

Peptoid Residues Make Diverse, Hyperstable Collagen Triple Helices

Julian L. Kessler¹, Grace Kang¹, Zhao Qin², Helen Kang¹, Frank G. Whitby³, Thomas E.

Cheatham III⁴, Christopher P. Hill³, Yang Li^{1,*}, and S. Michael Yu^{1,5}

¹Department of Biomedical Engineering, University of Utah, Salt Lake City, Utah 84112, USA

²Department of Civil & Environmental Engineering, College of Engineering & Computer Science, Syracuse University, Syracuse, New York 13244, USA

³Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112, USA

⁴Department of Medicinal Chemistry, College of Pharmacy, L. S. Skaggs Pharmacy Research Institute, University of Utah, Salt Lake City, Utah 84112, USA

⁵Department of Pharmaceutics and Pharmaceutical Chemistry, University of Utah, Salt Lake City, Utah 84112, USA

***Corresponding Author:** Yang Li (yang.d.li@utah.edu)

Abstract

The triple-helical structure of collagen, responsible for collagen's remarkable biological and mechanical properties, has inspired both basic and applied research in synthetic peptide mimetics for decades. Since non-proline amino acids weaken the triple helix, the cyclic structure of proline has been considered necessary, and functional collagen mimetic peptides (CMPs) with diverse sidechains have been difficult to produce. Here we show that *N*-substituted glycines (N-glys), also known as peptoid residues, exhibit a general triple-helical propensity similar to or greater than proline, allowing synthesis of thermally stable triple-helical CMPs with unprecedented sidechain diversity. We found that the N-glys stabilize the triple helix by sterically promoting the preorganization of individual CMP chains into the polyproline-II helix conformation. Our findings were supported by the crystal structures of two atomic-resolution N-gly-containing CMPs, as well as experimental and computational studies spanning more than 30 N-gly-containing peptides. We demonstrated that N-gly sidechains with diverse exotic moieties including a 'click'-able alkyne and a photo-sensitive sidechain can be incorporated into stable triple helices, enabling functional applications such as spatio-temporal control of cell adhesion and migration on a gelatin matrix. The folding principles discovered in this study open up opportunities for a new generation of collagen mimetic therapeutics and materials with extraordinary properties.

Introduction

Among the twenty canonical amino acids, proline (Pro), formally an imino acid, stands out as the single residue that features an *N*-substitution. Pro's unique cyclic sidechain mandates restricted backbone dihedral angles and, when incorporated into a protein chain, its tertiary amide group lacks the *N*-hydrogen atom to donate a hydrogen bond. As a result, Pro disrupts the regular secondary conformations favored by most amino acids, such as α -helices and β -sheets, and promotes distinct folding patterns, such as β -turns and polyproline helices. From a chemical perspective, there are numerous *N*-substituted α -amino acids as possible protein building blocks¹, of which evolution has sampled only *one* candidate for ribosomal protein expression: Pro. The rules governing the folding of *N*-substituted amino acids and their relationship to Pro's unique conformation are intriguing topics for both fundamental science and practical applications of *de novo* protein designs².

Collagen, the most abundant protein in vertebrates, best exemplifies a natural protein structure dictated by Pro's folding. The collagen chain comprises a repetitive sequence of Gly-X-Y triplets, where the X and Y positions are often occupied by Pro and its post-translational hydroxylation product, 4(*R*)-hydroxyproline (Hyp)³. As many as 22% of all residues in human collagen are either Pro or Hyp⁴ and, as a result, the defining structural motif of collagen is the intertwining of three polyproline II-type helices^{5,6}. This unique *triple helix* motif is responsible for collagen's remarkable properties, including higher-order assembly⁷, mechanical strength⁸, resistance to proteases, and binding with numerous cell receptors⁹. The triple helix has also inspired many synthetic designs aimed at recapitulating nature's supramolecular chemistry^{10,11}.

Because collagens are large, insoluble proteins that are difficult to study holistically, many research groups have turned to synthetic collagen mimetic peptides (CMPs)^{3,6,12}. CMPs featuring GlyProPro or GlyProHyp triplets have been synthesized as models of the collagen triple helix since the late 1960s¹³. Studies of CMPs revealed that all natural amino acids destabilize the triple helix when replacing Pro in the GlyProHyp triplet¹⁴, while hydroxylation of Pro at the Y position stabilizes the structure³. Throughout the 1990s¹² and more recently^{15,16}, a substantial portion of CMP studies have focused on uncovering and exploiting the stabilizing effects of the

post-translational modification of Pro on the triple helix by leveraging synthetic Pro derivatives^{3,15-19}.

Despite more than five decades of vigorous research, there have been few attempts to produce a stable collagen triple helix without Pro, and the structural forces imposed on the triple-helical conformation by unnatural *N*-substituted amino acids remain almost entirely unexplored. As probably the only precedent, Goodman and coworkers showed that replacing the central Pro with *N*-isobutylglycine (Nleu) within the GlyProPro triplet of a CMP results in a more stable triple helix²⁰. They suggested that the stabilization results from interchain interactions between the hydrophobic sidechains of Nleu and the adjacent Pro^{21,22}; however this hypothesis was not tested further. Meanwhile, peptoids, which are synthetic oligomers of *N*-substituted glycines (the simplest subset of *N*-substituted α -amino acids, e.g., Nleu), have been studied for decades as a promising class of synthetic peptidomimetics in biomedicine and material science^{23,24}. Repetitive peptoid sequences with various sidechains have been shown to form robust conformations resembling polyproline helices²⁵⁻²⁹, as well as a multitude of secondary and higher order structures including helical bundles³⁰, superhelices³¹, and two-dimensional sheets^{32,33}. Yet, so far, the structural influences of peptoid residues upon collagen triple helicity remain unclear.

Here we present a systematic study on the triple-helical folding propensity of *N*-substituted glycines (N-glys, **Fig. 1a**), and demonstrate their potential for making stable collagen triple helices with extraordinarily diverse sidechain structures. Using host-guest peptide systems for both collagen-mimetic and polyproline peptides, we examined more than 30 different peptides featuring N-glys with a spectrum of sidechain chemistries, for their triple-helical and polyproline-II-helix (PPII) propensities. We also acquired atomic-resolution crystal structures of two N-gly containing CMPs and performed molecular dynamics (MD) simulations, which together provide insights into the molecular structure of peptoid residues within collagen triple helices and their folding mechanism. Furthermore, to demonstrate utility, we present stable triple helices featuring exotic N-gly sidechains including light sensitive ones, which enabled spatiotemporal control of cell adhesion and migration on a collagen substrate. So far, construction of synthetic collagen mimetics has critically relied on Pro and its derivatives; the

design principles uncovered in this study drastically expand the library of residues with high triple-helical propensity, which has immense implications for a new generation of collagen-mimetic therapeutics and materials.

Results and Discussion

Triple-helical stability of CMPs with N-glys. To investigate the triple-helical folding propensity of peptoids residues, we inserted a series of N-gly guests into the central X position of a conventional CMP host peptide with the sequence: Ac-(GlyProHyp)₃-Gly-X-Hyp-(GlyProHyp)₃-NH₂ (designated as X-CMP, **Fig. 1b**, and supplementary Materials and Methods)³⁴. We measured the X-CMPs' triple-helical stability *via* thermal unfolding experiments monitored under circular dichroism (CD, **Fig. 1c**, Supplementary Section 3 and **Table S1**), first in a group of N-glys with sidechains selected from the canonical amino acids (**Fig. 1d**). It is known that Pro is the most stabilizing amino acid at position X, and substitution from Pro to another canonical amino acid clearly reduces T_m by 4-17 °C (**Fig. 1d**, column AA)¹⁴. Surprisingly, we found that many of the N-glys with canonical sidechains were at least as stable as Pro, and almost *all* N-gly residues were more stable than their amino acid counterparts (**Fig. 1d**, exception: Nval), with the biggest T_m difference seen between Nphe- and Phe-CMP (**Fig. 1c**). These results demonstrate that, in the X position of the CMP, shifting the sidechain from the C α carbon to the nitrogen (i.e., transforming a canonical amino acid to its peptoid analogue) may improve triple-helical stability.

Crystal structures of N-gly-CMPs. To confirm the triple-helical folding of N-gly-CMPs and to decipher the precise molecular structure of the N-glys within the peptide conformation, we determined the X-ray crystal structures of two compounds. We chose Nlys-CMP and Nphe-CMP for crystallographic analysis because of their high stabilities and differing N-gly sidechain moieties, and determined their structures at resolutions of 0.95 Å and 1.10 Å, respectively (**Fig. 2a, 2b**; **Supplementary Table S2, Fig. S1, S2**). The overall conformation in both crystal structures was consistent with those reported for conventional CMPs found in the Protein Data Bank (PDB, **SI Table S3**), with the average (ϕ , ψ) angles and standard deviations of all non-

terminal Pro, Hyp, and Gly residues being ($-70 \pm 6^\circ$, $162 \pm 6^\circ$), ($-58 \pm 5^\circ$, $147 \pm 7^\circ$), and ($-69 \pm 3^\circ$, $174 \pm 5^\circ$), respectively (**Fig. 2c**, **SI Fig. S3**). The dihedral angles of the two N-glys [Nlys: ($-86 \pm 3^\circ$, $176 \pm 3^\circ$), Nphe: ($-82 \pm 3^\circ$, $177 \pm 3^\circ$)] were also within the range populated by the X-position residues of collagen peptides in the PDB (**Fig. 2d**), though slightly shifted from the PDB's average angles ($\Delta\phi = -13^\circ$, $\Delta\psi = +21^\circ$). This deviation did not significantly alter the backbone structure and there was neither kinking nor bending of the triple helix near the N-gly in either crystal structure. Furthermore, the inter-strand hydrogen bonding pattern, formed along the length of the triple helix, agreed well with that seen in other collagen peptides in the PDB (**Fig. 2e**), although the hydrogen bonds formed with N-glys appeared to adopt a slightly less ideal geometry (**Fig. 2f**). This suggests that the triple-helical stability of N-gly-CMPs does not originate from improved hydrogen bonding³⁵ of N-glys.

Probing the structural factors for N-glys' triple-helical stability. Next, we introduced a series of guest N-glys into the X-CMPs to understand how sidechain structures affect triple-helical stability (**Fig. 3**). We first focused on hydrophobic sidechains, which previous reports suggested to be the main stabilizing factor in N-gly-CMPs^{21,22}. We found that the aliphatic Nchx as well as the aromatic Nphe produced triple helices with stability notably higher than Pro (**Fig. 3a**). In addition, Nphe derivatives having an electron-rich (Ntyr) or poor (Nnbz) aromatic ring generated triple helices with similar stability, suggesting that neither the aromaticity nor its electron density is crucial to triple helix stabilization. Meanwhile, although structurally similar, the hydrophilic Ndxn and the deprotonated Ntyr (pH 12) were both more stabilizing than Pro and only slightly less stabilizing than their hydrophobic analogs in the group (**Fig. 3a**). These results suggest that the hydrophobicity of the peptoid residue might not be the main driving force for the triple helix stabilization. Furthermore, we synthesized a pair of CMPs featuring central guest triplets of GlyNpheHyp and GlyNphelle to examine the hydrophobic interactions between adjacent residues (SI Table S4). Introduction of Ile did not provide any additional stabilizing effect, further de-emphasizing hydrophobicity in triple helix stabilization.

Next, we investigated the charged and hydrophilic peptoid sidechains (**Fig. 3b**). Although charged Nasp and Nlys destabilized the triple helices modestly through electrostatic repulsions,

when we altered the pH so their sidechains were uncharged, both X-CMPs were as stable as Pro-CMP. Also, we noted that peptoid residues with similar sidechain configurations conferred highly similar T_m values (**Fig. 3b**: Nasp, Nasn, and Nleu; **Fig. 3a**: Nchx, Nphe, Nnbz, Ntyr, Ndxn), while within each group, T_m values of the hydrophobic residues were only higher by 2-4 °C.

Lastly, we found that progressively increasing the size of the aliphatic side chains produced a concomitant increase in T_m of the corresponding X-CMP (**Fig. 3c**). Knowing that the hydrophobicity has modest stabilizing effects for triple helices, we reasoned that the large increase in stability from Gly-CMP (T_m , 37 °C) to Nchx-CMP (T_m , 63 °C) in this series cannot be explained by the hydrophobicity alone, and were prompted to look into other mechanisms by which peptoid residues stabilize the collagen triple helix.

Polyproline-II-helix (PPII) propensity of N-glys. It is generally accepted that Pro and Hyp residues stabilize the collagen triple helix by pre-organizing individual collagen strands into the PPII conformation³⁶. Therefore, we hypothesized that N-glys, with their Pro-like tertiary amide structure, could induce similar pre-organizational effects in N-gly-CMPs. To test this hypothesis, we first surveyed a large set of small molecule crystal structures containing N-glys from the Cambridge Structural Database (**SI Table S5, S6**). We found that the dihedral angles of these peptoid residues largely fall within the most probable range for Pro, with the highest density present in the area corresponding to a PPII helix (**Fig. 4a**). This suggests that similar to Pro, N-glys have a high natural propensity for the PPII conformation.

To systematically study the N-glys' propensity to adopt the PPII conformation, we employed another host-guest peptide sequence: Ac-GlyProPro-X-ProProGlyTyr-NH₂, designated as X-PP5, where X represents the N-gly or amino acid residue replacing the central Pro within the (Pro)₅ sequence. This short, proline-rich host peptide was previously utilized to measure the PPII propensity of the twenty canonical amino acids based on the CD intensity near 228 nm^{37,38}. Except for Sar-PP5, all N-gly-CMPs showed PPII CD intensity higher than Pro-PP5 (**Fig. 4b**) and their amino acid counterparts³⁷. Notably, the PPII propensities of Nphe, Nlys, and Nleu were higher than that of Hyp, which is known for having the highest PPII propensity among all

natural amino acids³⁹. This systematic investigation supports that N-glys promote the intrinsic PPII folding of individual peptide strands, which can contribute to triple-helical stabilization. Energy-wise, the free energy bonus arising from the PPII preorganization by just a few bulky N-glys in a triple helix (**Fig. 4b**, $\Delta \Delta G$) can add up to be roughly an extra amide-amide hydrogen bond (-2.0 kcal/mol)⁴⁰.

Previous studies have shown that peptoids with specific *N*-substitutions can form stable polyproline helices^{25-29,41}. So far, the frontier of peptoid research has largely focused on control over the *cis-trans* isomerization of peptoid amide bonds (i.e., adoptions of PPI or PPII conformation) through specific steric or stereoelectronic effects induced by *N*-aryl or *N*-C α -chiral sidechains^{25-29,42}. In this work, we systematically quantitated the general PPII propensity of peptoid residues. We learned that although Pro has the highest PPII propensity among all canonical amino acids, it is on the low side among the peptoid residues (**Fig. 4b**). The strong PPII propensity of peptoid residues as compared to the amino acids implies that Pro may be the an imperfect, biosynthetic surrogate for N-glys⁴³ where PPII folding is needed, but its moderate folding propensity can be improved *via* post-translational hydroxylation (e.g., in collagen).

MD simulation and steric effects of the *N*-substitution. We conducted MD simulations for triple-helical X-CMPs incorporating guest residues Gly, Sar, Nleu, Nchx, or Pro (**Fig. 4c**). We based our simulations on the crystal structures of Nphe- and Nlys-CMPs (**Fig. 2**) and chose N-glys with simple aliphatic side chains that lack potential for electrostatic and hydrogen bonding interactions. The simulation results demonstrated that the Nleu and Nchx residues displayed energy landscapes very similar to Pro, with a low-energy valley confined to the regions (φ , ψ angles) for PPII and the triple helix. In comparison, Sar which has a smaller sidechain, and to a greater extent Gly, showed more conformational flexibility (**Fig. 4c**). These simulation results are in agreement with our CD experiments, which showed that an N-gly's PPII propensity (i.e., Gly < Sar < Nleu) and the triple-helical stability of its corresponding X-CMP correlate with the size of its *N*-substitution (**Fig. 3c & 4c**). As a plausible explanation for this steric effect, a bulky sidechain can impose a steric restraint to the N-gly's backbone carbonyl, thereby hindering the rotation around its nitrogen-carbon bond and constraining the φ angle to values appropriate for

triple-helical folding (**Fig. 4d**). This is evident in the most flexible guest residue Gly, where addition of a single *N*-methyl group from Sar increases the T_m by 13 °C (**Fig. 3c**). Additionally, we believe that an N-gly's ψ angle is less sterically influenced by the *N*-substitution because it is located further from the sidechain, and that it is more likely to assume a value close to $\pm 180^\circ$ because the α -carbon has no sidechain.

The pyrrolidine ring of Pro has been assumed to be a structural requisite for collagen mimetics for decades^{3,44}. In an attempt to replace Pro in CMPs, Raines and coworkers found that converting Pro to *N*-methyl-L-alanine (NMe-Ala), which removes only the γ carbon and eliminates the ring structure, substantially destabilizes the triple helix⁴⁵ (**SI Fig. S4**). From this observation, they concluded that the conformational restrictions imposed by the Pro ring are more important for triple-helical stability than the presence of an *N*-substitution⁴⁵. In contrast, we discovered that N-glys with bulky sidechains can form well-folded collagen triple helices without Pro's cyclic sidechain. We believe that the low triple helicity of NMe-Ala⁴⁵ is due to its poor PPII propensity (**Fig. 4b**). Without the ring, steric repulsion between the two adjacent methyl groups of NMe-Ala precludes formation of the dihedral angles necessary to form the polyproline-II helix, thereby nullifying any stabilizing effect from the *N*-methyl group (**SI Fig. S4**).

***N*-C α chiral branching.** To understand why Nval, the only peptoid residue with an *N*-C α -branched sidechain in our library, has unexpected, low triple-helical stability (T_m : 36 °C, **Fig. 1d**), we created a group of X-CMPs featuring guest residues NEt, Nval, Nphe, Nspe and Nrpe (**Fig. 5**). The T_m differences (**Fig. 5a**) indicated that adding an *N*-C α methyl group to an N-gly destabilizes the triple helix, and that the triple-helical folding was greatly affected by the chirality of the *N*-C α -branching (**Fig. 5a**, Nspe versus Nrpe). For each chirality, 3D modeling based on the Nphe-CMP crystal structure revealed a specific location of possible steric clash between the branching methyl group and an adjacent peptide strand (**Fig. 5b,c**), which provides an explanation for Nval's low triple-helicity despite its higher-than-Pro PPII propensity (**Fig. 4b**). Although additional confirmatory studies are needed, this is the first time that a

stereoselective steric interaction at the *N*- α -carbon is presented in the context of collagen triple helix.

Sidechain functionalization. After a series of mechanistic investigations into triple-helical folding, we turned to leveraging the diverse pool of peptoid residues to produce functionalized collagen mimetics for potential applications. As one example, we demonstrated facile sidechain functionalization through ‘click’ chemistry, utilizing an alkyne-bearing residue (Nakn, **Fig. 6a**, **SI Methods** and **SI Discussions**). In another example, to demonstrate precise control over the structure and function of CMP-based molecules and materials, we synthesized Nnbz-CMP, an X-CMP featuring a UV-cleavable *N*-*o*-nitrobenzyl (nbz) sidechain⁴⁶. We hypothesized that Nnbz-CMP would form a robust triple helix, but UV irradiation would convert the stabilizing Nnbz residue into Gly, thereby dramatically weakening the triple helix (**Fig. 6b**). CD measurements confirmed this hypothesis (**Fig. 6c**). Moreover, characterizations of CMPs incorporating more than one Nnbz residues (e.g., Nnbz2,3, SI Table S7) indicated that (i) the stabilizing effect of N-glys may be additive, and that (ii) the increased number of Pro \rightarrow Nnbz substitutions resulted in even wider T_m differences before and after UV irradiation (**Fig. 6c**).

Next, we examined the Nnbz-substituted CMP’s capacity to hybridize to denatured collagen and tested their release afterwards by UV irradiation. Previously, we reported that single-strand CMPs [e.g., (GlyProHyp)₉] can form hybrid triple helices with collagens that are denatured by heat, proteases, or mechanical damage^{9,69}. Using Nnbz2-CMP labeled with fluorescein (F-Nnbz2-CMP), We found that Nnbz2-CMP displayed a drastically higher level of binding to gelatin (heat-denatured collagen) than the host peptide Pro-CMP, consistent with its stronger triple-helical propensity (**Fig. 6d**). Furthermore, upon brief UV irradiation, over 80% of the bound F-Nnbz2-CMP was released from the gelatin matrix (**Fig. 6d**).

Spatio-temporal modification of cell culture substrates. To further demonstrate X-CMP’s applications in biomaterials, we conjugated Nnbz2-CMP to a multi-arm poly(ethylene glycol) (MA-PEG) polymer (**SI Fig. S5a,b**). The resulting conjugate (MA-PEG-CMP) readily formed a hydrogel that was physically crosslinked *via* its Nnbz2-CMP triple helices; this stable hydrogel

showed the ability to dissolve and release its bound contents on demand as the CMP crosslinks unfold upon UV irradiation (SI Fig. S5). Next, taking advantage of PEG's ability to repel cell adhesion and Nnbz2-CMP's UV-triggered unbinding from gelatin (Fig. 6d), we demonstrated that a crosslinked gelatin film can be coated with MA-PEG-CMP to accurately photo-pattern cell attachment in selected areas (SI Fig. S6), or to trigger cell migration into defined regions at a designated time (Fig. 6e,f). These proof-of-concept experiments showcase the practical use of X-CMPs for creating complex cell culture substrates with spatio-temporal controls.

Billions of stable, Pro-free triple helices. Lastly, we synthesized **X7-CMP**, a (Gly-X-Hyp)₇ sequence where each X position was populated with a different peptoid residue (Fig. 6g). With a T_m 1 °C higher than (GlyProHyp)₇, **X7-CMP** represents a hyperstable, synthetic collagen triple-helix with the greatest sidechain-diversity to date. With hundreds of amines commercially available for efficient solid-phase peptoid synthesis, this work expands the amino-acid library for stable CMPs by more than an order of magnitude. Even with a modest library of 20 N-glycyls for X-substitution, for a short sequence of (Gly-X-Hyp)₇, over 1.2 billion (20^7) triple-helical peptides can be generated. The design principles discovered in this study open the door for not only deciphering new fundamental molecular interactions (e.g., $n \rightarrow \pi^*$)⁴⁷ in polyproline and triple-helical folding, but also developing a new class of collagen-mimetic therapeutics and biomaterials with remarkable functionalities. With the unprecedented structural diversity offered by X-CMPs, we anticipate the long-awaited new era of *de novo* design of functional collagen peptidomimetics.

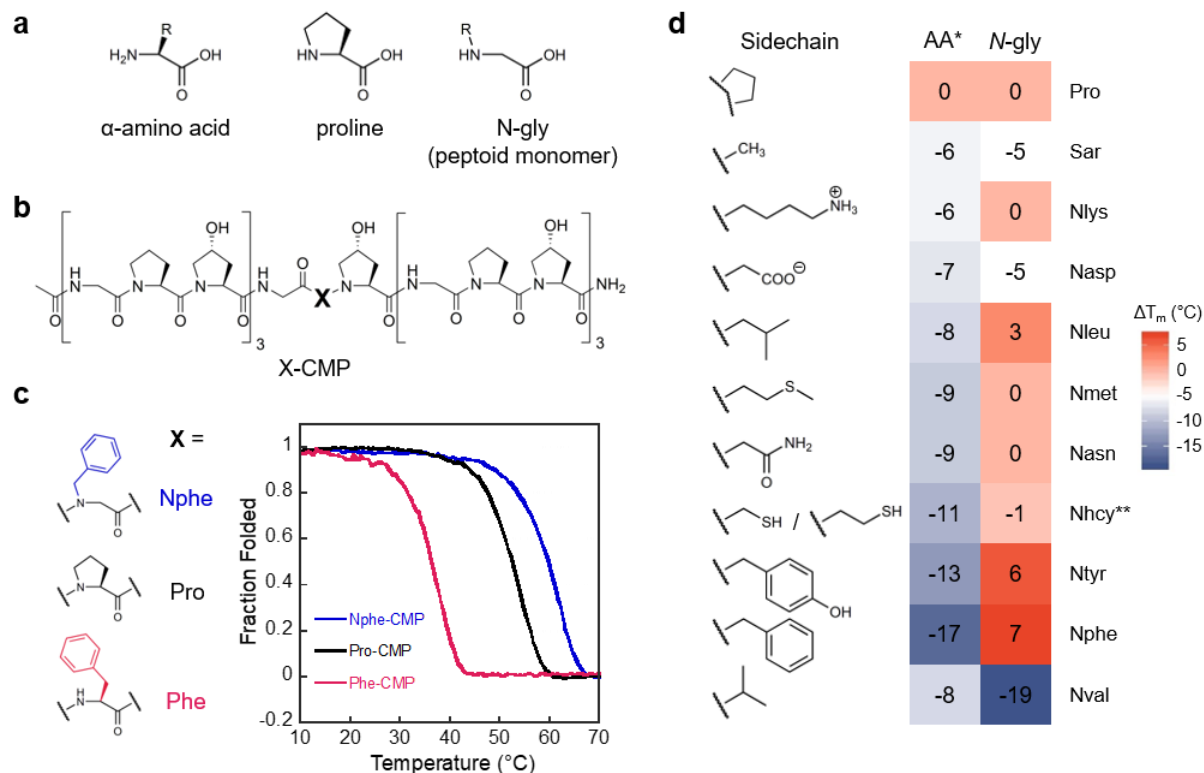


Fig. 1 | A single N-substituted glycine (N-gly), also known as a peptoid residue, in the central Pro position within (GlyProHyp)₇ shows high triple-helical propensity. **a**, General structures of a canonical amino acid, Pro, and an N-gly. **b**, Chemical structure of the X-CMP, a host-guest collagen mimetic peptide with the sequence of a (GlyProHyp)₇, where the Pro residue in the central X position is substituted, for example, with an N-gly residue. **c**, An X-CMP triple helix dissociates into single strands under gradual heating; monitored under CD, the middle point of this two-state transition is defined as the X-CMP's melting temperature (T_m). The Nphe sidechain at the X position resulted in a CMP triple helix considerably more stable than Pro (ΔT_m : +7 $^{\circ}\text{C}$). In comparison, the amino acid with the same sidechain was highly destabilizing (ΔT_m : -17 $^{\circ}\text{C}$). **d**, The triple helical stabilities of X-CMPs featuring N-glys with canonical sidechains, quantitated in ΔT_m (in comparison to the host peptide: Pro-CMP). Almost universally, peptoid residues (right column) have higher stability than their amino acid counterparts (left column) except for Nval, and many have stabilities similar to or better than Pro. * ΔT_m for amino acid residues taken from reference²¹. **Nhcy (homocysteine) is a similar but not exact equivalent to cysteine.

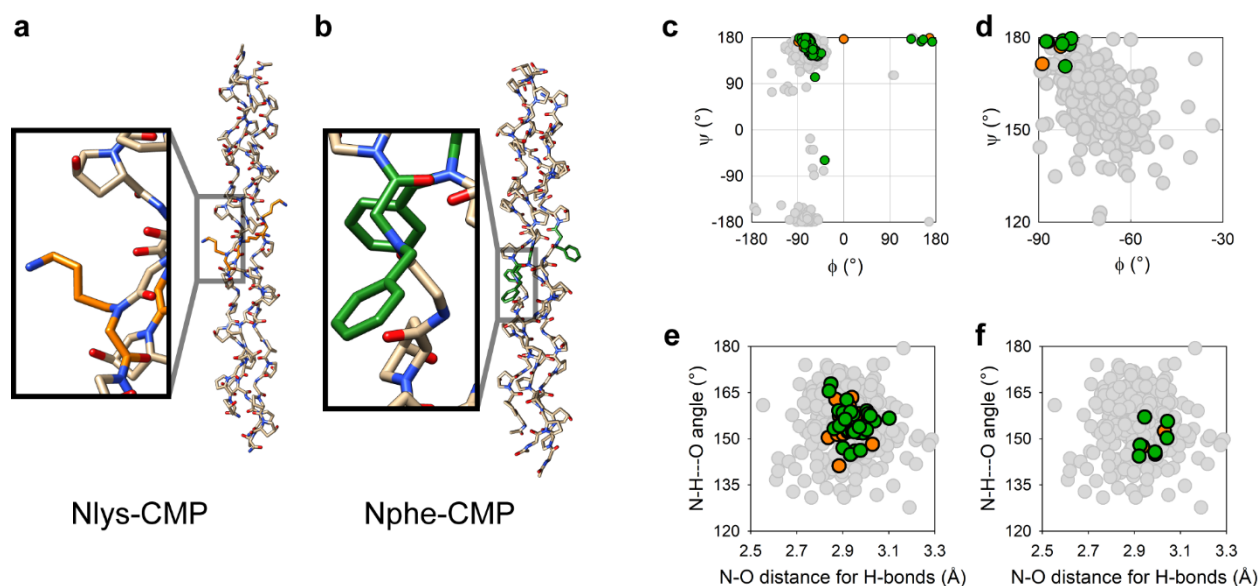


Fig. 2 | CMPs containing peptoid residues adopt a triple-helical conformation nearly identical to the native collagen. **a,b** Crystal structures of Nlys-CMP (**a**) and Nphe-CMP (**b**); each shows a triple helix with the N-gly residue uniquely colored (Nlys in orange and Nphe in green). **c**, A Ramachandran plot showing the ϕ and ψ angles of all amino acid residues in the crystal structure of Nlys-CMP (orange) and Nphe-CMP (green), overlaid with the ϕ and ψ angles from the reported CMP crystal structure in the PDB (grey, see SI for structure IDs used for comparison). **d**, A Ramachandran plot showing the ϕ and ψ angles for only the N-gly residues (Nlys in orange, Nphe in green), overlaid with the ϕ and ψ angles (in grey) from the X position of Gly-X-Y repeats of reported CMP crystal structures. The peptoid residues adopt a conformation common to the X position in previously reported Gly-X-Y collagen sequences, yet with slightly shifted average angles ($\Delta\phi$: -13° , $\Delta\psi$: $+21^\circ$). **e**, A plot of the N-H...O angle versus the N-O distance of all the interstrand amide-amide H-bonds in the crystal structures of Nlys-CMP (orange) and Nphe-CMP (green) overlaid with that (in grey) of all the H-bonds in the reported crystal structure of CMPs. **f**, A plot, same as **e** but showing H-bonds associated only with N-gly residues in the crystal structures (Nlys in orange, and Nphe in green). The H-bonds parameters for both residues are consistent with previously reported crystal structures of collagen peptides.

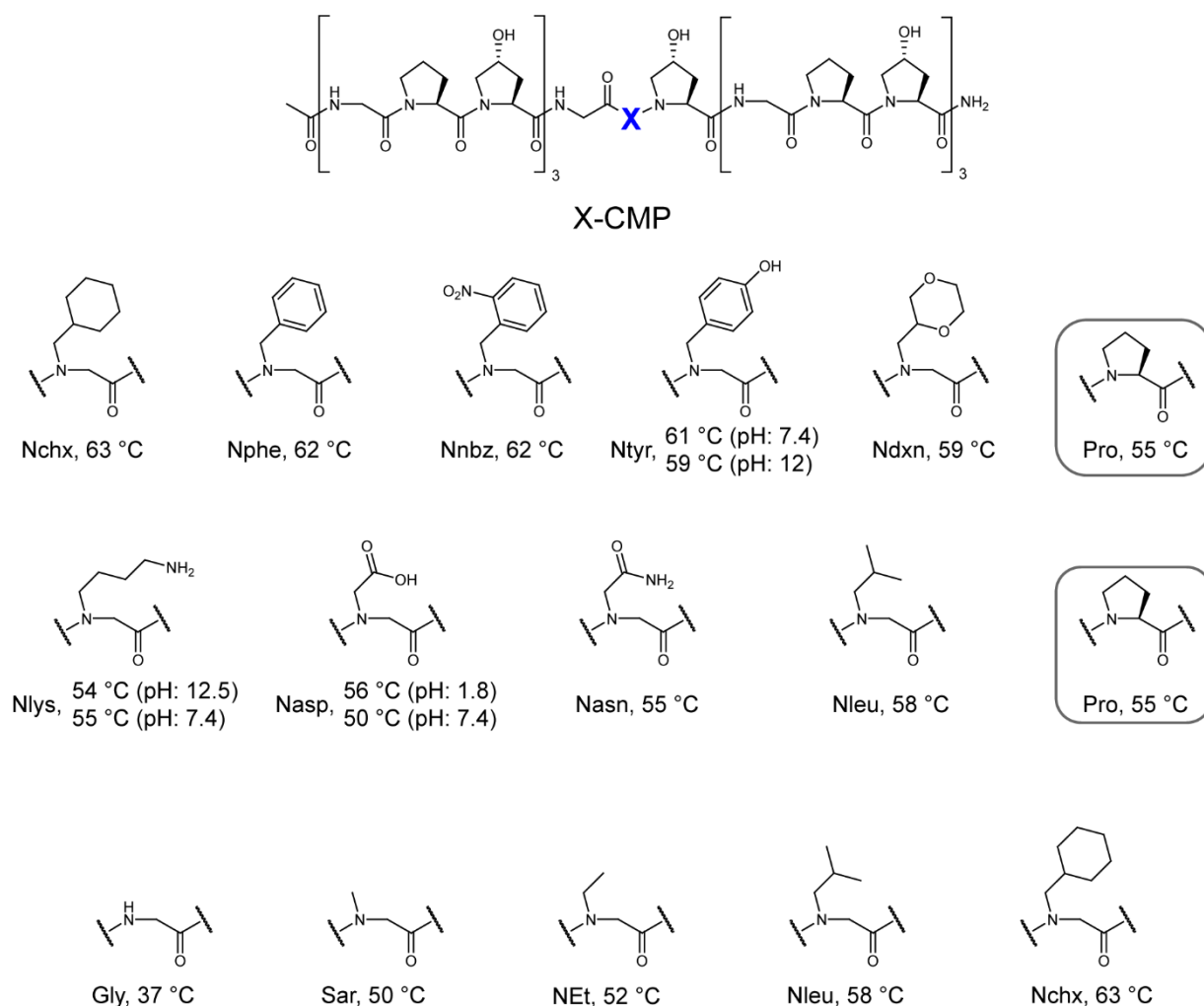


Fig. 3 | The structure of the peptoid sidechain affects triple-helical propensity. **a**, T_m values for a series of X-CMPs with N- α -substituted 6-membered rings in the sidechains. There is little difference in triple-helical stability among the aliphatic, aromatic, and hydrophilic rings, and every peptide had a T_m higher than the host: Pro-CMP. **b**, Although charged groups destabilized the triple helix at physiological pH, when the sidechains were uncharged, both Nlys and Nasp, as well as the hydrophilic Nasn, were as stabilizing as Pro. **c**, A series of X-CMPs with aliphatic sidechains of increasing size showing concomitantly increasing T_m values, implying an robust steric effect. Although the hydrophobicity in this series increases as the aliphatic sidechains increase in size, according to series **a** and **b**, hydrophobicity has only a minor effect on stability. The results suggest that the tertiary amide structure N-gly (which hinders the α helix and β sheet conformations) and the size of the N-substituted sidechain play the most critical roles in stabilization of N-gly-CMP triple helices.

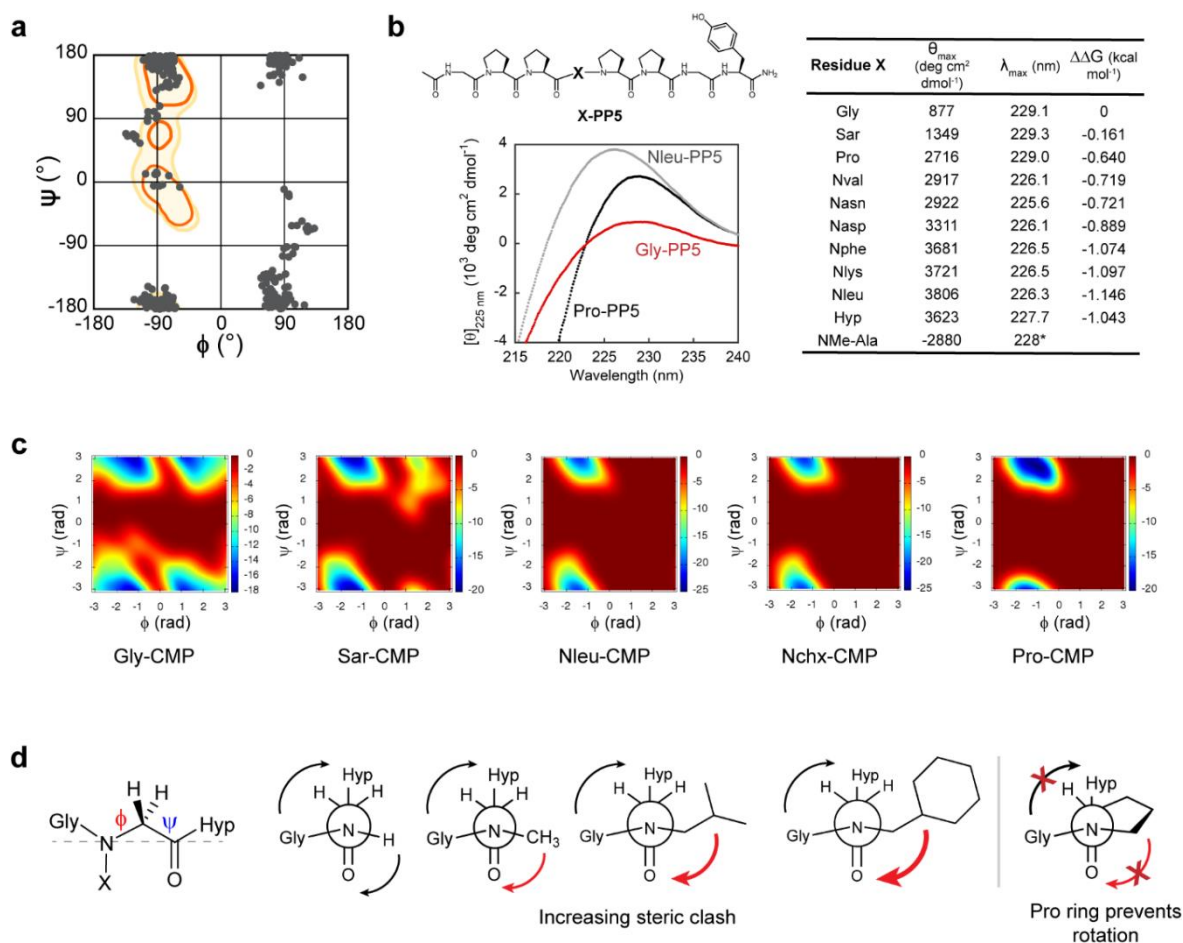


Fig. 4 | N-gly residues have a strong polyproline-II helix (PPII) propensity which improve the triple-helical stability of X-CMPs. **a**, A survey of ϕ and ψ angles of N-glys from high resolution crystal structures of small molecules (from Cambridge Structure Database). N-glys have a high natural propensity to fold into either the left- or right-handed polyproline conformation ($\phi \pm 90^\circ$, $\psi 180^\circ$). Watermarks show the regions highly populated by Pro residues. **b**, Structure of the PPII host-guest peptide **X-PP5** and the comparative CD spectra of Gly-PP5, Pro-PP5, and Nleu-PP5. The PPII propensity of each X residue within the X-PP5 model was assessed by the ellipticity value of the characteristic CD signal near 228 nm. All N-glys, with the exception of Sar, had a θ_{\max} higher than Pro indicating peptoids' high intrinsic PPII propensity. *No peak was observed near 228 nm for NMe-Ala (Supplementary data). **c**, Metadynamics calculations of the free energy landscape of the guest residues in Gly-, Sar-, Nleu-, Nchx-, and Pro-CMP. Each model yielded a native state consistent with the expected triple helix and PPII conformations (-70° , 170°); however, peptides with small guest residues (Gly and Sar) showed a secondary non-PPII energy well. **d**, Newman-like projections of N-glys with the grey dashed line indicating the line of sight for the projection. N-glys with large side chains can experience major steric clashes with its carbonyl group. Such reduced conformational flexibility of the N-gly may help preorganize the individual chains into a PPII conformation, thereby stabilizing the triple-helical assembly.

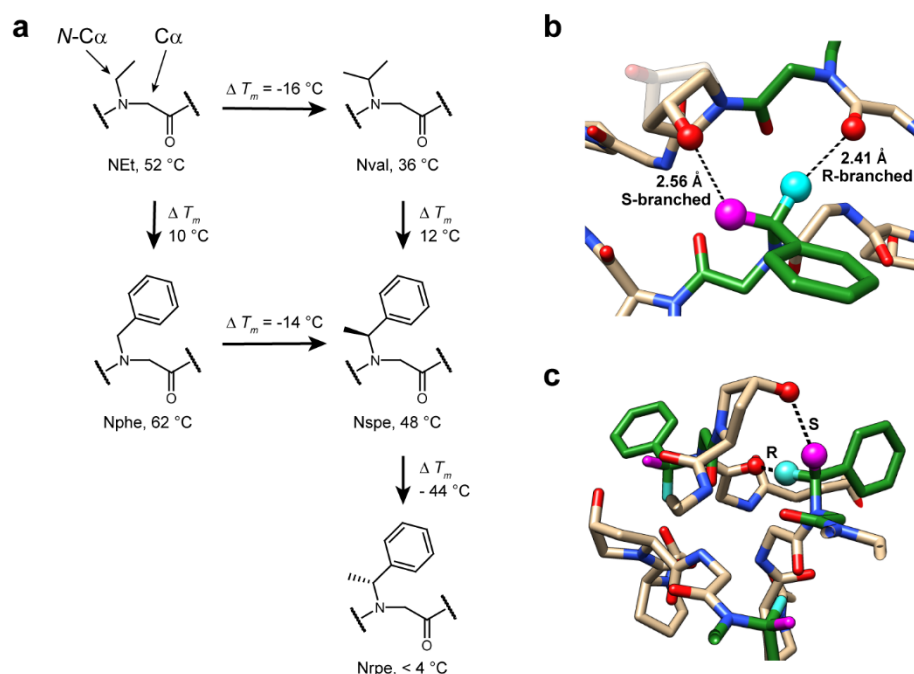


Fig. 5 | N-C α branching of peptoid residues affects triple-helical folding. **a**, With a N-C α branch, Nval- and Nspe-CMPs, had T_m values 14-16 °C lower than their unbranched analogs, NEt and Nphe (ΔT_m : -14 ~ -16 °C). Moreover, while residue Nspe conferred a triple helix with a T_m of 48 °C, its enantiomer, Nrpe completely abolished the triple-helical folding. **b,c**, Molecular modeling based on solved crystal structure of Nphe-CMP showing potential steric clashes within the triple helix introduced by the N-C α methyl branch of Nspe and Nrpe. The R-branch of Nrpe (cyan), pointing directly toward the inner core of the triple helix, may clash with the backbone carbonyl of the cross-chain Gly and interfere with proper backbone assembling; in contrast, the S-branch (magenta) in Nspe may clash with the hydroxyl group of a cross-chain Hyp, which may be less destabilizing since the steric repulsion may be alleviated by changing the Hyp's ring pucker (shown in half-transparency in **b**).

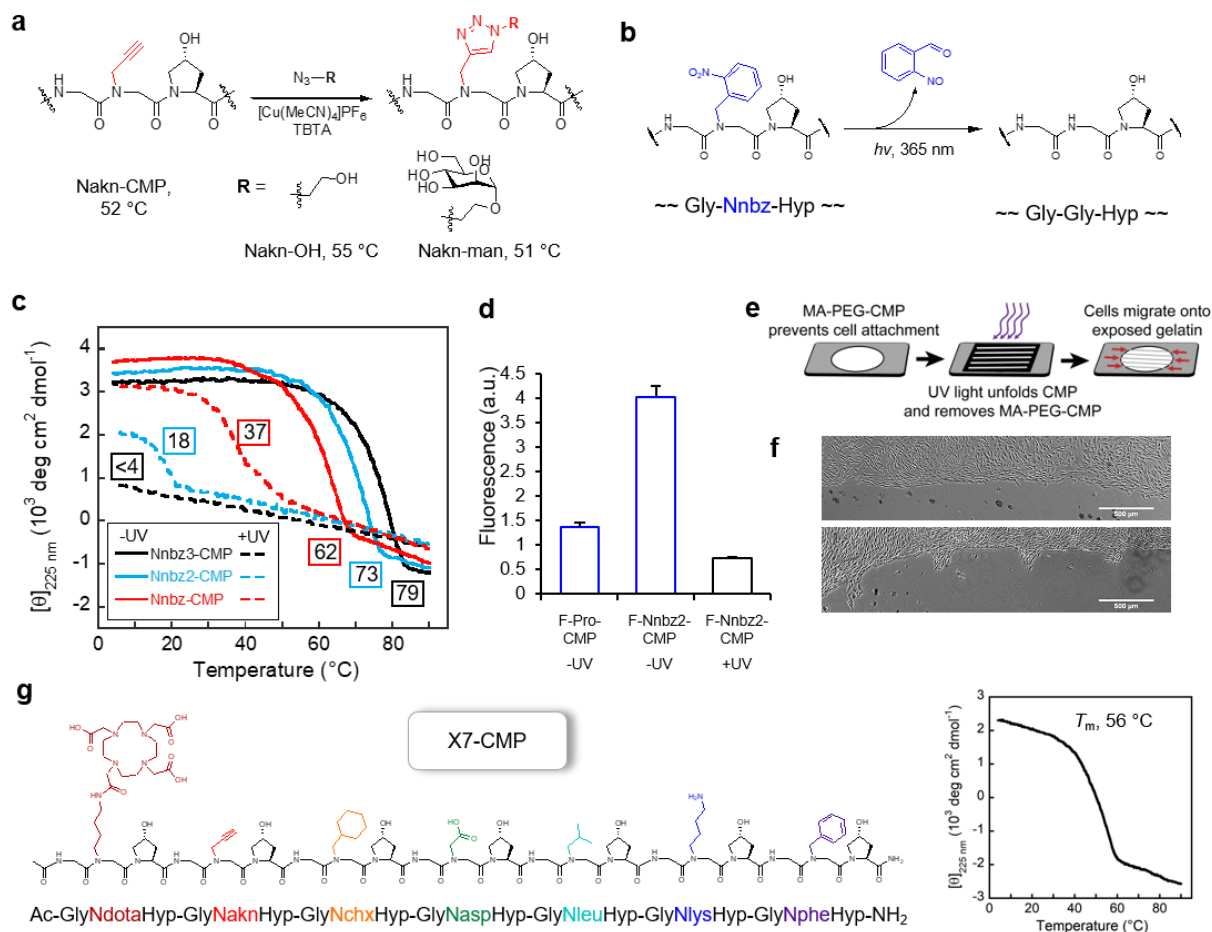


Fig. 6 | X-CMPs enable new functional design strategies. **a**, Facile modification of the alkyne-sidechain through alkyne-azide ‘click’ chemistry. Nakn-CMP retained strong triple-helical propensity after the conjugation, indicating that the N-substitution can well tolerate the spatially demanding triazole units within the triple helix (Supplementary Discussions). **b**, Reaction scheme of photo-conversion of the Nnbz residue to Gly by UV light. **c**, CD melting curves showing UV-induced triple helix destabilization of Nnbz-CMPs. A single Pro \rightarrow Nnbz substitution (Nnbz-CMP) in the central triad stabilized the triple helix (T_m : 55 $^{\circ}\text{C} \rightarrow$ 62 $^{\circ}\text{C}$), while photo-cleavage of the N-*o*-nitrobenzyl group destabilized the triple helix (T_m : 62 $^{\circ}\text{C} \rightarrow$ 37 $^{\circ}\text{C}$). Additional substitutions of Pro with Nnbz within the CMP (Nnbz2- and Nnbz3-CMPs, SI Table S7) further increased the ΔT_m of the peptide before and after UV irradiation. **d**, Comparative fluorescence intensity of gelatin films treated with fluorescently labeled X-CMPs. F-Nnbz2-CMP demonstrated higher affinity to gelatin than F-Pro-CMP, as well as triple-helical unfolding and releasing after UV irradiation. **e**, Schematics of UV-patterning on a gelatin substrate mediated by Nnbz2-CMP: MA-PEG-CMPs bound onto the crosslinked gelatin film prevented cell attachment; however, UV irradiation released the PEG-CMP conjugate, exposing gelatin that supports cell adhesion. **f**, Light micrographs of MDA-MB-231 cells grown to a confluency only in regions devoid of MA-PEG-CMP on a gelatin film (top). After exposing this

cell culture to UV light through a photomask (**e**), which unbound the MA-PEG-CMPs in selected areas, the cells began to infiltrate the patterned areas with exposed gelatin (bottom, taken 1 d after UV treatment). **g**, Pro-free X7-CMP hosting seven peptoid residues with diverse sidechain structures. Its high T_m of 56 °C demonstrates a hyper-stable triple-helical collagen peptide with the greatest sidechain diversity reported to date.

References

1. Horne, W.S. & Grossmann, T.N. Proteomimetics as protein-inspired scaffolds with defined tertiary folding patterns. *Nat. Chem.* **12**, 331-337 (2020).
2. Huang, P., Boyken, S.E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320-327 (2016).
3. Shoulders, M.D. & Raines, R.T. Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929-958 (2009).
4. Ramshaw, J.A.M., Shah, N.K. & Brodsky, B. Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J. Struct. Biol.* **122**, 86-91 (1998).
5. Ramachandran, G.N. & Kartha, G. Structure of collagen. *Nature* **176**, 593-595 (1955).
6. Bella, J., Eaton, M., Brodsky, B. & Berman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75 (1994).
7. Orgel, J.P.R.O., Irving, T.C., Miller, A. & Wess, T.J. Microfibrillar structure of type I collagen *in situ*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9001 (2006).
8. Zitnay, J.L. et al. Molecular level detection and localization of mechanical damage in collagen enabled by collagen hybridizing peptides. *Nature Communications* **8**, 14913 (2017).
9. Leitinger, B. Transmembrane collagen receptors. *Annu. Rev. Cell. Dev. Biol.* **27**, 265-290 (2011).
10. O'Leary, L.E.R., Fallas, J.A., Bakota, E.L., Kang, M.K. & Hartgerink, J.D. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nat. Chem.* **3**, 821 (2011).
11. Tanrikulu, I.C., Forticaux, A., Jin, S. & Raines, R.T. Peptide tessellation yields micrometre-scale collagen triple helices. *Nat. Chem.* **8**, 1008 (2016).
12. Holmgren, S.K., Taylor, K.M., Bretscher, L.E. & Raines, R.T. Code for collagen's stability deciphered. *Nature* **392**, 666-667 (1998).
13. Sakakibara, S., Kishida, Y., Kikuchi, Y., Sakai, R. & Kakiuchi, K. Synthesis of poly-(L-prolyl-L-prolylglycyl) of defined molecular weights. *Bull. Chem. Soc. Jpn.* **41**, 1273-1273 (1968).
14. Persikov, A.V., Ramshaw, J.A.M., Kirkpatrick, A. & Brodsky, B. Amino acid propensities for the collagen triple-helix. *Biochemistry* **39**, 14960-14967 (2000).
15. Maaßen, A. et al. Triple-helix-stabilizing effects in collagen model peptides containing PPII-helix-preorganized diproline modules. *Angew. Chem. Int. Ed.* **59**, 5747-5755 (2020).
16. Aronoff, M.R., Egli, J., Schmitt, A. & Wennemers, H. Alkylation of γ -Azaproline Creates Conformationally Adaptable Proline Derivatives for pH-Responsive Collagen Triple Helices. *Chemistry – A European Journal* **26**, 5070-5074 (2020).
17. Erdmann, R.S. & Wennemers, H. Functionalizable collagen model peptides. *J. Am. Chem. Soc.* **132**, 13957-13959 (2010).

18. Siebler, C., Erdmann, R.S. & Wennemers, H. Switchable proline derivatives: Tuning the conformational stability of the collagen triple helix by pH changes. *Angew. Chem. Int. Ed.* **53**, 10340-10344 (2014).
19. Hentzen, N.B., Smeenk, L.E.J., Witek, J., Riniker, S. & Wennemers, H. Cross-linked collagen triple helices by oxime ligation. *J. Am. Chem. Soc.* **139**, 12815-12820 (2017).
20. Goodman, M., Melacini, G. & Feng, Y. Collagen-like triple helices incorporating peptoid residues. *J. Am. Chem. Soc.* **118**, 10928-10929 (1996).
21. Feng, Y., Melacini, G. & Goodman, M. 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* **36**, 8716-8724 (1997).
22. Melacini, G., Feng, Y. & Goodman, M. Collagen-based structures containing the peptoid residue N-Isobutylglycine (Nleu): conformational analysis of gly-nleu-pro sequences by ¹H-NMR and molecular modeling. *Biochemistry* **36**, 8725-8732 (1997).
23. Simon, R.J. et al. Peptoids: a modular approach to drug discovery. *Proceedings of the National Academy of Sciences* **89**, 9367 (1992).
24. Robertson, E.J. et al. Design, synthesis, assembly, and engineering of peptoid nanosheets. *Acc. Chem. Res.* **49**, 379-389 (2016).
25. Wu, C.W., Sanborn, T.J., Huang, K., Zuckermann, R.N. & Barron, A.E. Peptoid oligomers with α -chiral, aromatic side chains: sequence requirements for the formation of stable peptoid helices. *J. Am. Chem. Soc.* **123**, 6778-6784 (2001).
26. Wu, C.W. et al. Structural and spectroscopic studies of peptoid oligomers with α -chiral aliphatic side chains. *J. Am. Chem. Soc.* **125**, 13525-13530 (2003).
27. Stringer, J.R., Crapster, J.A., Guzei, I.A. & Blackwell, H.E. Extraordinarily robust polyproline type I peptoid helices generated via the incorporation of α -chiral aromatic N-1-Naphthylethyl side chains. *J. Am. Chem. Soc.* **133**, 15559-15567 (2011).
28. Roy, O. et al. Homogeneous and robust polyproline type I helices from peptoids with nonaromatic α -chiral side chains. *J. Am. Chem. Soc.* **139**, 13533-13540 (2017).
29. Shah, N.H. et al. Oligo(N-aryl glycines): a new twist on structured peptoids. *J. Am. Chem. Soc.* **130**, 16622-16632 (2008).
30. Lee, B.-C., Zuckermann, R.N. & Dill, K.A. Folding a nonbiological polymer into a compact multihelical structure. *J. Am. Chem. Soc.* **127**, 10999-11009 (2005).
31. Murnen, H.K., Rosales, A.M., Jaworski, J.N., Segalman, R.A. & Zuckermann, R.N. Hierarchical self-assembly of a biomimetic diblock copolypeptoid into homochiral superhelices. *J. Am. Chem. Soc.* **132**, 16112-16119 (2010).
32. Mannige, R.V. et al. Peptoid nanosheets exhibit a new secondary-structure motif. *Nature* **526**, 415-420 (2015).
33. Nam, K.T. et al. Free-floating ultrathin two-dimensional crystals from sequence-specific peptoid polymers. *Nature Materials* **9**, 454-460 (2010).
34. Zuckermann, R.N., Kerr, J.M., Kent, S.B.H. & Moos, W.H. Efficient method for the preparation of peptoids [oligo(N-substituted glycines)] by submonomer solid-phase synthesis. *J. Am. Chem. Soc.* **114**, 10646-10647 (1992).
35. Zhang, Y., Malamakal, R.M. & Chenoweth, D.M. Aza-glycine induces collagen hyperstability. *J. Am. Chem. Soc.* **137**, 12422-12425 (2015).
36. Adzhubei, A.A., Sternberg, M.J.E. & Makarov, A.A. Polyproline-II helix in proteins: Structure and function. *J. Mol. Biol.* **425**, 2100-2132 (2013).
37. Brown, A.M. & Zondlo, N.J. 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* **51**, 5041-5051 (2012).

38. Pandey, A.K., Thomas, K.M., Forbes, C.R. & Zondlo, N.J. Tunable control of polyproline helix (PPII) structure via aromatic electronic effects: an electronic switch of polyproline helix. *Biochemistry* **53**, 5307-5314 (2014).
39. Horng, J.-C. & Raines, R.T. Stereoelectronic effects on polyproline conformation. *Protein Sci.* **15**, 74-83 (2006).
40. Jenkins, C.L., Vasbinder, M.M., Miller, S.J. & Raines, R.T. Peptide bond isosteres: Ester or (E)-alkene in the backbone of the collagen triple helix. *Org. Lett.* **7**, 2619-2622 (2005).
41. Butterfoss, G.L., Renfrew, P.D., Kuhlman, B., Kirshenbaum, K. & Bonneau, R. A preliminary survey of the peptoid folding landscape. *J. Am. Chem. Soc.* **131**, 16798-16807 (2009).
42. Crapster, J.A., Guzei, I.A. & Blackwell, H.E. A peptoid ribbon secondary structure. *Angewandte Chemie (International ed. in English)* **52**, 5079-5084 (2013).
43. Delauney, A.J. & Verma, D.P.S. Proline biosynthesis and osmoregulation in plants. *The Plant Journal* **4**, 215-223 (1993).
44. Egli, J., Schnitzer, T., Dietschreit, J.C.B., Ochsenfeld, C. & Wennemers, H. Why proline? Influence of ring-size on the collagen triple helix. *Org. Lett.* **22**, 348-351 (2020).
45. Cram, D.J. The design of molecular hosts, guests, and their complexes. *Science* **240**, 760 (1988).
46. Li, Y. et al. Targeting collagen strands by photo-triggered triple-helix hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14767-14772 (2012).
47. Gorske, B.C., Stringer, J.R., Bastian, B.L., Fowler, S.A. & Blackwell, H.E. New strategies for the design of folded peptoids revealed by a survey of noncovalent interactions in model systems. *J. Am. Chem. Soc.* **131**, 16555-16567 (2009).

Acknowledgements

The authors thank Rodrigo Galindo-Murillo and Xiaolei Zhu for consultation on the simulations of X-CMP structures and Hendra Wahyudi for assistance in the synthesis of Fmoc-Nnbz-OH. This research was funded by grants from the National Institutes of Health (R01AR071358, R21EY029430, and R21OD026618) awarded to SMY.

Author contributions

Y.L. and J.L.K. designed the studies. J.L.K., Y.L., G.K., and H.K. performed the syntheses. J.L.K and Y.L. performed the experiments. J.L.K. and F.G.W. solved the crystal structures. C.P.H. oversaw the crystallography. Z.Q. and T.E.C. performed the simulations. J.L.K., S.M.Y., and Y.L. wrote the paper. All authors were involved in the completion of this work.

Competing interests

The authors declare no competing interests.