Shoulders, M.D., I.A. Guzei, and R.T. Raines, *4-Chloroprolines: Synthesis, conformational analysis, and effect on the collagen triple helix.* Biopolymers, 2008. **89**(5): p. 443-454.
19. Zhang, Y., R.M. Malamakal, and D.M. Chenoweth, *Aza-Glycine Induces Collagen Hyperstability.* Journal of the American Chemical Society, 2015. **137**(39): p. 12422-12425.
20. Zhang, Y., M. Herling, and D.M. Chenoweth, *General Solution for Stabilizing Triple Helical Collagen.* Journal of the American Chemical Society, 2016. **138**(31): p. 9751-9754.
21. Kasznel, A.J., et al., *Structural Basis for Aza-Glycine Stabilization of Collagen.* Journal of the American Chemical Society, 2017. **139**(28): p. 9427-9430.
22. Harris, T. and D.M. Chenoweth, *Sterics and Stereoelectronics in Aza-Glycine: Impact of Aza-Glycine Preorganization in Triple Helical Collagen.* Journal of the American Chemical Society, 2019. **141**(45): p. 18021-18029.

23. Melton, S.D., et al., *Rules for the design of aza-glycine stabilized triple-helical collagen peptides.* Chemical Science, 2020. **11**(39): p. 10638-10646.
24. Egli, J., et al., *Hydrophobic Moieties Bestow Fast-Folding and Hyperstability on Collagen Triple Helices.* Journal of the American Chemical Society, 2019. **141**(14): p. 5607-5611.
25. Egli, J., et al., *Influence of Lipidation on the Folding and Stability of Collagen Triple Helices—An Experimental and Theoretical Study.* Journal of the American Chemical Society, 2021. **143**(15): p. 5937-5942.
26. Russell, L.E., J.A. Fallas, and J.D. Hartgerink, *Selective Assembly of a High Stability AAB Collagen Heterotrimer.* Journal of the American Chemical Society, 2010. **132**(10): p. 3242-3243.
27. O'Leary, L.E.R., J.A. Fallas, and J.D. Hartgerink, *Positive and Negative Design Leads to Compositional Control in AAB Collagen Heterotrimers.* Journal of the American Chemical Society, 2011. **133**(14): p. 5432-5443.
28. Fallas, J.A., et al., *Structural Insights into Charge Pair Interactions in Triple Helical Collagen-like Proteins*.* Journal of Biological Chemistry, 2012. **287**(11): p. 8039-8047.
29. Wei, F., J.A. Fallas, and J.D. Hartgerink, *Sequence Position and Side Chain Length Dependence of Charge Pair Interactions in Collagen Triple Helices.* Macromolecular Rapid Communications, 2012. **33**(17): p. 1445-1452.
30. Jalan, A.A. and J.D. Hartgerink, *Simultaneous Control of Composition and Register of an AAB-Type Collagen Heterotrimer.* Biomacromolecules, 2013. **14**(1): p. 179-185.
31. Hulgan, S.A.H., et al., *Covalent Capture of Collagen Triple Helices Using Lysine–Aspartate and Lysine–Glutamate Pairs.* Biomacromolecules, 2020. **21**(9): p. 3772-3781.
32. Feng, Y., G. Melacini, and M. Goodman, *Collagen-Based Structures Containing the Peptoid Residue N-Isobutylglycine (Nleu): Synthesis and Biophysical Studies of Gly-Nleu-Pro Sequences by Circular Dichroism and Optical Rotation.* Biochemistry, 1997. **36**(29): p. 8716-8724.

33. Melacini, G., Y. Feng, and M. Goodman, *Collagen-Based Structures Containing the Peptoid Residue N-Isobutylglycine (Nleu): Conformational Analysis of Gly-Nleu-Pro Sequences by 1H-NMR and Molecular Modeling.* Biochemistry, 1997. **36**(29): p. 8725-8732.

34. Kwak, J., et al., *Triple Helical Stabilities of Guest–Host Collagen Mimetic Structures.* Bioorganic & Medicinal Chemistry, 1999. **7**(1): p. 153-160.

35. Kessler, J.L., et al., *Peptoid Residues Make Diverse, Hyperstable Collagen Triple-Helices.* Journal of the American Chemical Society, 2021. **143**(29): p. 10910-10919.

36. Improta, R., C. Benzi, and V. Barone, *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution.* Journal of the American Chemical Society, 2001. **123**(50): p. 12568-12577.

37. Improta, R., et al., *Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 2. A Quantum Mechanical/Molecular Mechanical Study of (Proline-Proline-Glycine)n Polypeptides.* Journal of the American Chemical Society, 2002. **124**(26): p. 7857-7865.

38. Soltis, S.M., et al., *New paradigm for macromolecular crystallography experiments at SSRL: automated crystal screening and remote data collection.* Acta Crystallographica Section D, 2008. **64**(12): p. 1210-1221.

39. McPhillips, T.M., et al., *Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines.* Journal of Synchrotron Radiation, 2002. **9**(6): p. 401-406.

40. Cohen, A.E., et al., *An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot.* Journal of Applied Crystallography, 2002. **35**(6): p. 720-726.

41. Russi, S., et al., *The Stanford Automated Mounter: pushing the limits of sample exchange at the SSRL macromolecular crystallography beamlines.* Journal of Applied Crystallography, 2016. **49**(2): p. 622-626.

42. Gonzalez, A. and Y. Tsai, *AutoXDS*. 2010, Stanford University, Stanford, CA.
43. Kabsch, W., *XDS.* Acta Crystallographica Section D, 2010. **66**(2): p. 125-132.
44. Evans, P., *An introduction to data reduction: space-group determination, scaling and intensity statistics.* Acta Crystallographica Section D, 2011. **67**(4): p. 282-292.
45. McCoy, A.J., et al., *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**(4): p. 658-674.
46. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallographica Section D, 2004. **60**(12): p. 2126-2132.
47. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallographica Section D, 2012. **68**(4): p. 352-367.
48. Pettersen, E.F., et al., *UCSF Chimera—A visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.
49. Barducci, A., G. Bussi, and M. Parrinello, *Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method.* Physical Review Letters, 2008. **100**(2): p. 020603.
50. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
51. Bonomi, M., et al., *PLUMED: A portable plugin for free-energy calculations with molecular dynamics.* Computer Physics Communications, 2009. **180**(10): p. 1961-1972.
52. Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* Journal of Computational Chemistry, 2010. **31**(4): p. 671-690.
53. MacKerell, A.D., et al., *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins.* The Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
54. Riniker, S. and G.A. Landrum, *Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation.* Journal of Chemical Information and Modeling, 2015. **55**(12): p. 2562-2574.

55. Jakalian, A., et al., *Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method.* Journal of Computational Chemistry, 2000. **21**(2): p. 132-146.

56. Wang, J., et al., *Development and testing of a general amber force field.* Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.

57. Tian, C., et al., *ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution.* Journal of Chemical Theory and Computation, 2020. **16**(1): p. 528-552.

58. Liu, D.C. and J. Nocedal, *On the limited memory BFGS method for large scale optimization.* Mathematical Programming, 1989. **45**(1): p. 503-528.

59. Onufriev, A., D. Bashford, and D.A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model.* Proteins: Structure, Function, and Bioinformatics, 2004. **55**(2): p. 383-394.

60. Kovacs, J.M., C.T. Mant, and R.S. Hodges, *Determination of intrinsic hydrophilicity/hydrophobicity of amino acid side chains in peptides in the absence of nearest-neighbor or conformational effects.* Peptide Science, 2006. **84**(3): p. 283-297.

61. Nishi, Y., et al., *Different Effects of 4-Hydroxyproline and 4-Fluoroproline on the Stability of Collagen Triple Helix.* Biochemistry, 2005. **44**(16): p. 6034-6042.

62. Hollingsworth, S.A. and P.A. Karplus, *A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins.* 2010. **1**(3-4): p. 271-283.

63. Brown, A.M. and N.J. Zondlo, *A propensity scale for type II polyproline helices (PPII): aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline-aromatic interactions.* Biochemistry, 2012. **51**(25): p. 5041-51.

64. Qiu, R., et al., *Cis-trans isomerization of peptoid residues in the collagen triple-helix.* Nature Communications, 2023. **14**(1): p. 7571.

65. Zondlo, N.J., *Aromatic–Proline Interactions: Electronically Tunable CH/π Interactions.* Accounts of Chemical Research, 2013. **46**(4): p. 1039-1049.

66. Eastwood, J.R.B., et al., *Guidelines for designing peptoid structures: Insights from the Peptoid Data Bank.* Peptide Science, 2023. **115**(3).