I. Collagen-like polypeptides.
II. Helix-turn-helix peptides and turn mimetics.

By

Nan Dai

Dissertation submitted to the faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Chemistry

Dr. Felicia A. Etzkorn, Committee Chair
Dr. Paul R. Carlier
Dr. Richard D. Gandour
Dr. Webster L. Santos
Dr. Larry T. Taylor

July 25th, 2008
Blacksburg, Virginia

Keywords: collagen, triple helix, polypeptides, conformationally locked isostere, stability, helix-turn-helix, helicity, HTH-turn mimic

I. Collagen-like polypeptides.
II. Helix-turn-helix peptides and turn mimetics.

Nan Dai

ABSTRACT

Collagen is one of the most important and abundant proteins in mammals. It consists of three left-handed PPII helixes coiled along a common axis to form a very compact right-handed super helix. The primary structure is shown to be $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats with high content of prolyl residues at both Xaa and Yaa positions. *Cis-trans* isomerization of the prolyl amide bonds is one of the rate-limiting steps during collagen triple helix folding.

The conformationally locked alkene isosteres Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OH and Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH were designed and synthesized. The synthesis of the Gly-Pro isostere had no stereo-control, and the two diastereomers of the tripeptide isostere Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OBn were separated by normal phase HPLC. Although the stereoselectivity of the asymmetric reduction was not good for the Pro-Pro isostere, the resulting diastereomers was separable by flash chromatography, and the absolute stereochemistry of the two diastereomers was determined by Mosher's method.

The Gly-Pro alkenyl peptides, and their control peptide $Ac\text{-}(Gly\text{-}Pro\text{-}Hyp)_8\text{-}Gly\text{-}Gly\text{-}Tyr\text{-}NH_2$ were synthesized and purified. All three peptides showed a maximum around 225 nm and a minimum close to 200 nm in the CD spectra, which indicated the formation of PPII helixes. The T_m value of the control peptide was determined to be 50.0 °C. The peptide with Gly-Ψ[(*E*)CH=C]-L-Pro-Hyp as the guest triplet formed a stable triple helix with a T_m value

of 28.3 °C. The peptide with Gly-Ψ[(*E*)CH=C]-D-Pro-Hyp as the guest triplet showed a linear decrease in the ellipticity with increasing temperature, which indicated that no triple helix was formed.

The Pro-Pro alkenyl peptide and its control peptide H-(Pro-Pro-Gly)$_{10}$-OH were synthesized and purified. The T_m value of control peptide was determined to be 31.6 °C by extrapolation to 0 M TMAO in PBS buffer, which was very close to the measured value of 31.5 °C. The Pro-Pro alkenyl peptide began to show a maximum around 225 nm in the CD spectra when the concentration of TMAO was higher than 2.5 M. After extrapolation to 0 M TMAO, the T_m value was determined to be –22.0 °C. These results indicate that the backbone inter-chain hydrogen bond is one of the major forces in stabilizing the collagen triple helix, while *cis-trans* isomerization has limited contribution. The intrinsic properties of the amide bond may have huge influence on the stability of the collagen triple helix.

The helix-turn-helix motif is an important tertiary structure in DNA-binding proteins. Stepwise modifications of the Antennapedia HTH peptide (27-55) were performed to improve the helicity and stability. The peptide with more side-chain ion-pairs was over 4 times more helical than the native Antp peptide, while the Ala-based peptide was over 9 times more helical than the native peptide.

A 12-membered ring, Fmoc-protected HTH-turn mimic was designed and synthesized, and was ready for solid phase peptide synthesis. The solubility of the cyclic peptide was very poor, and the purification of the final product was very difficult. The solubility problem might also affect solid phase peptide synthesis in the future.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Dr. Felicia A. Etzkorn, for her constant encouragement, intellectual support, consideration for our academic and personal lives, and providing excellent working environment for her students. I am very lucky to have an excellent advisor like her, and this experience will be a precious treasure for my whole life. I would also like to express my sincere appreciation to my committee members, Dr. Paul R. Carlier, Dr. Richard D. Gandour, Dr. Webster L. Santos, and Dr. Larry T. Taylor for their excellent teaching, precious advice and help. I thank all of the faculty, administrators and staff for providing the excellent environment in the Chemistry Department at Virginia Tech.

I want to thank all the people that provided advice, help and support to my reasearch. I want to thank Dr. Paul R. Carlier for the advice with regards to the asymmetric synthesis, Dr. Richard D. Gandour for computational advice, Dr. David G. Kingston for allowing me to use his HPLC and solvent distillation systems, Prof. Robert H. White and Kim Harich (Biochemistry Department, Virginia Tech) for allowing me to use their CD and UV-vis spectrometer. Dr. Shugeng Cao for the advice for normal phase HPLC purification, Dr. Carla Slebodnick for the advice regarding asymmetric crystallization, Dr. Mehdi Ashraf-Khorassani for the LC-MS analysis, Dr. Bing Fang (Virginia Bioinformatics Institute, Virginia Tech) for the MALDI-TOF analysis. Mr. Mathew D. Shoulders (Department of Chemistry and Biochemistry, University of Wisconsin-Madison) for Data process and software help.

I also deeply thank my former and current lab colleagues: Dr. Galina Kapustin, Dr. Bailing Xu, Dr. Xiaodong Wang, Dr. Tao Liu, Dr. Song Zhao, Ms. Guoyan Xu, Ms. Ashley

Mullins, Ms. Ana Y. Mercedes-Camacho, Mr. Xingguo Chen, Ms. Jiajia Li, Mr. Chao Yang, Mr. Jun Qi, Dr. Changhui Liu, Dr. Yiqun Zhang, Dr. Danny Hsu and Dr. Ming Ma for helping with all kinds of experimental details and invaluable discussion about a variety of scientific topics. Financial supports from Virginia Tech, NIH and ACS are appreciated.

Most of my thanks are reserved for my parents, who have provided consistent encouragement and support all my life. As their only son, I feel sorry that I could not be with them for the past six years. Finally but most importantly, words cannot express my gratitude to my wife, Liangming Hu. I greatly appreciate the love and support she has given me. Marrying her is the most important achievement and the best decision of my life.

Table of Contents

List of Figures

List of Schemes

List of Tables

List of Abbreviations

Amino Acids

Ala, A	Alanine
Asn, N	Asparagine
Asp, D	Aspartic acid
Arg, R	Arginine
Cha	Cyclohexyl alanine
Clp	4(*R*)-Chloroproline
clp	4(*S*)-Chloroproline
Cys, C	Cysteine
Flp	4(*R*)-Fluoroproline
flp	4(*S*)-Fluoroproline
Gln, Q	Glutamine
Gly, G	Glycine
His, H	Histidine
Hyl	Hydroxylysine
Hyp	4(*R*)-Hydroxyproline
hyp	4(*S*)-Hydroxyproline
Ile, I	Isoleucine
Leu, L	Leucine
Lys, L	Lysine
Met, M	Methionine
Mep	4(*S*)-Methyl proline
mep	4(*R*)-Methyl proline
Phe, F	Phenyalanine
Pro, P	Proline
Ser, S	Serine
Thr, T	Threonine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
Val, V	Valine
Xaa	Amino acids at the X position
Yaa	Amino acids at the Y position
pSer	Phosphoserine
pThr	Phosphothreonine
AlloThr	*allo*-threonine

Abbreviations

AA	Amino acid
Ac	Acetyl
Antp	Antennapedia type homeodomain
aq	Aqueous
BINOL	1,1'-Binaphthalene-2,2'-diol

BM	Basement membrane
Bn	Benzyl
Boc	*tert*-Butyloxycarbonyl
BOP	Benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate
Bu	Butyl
Cbz	Carbobenzoxy
CD	Circular dichroism
CDI	1,1'-Carbonyldiimidazole
CE	Capillary electrophoresis
COL2A1	Collagen, type II, alpha 1chain
COSY	Correlation spectroscopy
Cy	Cyclohexyl
DCC	1,3-dicyclohexylcarbodiimide
DCE	Dynamic capillary electrophoresis
DCM	Dichloromethane
DCU	dicyclohexylurea
DIEA	Diisopropylethylamine
DMA	*N*,*N*-Dimethyl aniline
DMAP	*N*,*N*-dimethyl-4-amino pyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride
ER	Endoplasmic reticulum
Et	Ethyl
Fmoc	9-Fluorenylmethoxycarbonyl
h	Hour
HATU	2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	*O*-Benzotriazole-*N*,*N*,*N'*,*N'*-tetramethyl-uronium-hexafluoro-phosphate
HOAt	1-Hydroxy-7-Azabenzotriazole
HOBt	*N*-Hydroxybenzotriazole
HSQC	Heteronuclear single quantum coherence
HTH	Helix-turn-helix
HPLC	High performance liquid chromatography
IR	Infrared spectroscopy
$K_{cis/trans}$	Equilibrium constant between the *cis*- and *trans*-conformations
LDA	Lithium diisopropylamide
MALDI	Matrix-assisted laser desorption/ionization
Me	Methyl
min	Minute
MS	Mass spectroscopy
NB-Enantride	Lithium hydrido(9-BBN-nopol benzyl ether adduct)

NMP	*N*-Methyl-2-Pyrrolidone
NMR	Nuclear magnetic resonance spectroscopy
NOESY	Nuclear Overhauser Enhancement Spectroscopy
NP	*p*-Nitrophenyl
PBS	Phosphate buffered saline
Ph	Phenyl
PPI	Polyproline type **I**
PPII	Polyproline type **II**
Pr	Propyl
PTM	Post-translational modification
RCM	Ring-closing metathesis
rt	Room temperature
SPPS	Solid phase peptide synthesis
Su	Succinimide
TBAF	Tetrabutylammonium fluoride
TBS	*tert*-Butyldimethylsilyl
TEA	Triethylamine
TFA	Trifluoroacetic acid
THF	Tetrahedrofuran
TIPSH	Triisopropylsilane
TLC	Thin layer chromatography
T_m	Melting point (mid-point)
TMAO	Trimethylamino *N*-oxide
TMEDA	Tetramethylethylenediamine
TMS	Trimethylsilyl
TOF	Time-of-flight mass spectrometer
SAR	Structure activity relationship
UV-vis	Ultraviolet-visible spectroscopy

Chapter 1 Introduction to Collagen

Collagen is one of the most important and abundant proteins in mammals.[1] It is the most important structural protein, especially in connective tissues. It accounts for more than 25 % of all proteins in the body, about 30 % of all vertebrate body proteins, more than 90 % of the extracellular proteins in the tendon and bone, and more than 50 % of the proteins in the skin.[1,2] Collagen acts as a scaffolding material to maintain the strength and elasticity for our bodies.[3] It is responsible for the structural integrity of tissues and organs, and also regulates many biological events, such as cell attachment, migration and differentiation, tissue regeneration and animal development.[4]

1.1 Structure of collagen

1.1.1 Primary structure of collagen

Collagen usually refers to a family of proteins with at least one triple helix region.[5] The primary structure of collagen can be shown as a polymer of $(Gly\text{-}Xaa\text{-}Yaa)_n$, in which Gly-Xaa-Yaa is the repeating unit. There is a glycine (Gly) residue every three residues in the sequence, which is one of the most important structural features for the collagen triple helix. Gly is the only amino acid without a side-chain, so it can fit into the highly compacted triple helix at the specific position. Mutation of Gly to any other residues, even alanine (Ala), the second smallest residue, alters or even breaks the triple helical conformation.[6]

The amino acid residues at the Xaa and Yaa positions may vary, but high occurrences of certain amino acid residues at these positions are observed. At the Xaa position, approximately 10 % of the residues are proline (Pro). At the Yaa position, 10-12 % of the

residues are 4(*R*)-hydoxyproline (Hyp),[7,8] a non-natural amino acid that is derived from proline during post-translational hydroxylation by 4-prolyl hydroxylase.[9] The cyclic side-chains of Pro and Hyp are also believed to contribute to the folding of the collagen triple helical structure.[10] Pro and Hyp are preferred, but not required, to form a collagen triple helix.

1.1.2 Polyproline type II helix

Each chain in the collagen triple helix forms a left-handed polyproline type II (PPII).[11,12] The PPII helixes are widely found in both folded[13,14] and unfolded proteins.[15] The PPII helix is very important in biological signal transduction, transcription, cell mobility and immune responses.[16-18] It is an important component in the protein unfolded state[19] and contributes to the stability and ensemble of unfolded proteins.[17,20]

The PPII helix has high water content, and is relatively short and loose structure. About 90 % of the PPII helixes are less than 5 turns, the PPII helixes with more than 10 turns are very rare.[21] The average main-chain dihedral angles Φ, Ψ and ω are –75°, 145° and 180° respectively.[13,22,23] The helical pitch is about 9.3 Å per turn and 3.0 residues per turn, thus the rise per residue is approximately 3.1 Å (Figure 1.1).[24]

Unlike the α-helix, the PPII helix has no main-chain to main-chain hydrogen bonding because of the long distance and incorrect orientation of the corresponding amide oxygen and nitrogen atoms.[25,26] Moreover, because of the high occurrence of Pro in the PPII helix, the PPII helix chain lacks H-bond donors, and intra-chain hydrogen bonding is disfavored.

The PPII helix is more surface exposed than any other structure.[13] Stapley and Creamer showed that polar residues in the PPII helixes are 60 % more solvent exposed, and the hydrophobic residues are 50 % more solvent exposed than the average for all residues in

proteins.[21] Backbone solvation is essential for the stability of the PPII helical structure.[27-29] The water molecule is a good hydrogen bond donor as well as a good hydrogen bond acceptor and it acts as a bridge to associate amide nitrogen and oxygen atoms on the PPII backbone.[18,30] Other factor that stabilize the PPII helix are: 1) side-chain-side-chain hydrophobic packing, 2) hydrogen bonding between *i* and *i* + 3 residues,[21,31] 3) side-chain-main-chain hydrogen bonding between the *i* and *i* + 1 residues,[21] 4) inter-chain hydrogen bonding, especially in collagen.

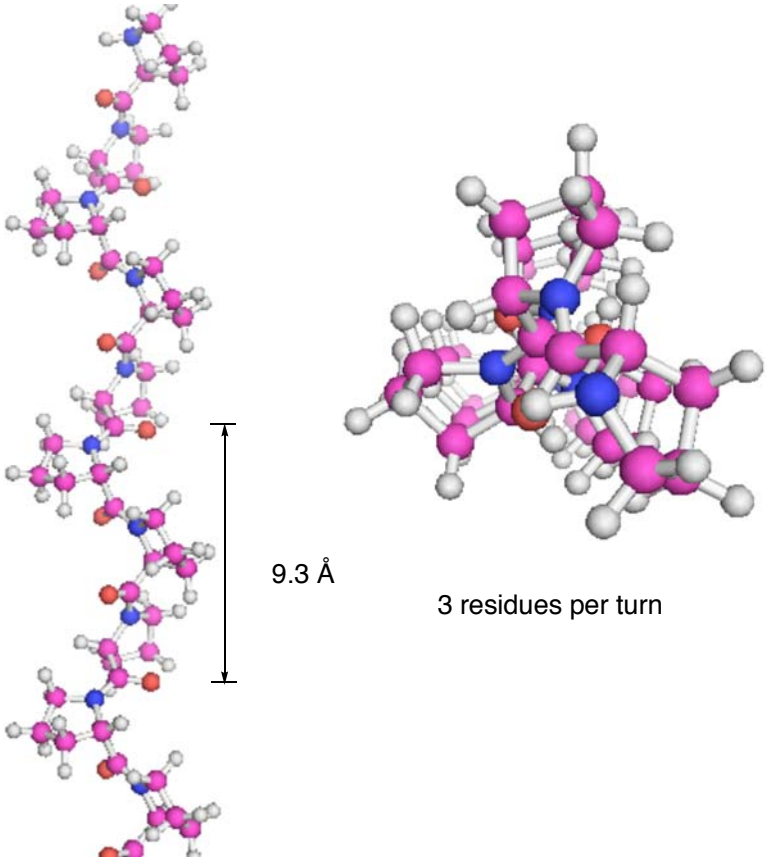


Figure 1.1 Structure of polyproline type II helix. (Generated from ppII.pdb with MacPyMOL 1.0)

All the peptide amide bonds in the PPII helix adopt the *trans*-conformation,[25,26] which is the major difference from the relatively compact polyproline type I (PPI) helix. The PPI

helix is a right-handed structure with 3.3 residues per turn. The backbone dihedral angles Φ, Ψ and ω are –75°, 160° and 0° respectively.[18,32] The amide bonds in PPI helix adopt the *cis*-conformation.[32]

The PPII helix has a high preference for Pro and Pro derivatives, although Pro is not required to form the PPII helix. An all-Pro PPII helix is found only in a few proteins.[13,21] About 25 % of the PPII helixes have no Pro in the sequence at all.[21] Positively charged residues, such as glutamine and lysine, are preferred in the PPII sequence at the $i + 1$ position because they can form the side-chain to main-chain hydrogen bond with the oxygen of the C=O at the i position to stabilize the PPII structure.[21] It is notable that Gly does not have high occurrence in the PPII helix sequence,[21] while it is essential to form the collagen triple helix structure. Not all of the PPII helixes can form a collagen triple helix.

1.1.3 Collagen triple helix

The collagen triple helix model was determined from the X-ray diffraction patterns the collagen fibers in tendons.[33-36] This model has three left-handed PPII helixes coiled with each other along a common axis, and forms a right-handed super helical structure with the Gly-Xaa-Yaa repeating unit.[35,37] All of the peptide bonds in the collagen triple helix adopt the *trans*-conformation, as shown in Figure 1.2.

The structure of the PPII helix in collagen is slightly different from the free single chain PPII helix. In the collagen triple helix, each PPII chain has 3.33 residues per turn and the axis repeat is 2.86 Å per residue,[24,33] while the number is exactly three residues per turn and the axis repeat is 3.1 Å per residue in the free single-chain PPII helix.[25] The axial repeat is 28.6 Å for the right-handed triple helix. The helical structure moves to a new start after ten residues, and the new start point is located on a different chain. Sequence dependent

structural variations were observed crystallographically. The strong intermolecular interactions may also change the structural parameters of the collagen triple helix.[38]

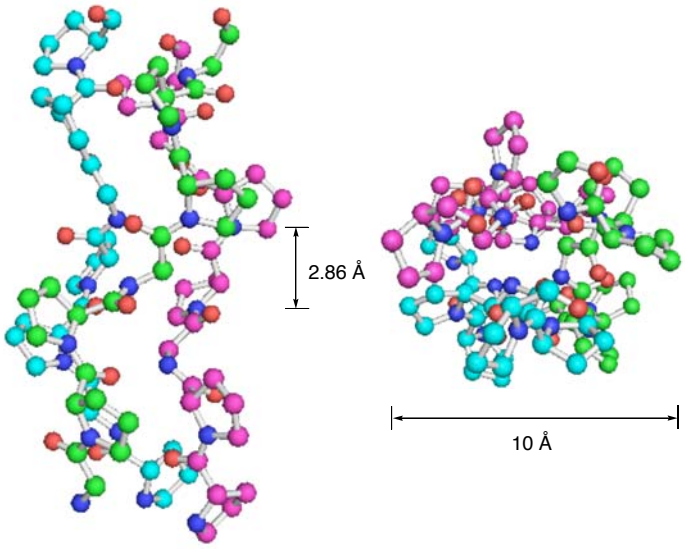


Figure 1.2 Structure of the synthetic collagen-like peptide (Pro-Pro-Gly)$_9$.[39] (Generated from PDB file 1ITT with MacPyMOL 1.0)

In the collagen triple helix, a Gly residue faces the Xaa residue of a neighboring chain, and the Xaa residue faces the Yaa residue of the third chain, as shown in Figure 1.2. This configuration is called staggered arrangement.[40] The staggered arrangement of the collagen triple helix is the major difference from other multi-stranded helixes such as the DNA coiled-coil double helical structure.[40] The staggered arrangement is especially important for most natural collagen, because the six possible alignments of the three chains provide an array of residues from different chains for recognition during the folding processes.[41]

To form the very compact collagen triple helix, the hydrogen atoms of the Gly residues point inside the structure.[40] The replacement of a hydrogen atom of Gly to a larger side-chain at the α-carbon breaks the main-chain-main-chain hydrogen bond between the N-H of the Gly residue and the C=O of the Xaa residue of another chain.[6,35] Bella *et al.*

confirmed the existence of this hydrogen-bonding pattern by studying the crystal structure of a synthetic Gly→Ala polypeptide.[6] They also pointed out that the mutation of Gly to Ala altered the triple helical structure and interrupted the inter-chain hydrogen bonding patterns.[6] Computational studies suggest that replacing Gly with D-alanine or D-serine residues might stabilize the triple helix,[42] but the synthetic peptides (Pro-Hyp-D-Ala)$_7$ and (Pro-Hyp-D-Ser)$_7$ both failed to form a stable triple helix.[43] The isolated collagen PPII helixes are not stable if the polypeptide chain contains other residues except for Pro or Hyp.[40]

Pro is preferred at the Xaa position, while Hyp has the highest preference at the Yaa position.[7,8] Other residues are also possible at these positions, although the stability of the resulting collagen triple helix may decrease.[44,45] The side-chains of the Xaa and Yaa residues are pointed out of the helix and exposed to the solvent, thus freely accessible to water molecules and for forming binding interactions, which is very important in collagen fibril formation.[40] The interactions between oppositely charged residues and hydrophobic interactions between residues of different chains also contribute to the fibril formation.[46,47]

The lateral association of triple helixes, which is important in fibril formation, is determined by both the sequence of all three chains and the symmetry of the helixes.[40] The "ten residues forming three turns" collagen triple helical structure in the fibers in tendons has a 10/3 (or 10_7) symmetry, and most natural collagen fibers adopt a 10/3 symmetry.[35,48] Crystallographic studies of synthetic model peptides with Gly-Pro-Hyp and related sequences were used to reveal detailed structural information about the collagen triple helix. In 1981, Okuyama *et al.* showed the first atomic resolution X-ray diffraction pattern of a synthetic collagen polypeptide (Pro-Pro-Gly)$_{10}$.[49] They reported the model peptide has a 7/2 (or 7_5) symmetry, with 3.5 residues per turn and an axial repeat of 2 nm.[49] This result is different

from the 10/3 symmetry obtained by X-ray diffraction of natural fibers,[50] but it is very close to the structure Cohen and Bear proposed.[51] In 1994, Bella *et al.* reported the crystal structure of a synthetic Gly→Ala polypeptide at 1.9 Å resolution.[6] They found that the polypeptide forms a triple helix structure with a length of 8.7 nm and a diameter of 1.0 nm, and the polypeptide again showed a 7/2 symmetry.[6] They believed that imino acid (Pro or Hyp) rich peptides have a 7/2 symmetry, while imino acid poor regions adopt a 10/3 symmetry.[6] Kramer *et al.* studied the X-ray crystal structure of a synthetic peptide with an imino acid poor region, and found that in this region, the structure of the collagen triple helix is close to a 10/3 helix, while in the imino acid rich region, the triple helical structure is close to a 7/2 helix.[12,52]

Additional structural information was revealed from studies of the crystal structures of $(Pro\text{-}Pro\text{-}Gly)_{10}$ and related synthetic peptides.[39,53,54] The crystal structure of $(Pro\text{-}Pro\text{-}Gly)_{10}$ obtained at 1.3 Å resolution showed that the preferential distribution of Pro backbone and side-chain conformations are position dependent.[55-57] The Pro residues at the Xaa position exhibit an average main-chain torsion angle of $\Phi = -75°$ and a positive side-chain angle (down puckering of the Pro ring).[55,56] The Pro residues at the Yaa position displayed a significantly smaller main-chain torsion angle of $\Phi = -60°$ and a negative side-chain angle (up puckering of the Pro ring).[55,56] These results also explained the stabilizing effect of Hyp at different positions, which will be discussed later.

Table 1.1 Dihedral angles for the " ideal" 10/3 and 7/2 collagen triple helixes. [49,50,58]

	Gly Φ (°)	Gly Ψ (°)	Xaa Φ (°)	Xaa Ψ (°)	Yaa Φ (°)	Yaa Ψ (°)
7/2 helix	–70.2	175.4	–75.5	152.0	–62.6	147.2
10/3 helix	–67.6	151.4	–72.1	164.3	–75.0	155.8

1.2 Collagen types and biological functions

1.2.1 Classification of natural collagens

Collagen is important in the development of large complex organisms. It provides an insoluble scaffold to maintain the shape and form of organisms; it is also responsible for the attachment of large macromolecules, glycoproteins, hydrated polymers and inorganic ions, and even cell attachments.[59] Collagens are different in morphology and function, thus play different roles in maintaining the differentiation of various biological features.[60] They are found from supramolecular assembly in rope-like fibrils, which provide the fibrous scaffold to maintain the integrity of tendons, ligaments, and bone, to net-like sheets in the basement membranes, which support the epithelial and endothelial cells. Collagens are also found as fine filaments from anchor basement membranes to specialized structure in the deep dermis, and as hexagonal lattices in cartilage and Descemet's membranes, which provide the substrates for cell differentiation, proliferation, migration and support.[59]

Based on the difference in the helix length, the nature and the length of the non-helical parts, collagens are categorized into different types,[9,61] which were assigned with Roman numerals (**I**, **II**, **III** etc.) in the order of discovery. More and more types of collagens have been isolated and identified.[60,62,63] To date, 28 different types of collagen have been isolated and found to exist independently or co-existent in various connective tissues.[64] Every type of collagen has the unique triple helical structure in common, consisting of three polypeptide chains (α-chain). A natural collagen triple helix can be a homotrimer consisting of three identical α-chains, which is shown to be $[\alpha 1(\mathbf{V})]_3$, or a heterotrimer with two identical α-chains and one dissimilar chain, which is shown to be $[\alpha 1(\mathbf{V})]_2\alpha 2(\mathbf{V})]$, or three dissimilar chains, which is shown to be $\alpha 1(\mathbf{V})\alpha 2(\mathbf{V})\alpha 3(\mathbf{V})$.

Based on the molecular and supramolecular structures, collagen can be classified into six subfamilies.[65] The classification, chain composition and tissue distribution of selected collagen types are showed in Table 1.2.

1.2.2 Biological functions of common natural collagens

The fibril-forming collagens are the most abundant subfamily of all collagens. Especially for Type **I**, **II** and **III** collagens, which have large sections of homologous sequences that are independent of the species.[66] About 80-90 % of all collagens are either Type **I**, **II** or **III**.[63] Other collagen types exist in very small quantities and are found only in specific biological structures. For example, Type **IV** collagen is a loose two-dimensional fibrillar network which is found only in the basement membrane.

Type **I** is the most abundant collagen form. It is found in almost all connective issues except for hyaline cartilage.[67] It accounts for about 95 % of the entire collagen content and about 80 % of the proteins in bone.[68] It is also the predominant protein in skin, tendons, ligaments, sclera, cornea, and blood vessels. Type **I** collagen is usually a heterotrimer with two of them are identical, termed as α1(**I**), and the third chain, termed as α2(**I**), has a different chemical composition. A homotrimer type **I** collagen consisting of three identical α1(**I**) chains is found in embryonic tissues with very small quantities.[60] Type **I** collagen is a fibril-forming collagen, and its major biological function is acting as a scaffold for the extracellular matrix of connective tissues.[59] Type **I** collagen can incorporate the minerals during the formation of bone, which is unique to other types of collagens.[59,69-72] This is important for the integrity and strength of the bones, and mutations in type **I** collagen cause brittle bones.[59] Moreover, the mutations of the genes *COL1A1* and *COL1A2*, which encode

Table 1.2 Classification, chain composition, and body distribution of collagen.[5,60,63,65,73,74]

Class	Type	Chain composition [a]	Tissue distribution
Fibril-forming collagens	**I**	$[\alpha1(\mathbf{I})]_2\alpha2(\mathbf{I})$, $[\alpha1(\mathbf{I})]_3$	Skin, tendon, bone, cornea, dentin, fibrocartilage, large vessels, intestine, uterus, dermis
	II	$[\alpha1(\mathbf{II})]_3$	Hyaline cartilage, vitreous, nucleus pulposus, notochord
	III	$[\alpha1(\mathbf{III})]_3$	Large vessels, uterine wall, dermis, intestine, heart valve, gingiva (Usually coexists with Type **I**)
	V	$\alpha1(\mathbf{V})\alpha2(\mathbf{V})\alpha3(\mathbf{V})$, $[\alpha1(\mathbf{V})]_2\alpha2(\mathbf{V})$, $[\alpha1(\mathbf{V})]_3$	Cornea, placental membranes, bone, large vessels, hyaline cartilage, gingival
	XI	$\alpha1(\mathbf{XI})\alpha2(\mathbf{XI})\alpha3(\mathbf{XI})$	Cartilage, intervertebral disc, vitreous humour
	XXIV	$[\alpha1(\mathbf{XXIV})]_3$	Bone, cornea
	XXVII	$[\alpha1(\mathbf{XXVII})]_3$	Cartilage
Network-forming collagens	**IV**	$[\alpha1(\mathbf{IV})]_2\alpha2(\mathbf{IV})$	Basement membrane (BM)
	VI	$\alpha1(\mathbf{VI})\alpha2(\mathbf{VI})\alpha3(\mathbf{VI})$	Descemet's membrane, skin, nucleus pulposus, heart muscle
	VIII	$[\alpha1(\mathbf{VIII})]_2\alpha2(\mathbf{VIII})$, $[\alpha1(\mathbf{VIII})]_3$	Produced by endothelial cells, descemet's membrane
	X	$[\alpha1(\mathbf{X})]_3$	Hypertrophic and mineralizing cartilage
Fibril-associated collagens with interrupted triple helixes	**IX**	$\alpha1(\mathbf{IX})\alpha2(\mathbf{IX})\alpha3(\mathbf{IX})$	Cartilage, cornea
	XII	$[\alpha1(\mathbf{XII})]_3$	Chicken embryo tendon, bovine periodontal ligament
	XIV	$[\alpha1(\mathbf{XIV})]_3$	Associated with type **I** in fibril surface
	XVI	$[\alpha1(\mathbf{XVI})]_3$	Hyaline cartilage, skin
	XIX	$[\alpha1(\mathbf{XIX})]_3$	Endothelial, neuronal, mesenchymal epithelial BM zone
	XX	$[\alpha1(\mathbf{XX})]_3$	Corneal epithelium
	XXI	$[\alpha1(\mathbf{XXI})]_3$	Heart, stomach, kidney, placenta
	XXII	$[\alpha1(\mathbf{XXII})]_3$	Cartilage-synovial junctions, confined zone between anagen hair follicle and dermis
	XXVI	$[\alpha1(\mathbf{XXVI})]_3$	Testis, ovary
Trans-membrane collagens	**XIII**	$[\alpha1(\mathbf{XIII})]_3$	Cetal skin, bone, intestinal mucosa
	XVII	$[\alpha1(\mathbf{XVII})]_3$	Skin, intestinal epithelia
	XXIII	$[\alpha1(\mathbf{XXIII})]_3$	Human heart and retina
	XXV	$[\alpha1(\mathbf{XXV})]_3$	Neurons
Anchoring fibrils	**VII**	$[\alpha1(\mathbf{VII})]_3$	Skin, placenta, lung, cartilage, cornea
Multiplxin	**XV**	$[\alpha1(\mathbf{XV})]_3$	BM zones
	XVIII	$[\alpha1(\mathbf{XVIII})]_3$	Perivascular BM zones

a) The composition of the polypeptide chains (α-chain) of the collagen triple helix is shown as: α: each peptide chain; Arabic numerals: number of different chains in one triple helix; Roman numerals: type of the collagen.

the propeptide of type **I** collagen proα1(**I**) and proα2(**I**) chains, respectively, have been identified as the molecular reason of different types of osteogenesis imperfecta.[59]

Type **II** collagen is a homotrimer consisting of three identical α1(**II**) chains, and the structure of the α1(**II**) subunit is very close to that of α1(**I**).[65] Type **II** collagen is essentially unique to hyaline cartilage, and is the most abundant collagen form in cartilage, inter-vertebral disc, notochord and vitreous humour and forms narrow *D*-periodic fibrils.[59] Type **II** collagen exists in two forms due to the difference in connection, and the α1(**II**) chain appears to be encoded with the same gene as α3(**XI**) chain.[75] Type **IX** collagen is covalently cross-linked to type **II** collagen in cartilage in an anti-parallel orientation.[76] Type **II** collagen is the main target of the progressive destruction in rheumatoid arthritis. Sequence polymorphisms of type **II** collagen make the cartilage matrix susceptible to autoimmune attack or subsequent degradation, and either can result in rheumatoid arthritis.[77,78] Moreover, mutation in *COL2A1*, a gene encoding type **II** collagen, can result in the Stickler syndrome, the Goldblatt syndrome, the Wagner syndrome, hypochondrogenesis, osteoarthritis and spondyloepiphyseal dysplasia.[59]

Type **III** collagen is also a homotrimer consisting of three identical α1(**III**) chains.[65] It is abundant in hollow organs, dermis, placenta and uterus, and forms fine banded, *D*-periodic fibrils.[59] Blood vessels consist predominantly of type **III** collagen.[65] Compared to type **I** collagen, type **III** collagen is very limited in quantity in nature (about 10 %).[65] Type **III** collagen usually co-exists with type **I** collagen. A partially processed p*N*collagen containing N-terminal propeptides, but no *C*-propeptides, is often observed in type **III** collagen.[59] Mutation in *COL3A1*, a gene encoding type **III** collagen, can result in the Ehlers-Danlos syndrome type IV (EDS IV) and aortic aneurysm.[59]

Other fibril-forming collagens, such as type **V** and type **XI** collagens are much smaller in quantities. Type **V** collagen has very complex chemical composition. It can be a homotrimer with three identical α1(**V**) chains, or a heterotrimer with two identical α-chains and one dissimilar chain $[\alpha1(\mathbf{V})]_2\alpha2(\mathbf{V})$, or with three dissimilar chains α1(**V**)α2(**V**)α3(**V**).[59,65] Moreover, the proα1(**XI**) chain can substitute for the proα1(**V**) chain to form a hybrid triple helix.[79] Type **V** collagen is usually observed as heterotypic fibrils with type **I** collagen in cornea and placenta, and it binds to thrombospondin more than other fibril-forming collagens.[80,81] Type **V** collagen plays an critical role in determining the diameters of heterotypic collagen fibrils in tissues.[82] Unlike type **I**, **II**, **III** collagens, type **V** collagen has a higher molecular weight *in vivo* than that of the pepsin-extracted material, which indicates that the molecule is only partially processed, and the non-helical domains are still preserved.[83] Abnormal increasing in the type **V** collagen content can cause the bone brittleness by interfering with the process of mineralization, and result in a lethal brittle bone disease.[84] Mutation of *COL5A1* or *COL5A2*, which encode α1(**V**) and α2(**V**) respectively, can cause type I or type II Ehlers-Danlos syndrome.[85-88]

Type **XI** collagen is observed as heterotypic fibrils with type **II** and **IX** collagens in several tissues. It is a heterotrimer consisting of three dissimilar chains α1(**XI**)α2(**XI**)α3(**XI**). The α3(**XI**) chain is similar to the α1(**II**) chain, and encoded with the same *COL2A1* gene.[59] However, the proα3(**XI**) chain is more glycosylated than the proα1(**II**) chain, and the N-propeptide of the proα3(**XI**) chain can only be partially cleaved.[59]

Type **IV** collagen is a highly specialized collagen form found only in basement membranes, which are thin sheet-like extracellular structures that compartmentalize tissues, provide substrata for organ cells, and important signals for the differentiation, maintenance

and shape of tissues.[89,90] Type **IV** collagen is one of the major components of all basement membranes.[89] The chain composition of type **IV** collagen is very complicated, and at least six genetically distinct chains have been discovered.[90] However, most of type **IV** collagen is heterotrimer consisting of two distinct chains $[\alpha1(\mathbf{IV})]_2\alpha2(\mathbf{IV})$, a homotrimer with three identical $\alpha(\mathbf{IV})$-chains was also confirmed.[91,92] Studies showed that the $\alpha1(\mathbf{IV})$ and $\alpha2(\mathbf{IV})$ chains are ubiquitous, while the rest four chains are limited in quantities and restricted in distribution.[93] Many diseases have been linked to the mutation of type **IV** collagen. The mutation in the $\alpha3(\mathbf{IV})$ chain may cause the Goodpasture Syndrome;[92,94] the Alport Syndrome is primarily caused by the mutation of *COL4A5*,[95-97] which encoding the $\alpha5(\mathbf{IV})$ chain; the recessive mutations of both *COL4A3* and *COL4A4* genes result in autosomal inherited forms of the Alport syndrome.[98] The Diffuse Esophageal Lewmyomatosis is caused by the deletions involving the 5' end of *COL4A5* extending into the second intron of the adjacent *COL4A6* gene.[99]

Other types of natural collagens are very limited in quantities in nature, and the biological properties of some newly discovered types are still under study. The details will not be discussed here.

1.3 The stability of collagen

1.3.1 Hydrogen bonding patterns

The backbone intra-chain hydrogen bond is not favorable in collagen because of the incorrect orientation and lack of hydrogen bond donor in the PPII helix.[25,26] The presence of inter-chain backbone hydrogen bonding patterns have been confirmed by X-ray crystallography.[6] There are two hydrogen bond types found in the collagen triple helix,

namely, direct backbone hydrogen bonds between different chains, and water mediated hydrogen bonds.[100]

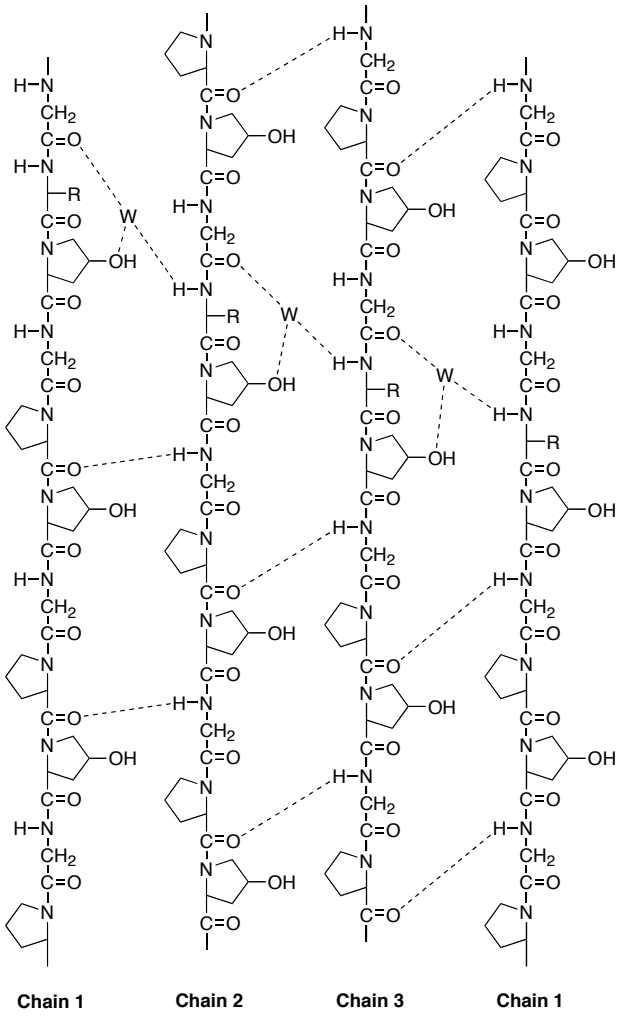


Figure 1.3 Backbone-backbone hydrogen bonding patterns in the collagen triple helix. (Reprinted from Brodsky *et al.*, *Adv. Protein Chem.* **2005**, Copyright 2005, with permission from Elsevier.)

The most common hydrogen bonding pattern observed is the interchain backbone direct hydrogen bonding. Bella *et al.* confirmed that this hydrogen bond is between the N-H of the Gly residue of one chain and the C=O of the Xaa residue (Pro) in an adjacent chain.[6] The interchain hydrogen bond connectivity is in the same direction as N-H···O=C, which can be described as 1→2→3→1 (the number designates different chains in the triple helix). The

carbonyl groups of the Gly and Yaa residues (Hyp) have no hydrogen bond partner, and are therefore not involved in a hydrogen bond.[6] Even if the Xaa or Yaa position is occupied by residues other than Pro and Hyp, and the amide N-H is available for hydrogen bonding, no carbonyl group in the triple helix is within hydrogen bonding distance, and therefore the N-H cannot form a hydrogen bond.[100] No other direct backbone interchain hydrogen bonding patterns have been reported.

The direct hydrogen bonding pattern also explains why the Gly residue is required every third amino acid residues in the collagen triple helix. It is believed that the hydrogen bond between the backbone N-H of Gly and the backbone C=O of the Xaa residue of a neighboring chain is the major source of the stability for the collagen triple helix.[40] Structurally, the two hydrogen atoms of Gly face inside the triple helix. The mutation of Gly to Ala will increase the distance to the hydrogen bond acceptor, and weaken or break the hydrogen bond.[6]

When the Xaa position is occupied by residues other than Pro, an interchain hydrogen bond mediated by one water molecule forms between the amide N-H of the Xaa residue and the C=O group of Gly in another chain. This hydrogen bonding pattern links two chains with a hydrogen bond opposite to the direct backbone-backbone N-H···O=C hydrogen bonding.[12,101-103] The water molecule involved in the hydrogen bond can also form an additional hydrogen bond with the hydroxyl group of Hyp or the side-chains of other residues to form a hydrogen bond network and thus stabilize the triple helix.[104]

The side-chains of Pro and Hyp extend out of the triple helix and cannot form direct hydrogen bond with any other group within the molecule. An extensive and ordered water mediated hydrogen bonding network has been observed.[6,49,105] Water mediated hydrogen

bonds can be formed between all available backbone carbonyl groups and the hydroxyl group of Hyp.[7,106] Okuyama studied the crystal structure of $(Pro\text{-}Pro\text{-}Gly)_{10}$ and indicated that the backbone carbonyl groups are the possible hydration sites.[49] Bella *et al.* reported five types of water mediated hydrogen bonds in their study of the crystal structure of $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Hyp-Ala-$(Pro\text{-}Hyp\text{-}Gly)_5$.[105] The nomenclature is shown in Table 1.3.

Table 1.3 Nomenclature of water mediated hydrogen bond. (Reprinted from Bella *et al.*, *Structure* **1995**, Copyright 1995, with permission from Elsevier.)[a]

Water Bridge	Connectivity	Name	Residue requirement	
			Residue 1	Residue 2
C=O···W_n···O=C	Intra-chain Inter-chain	α_n β_n	None	None
O-H···W_n···O=C	Intra-chain Inter-chain	γ_n δ_n	Hyp	None
N-H···W_n···O=C	Intra-chain Inter-chain	ε_n ζ_n	No Pro or Hyp	None
N-H···W_n···O-H	Inter-chain	η_n	No Pro or Hyp	Hyp
Any···W_n···Any	Between triple helixes	ω_n	None	None

a) In which W denotes a water molecule, n is the number of water molecules that participate in the water bridge.

Water molecules can form water bridges to link two carbonyl groups or one carbonyl group and one hydroxyl group of a Hyp residue. It is possible that multiple water molecules participate in the water bridge.[105] Brodsky reported the presence of water bridges with up to 5 water molecules.[100] It is also possible for one hydroxyl group of Hyp to connect with different carbonyl groups and forms intra-chain and inter-chain water mediated hydrogen bonds at the same time.[105]

Other than Hyp, the side-chains of some other residues can also form hydrogen bonds to stabilize the collagen triple helical structure. Kramer *et al.* reported that in the crystal structure of the host-guest peptide with a Gly-Glu-Lys sequence in the middle, no direct ion-pair interaction formed between the side-chains of the positively charged Lys and negatively

charged Glu. Instead, the Lys side-chain formed a hydrogen bond with the carbonyl group of the residue at the Yaa position, and the Glu side-chain formed a hydrogen bond with the Hyp hydroxyl group in the neighboring chain.[102] A similar hydrogen bonding pattern was also reported in peptide T3-785 with a ventral region of Gly-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala, where the Arg side-chain formed a hydrogen bond with a backbone carbonyl group of the neighboring chain.[52,107] Mechling and Bachinger reported that the polypeptide (Gly-Glu-Arg)$_{15}$-Gly-Pro-Cys-Cys-Gly formed a pH-dependent collagen-like triple helix at acidic or basic conditions, but it was insoluble at neutral conditions.[108] The Thr and Ser side-chains, like Hyp, can also form various water-mediated hydrogen bonds,[109] but the stability (T_m = 18 °C for (Gly-Hyp-Thr)$_{10}$)[110] is much lower than that of Hyp occupying the corresponding position in the triple helixes (T_m = 64.6 °C for (Hyp-Hyp-Gly)$_{10}$).[111]

There are also possible hydrogen bonds between the hydrogen of the α-carbons and the oxygen of the carbonyl groups, which were first described in the PPII helix.[48,112] Bella and Berman studied the crystal structure of the singly Gly→Ala peptide and found two possible C^{α}-H···O=C hydrogen bonding pairs: (1) the C^{α}-H of Gly from one chain and the C=O of Gly and C=O of Pro from the other two chains; and (2) the C^{α}-H of Hyp from one chain and the C=O of Pro from the neighboring chain.[113] This C^{α}-H···O=C hydrogen bond is very weak, and makes limited contributions to the overall stability of the collagen triple helix.[104]

1.3.2 Amino acid sequence

Chain length

Natural collagen peptides have different chain lengths, amino acid sequences, and conformations. Therefore, it is very difficult to compare the stability of different types of

natural collagens. Synthetic collagen peptides have controlled sequences and chain lengths. They are ideal targets to reveal detailed information about the relationship between amino acid sequences and the stability of the collagen triple helix.

For natural collagens, the number of the repeating unit of Gly-Xaa-Yaa is usually very large, and the chain length does not cause a problem for the stability of collagen globular domain. For synthetic collagen peptides, the chain is usually short, and the chain length will affect for the stability of the triple helical structure.

Synthetic collagen has a 7/2 symmetry, with 3.5 residues per turn.[6,49] A peptide with only one or two repeating units cannot form a triple helical structure no matter what the sequence is. Peptides with the sequences $(Pro\text{-}Pro\text{-}Gly)_n$ and $(Pro\text{-}Hyp\text{-}Gly)_n$ have been used to study the effect of chain length on the stability of the collagen triple helix. Melting point (or mid-point) T_m, defined as the temperature at which half of the collagen triple helix disassociates to random coils, has been used to quantitatively describe the stability of the triple helical structure.[40] A higher T_m value suggests greater stability of the specific triple helix.

Table 1.4 Melting points of $(Pro\text{-}Pro\text{-}Gly)_n$ and $(Pro\text{-}Hyp\text{-}Gly)_n$.

$(Pro\text{-}Pro\text{-}Gly)_n$		$(Pro\text{-}Hyp\text{-}Gly)_n$	
n	T_m (°C)	n	T_m (°C)
5	n. d. [a]	3	n. d. [d]
7	6-7 [a]	5	5,[a] 18 [d]
8	18 [a]	6	36-37 [d]
10	26,[a] 41,[b] 32.6 [c]	7	36 [b]
15	52 [a]	8	47 [e]
20	72 [a]	9	67 [d]
		10	69,[b] 60 [c]

a) In 10 % AcOH solution;[114-116] b) in 50 mM AcOH solution;[117,118] c) in 10 mM PBS buffer (pH 7.0);[45] d) measured as Ac-$(Gly\text{-}Pro\text{-}Hyp)_n$-$NH_2$ in H_2O;[119] e) measured as Ac-$(Gly\text{-}Pro\text{-}Hyp)_n$-Gly-Gly-$NH_2$ in 10 mM PBS buffer (pH 7.0).[120]

The stability of the collagen triple helix increases with the increase of the chain length. The melting point will not increase with the lengthening of the peptide chain indefinitely. Shaw and Schurr studied the thermodynamics of the collagen polypeptides and predicted that the melting point of $(Pro\text{-}Pro\text{-}Gly)_n$ with an infinite chain length was about 400 °K.[115] It is important to choose an appropriate chain length to ensure that the T_m value is in a measurable range. The triple helical structure is more stable with Gly-Pro-Hyp repeats than with Gly-Pro-Pro repeats (at the same length) because of the intrinsic stabilizing ability of Hyp.[121,122] The effect of Pro and Hyp puckering will be discussed later.

Sequence

The use of a "host-guest" peptide system makes studying the effect of a single amino acid on the overall structural stability possible. O'Neil and DeGrado,[123] and Smith *et al.*[124] used this system to study the stability of the α-helix and β-sheet respectively. Brodsky and co-workers designed a host-guest peptide, $Ac\text{-}(Gly\text{-}Pro\text{-}Hyp)_3\text{-}Gly\text{-}Xaa\text{-}Yaa\text{-}(Gly\text{-}Pro\text{-}Hyp)_4\text{-}Gly\text{-}Gly\text{-}NH_2$, and used this system in collagen stability studies.[44,45,125-130] The melting point results are shown in Table 1.5.

Persikov *et al.* analyzed the correlation between the amino acid sequence and the T_m value, and found that imino acids such as Pro, Hyp and Flp have a stabilizing effect at both Xaa and Yaa positions, while Gly and aromatic residues show a destabilizing effect at both positions.[120] Many residues show different preference at the Xaa or Yaa position because of the difference of inter-chain interactions and solvent exposure abilities. Charged residues at the Xaa position have very little effect on the stability of the triple helix compared to Pro, and the destabilizing effect is much greater if the charged residues are placed at the Yaa position. This can be explained by the solvent accessibility difference between the Xaa and Yaa

positions. The side-chain of the Yaa residue is less solvent accessible and closer to the side-chains at the neighboring chains.[131] A high propensity of charged residues at the Xaa positions can improve direct or water mediated inter-chain hydrogen bonding. The low propensity of the amino residues with branched side-chains at the Yaa position is due to the decrease in solvent accessibility of the backbone carbonyl groups.[120] It has also been reported that the charged residues can stabilize the PPII helix in the unfolded state,[21] and the ability of unfolded polypeptide chains to adopt a PPII helical structure also contributes to the triple helix stability.

Table 1.5 The melting point of the host-guest peptides. (Reprinted with permission from Persikov, A. V. *et al.*; *Biochemistry* **2000**, Copyright 2000, American Chemical Society.)[44,45,125-129]

Gly-Xaa-Hyp		Gly-Pro-Yaa	
Xaa residue	T_m (°C)	Yaa residue	T_m (°C)
Pro	47.3	Hyp	47.3
Hyp	47.3	Arg	47.2
Glu	42.9	Pro	45.5
Ala	41.7	Flp	43.7
Lys	41.5	Met	42.6
Arg	40.6	Ile	41.5
Gln	40.4	Gln	41.3
Asp	40.1	Ala	40.9
Leu	39.0	Val	40.0
Val	38.9	Glu	39.7
Met	38.6	Thr	39.7
Ile	38.4	Cys	37.7
Asn	38.3	Lys	36.8
Ser	38.0	His	35.7
His	36.5	Ser	35.0
Thr	36.2	Asp	34.0
Cys	36.1	Gly	32.7
Tyr	34.3	Leu	31.7
Phe	33.5	Asn	30.3
Gly	33.2	Tyr	30.2
Trp	31.9	Phe	38.3
		Trp	26.1

T_m values were measured at concentration of 1 mg/mL in 10 mM pH PBS buffer (0.15 M NaCl, 10 mM phosphate buffer, pH 7.0)

It is notable that the stability of a host-guest system cannot be translated into the stability of the peptide when the whole sequence is homogenous. For example, the host-guest peptide with Gly-Pro-Flp in th middle of the sequence (T_m = 43.7 °C) has a similar stability to the peptide with Gly-Hyp-Pro (T_m = 43.0 °C).[45] The peptide (Pro-Flp-Gly)$_{10}$ forms the most stable triple helical structure with a T_m of 87.0 °C,[118] even more stable than (Pro-Hyp-Gly)$_{10}$. While (Hyp-Pro-Gly)$_{10}$ does not form a stable triple helical structure.[132]

1.3.3 Stabilizing effect of Pro and Hyp

The puckering of the prolyl 5-membered ring

It is very interesting that (Pro-Hyp-Gly)$_{10}$ forms a very stable triple helix with a T_m value of 60 °C,[121,133] while (Hyp-Pro-Gly)$_{10}$ cannot form a triple helical structure at all.[132] Many studies were performed to reveal the role of Pro and Hyp in the overall stability of the collagen triple helix. It is known that Hyp at the Yaa position has a special stabilizing effect on the overall triple helix structure,[121,134] and this effect is not because of the direct hydrogen bonding between the hydroxyl group of Hyp and backbone carbonyl group within the same molecule. Hydrogen bonding through water bridges involving the hydroxyl group of Hyp was proposed as a stabilizing factor.[50] This hypothesis was confirmed by studies of the X-ray crystal structure, which showed the presence of different types of water mediated hydrogen bonds.[6,105] However, Inouye *et al.* found that both (Pro-4(*S*)-Hyp-Gly)$_{10}$ and (Hyp-Pro-Gly)$_{10}$, cannot form a stable triple helix, although the hydroxyl group of Hyp is available for hydrogen bonding.[132,133] These results weakened the hypothesis that water-mediated hydrogen bonds involving Hyp provide the major stabilizing force.

Raines and co-workers used 4(*R*)-fluoroproline (Flp) to replace 4(*R*)-Hyp in the sequence, and obtained a more stable triple helical structure.[135] The T_m value is 87.0 °C,

which is higher than for any other sequence with 10 repeats, including $(Pro\text{-}Hyp\text{-}Gly)_{10}$ (T_m = 60 °C).[135] Since the fluorine atom has a very low tendency to form a hydrogen bond,[136] they believed that the inductive effect generated by the electronegative oxygen of the hydroxyl group is more important in stabilizing the collagen triple helix than water bridges. The inductive effect of a 4(*R*)-Hyp hydroxyl group helps Hyp maintain the required *trans*-conformation of the amide bond preceding Hyp.[118,135] Later, they also found that 4(*S*)-Hyp favors the *cis*-conformation and destabilizes the triple helix when placed at either the Xaa or the Yaa positions.[117] Vitagliano *et al.* studied the crystal structure of $(Pro\text{-}Pro\text{-}Gly)_{10}$ and found that the Pro residues at the Xaa positions have a positive side-chain angle (down puckering),[55,56] while the Pro residues at the Yaa positions have a negative side-chain angle (up puckering).[55,56] These results were used to explain the stabilizing and destabilizing effects of the Hyp and Flp in triple helixes.

Yaa position

To achieve a stable triple helical structure, the imino acid residues should be allocated in the triple helix without significant strain.[137] 4(*R*)-Hyp adopts exclusively C^{γ}-*exo* conformation (up puckering) due to the inductive effect.[135] The imino acid ring at the Yaa position also prefers the C^{γ}-*exo* pucker, so 4(*R*)-Hyp fits well at the Yaa position. Moreover, the dihedral angles adopted by 4(*R*)-Hyp are similar to those adopted by Pro at the Yaa position, but significantly different from those at the Xaa position.[56,57] 4(*S*)-Hyp prefers the C^{γ}-*endo* pucker (down puckering); it does not fit well at the Yaa position, and therefore destabilizes the triple helical structure when located at this position.[117]

The fluorine atom is more electronegative than the hydroxyl group, so the inductive effect in 4(*R*)-Flp is greater than that in 4(*R*)-Hyp, which was demonstrated by the lower

NH_2^+ pK_a of the conjugated acid of 4(*R*)-Flp.[138] Like 4(*R*)-Hyp, 4(*R*)-Flp adopts exclusively the C$^\gamma$-*exo* pucker. The torsion angle Ψ in crystalline Ac-Flp-OMe is about 141°,[139] which is very close to the torsion angle of the Yaa residues (150°) in collagen crystals.[6] The 4(*R*)-fluorine substitution at the Yaa position also helps the amide bond preceding 4(*R*)-Flp to adopt the *trans*-conformation,[138] and increases the overall stability of the (Pro-Flp-Gly)$_{10}$ triple helical structure.[118,135] Conversely, 4(*S*)-Flp prefers the C$^\gamma$-*endo* pucker increasing the ratio of the *cis*-conformation,[140] and decreasing the stability of the collagen triple helix when placed at the Yaa position.[117,133]

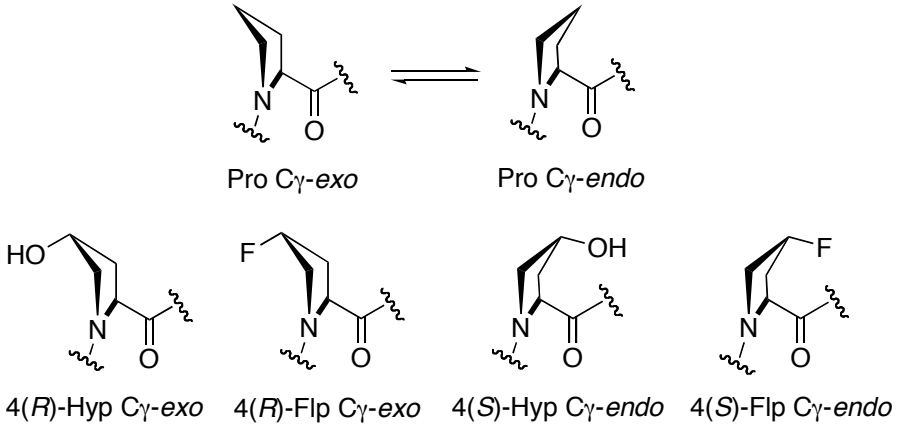


Figure 1.4 The puckering of the prolyl 5-membered ring.

Table 1.6 Melting points of the peptides with 4-substituted prolyl amino acid residues.

(Xaa-Yaa-Gly)$_7$		(Xaa-Yaa-Gly)$_{10}$	
Sequence	T_m (°C)	Sequence	T_m (°C)
Pro-4(*R*)-Flp-Gly	45 [117]	Pro-4(*R*)-Flp-Gly	77 [141]
Pro-4(*R*)-Hyp-Gly	36 [117]	4(*R*)-Hyp-4(*R*)-Hyp-Gly	64.6 [111]
4(*R*)-mep-4(*S*)-Mep-Gly	36 [142]	Pro-4(*R*)-Hyp-Gly	62.2 [141]
4(*S*)-Flp-Pro-Gly	33 [143]	4(*S*)-Flp-Pro-Gly	54.5 [141]
Pro-4(*S*)-Mep-Gly	29 [142]	Pro-4(*R*)-Clp-Gly	52[144]
Pro-4(*R*)-Clp-Gly	23[144]	4(*S*)-Clp-Pro-Gly	33[144]
4(*R*)-mep-Pro-Gly	13 [142]	Pro-Pro-Gly	31.4 [141]
Pro-Pro-Gly	6-7 [115]	4(*S*)-Flp-4(*R*)-Flp-Gly	29.9 [111]
		4(*S*)-Hyp-4(*R*)-Hyp-Gly	9.0 [111]

In summary, the stabilizing effect of 4(*R*)-Hyp at the Yaa position can be attributed to the inductive effect of the hydroxyl group of 4(*R*)-Hyp, which reduces the conformational strain of 4(*R*)-Hyp and helps the Hyp residue to adopt the intrinsically correct main-chain dihedral angles.[56]

Xaa position

The same method was used to explain the destabilizing effect of 4(*R*)-Hyp and 4(*R*)-Flp at the Xaa position. The 5-membered ring of Pro prefers the C^{γ}-*endo* pucker at the Xaa position.[55,56] 4(*R*)-Hyp and 4(*R*)-Flp both adopt exclusively the C^{γ}-*exo* pucker, which is not preferred at the Xaa position, while 4(*S*)-Hyp and 4(*S*)-Flp both adopt exclusively the C^{γ}-*endo* pucker which is preferable at this position. Renner *et al.* incorporated 4(*R*)-Flp and 4(*S*)-Flp residues in place of a Pro residue with a *cis*-amide bond in barstar, a single domain extracellular RNAse inhibitor protein. They found that the mutant with 4(*S*)-Flp has a higher T_m, and the mutant with 4(*R*)-Flp has a lower T_m than the original protein.[145] Raines and co-workers synthesized the peptides with 4(*R*)-Flp and 4(*S*)-Flp at the Xaa positions in the sequence of $(\text{Pro-Pro-Gly})_7$, the order of the stability of the three peptides is $(4(S)\text{-Flp-Pro-Gly})_7 > (\text{Pro-Pro-Gly})_7 >> (4(R)\text{-Flp-Pro-Gly})_7$.[143] These results agree with the hypothesis that the C^{γ}-*endo* pucker of 4(*S*)-Flp is favored at the Xaa position and stabilizes the collagen triple helix.

Considering the dihedral angles Φ, Ψ and ω of 4(*S*)-Flp, unlike the stabilizing effect of 4(*R*)-Flp at the Yaa position in which all three dihedral angles Φ, Ψ and ω are satisfied, the C^{γ}-*endo* pucker of 4(*S*)-Flp satisfies the torsion angles of Φ and Ψ. But 4(*S*)-Flp prefers the *cis*-conformation (ω = 0°), which is opposite to the main-chain dihedral angle of the collagen triple helix with the *trans*-conformation (ω = 180°).[139,143] This explains why (Pro-4(*R*)-Flp-

$Gly)_7$ is more stable than $(4(S)$-Flp-Pro-$Gly)_7$. For peptides with 4(*S*)-Flp at the Xaa position, pre-organization of Φ and Ψ at the Xaa position is more important than the pre-organization of ω.[143]

Barth *et al.* studied the stability of peptides with the sequence Ac-$(Xaa\text{-}Hyp\text{-}Gly)_5$-Pro-Cys($S^tBu$)-$(Gly)_3$-$NH_2$. They found that when Xaa is Pro, the peptide forms a triple helical structure; when Xaa is 4(*R*)-Flp, the peptide fails to form a triple helix, but each chain forms a PPII helix; when Xaa is 4(*S*)-Flp, no triple helix or PPII helix was formed and each chain is a random coil.[146] They believed that 4(*S*)-Flp prevented the formation of the triple helix, while 4(*R*)-Flp helped the stability of the PPII helix. The formation of a single-chain PPII helix does not correlate with the formation of the triple helix, due to numerous factors affecting triple helix folding, such as van der Waals contacts, electrostatic interactions, and stereo-electronic effects.[146]

The results are very complicated when Hyp is placed at the Xaa position. $(4(R)$-Hyp-$Pro\text{-}Gly)_{10}$ does not form a triple helix as expected,[132,133] while $(4(S)$-Hyp-Pro-$Gly)_{10}$ also fails to form a stable triple helix,[133] which does not agree with the above mentioned hypothesis. It is very interesting that Ac-(Gly-4(*R*)-Hyp-$4(R)\text{-}Hyp)_{10}$-NH_2, which has 4(*R*)-Hyp at both the Xaa and Yaa positions, forms a more stable triple helix than Ac-(Gly-Pro-$4(R)\text{-}Hyp)_{10}$-NH_2.[147] Schumacher *et al.* studied the X-ray crystal structure of (Gly-4(*R*)-Hyp-$4(R)\text{-}Hyp)_9$ and found that 4(*R*)-Hyp residues at both Xaa and Yaa positions show up-puckering (C^γ-*exo* pucker).[58] Although the Xaa position does not favor a C^γ-*exo* pucker, the triple helix still adopts a 7/2 symmetry.[58]

In peptides with the sequence Ac-$(Gly\text{-}4(R)\text{-}Hyp\text{-}Yaa)_{10}$-$NH_2$, when the Yaa position is occupied by valine (Val) and threonine (Thr), the peptides are able to form a stable triple

helix in water. When the Yaa position is occupied by *allo*-threonine (alloThr), serine (Ser) and alanine (Ala), the peptides form stable triple helixes in propanediol.[148] The peptide with Ser at the Xaa position forms the least stable triple helical structure, which indicates that, except for the property and position of the methyl group and the hydroxyl group, the stereochemical configuration is also important for the triple helix stability.[148]

Kobayashi and co-workers studied the crystal structure and thermodynamics of the triple helix formed by polypeptides containing Hyp and Flp, and they believed that Hyp and Flp have different effects on the stability of the collagen triple helix.[141,149,150] Compared to $(Pro\text{-}Pro\text{-}Gly)_{10}$, $(Pro\text{-}4(R)\text{-}Hyp\text{-}Gly)_{10}$ has a greater degree of hydration, which leads to a larger enthalpy of folding, but $(Pro\text{-}4(R)\text{-}Flp\text{-}Gly)_{10}$ has a lower degree of hydration, which results in a smaller enthalpy of folding.[141] They concluded that two parameters are responsible for the enhancement of the triple helix stability, 4(*R*)-Flp and 4(*S*)-Flp increase triple helix stability entropically, while 4(*R*)-Hyp increases triple helix stability enthalpically.[141] The X-ray crystal structures confirmed that $(4(R)\text{-}Hyp\text{-}4(R)\text{-}Hyp\text{-}Gly)_{10}$ has a similar degree of hydration to $(Pro\text{-}4(R)\text{-}Hyp\text{-}Gly)_{10}$ in the collagen triple helix. $(4(R)\text{-}Hyp\text{-}4(R)\text{-}Hyp\text{-}Gly)_{10}$ is more highly hydrated than $(Pro\text{-}4(R)\text{-}Hyp\text{-}Gly)_{10}$ in the single-coil state, which means a lower enthalpy barrier during the formation of the triple helix than $(Pro\text{-}4(R)\text{-}Hyp\text{-}Gly)_{10}$.[149] This explains why $(4(R)\text{-}Hyp\text{-}4(R)\text{-}Hyp\text{-}Gly)_{10}$ forms a more stable triple helix.

Most recently, Shoulders *et al.* replaced Hyp with 4(*S*)-methylproline (Mep) and 4(*R*)-methylproline (mep) to study the stability of the triple helix.[142] They found that $(mep\text{-}Mep\text{-}Gly)_7$, $(mep\text{-}Pro\text{-}Gly)_7$ and $(Pro\text{-}Mep\text{-}Gly)_7$ all form stable triple helixes. $(Pro\text{-}Mep\text{-}Gly)_7$ has a higher T_m value than $(mep\text{-}Pro\text{-}Gly)_7$ because of the more favorable C^{γ}-*exo*

pucker.[142] Moreover, (mep-Mep-Gly)$_7$ has a similar T_m value to (Pro-Hyp-Gly)$_7$.[142] This indicated that side-chain heteroatoms and side-chain solvation are not necessary for the triple helix formation. The steric effect is more important than the inductive stereo-electronic effect in triple helix folding.[142]

Some other Hyp and Flp derivatives were placed at the Xaa or Yaa positions to study their roles in stabilizing the triple helix. Neither Ac-(Gly-3(*S*)-Hyp-4(*R*)-Hyp)$_{10}$-NH_2 nor Ac-(Gly-Pro-3(*S*)-Hyp)$_{10}$-NH_2 forms a stable triple helix in water.[151] (3(*S*)-Flp-4(*R*)-Flp-Gly)$_7$ forms a triple helical structure, but the stability is lower than when Pro is at the Xaa position because of an unfavorable inductive effect that affects the direct inter-chain hydrogen bonding.[152] The crystal structure of (Gly-Pro-4(*R*)-Hyp)$_3$(Gly-3(*S*)-Hyp-4(*R*)-Hyp)$_2$(Gly-Pro-4(*R*)-Hyp)$_4$ also confirmed the C^{γ}-*endo* pucker of 3(*S*)-Hyp in the collagen triple helix.[153]

Conclusion of the stabilizing effects of Hyp and Flp

The stability of a collagen triple helix is determined by two parameters, the enthalpy and entropy during the peptide folding. A high degree of hydration and water mediated hydrogen bonding can increase the enthalpy of the single-strand state, resulting in a decrease in the enthalpy barrier during the folding process. Favorable inductive effects can pre-organize the single peptide chain, which decreases the entropy during the peptide folding.

The enthalpy and entropy are different in importance for Hyp and Flp. The fluorine atom is unlikely to form a hydrogen bond, while it has greater inductive effect. So Flp tends to pre-organize the single PPII helix chain. The entropic effect is therefore more important for Flp. The hydroxyl group in Hyp can form various water-mediated hydrogen bonds, and Hyp tends to form water-mediated hydrogen bonds to lower the enthalpy barrier during the process of triple helix folding. Therefore, enthalpy is more important for Hyp.

The stabilizing effect is different at the Xaa and Yaa positions. At the Yaa position, the collagen triple helix favors the C^{γ}-*exo* pucker. The preferred 4(*R*)-Hyp and 4(*R*)-Flp induce the *trans*-conformation of the amide bond, and all three dihedral angles: Φ, Ψ and ω are satisfied at the same time. At the Xaa position, the prolyl 5-membered ring prefers the C^{γ}-*endo* pucker. The preferred 4(*S*)-Hyp and 4(*S*)-Flp favor the *cis*-conformation which conflicts with the all *trans*-conformation of the collagen triple helix. Stabilization or destabilization by the inductive effect should be considered together with other factors, such as steric hindrance and solvation effects.

1.3.4 *Cis-trans* isomerization

The peptide amide has partial double bond character and is planar.[154] The amide bond can adopt either the *cis* (ω = 0°) or the *trans* (ω = 180°) conformation. Other than Pro and prolyl derivatives, the *cis*-content of a specific amide bond in unfolded proteins and peptides is very low. Secondary amides, *N*-methylformamide and *N*-methylacetamide, were used as models to study the equilibrium constants of *cis/trans* isomerization ($K_{cis/trans}$) in various solvents. The percentage of structure exhibiting the *cis*-conformation is in the range of 1.4 % to 10.5 % with the lower numbers being observed in aqueous solutions.[155-158] Scherer *et al.* studied the peptides with Gly, Ala, Phe, and Tyr in aqueous solution and found that only 0.11 % to 0.48 % of the amide bonds are in the *cis*-conformation.[159]

For the amide bond preceding the Pro residue, the *cis*-fraction is much higher. It was reported that the *cis*-content can be as high as 10 % to 30 % in short unfolded Pro containing peptides, which means that the equilibrium constant between the *cis* and *trans*-conformation ($K_{cis/trans}$) is in the range of 0.11 to 0.43.[160,161] Schoetz *et al.* reported that the *cis*-content of the amide bond can be as high as 46 % in the dipeptide Phe-Pro as determined by a dynamic

capillary electrophoresis (DCE) study.[162] The equilibrium constant was also measured for model compounds containing 4-substituted prolyl derivatives in D_2O by NMR, and the values of $K_{cis/trans}$ are in the reported range.[117,145] The $K_{cis/trans}$ value is 0.19 when a residue is attached to Pro, corresponding to 16 % cis, and is 0.087 when a residue is attached to Hyp in unfolded type **I** collagen.[163]

The stability of the collagen triple helix is determined by the free energy difference (ΔG^0) between the unfolded state and the triple helix, which means that changes in the unfolded state contribute to the stability of the triple helix. The collagen triple helix requires all peptide bonds to be in the *trans*-conformation, so a change in the *cis* to *trans* ratio in the unfolded state will directly affect the triple helix stability. For example, a cuticle collagen from a deep sea hydrothermal vent worm *Riftia pachyptila* has a very low content of Pro and Hyp residues, but a high T_m of 37 °C.[164,165] The Yaa position in the sequence of this cuticle collagen is frequently occupied by threonine (Thr). The Thr residue at the Yaa position was frequently galactosylated to galactosylated threonine Thr(β-Gal). The peptide Ac-(Gly-Pro-Thr)$_{10}$-NH_2 was unable to form a triple helix, while the galactosylated version Ac-(Gly-Pro-Thr(β-Gal))$_{10}$-NH_2 has a T_m value of 39 °C.[166] The stabilizing effect of galactosyl threonine can be explained by the occlusion of water molecules and an increase of hydrogen bonds.[167] In addition, galactosylation of the Thr residues also restricts the available conformations in the unfolded state and stabilizes the triple helical structure.[168]

Cis-trans isomerization of the amide bond preceding prolyl residues is regarded as one of the rate-determining steps for the refolding of many denatured proteins.[169-173] Bachinger and Bruckner also indicated that *cis-trans* isomerization is important in the formation of the collagen triple helix.[174-177] Collagen has high content of Pro and Hyp, and

each chain has a significant number of peptide bonds in the *cis*-conformation in the unfolded state. About 16 % of the Pro and 8 % of the Hyp residues are in the *cis*-conformation in thermally unfolded type **I** collagen.[163] In the collagen triple helical structure, all the amide bonds in the collagen triple helix adopt the *trans*-conformation,[178] which means that, in the transition from the unfolded state to the triple helix, all of the *cis*-amide bonds must convert to the *trans*, because the *cis*-amide bond cannot be incorporated into the compact triple helical structure.[40] The ratio of *cis* to *trans* conformations in the unfolded state directly affects the stability of the triple helix.

The energy difference between the *cis*- and *trans*-isomers of Pro and 4-substituted prolyl derivatives in the unfolded state is relatively small,[145] because of the similar internal van der Waals interactions in the *cis* and *trans* conformations and the extra alkyl substituent of the tertiary amide bond.[40] However, the activation energy of *cis*-*trans* isomerization is usually fairly high with an average value of 85 kJ/mol.[169,179-181] The activation energy is extremely temperature dependent and decreases significantly with an increase in the temperature. Bachinger *et al.* reported the activation energy for the peptide Col 1-3, which contains a precursor specific region for bovine type III procollagen, is 70 kJ/mol at 5-15 °C; it lowers to 44 kJ/mol at 30-40 °C.[174] It is clear that the process of *cis*-*trans* isomerization is kinetically controlled.

Favorable inductive effects by the electronegative hydroxyl or fluoro group of 4(*R*)-Hyp and 4(*R*)-Flp can significantly increase the percentage of the amide bonds with the *trans*-conformation in the unfolded state.[117] The pre-organized single chain reduces the entropy of the unfolded state, lowers the energy barrier during the folding process, and

increases the stability of the collagen triple helix. The kinetics of triple helix folding will be discussed later.

Table 1.7 $K_{cis/trans}$ values for model compounds Ac-Yaa-OMe.[117,142,144]

Ac-Yaa-OMe	Pro	Hyp	hyp	Flp	flp	Mep	mep	Clp	clp
$K_{cis/trans}$	0.217	0.163	0.417	0.149	0.400	0.139	0.278	0.185	0.455
% *cis*	17.8	14.0	29.4	13.0	28.6	12.2	21.8	15.6	31.3

1.4 Folding of the collagen triple helix

Folding of the natural collagen triple helix is a multi-step, coordinated process *in vivo*.[182,183] The collagen peptide is synthesized in the rough endoplasmic reticulum (ER) membrane in a precursor form, procollagen, which contains *C*- and *N*-terminal propeptides, and the central region is $(Gly\text{-}Xaa\text{-}Yaa)_n$.[184-186] Then the synthesized chain is modified during the post-translational modification step by enzymes, such as prolyl and lysyl hydroxylase, to prepare the unfolded chain with Hyp or Hyl (hydroxylysine) in the sequence. The *C*-terminal propeptides of the three chains associate with each other and form a *C*-terminal trimer. The trimer propagates in a "Zipper-like" mechanism and forms a triple helical structure. After the formation of the triple helix, the propeptides at the *C*- and *N*-terminals are cleaved, and the triple helix assembles to form cross-linked fibrils.

When considering this process, the nucleation and propagation steps hold the most interest to researchers because these two steps are essential for both natural and synthetic collagen peptide folding. Short peptides that have a high content of imino acids (Pro, Hyp) can adopt a collagen-like triple helix. The triple helix can be stabilized by either the presence of repeating Gly-Pro-Hyp units at one, or both ends of the peptide chain, or covalent bonds crossing-linking the three chains.[187-189] These peptides provide important models for the

studies of triple helix conformation, thermodynamics, and the kinetics of collagen triple helix folding.[183]

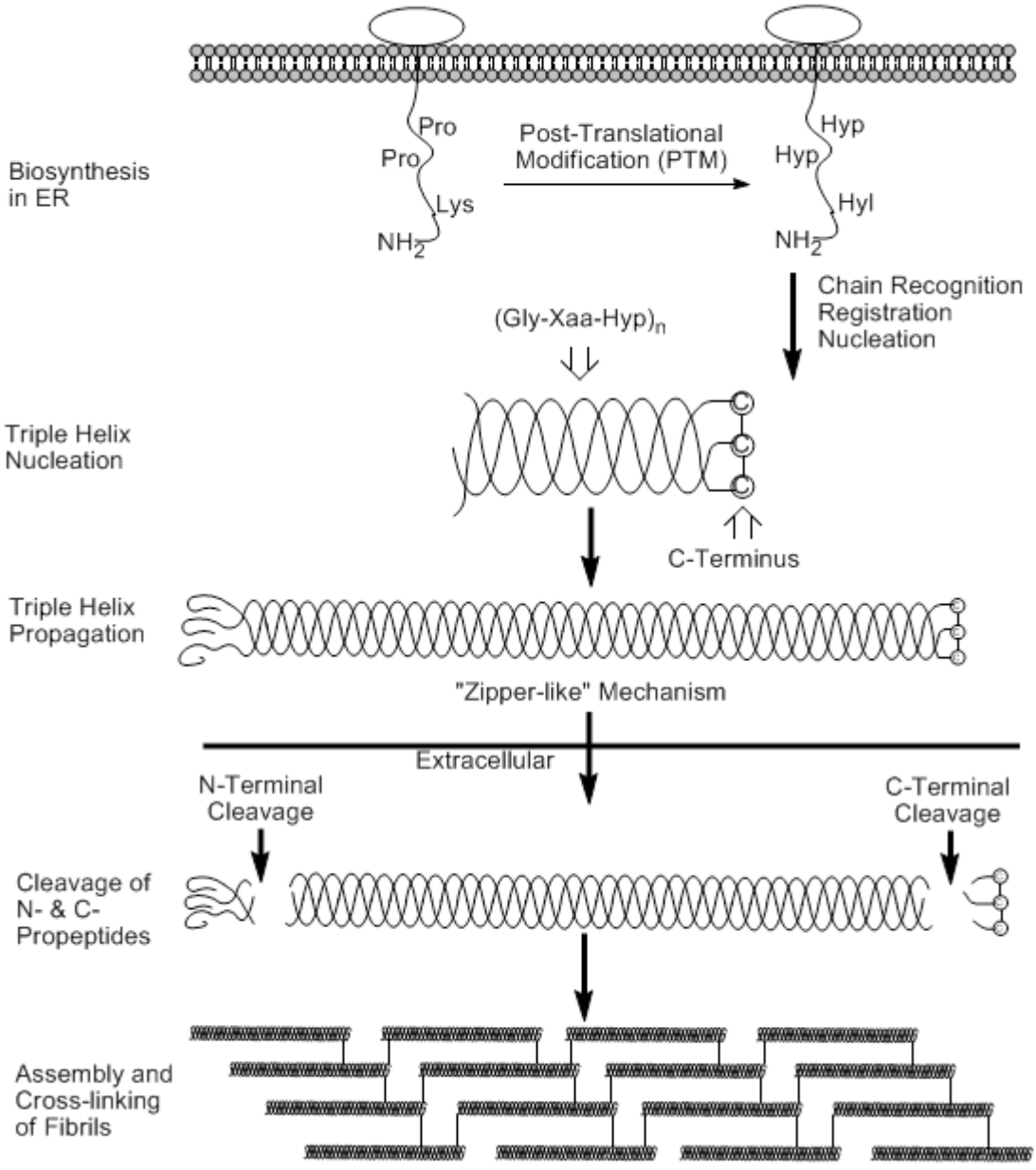


Figure 1.5 Natural collagen folding process. (Reprinted from Baum and Brodsky, *Curr. Opin. Strut. Biol.* **1999**, Copyright 1999, with permission from Elsevier.)

1.4.1 Nucleation

In natural collagen triple helix folding, nucleation is the step following the enzymatic post-translational modification. After the formation of Hyp and other hydrolyzed residues, which is catalyzed by prolyl hydroxylase and other enzymes, three chains are brought into register by association of the globular *C*-terminal domains.[184] If the collagen consists of different chains, the correct polypeptide chains must be selected before registration. Then the three registered chains are connected by various inter-chain interactions, such as covalent disulfide bonding, inter-chain hydrogen bonding, and side-chain interactions, and finally form a nucleus of a triple helical conformation at the *C*-terminus.[183,184]

The short globular domain formed by nucleation is very important for collagen folding. Unlike the α-helix and other proteins, collagen peptides have no backbone intra-chain hydrogen bonds. A single peptide chain cannot fold and it requires the nucleation at the *C*-terminus to bring three chains close enough for inter-chain interactions to form.[184] The formation of a *C*-terminal nucleated trimer is believed to be one of the rate-determining steps of the collagen triple helix folding process.[128,190] The nucleation domain serves to register the three chains at the nucleated terminus. They are also involved in the selection of different chains when the collagen is a heterotrimer. In addition, many collagen nucleation domains contain cross-linked disulfide bonds between the three chains. For example, type **III** collagen has a disulfide knot of six cysteines (two per chain) that links the three chains at the *C*-terminus. During the physiological folding process, the earlier non-covalent interactions between adjacent domains determine the formation of this knot, and once this knot is formed, it serves as an ideal registration and nucleation site to replace the nucleation domain.[175]

The hydroxylation of the Pro to Hyp residues at the Yaa position is important for nucleation at the *C*-terminus. It creates a unique region with 5 to 6 consecutive Gly-Xaa-Hyp units, which is believed to act as a nucleation site for the registration of the three chains. It also helps to arrange each chain with the correct dihedral angles, and form the proper inter-chain hydrogen bonds.[191] The deletion studies on Type **III** collagen showed that this Hyp-rich domain is required for nucleation, but two consecutive Gly-Xaa-Hyp units at the *C*-terminus may be enough for nucleation.[191]

Areida *et al.* reported that the nucleus of the triple helix can form at the *N*-terminus, and the propagation direction is from the *N*- to *C*-terminus in the ectodomain of type **XVII** collagen.[192] Frank *et al.* also found evidence that nucleation is possible at both *C*-terminus and *N*-terminus.[193] A kinetic study showed that the location of the nucleation affects the stability of the triple helix, but it does not affect the folding rate of the triple helix.[193]

1.4.2 Propagation

Propagation follows the nucleation step. After the formation of the *C*-terminal triple helical nucleus, the propagation of the collagen triple helix proceeds in a *C*- to *N*-terminus direction with a "zipper-like" mechanism.[184] The "zipper-like" folding mechanism of collagen is unique, because no other globular proteins that fold through this mechanism have to overcome a large number of consecutive unfavorable *cis-trans* isomerization processes.[184]

In the "zipper-like" propagation mechanism, the Gly-Xaa-Yaa repeating units have a similar ability to fold into a triple helical structure.[184] If the correct registration of the three chains is not maintained, misaligned helixes, in which incorrect tripeptide units form short triple helical segments, are likely to be generated.[184] So the correct registration is very

important to maintain the right collagen structure, especially when the collagen monomers have no disulfide bonds.[184]

Bachinger and co-workers studied the kinetics of triple helix folding, and found that only a very short segment (about 25 tripeptide units) between the nucleus and the first *cis*-amide bond can fold at a very fast rate.[174] Then the high activation energy of prolyl *cis-trans* isomerization, which is about 78 kJ/mol, will slow down the "zipper-like" folding process.[174] They indicated that the rate-limiting step of propagation is *cis-trans* isomerization of the peptide bonds of the imino acids.[174,175,194] They also suggested that the natural collagen folding process requires a specific *cis-trans* isomerase.[195] Bachmann *et al.* studied the triple helix formation of the peptide with all-*trans* amide bonds, and indicated that the "zipper-like" folding mechanism is entropy controlled.[196]

1.4.3 Thermodynamics and kinetics of folding of the collagen triple helix

Thermodynamics of folding of the collagen triple helix

The collagen triple helix is a thermodynamically stable structure, while establishment of the coil-to-triple helix equilibrium is relatively slow.[197,198] When heating the collagen triple helix in solution, the transition from triple helix to the unfolded state occurs within a very narrow temperature range, which indicates that the unfolding process is highly cooperative. The temperature at the midpoint of this transition (T_m) is used to describe the stability of the collagen triple helix, and T_m is specific for different collagens from different species.[40] The T_m value of natural collagens in solutions is only 3-4 °C above the body temperature of the organism from which it is isolated.[178,184]

Most natural collagens show an irreversible denaturation curve, because the molecules were proteolytically cleaved in the process, and some important domains required

for folding are missing.[168] The T_m value was found to be dependent on the rate of temperature increase and the concentration of the peptide in solution. It increases with an increase in concentration, although the concentration does not significantly influence the unfolding transition.[168] However, concentration-dependent hysteresis is obvious during the refolding process.[168,199]

The enthalpy change ΔH^0 associated with the conversion from the coil state to triple helix for collagens is about –15 to –18 kJ/mol per tripeptide unit, and is independent of temperature. The value of ΔH^0 increases with an increase in the imino acid content of the peptide.[134] The enthalpy change ΔH^0 per residue of the triple helix formation is significantly larger than that for globular proteins, but the factors that cause that big enthalpy change are still unknown.[168]

It is notable that the calorimetric enthalpy and the van't Hoff enthalpy are significantly different in the thermodynamic analysis of the triple coil to triple helix transition.[134] The ratio of the van't Hoff enthalpy and the calorimetric enthalpy can be used to determine the cooperative length of the collagen triple helix, in which a tripeptide unit has cooperative interactions with adjacent triplets.[200] Davis and Bachinger studied the fibrillar collagen types **I**, **II** and **III** and determined that the cooperative length is 80-100 tripeptide units, which corresponds to about one-tenth of the total length of the triple helix.[200] It is notable that the cooperative length is only an average parameter for the whole triple helix and does not consider the sequence variations at specific positions.[168]

In the thermodynamic studies of fibrillar collagens, no equilibrium intermediates were observed during the unfolding transition for the triple helixes of type **I**, **II** and **III** collagens, although these types of collagens may have different chain compositions in a

single triple helix. In contrast, some stable intermediates were found in the unfolding process for type **V** and **XI** collagens.[201] It is believed that sequence-dependent stability is responsible for the presence of stable intermediates.[201]

Synthetic collagen-like peptides provide excellent models for revealing the details of the stability of the collagen triple helix. Of all the classical factors that affect protein stability, the electrostatic interactions are thought to play only a minor role in stabilizing the collagen triple helix.[168] The hydrophobic effect may contribute to the stability of the triple helical structure, but its contribution is not as significant as it is to the stability of globular proteins.[168]

The stabilizing effects at the molecular level were discussed in Section 1.3. The stabilizing or destabilizing mechanisms are very complex and many factors simultaneously affect triple helix folding. There is no way to quantitatively measure the contribution of each factor to the overall stability of the triple helix. It is still very difficult to predict whether a single structural change will have positive or negative contribution to the overall stability of the triple helix.

Kinetics of folding of the collagen triple helix

The simplest model used in the study of the thermodynamics and kinetics of collagen triple helix folding is called the "all-or-none" model, and the mechanism can be described as follows,[122]

$$3C \overset{\beta s^x}{\longleftrightarrow} H_x \overset{s}{\longleftrightarrow} \cdots \overset{s}{\longleftrightarrow} H_i \overset{s}{\longleftrightarrow} \cdots \overset{s}{\longleftrightarrow} H_{3n-2} \qquad (1)$$

where C is a coiled chain, H_i is a triple helix with i tripeptide units already in the triple helical structure, β is the cooperative parameter for the nucleation step, βs^x is the equilibrium constant for the nucleation step to form the nucleus H_x, and s is the equilibrium constant for

adding a tripeptide unit in the propagation steps. Only 3n-2 tripeptide units can form inter-chain hydrogen bonds due to the staggered arrangement of the triple helix and the complete triple helical structure is denoted H_{3n-2}. The "all-or-none" model makes an assumption that the nucleation step is rate-limiting. The triple helix is completed in the propagation steps immediately after the nucleation step. This model also supposes that all the propagation steps are identical with the same equilibrium constant. Under these conditions, the concentration of H_i is negligible compared to the concentrations of C and H_{3n-2}. Then the equilibrium constant K can be shown to be:

$$K = \frac{[H_{3n-2}]}{[C]^3} = \beta s^{3n-2} = \frac{F}{3c_0^2(1-F)^3} \quad (2)$$

where c_0 is the initial total concentration of coiled chains, F is the fraction of chains or tripeptide units in the triple helical state, which means $F = 3[H]/ c_0$.[168]

The length of the triple helix is the major factor to determine the complexity of the kinetics of the collagen peptide folding. The "all-or-none" mechanism used in kinetic studies is only good for short collagens and model peptides, and the kinetics of a system with up to 45 tripeptide repeating units per molecules (15 units per chain) have been successfully treated using this approximation.[202,203] This model cannot provide a good approximation for long triple helixes, in which the kinetics are more complex than those of short naturally occurring or designed systems.[168]

Bachinger and co-workers studied the folding kinetics of the synthetic model peptides $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Pro\text{-}Hyp\text{-}Gly)_{10}$.[204] They found that the kinetics of the refolding process were extremely concentration dependent as expected. They also found that it was impossible to fit the experimental kinetic curve to theoretical integrated equations of even reaction orders.[204] These data implied that the reaction order changed as the folding process

progressed. The initial reaction orders were close to 3 at very low concentrations but dropped with increasing concentration. Most of the time, the reaction orders were between 2 and 3, much closer to 2 than to 1 or 3.[168,204]

The simplest collagen folding mechanism fails to explain the high reaction order and its concentration dependence, so a new folding mechanism was proposed.[204] In this mechanism, the simultaneous collision of three chains in solution phase is extremely rare, and the direct formation of a triple helical nucleus H^* can be neglected. A very unstable dimeric nucleus D^* is proposed to be the intermediate which is in fast pre-equilibrium with two single chains. D^* forms in a bimolecular reaction and H^* is formed from it by combination with a third chain.[204]

$$2C \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \begin{matrix} C \\ + \\ D^* \end{matrix} \xrightarrow{k_3} H^* \xrightarrow{k_4} H \tag{3}$$

In this mechanism, either nucleation or propagation can be the rate-limiting step. The formation of D^* is a fast reaction, and the formation of the triple helical nucleus H^* is relatively slow.[204] Then the steady-state equilibrium is achieved as detailed in Eq. 4.[204]

$$\frac{d[D^*]}{dt} = k_1[C]^2 - k_2[D^*] - k_3[D^*][C] = 0 \tag{4}$$

The rate of dissociation of D^* is much faster than the formation of H^*, which means $k_2 >> k_3[C]$. Then the rate of formation of H^* is

$$\frac{[D^*]}{[C]^2} = \frac{k_1}{k_2} = K$$

$$\frac{d[H^*]}{dt} = k_3[D^*][C] = Kk_3[C]^3 = k_a[C]^3 \tag{5}$$

It is very clear that the rate constant k_a is a third-order rate constant, and it is derived from the equilibrium constant K, and a second order rate constant k_3. At very low concentrations, the formation of H^* is the rate-limiting step, and the rate of triple helix formation was estimated to be close to the rate of nucleus formation as shown in Eq. 5. The rate of helix propagation is much faster than the rate of nucleation. The initial rate is given in Eq. 6.

$$\left(\frac{dF}{dt}\right)_{t=0} = 3k_a\left[C_0\right]^2 \tag{6}$$

It is clear that the initial triple helix folding process is a second-order reaction. At sufficiently high concentrations, at which the nucleation is not a problem, the propagation then becomes the rate-limiting step and this process is a first-order reaction.

$$\left(\frac{dF}{dt}\right)_{t=0} = k_4\left[C_0\right] \tag{7}$$

Real collagen peptide folding is so complicated that it is difficult to decide which step is rate limiting. From the kinetic study of the synthetic model peptides, $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Pro\text{-}Hyp\text{-}Gly)_{10}$, nucleation is the rate-limiting step for $(Pro\text{-}Pro\text{-}Gly)_{10}$, whose reaction order is 2.8-3 at the low concentrations, and about 2.5 for the average of all data, while $(Pro\text{-}Hyp\text{-}Gly)_{10}$ has a average reaction order of 1.5 which means that the propagation step is more rate-limiting.[204] The value of k_a differs by a factor of 1000 between $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Pro\text{-}Hyp\text{-}Gly)_{10}$, while the propagation rate constant k_4 only differs by a factor of 6.[204] These results confirmed that 4(*R*)-Hyp stabilizes the triple helix by pre-organizing the chains in the unfolded state, and increases the reaction rate at the nucleation step.

The kinetics of many other natural or synthetic collagen peptides have been studied.[161,205,206] The results showed that small sequence differences can cause significant

changes in the folding process. The above folding mechanism provides a good approximation for collagen folding based on the kinetic and thermodynamic studies, which are also supported by computational studies.[207]

Most of the natural collagens have a nucleation domain that is located at the *N*-terminus or *C*-terminus of the triple helix. The presence of a cross-link between the three chains in a collagen molecule largely influences the kinetics.[175,200] Usually, Cys-disulfide bonding at the *C*- or *N*-terminus helps the nucleation of the triple helix. The kinetic study of the fragment Col1-3, which contains an *N*-terminal propeptide, showed that *cis-trans* isomerization is the rate determining step, and the refolding process is a first order reaction.[174] It was also found that the refolding process of the Col1-3 consists of two phases, and the fast phase was not resolved.[174] The kinetics of folding of the cross-linked collagen triple helixes can be very complex and two more models were built to reveal the details of the folding process. Bachinger and Engel described detailed mechanisms, and provided an excellent explanation for the kinetics of various cross-linked triple helixes.[168]

1.5 Summary

Collagen is one of the most abundant and important proteins in mammals. It comprises at least 28 families. Collagen acts as a scaffolding material for our bodies, controls cell shape and differentiation, and is responsible for the elasticity and strength of the body.

Collagen is a super-helical structure. It consists of three left-handed PPII helixes coiled along a common axis to form a very compact right-handed super helix. The primary structure is shown to be $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats with 10 % of Xaa residue being Pro and 10-12 % of Yaa residues being Hyp. For most of the fibrillar collagens, n is greater than 300, but the number varies for different collagen types.

Inter-chain backbone hydrogen bonding is one major source of stabilization for the collagen triple helix. Various water-mediated hydrogen bonds also play an important role in stabilizing the triple helical structure. Hyp residue can form water-mediated hydrogen bonds. The hydroxyl group can also pre-organize the amide bond to the favorable *trans*-conformation to stabilize the triple helix by its inductive effect. *Cis-trans* isomerization of the prolyl amide bonds also contributes to the stability of the collagen triple helixes in either a positive or negative manner.

Nucleation and *cis-trans* isomerization are the rate-limiting steps during collagen triple helix folding. The kinetics change during the process of triple helix folding. The initial reaction orders were close to 3 at very low concentrations, but dropped with increasing concentration. Most of the time, the reaction orders were between 2 and 3, much closer to 2 than to 1 or 3.

Chapter 2 Design and synthesis of conformationally locked Gly-Pro and Pro-Pro alkene isosteres

2.1 Design and Synthesis of the Gly-Ψ[(E)CH=C]-Pro-Hyp isostere

2.1.1 Theoretical base for the design of the alkene isosteres

Collagen has low immunogenicity compared to other biomaterials and can be changed into different forms, such as sheets, tubes, sponges, powders, injectable solutions and dispersions, to fulfill different clinical requirements.[208-210] It is widely used as a biomaterial in tissue engineering and drug delivery systems.[60] But many problems regarding the inherent properties of natural collagens, such as poor mechanical strength, insufficient supply and ineffectiveness in the management of the infected sites limit the usage of the native collagens in practical applications.[60] The most common source of clinical collagen is *Bos Taurus*, a domestic cow that can be found throughout the world. But bovine collagen has significant immunological and pathological side effects to humans.[211-213] Moreover, natural collagen molecules also cause problems in collagen biological studies due to their poor solubility in most physiological buffers, and due to the difficulties in expression and purification of recombinant collagen and generation of shorter collagen fragments.[214]

Many collagen-like polypeptides of various lengths with repeating units Gly-Xaa-Yaa have been chemically synthesized and characterized since 1968.[215] Synthetic collagen polypeptides are already regarded as biomaterial candidates to improve and replace natural collagens.[216-219] But synthetic peptides have limited lengths, which are usually less than 10 nm, much shorter than natural collagens. Koide and Raines found that synthetic short

peptides with cysteine-knots at the end could spontaneously self-assemble to collagen-like triple helical supermolecules.[220,221] Kotch and Raines used this method to make stable collagen-like fibrils which were over 400 nm in length, longer than any known collagen.[220] These discoveries made synthetic collagen-like polypeptides a possible replacement for natural collagens in practical applications.

Structural details about the collagen triple helix have been elucidated from X-ray diffraction patterns.[24,33,34] High resolution X-ray diffraction from collagen single crystals enables the study of hydrogen bonding patterns, single-residue conformations and mutation effects.[6,57] Real-time NMR and differential scanning calorimetry (DSC) illustrate the kinetics and thermodynamics of collagen triple helix folding.[141,222] These techniques help us to understand the structure and stability of collagen-like peptides and improve our ability to design and synthesize more stable collagen-like biomaterials.

The crystal structure of $(Pro\text{-}Pro\text{-}Gly)_{10}$ showed that the Pro residues at the Xaa position adopt the C^{γ}-*endo* pucker, while the Pro residues at the Yaa position prefer the C^{γ}-*exo* pucker.[55,56] Raines used 4(*R*)-Flp to replace 4(*R*)-Hyp and found that the preference of 4(*R*)-Hyp and 4(*R*)-Flp to adopt the C^{γ}-*exo* pucker is responsible for the additional stabilizing effect of Hyp and Flp.[138] Kobayashi and co-workers believed that the high degree of hydration of Hyp and water-mediated hydrogen bonds are the reason that Hyp can stabilize the collagen triple helix.[141,149,150] 4(*R*)-Flp and 4(*S*)-Flp increase the triple helix stability entropically, while 4(*R*)-Hyp increases the triple helix stability enthalpically (Section 1.3.3).[141] Pre-organization of the peptide chains in the unfolded state can lower the energy barriers from the coil state to the triple helix transition, and increase the stability of the resulting collagen triple helix.[117,196]

The peptide amide bond has partial double bond character and is planar. The amide bond preceding proline has a much greater *cis*-fraction than preceding other residues,[117,145] while all of the peptide amide bonds in the collagen triple helixes are in the *trans*-conformation.[178] In the transition from unfolded peptide chains to the collagen triple helix, all of the *cis*-amide bonds must be converted to the *trans*-conformation because of conformational restraints.[178] Bachinger *et al.* showed that *cis-trans* isomerization of the prolyl amide bond is the rate-limiting step of collagen peptide folding.[174,195] For a common length type I collagen $(Xaa\text{-}Yaa\text{-}Gly)_{300}$, there will be a significant number of amide bonds in the *cis*-conformation in the unfolded state.[163] So there will be a huge entropic barrier during the collagen folding process. The ratio of the *cis* to *trans* conformations in the unfolded state directly affects the stability of the triple helix.[117] Locking amide bonds to the *trans*-conformation can lower the entropic energy barrier during the collagen folding process, thus increase the stability of the collagen triple helix.[117,196]

The direct backbone interchain hydrogen bonding between the C=O of the Xaa residue (usually Pro) and the N-H of Gly stabilizes the collagen triple helix.[54] Even with the Gly→Ala mutation, the direct interchain hydrogen bonding is still between the C=O of the Xaa residue and the N-H of Gly (or Ala).[6] There is no report demonstrating that the C=O of Gly participates in any direct interchain hydrogen bonding, although it is a candidate for water mediated hydrogen bonding.[57] Replacement of one Pro-Gly amide bond at the Yaa position in the middle of the sequence of $(Pro\text{-}Pro\text{-}Gly)_{10}$ with an ester or alkene bond was found to destabilize the triple helical structure.[223]

Based on previous studies,[224] we designed a conformationally locked collagen Gly-Pro-Hyp tripeptide isostere (Figure 2.1). Isostere **1** was the Gly-Pro-Hyp tripepitde building

block. In isostere **2**, the amide bond between Gly and Pro was replaced with an (*E*)-alkene bond to stabilize the collagen triple helix. The alkene bond locked Gly-Pro in the *trans*-conformation to prevent *cis-trans* isomerization, which should reduce the conformational entropy of collagen triple helix folding. Meanwhile, the C=O of Pro and N-H of Gly were left intact to maintain the interchain hydrogen bonding. We expected this substitution to lead to an overall stabilization of the collagen triple helix.

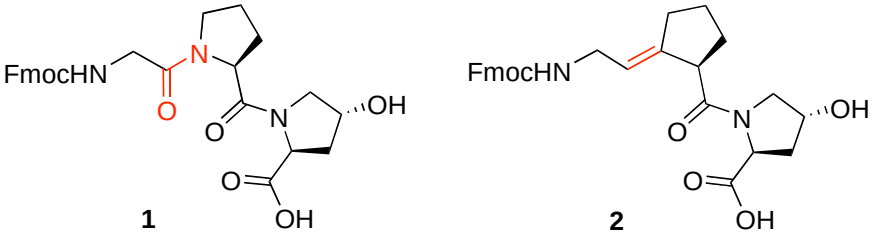


Figure 2.1 Design of collagen-like tripeptide isostere **2**.

2.1.2 Synthesis of Fmoc-Gly-Pro-Hyp-OH tripeptide mimic

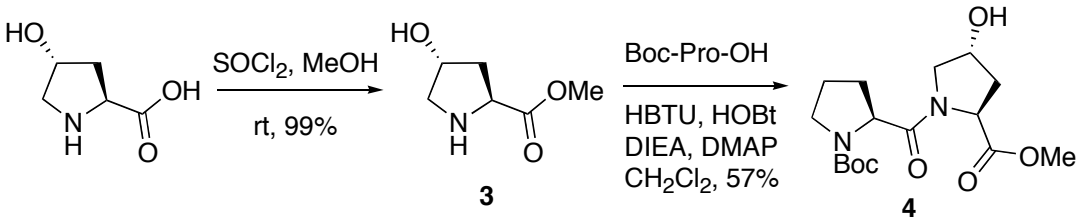
The tripeptide repeat Gly-Pro-Hyp is the most common repeating unit used in collagen studies, and the properties and structure of collagen-like polypeptides containing Gly-Pro-Hyp have been studied thoroughly. In collagen host-guest peptide systems, the Gly-Pro-Hyp tripeptide monomer is the building block of the sequence. Moreover, collagen peptides are difficult to purify even by HPLC. Solid phase peptide synthesis with single amino acid will yield side-products with minus one or minus two residues sequences, which are very difficult to separate from the desired peptide. Coupling tripeptides can make the mass of side products more different from the target, and minimize the problems during peptide purification. To study the stability of the peptides containing the conformationally locked alkene isosteres, the Gly-Pro-Hyp tripeptide building block was synthesized first.

9-Fluorenylmethoxycarbonyl (Fmoc) protected amino acids are widely used in solid phase peptide synthesis.[225,226] Compared to Merrifield's carbobenzoxy (Cbz) or *tert*-butyloxycarbonyl (Boc) methods, Fmoc chemistry has milder deprotection conditions, is easier to handle, and does not require the use of highly toxic HF in the final cleavage step.[227,228] So the Fmoc method was adopted and tripeptide Fmoc-Gly-Pro-Hyp-OH was synthesized to be used as the building block for solid phase peptide synthesis.

Methyl protected Gly-Pro-Hyp tripeptide building block

To synthesize the Gly-Pro-Hyp tripeptide building block, a methyl group was used as the protecting group at the *C*-terminal end of the tripeptide. Starting from commercially available 4(*R*)-hydroxyproline, methyl ester **3** was synthesized by Fisher esterification and crystallized as a clear plate-like crystal from a MeOH-hexanes mixture, or precipitated as a white solid from diethyl ether (Scheme 2.1).[229]

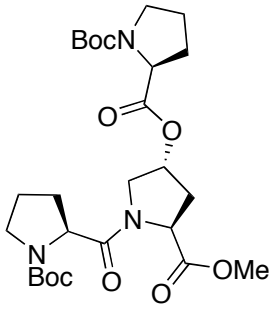
Scheme 2.1 Synthesis of dipeptide Boc-Pro-Hyp-OMe **4**.



Hydroxyproline methyl ester **3** was coupled with Boc-proline to afford dipeptide **4**. HBTU was used as the coupling reagent because the yield was usually higher than when EDC was used. The secondary amine group of Hyp should be more reactive than the secondary hydroxyl group, but when two equivalents of Boc-proline were used, only about 10 %-15 % of the desired dipeptide **4** was obtained. The hydroxyl group was also coupled with Boc-proline and a double-coupled side product **5** was obtained in about 40 % yield (Figure 2.2). The ratio of 4-hydroxyproline methyl ester **3** to Boc-Pro-OH was then strictly

controlled to 1 to 1 and the yield increased to 57 %. When 4-hydroxyproline methyl ester **4** was in excess, there was no significant improvement in the yield. Using a more expensive coupling reagent, benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluoro-phosphate (BOP), gave only the desired dipeptide Boc-Pro-Hyp(OH)-OMe **4**, but the yield was also not very high (68 %).[230]

The poor selectivity between the hydroxyl group in Hyp and the secondary amine in Pro was observed when synthesizing tripeptide Fmoc-Gly-Pro-Hyp-OMe by direct EDC or HBTU coupling. The yield was less than 10 %, and Fmoc-Gly-OH was coupled at both the amine group of Pro and the hydroxyl group of Hyp, even when the ratio of the two reactants was controlled. A new strategy had to be used to improve the yield. The *tert*-butyldimethylsilyl (TBS) group was used to protect the hydroxyl group before the coupling reaction was performed.



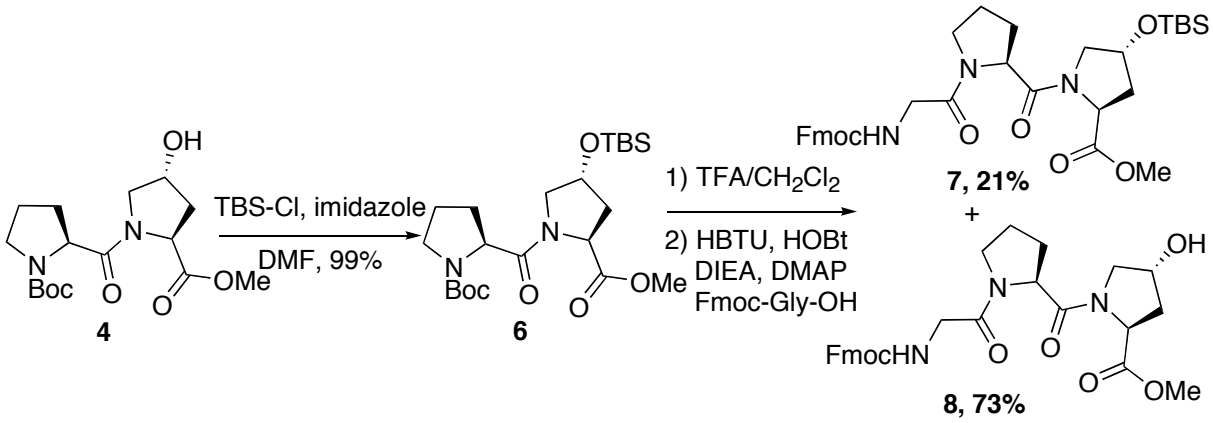
5

Figure 2.2 Double-coupled side product Boc-Pro-Hyp(Boc-Pro)-OMe **5**.

Starting from dipeptide **4**, TBS-Cl was used to provide TBS protected tripeptide **6** in the presence of imidazole (Scheme 2.2). Then the Boc-group was removed using TFA, and the free amine was coupled with Fmoc-Gly-OH with HBTU as the coupling reagent. After the reaction, both protected tripeptide **7** and unprotected tripeptide **8** were found. The ratio of these two products varied, but usually unprotected tripeptide **8** dominated. This indicated that the TBS group was also partially cleaved during the TFA deprotection of the Boc group.

Deprotecting the Boc group first, then attaching the TBS protecting group might solve this problem. But the resulting amine product has very poor visibility under UV, in ninhydrin or $KMnO_4$, and it is very difficult to obtain a pure intermediate for the next coupling reaction. This method was not used.

Scheme 2.2 Synthesis of tripeptide Fmoc-Gly-Pro-Hyp-OMe **8**.

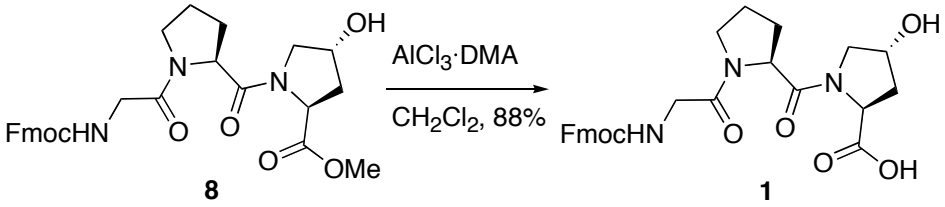


The most common deprotection method for the methyl group is hydrolysis under basic conditions, such as in aqueous LiOH solution.[231] But these basic conditions can also simultaneously cleave the Fmoc group. A new method to deprotect methyl esters using the Lewis acid $AlCl_3$ and *N*,*N*-dimethyl aniline (DMA), was recently reported.[232] The activity of the catalyst was extremely important for this reaction. $AlCl_3$ had be very dry and well dissolved in DMA, although $AlCl_3$ dissolved very slowly in DMA. When it was ready for use, the solution should be clear dark blue or green. This method could remove the methyl group without affecting the Fmoc group (Scheme 2.3). Even if the reaction did not work, nearly 100 % of the starting material could be recovered.

Unfortunately, the desired product tripeptide **1** had a very similar polarity to the catalyst complex, and it was very difficult to separate them by flash chromatography. Even after HPLC purification, there was still a small amount of the catalyst complex present which

made the solid product light blue. A significant amount of the product was lost during the HPLC purification. A new protecting group was used in place of the methyl protecting group to improve the overall yield.

Scheme 2.3 Synthesis of the tripeptide building block Fmoc-Gly-Pro-Hyp-OH **1**.



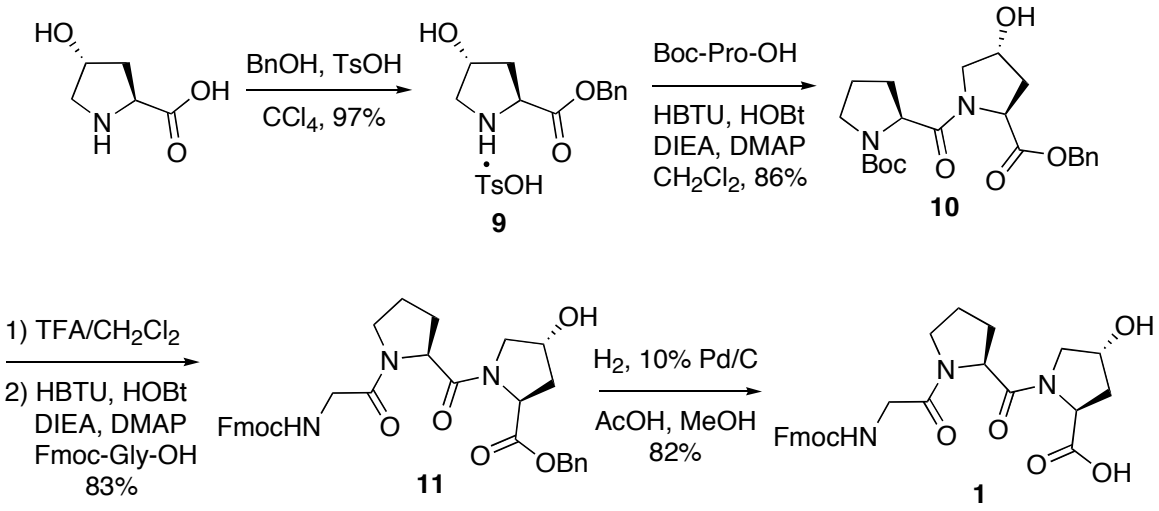
Benzyl protected Gly-Pro-Hyp tripeptide building block

Hodges and Raines reported that H_2 on Pd/C in a mixture of MeOH and AcOH could remove a benzyl group without affecting the Fmoc protection.[152] A benzyl group was used to protect the *C*-terminal end of tripeptides, and the synthetic route was shown in Scheme 2.4. Starting from commercially available 4-hydroxyproline, 4-hydroxyproline benzyl ester **9** was synthesized by heating at reflux in a mixture of CCl_4 and BnOH.[233] Dipeptide Boc-Gly-Hyp-OBn **10** was synthesized by HBTU peptide coupling. The Boc protecting group was then removed with 25 % TFA in CH_2Cl_2, and tripeptide **11** was synthesized with HBTU as the coupling reagent. The benzyl group was removed with H_2 on 10 % Pd/C in 50 % AcOH in MeOH to produce the tripeptide building block **1**.

It is noticeable that very little double-coupled products were observed during the synthesis with the benzyl protecting group. Benzyl and methyl groups have similar, very weak inductive effects on the 4-hydroxyl group of Hyp, and the electronic effect should not be a major factor in explaining this result. I suspect that the size difference between the benzyl and methyl protecting groups is the reason why the benzyl protected Hyp resulted in fewer side products during peptide coupling reactions. The benzyl protecting group may have

long-range interactions with the hydroxyl side-chain, and may partially block the hydroxyl group during the reaction. Therefore, the hydroxyl group of Hyp is much less reactive, and the desired product is favored.

Scheme 2.4 Alternative synthetic route for the tripeptide building block Fmoc-Gly-Pro-Hyp-OH **1**.

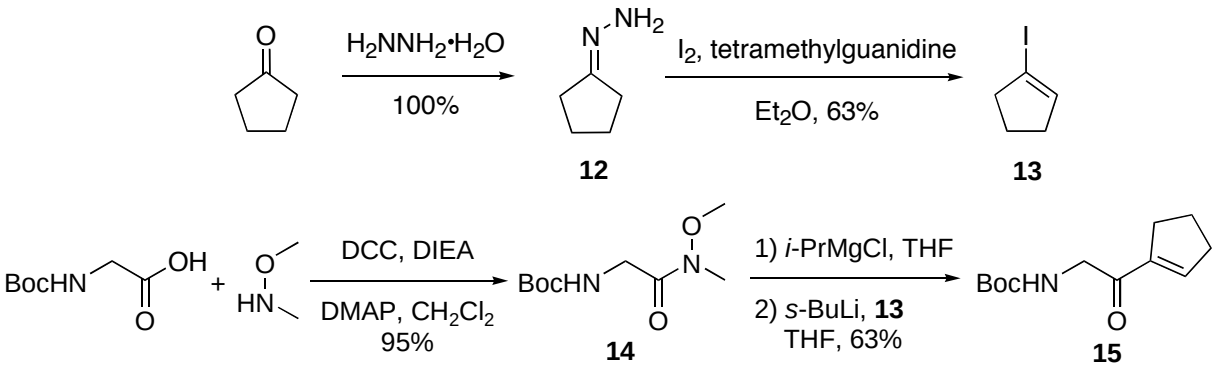


2.1.3 Synthesis of Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp-OH isostere

Wang *et al.* published a method for the synthesis of the alkene isostere Ser-Ψ[(*E*)CH=C]-Pro,[234] which was modified to synthesize the Gly-Ψ[(*E*)CH=C]-Pro alkene isostere. Starting from commercially available cyclopentanone, hydrazone **12** was synthesized in a high yield (Scheme 2.5). After work-up and vacuum drying, hydrazone **12** was used to make 1-iodocyclopentene **13** without further purification. Iodocyclopentene **13** was not very stable. It was stored under Ar in a freezer and usually used within one week. It could be stored much longer (up to two months) after careful removal of moisture and protected from light at –20 °C. Since iodocyclopentene **13** was highly volatile, no heat was allowed during the work-up, and it was usually used without column purification and vacuum

drying. If there was too much water in the system, it could be purified with Et_2O by flash chromatography.

Scheme 2.5 Synthesis of α,β-unsaturated ketone **15**.

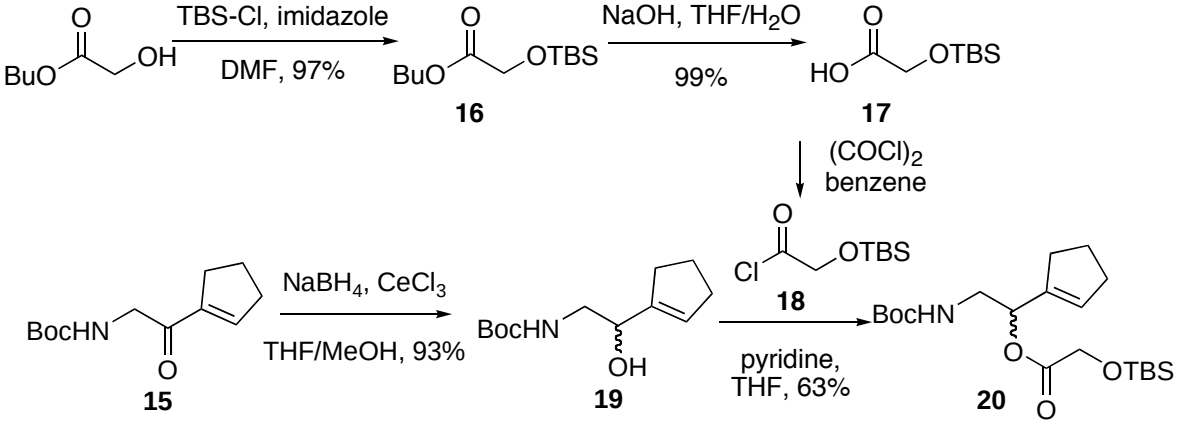


The Boc-Gly Weinreb amide **14** was synthesized by DCC coupling with Boc-Gly-OH and *N,O*-dimethyl hydroxylamine (Scheme 2.5). The product was crystallized as colorless needle-like crystals from EtOAc-hexanes or ether binary solvent system. To synthesize the Boc-Ser Weinreb amide, 50 % DMF in CH_2Cl_2 was used as the solvent.[234] When synthesizing the Boc-Gly Weinreb amide **14**, the yield was much lower in a DMF-CH_2Cl_2 binary solvent system was used than only in CH_2Cl_2. Iodocyclopentene **13** was reacted with *s*-BuLi to generate cyclopentenyl lithium, which was then reacted with the Boc-Gly Weinreb amide **14** to synthesize α,β-unsaturated ketone **15**. The proton on the nitrogen atom of the Boc-Gly Weinreb amide is the most reactive site within the molecule, and cyclopentenyl lithium would react with this proton first. About one equivalent of *i*-PrMgCl was used to remove the proton on the N-H first to obtain a higher yield, and reduce the usage of *s*-BuLi and iodocyclopentene **13**.

The synthesis of acid chloride **18** has been reported by Bischofberger *et al.*[235] TBS-protected butyl glycolate **16** was synthesized by mixing TBS-Cl and *n*-butyl glycolate in the absence of solvent (Scheme 2.6). The product was purified by vacuum distillation at 85-

90 °C. One modification was made to improve the yield of acid **17**. One equivalent of NaOH was used to remove the *n*-butyl group instead of KOH, and a much higher yield was obtained (99 % compared to 80 %). The increase of the yield might be because NaOH had a higher purity than KOH. The use of one equivalent of base was important. If an excess of base was used, acid **17** was very difficult to extract back into the organic layer after an acid wash, and this greatly decreases the yield.

Scheme 2.6 Synthesis of ester **20**.

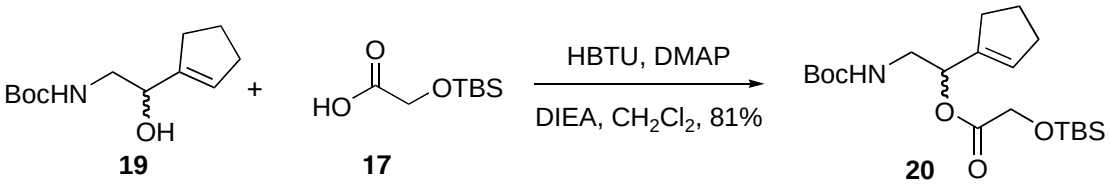


Ketone **15** was reduced by $NaBH_4 \cdot CeCl_3$, and alcohol **19** was obtained (Scheme 2.6). Unlike the synthesis of the Ser-Ψ[(*E*)CH=C]-Pro isostere, the reduction has no stereoselectivity because Gly has no C^{α}-side-chain, and both enantiomers were obtained. Acid chloride **18** was synthesized by reflux with $(COCl)_2$ in benzene. Because acid chloride **18** was susceptible to moisture, it was used without work-up or purification. After removing the solvent benzene by distillation, acid chloride **18** was dissolved in THF and mixed with pyridine before addition to the alcohol **19**-THF solution. The resulting ester **20** was purified by flash chromatography and dried in vacuo.

Ester **20** could also be synthesized by coupling acid **17** with alcohol **19** using peptide coupling reagents such as EDC or HBTU. The HBTU coupling method gave a higher yield

of 81 %, but the starting materials could not be recovered and the coupling reagent was more expensive than $(COCl)_2$. The acid chloride method was clean, and no side-product was formed during the reaction. Although the yield was lower, nearly all of the unreacted starting material alcohol **19** could be recovered. It was not necessary to use the HBTU coupling method for the Gly-Pro isostere. When synthesizing the Pro-Pro isostere, the HBTU coupling method was necessary due to the extremely low yield of the acid chloride method (Section 2.2.3).

Scheme 2.7 HBTU coupling method to synthesize ester **20**.



The Ireland-Claisen rearrangement is the key step in the synthesis procedure. Through a chair-form 6-membered ring transition state, the stereochemistry was controlled so the alkene in the product adopted the (*E*)-configuration (Figure 2.3),[234,236,237] which was confirmed by NOE spectroscopy later.

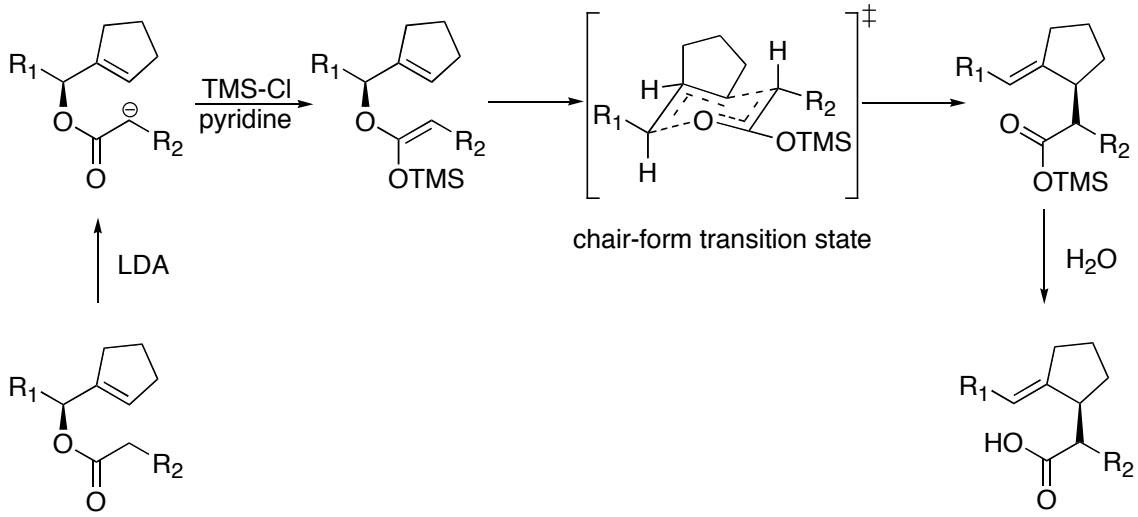
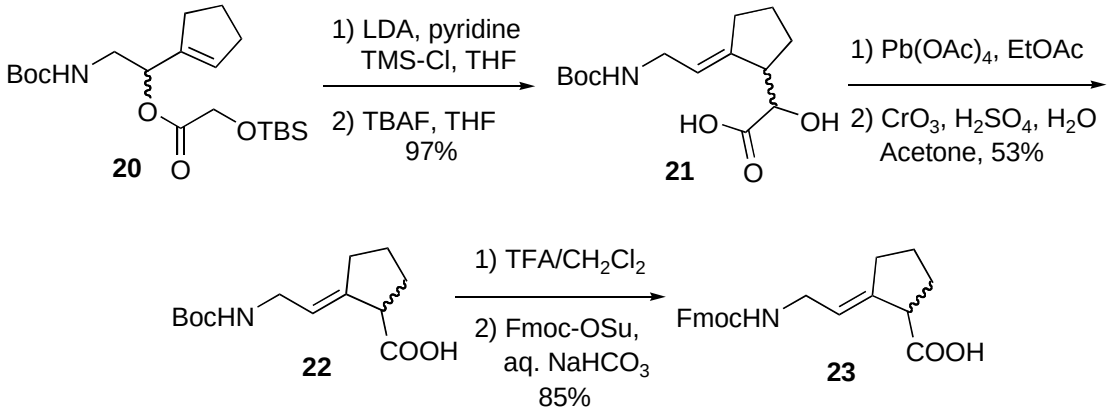


Figure 2.3 Mechanism of the Ireland-Claisen rearrangement.

Because ester **20** was a mixture of two enantiomers, the product after the rearrangement was still a racemic mixture. This reaction was extremely moisture sensitive, and all of the glassware was flame dried prior to use. The quality of the reagents used was also very important. The solvents had to be anhydrous and sealed from air at all times. Lithium diisopropylamide (LDA) was freshly prepared and contained no solid material. Poor quality LDA can cleave the TBS protecting group and stop the rearrangement from occurring. If the rearrangement did not happen and the starting material was recovered, then it was necessary to check the quality of TMS-Cl and pyridine. The rearrangement did not occur when a white precipitate appeared after mixing TMS-Cl and pyridine. So it was important to check TMS-Cl and pyridine before setting up the reaction.

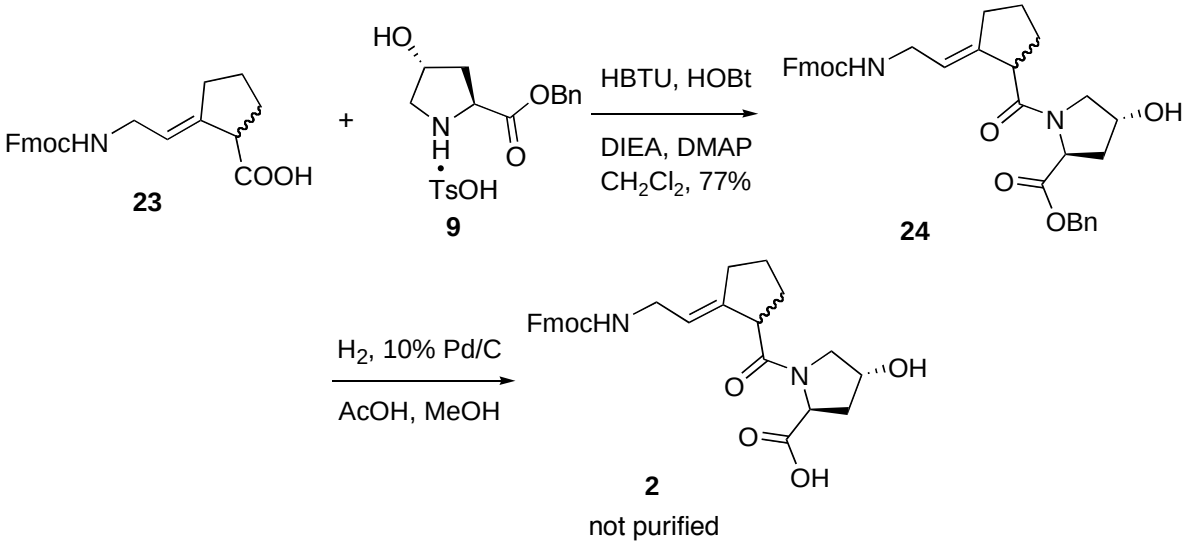
After the Ireland-Claisen rearrangement, the resulting intermediate was treated with tetrabutylammonium fluoride (TBAF) to cleave the TBS group, and α-hydroxy acid **21** was obtained (Scheme 2.8). α-Hydroxy acid **21** was decarboxylated with $Pb(OAc)_4$ in EtOAc to give an aldehyde. The resulting aldehyde was not stable, and was oxidized to acid **22** with freshly prepared Jones reagent without further purification.

Scheme 2.8 Synthesis of Fmoc-protected acid **23**.



The Boc protecting group was converted to the Fmoc protecting group to make the final product compatible with solid phase synthesis. The Boc group was cleaved with 25 % TFA in CH_2Cl_2 (Scheme 2.8). Then the isostere was protected with Fmoc-OSu in $NaHCO_3$. Fmoc-Cl was used to add the Fmoc protecting group to the amine group first, but the yield was only 30-35 %. A similar yield was reported with the same substrate (Dr. Xiaodong J. Wang, unpublished results). When using Fmoc-OSu to add the Fmoc protecting group, the yield increased to 70-85 %. This result was opposite to the Ser-Pro isostere case. When changing the protecting group during the synthesis of Ser-Ψ[(*E*)CH=C]-Pro isostere, Fmoc-OSu gave only 20-30 % yield, while Fmoc-Cl provided a yield of more than 80 % (Dr. Song Zhao, unpublished results). The reason for this difference was unknown.

Scheme 2.9 Synthesis of tripeptide isostere **2**.



Tripeptide **24** was synthesized by HBTU peptide coupling of Fmoc-protected acid **23** with hydroxyproline benzyl ester **9** (Scheme 2.9). The benzyl protecting group was removed by hydrogenation over a 10 % Pd/C catalyst in 50 % AcOH in MeOH. The ^{1}H NMR spectrum showed that the alkene proton peak at 5.4 ppm was present in the product (Appendix). The tripeptide alkene isostere **2** was still a mixture of two diastereomers with

about 3:2 ratio, and separation by HPLC was planned. A model reaction of solid phase peptide synthesis of Ac-(Gly-Pro-Hyp)$_7$-Gly-Gly-NH_2 with the tripeptide building block Fmoc-Gly-Pro-Hyp-OH **1** showed that the hydroxyl group of Hyp had to be protected, and it was very difficult to add a protecting group to hydroxyl at this stage without affecting the Fmoc-protecting group. So the side-chain unprotected alkene isostere **2** was synthesized, but not purified by HPLC.

2.1.4 Synthesis of Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH tripeptide mimic

A model solid phase peptide synthesis with the tripeptide building block Fmoc-Gly-Pro-Hyp-OH **1** was performed on Rink amide MBHA resin. Although the desired product could be found by LC-MS, the crude product from solid phase synthesis was very impure and difficult to purify by HPLC. The free hydroxyl group on the Hyp might have been the cause of this problem. The hydroxyl group is unlikely to couple with the tripeptide acid building block during solid phase synthesis due to the low reactivity of a secondary alcohol and the steric effect, but the final capping with Ac_2 also reacts with the free hydroxyl group. This was proven by LC-MS; in addition to the desired $MH_2^{2+}/2$ fraction, a significant amount of the $[M+45]^{2+}/2$ fraction was also found indicating one more acetyl group. Therefore, side-chain unprotected tripeptide building blocks could not yield the desired polypeptide with good purity and yield.

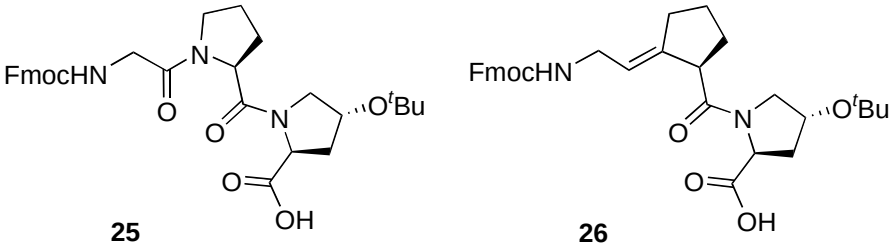
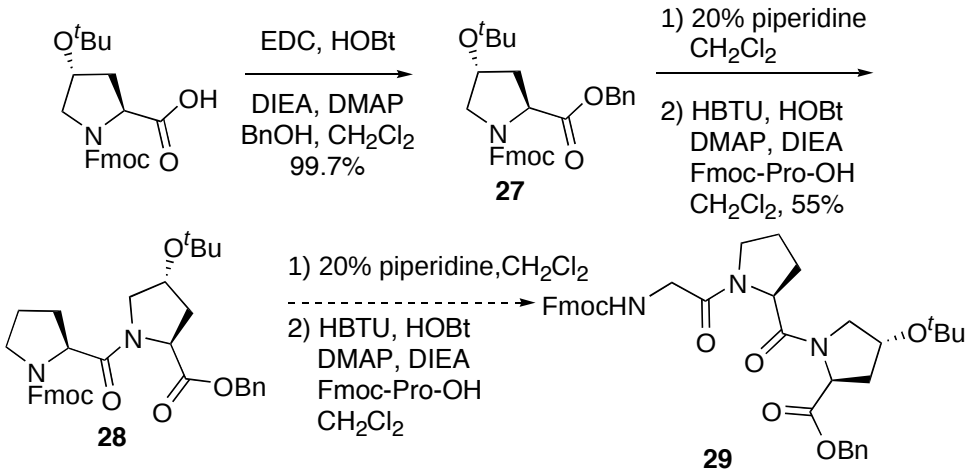


Figure 2.4 The tripeptide building block 25 and tripeptide isostere **26**.

This result meant that the hydroxyl group of Hyp had to be protected. The *tert*-butyl group was chosen to protect the hydroxyl of Hyp because it could be removed simultaneously during the cleavage of peptides from the resin by TFA (Figure 2.4). The TBS group can also be removed by TFA, but the TBS protecting group is not as stable as the *tert*-butyl protection.

The side-chain protected tripeptide building block Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH **25** was designed. There are two ways to synthesize a polypeptide: 1) from the *C*-terminus to the *N*-terminus (right to left); 2) from the *N*-terminus to the *C*-terminus (left to right). The first method is preferred because it has less of an epimerization problem.

Scheme 2.10 Synthesis of building block Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH from right to left.



The *tert*-butyl protected hydroxyproline benzyl ester could not be synthesized by Fisher esterification with BnOH and acid, because the *tert*-butyl protecting group would be cleaved during the reaction. Moreover, since the *tert*-butyl group would be cleaved by TFA, the Boc protecting group could not be used to protect the *N*-terminus in this situation. Starting from commercially available Fmoc-Hyp(^{t}Bu)-OH, benzyl ester **27** was synthesized by EDC coupling (Scheme 2.10). The Fmoc-group was cleaved by piperidine and then

coupled with Fmoc-Pro-OH to make dipeptide **28**. Then the Fmoc-group was cleaved again, and the dipeptide was coupled with Fmoc-Gly-OH to make tripeptide **29**.

Some problems occurred during the deprotection and the following coupling steps. After adding piperidine to remove the Fmoc-group, excess piperidine could not be completely removed by chasing with $CHCl_3$ and MeOH, or under high vacuum. A side product formed as shown in Figure 2.5.[238] This side-product cleaved the Fmoc protecting group during the next coupling reaction and made the product very impure. Unprotected proline derivatives were invisible under UV or in I_2, and had very poor visibility in ninhydrin and H_2SO_4 solution, and some impurities were not visible in many commonly used visualization reagents. It was very difficult to obtain a pure product of unprotected proline derivatives from flash chromatography.

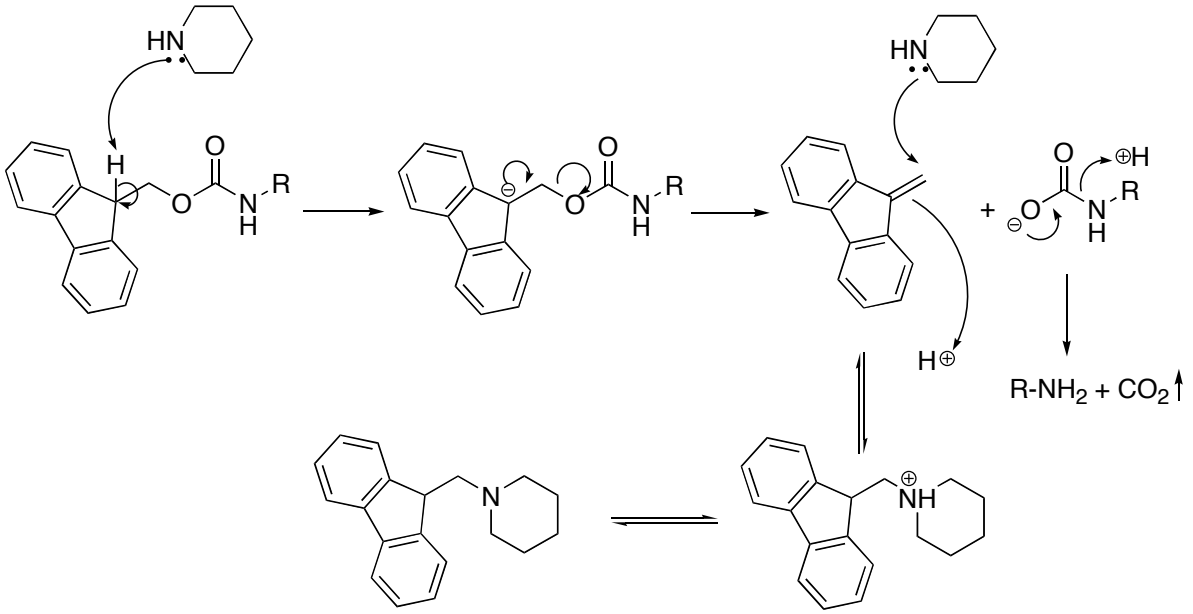


Figure 2.5 Mechanism of Fmoc deprotection by piperidine.[238]

The yield of benzyl ester **27** was 99.7 %, and the yield from the synthesis of dipeptide **28** was only about 55 %. But when an attempt was made to synthesize tripeptide **29** by

HBTU peptide coupling, no desired product was found after column purification. This method required column purification after each deprotection and coupling reaction, and it was not a good method to make the tripeptide building block in large scales.

Ottl *et al.* published a method to synthesize the tripeptide building block Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH from the *N*-terminal end to the *C*-terminal end.[239] Usually, the activated ester method causes epimerization, but in the specific Gly-Pro-Hyp sequence, it seems that epimerization is not a problem. Glycine has no chiral center, and protected proline and hydroxyproline are both secondary amines and are difficult to epimerize (Figure 2.6).[240] For prolyl amino acids, during the azlactone epimerization process, one intermediate is not stable, and difficult to form due to the property of tertiary amides. So prolyl residues are difficult to epimerize.

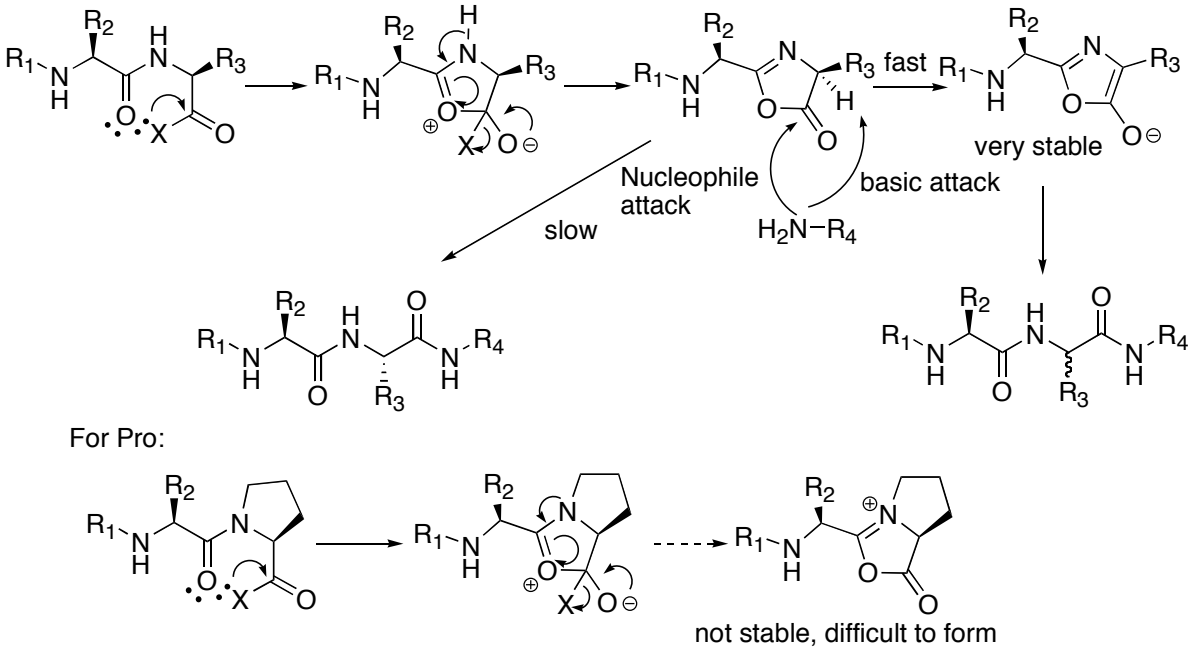


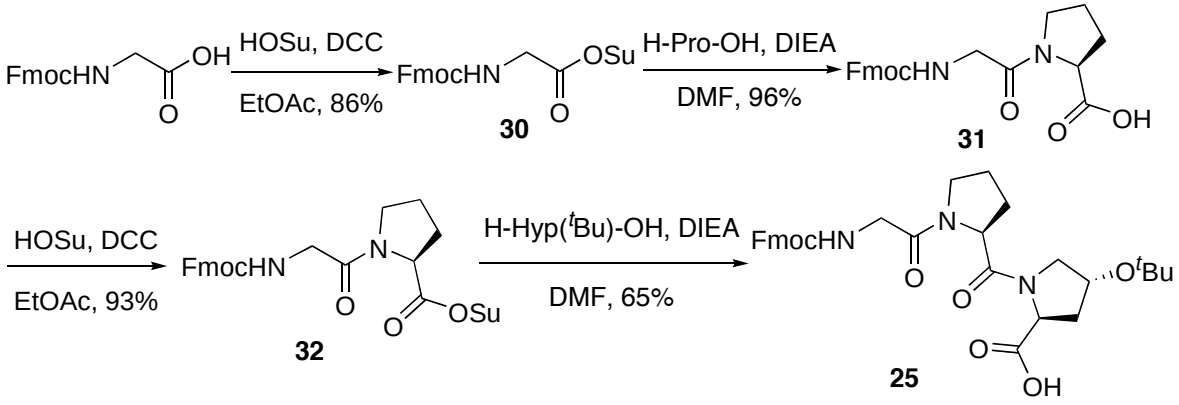
Figure 2.6 Epimerization mechanism of active esters through the azlactone.[240]

Starting from Fmoc-Gly-OH, activated ester **30** was synthesized by DCC coupling in EtOAc, and purified by precipitating from EtOAc-hexanes.[241] The reaction could also be

carried by EDC coupling in a mixture of dioxane, CH_3CN and $CHCl_3$.[242] The yield was slightly higher (93.8 % compared to 86.0 %), but considering that EDC is much more expensive than DCC, the latter coupling reagent was adopted.

Activated ester **30** and H-Pro-OH were stirred with DIEA in DMF, and dipeptide acid **31** was obtained in 96 % yield (Scheme 2.11). No column purification was needed and the acid was precipitated from EtOAc-hexanes.[239] Dipeptide activated ester **32** was obtained by DCC coupling. The desired tripeptide building block **26** was obtained by stirring activated ester **32** and H-Hyp(^{t}Bu)-OH under basic conditions. Although the literature stated that no column purification was needed, analytical HPLC showed that the product was not pure and the model SPPS generated a very impure crude polypeptide. The crude product was purified with $CHCl_3$:MeOH:AcOH in a 100:2:0.5 ratio followed by precipitation from EtOAc-hexanes. The side-chain protected tripeptide building block Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH **25** was obtained as a white solid which was easy to handle during solid phase peptide synthesis.

Scheme 2.11 Synthesis of building block Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH from left to right.

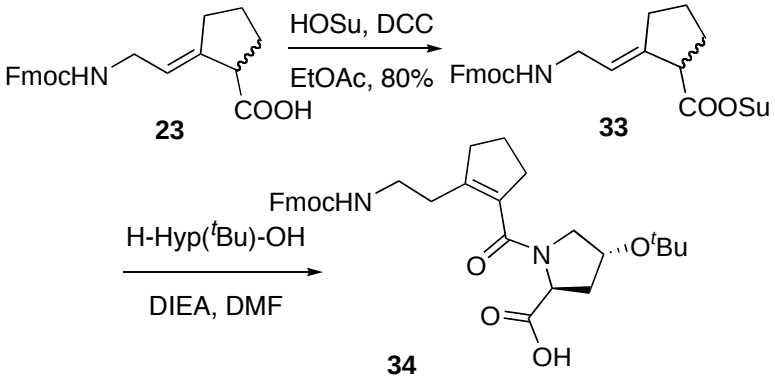


All the reactions in this method were very fast, and it was easy to observe the completion of the reactions. No column purification was required in this method, which made it ideal for large-scale synthesis.

2.1.5 Synthesis of Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OH isostere

The alkene isosteres Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OH **26** and Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp-OH **2** differ only in the last amino acid. The synthetic routes to Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23** in both syntheses were identical.

Scheme 2.12 Formation of isomerized alkene isostere **34**.



The OSu activated ester method was tried first. Starting from the Gly-Pro alkene isostere **23**, activated ester **33** was synthesized by DCC coupling (Scheme 2.12). Then H-Hyp(^{t}Bu)-OH was added and stirred with 0.5 equivalent DIEA at rt for 4 h. Unfortunately, no desired product was obtained. After characterization, the resulting product was identified as the isomerized alkene isostere **34**. The activated ester increased the acidity of the hydrogen atom at the α-position. In the presence of base, the α-proton was deprotonated, and the double bond shifted into the ring to form a more stable α,β-unsaturated structure (Figure 2.7).[224]

This result showed that in the presence of both activated ester and base, the double bond of activated ester **33** isomerized into the 5-membered ring to form a more stable α,β-unsaturated structure. Wang *et al.* reported a similar phenomenon during SPPS of Ser-Ψ[(*E*)CH=C]-Pro isostere using HBTU-HOBt as the coupling reagent and DIEA as the base.[224] Changing the coupling reagent to HATU-HOAt and base to collidine reduced the

possibility of double bond isomerization.[224] Collidine was used to replace DIEA in the OSu activated ester method, but no reaction was observed and activated ester **33** was recovered.

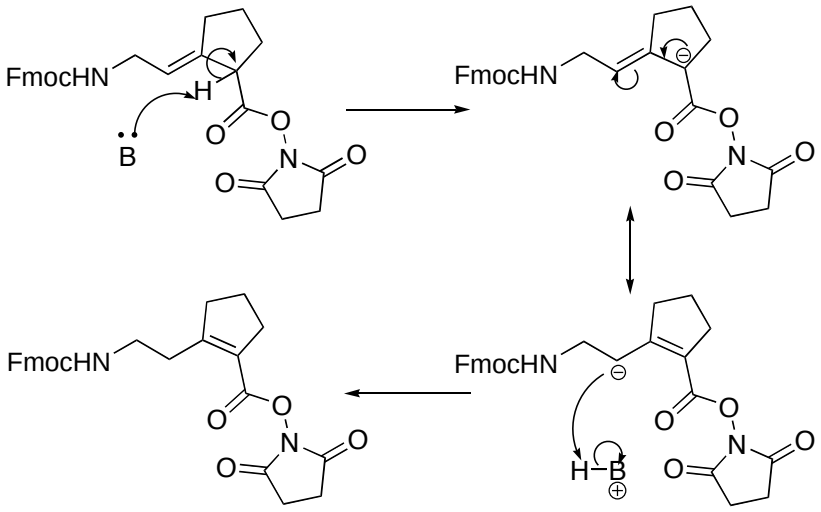
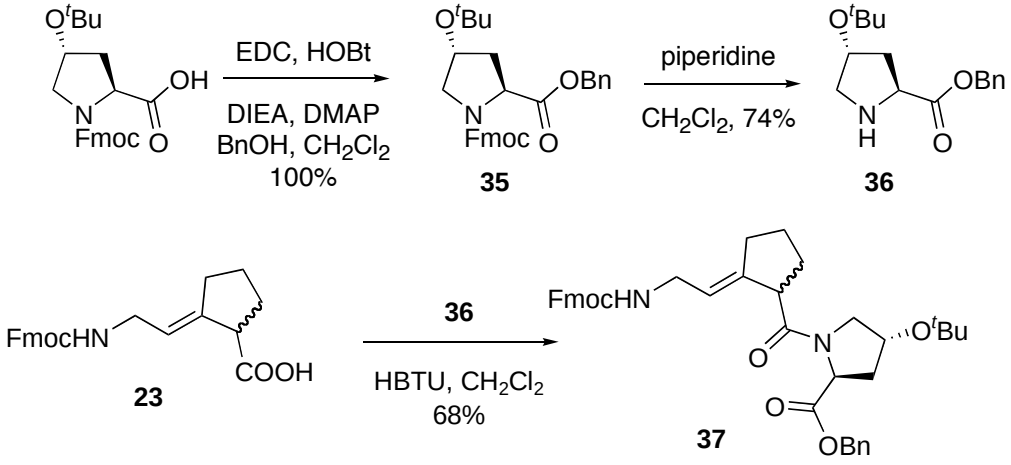


Figure 2.7 Mechanism of double bond isomerization.

A new synthetic route was designed (Scheme 2.13). Starting from commercially available Fmoc-Hyp(^{t}Bu)-OH, benzyl alcohol was coupled with EDC to protect the *C*-terminus. Then the Fmoc protecting group was cleaved with piperidine. Column chromatography was used to remove the piperidine containing side product. The resulting amine **36** was coupled to Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23** with HBTU. To prevent the double bond migration, no base or HOBt was used. The tripeptide alkene isostere **37** was synthesized as a clear oil, while ^{1}H NMR and HSQC showed that the double bond had not shifted into the ring. HATU and HOAt were also used as coupling reagents to synthesize isostere **37**, but no significant yield improvement was observed. Moreover, the resulting HOAt caused some purification problems. HATU is more expensive than HBTU, so HBTU in the absence of base was used as the standard method to synthesize the tripeptide alkene isostere **37**.

Scheme 2.13 Synthesis of tripeptide isostere **37**.



2.1.6 Introduction of the chiral center

Until then, all of the reactions involved no stereoselectivity, and diastereomers with both *R*- and *S*-configurations were synthesized. All the natural amino acids are in the L-configuration, so only the tripeptide alkene isostere with the L-Pro mimetic structure was desired. Throughout the synthesis, a chiral center could be introduced at any one of four stages (Figure 2.8).

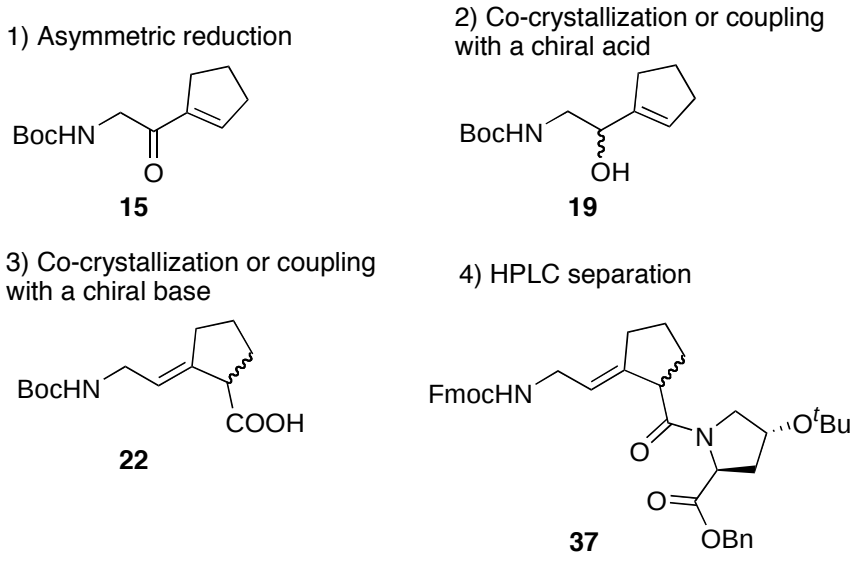
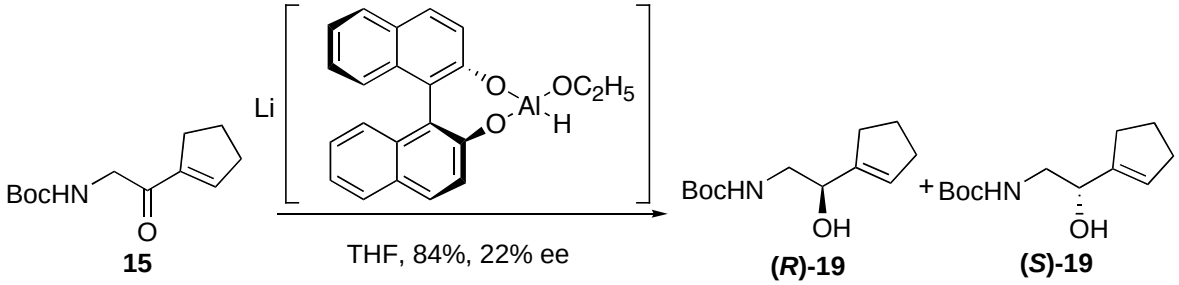


Figure 2.8 Four feasible stages to introduce a chiral center.

Asymmetric reduction

The first feasible opportunity to introduce a chiral center is through asymmetric reduction of α,β-unsaturated ketone **15** to form a chiral alcohol. Glycine is the only natural amino acid without a chiral center, which means that substrate control strategies[243] will not work, and only reagent control strategies need to be considered. Vabeno *et al.* showed that $LiAlH(O^tBu)_3$-EtOH or NB-Enantride-THF systems at –78 °C provided very good selectivity, but this result was only obtained in diastereoselective reduction systems under Felkin-Ahn control.[244] Corey *et al.* reported that the oxazaborolidines were effective catalysts in asymmetric reduction with relatively high yields.[245,246] These catalysts are not commercially available, but they can be easily synthesized by addition of phenyl Grignard to an amino acid. They only produce good stereoselectivity when the carbonyl is attached to an aryl group.

Scheme 2.14 Asymmetric reduction of α,β-unsaturated ketone **15**.

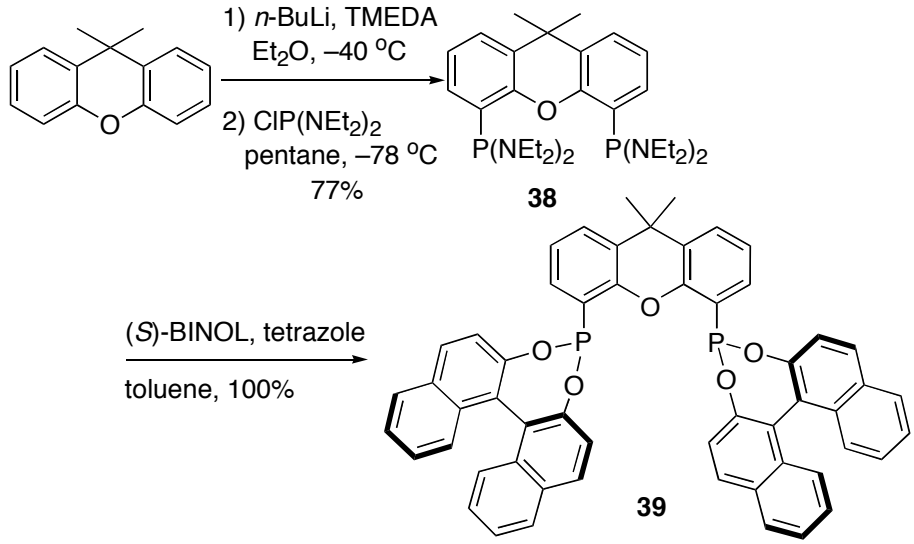


Noyori *et al.* reported that stoichiometric amount of (*S*)-(–)-binaphthol-modified $LiAlH_4$ yielded only one major enantiomer.[247,248] Previous studies in our group showed that the reduction worked with a 53 % yield, but stereoselectivity was not obtained.[249] A modified Noyori asymmetric reduction method was performed at –100 °C in THF (Scheme 2.14).[250] The catalyst (*S*)-binaphthol has similar polarity to alcohol **19**, which caused difficulty during

purification. Chiral staionary phase HPLC showed that both enantiomers were obtained, and the enantiomeric excess (e.e.) value was only 22 %.

Reetz and Li reported an asymmetric transfer hydrogenation reaction catalyzed by a Ru-catalyst in the presence of BINOL-derived diphosphonite ligands, which provided high ee % values for the asymmetric reduction of alkyl/alkyl ketones.[251] The diphosphonite ligand **39** was not commercially available, but it could be synthesized through a three-step synthesis (Scheme 2.15).[252,253]

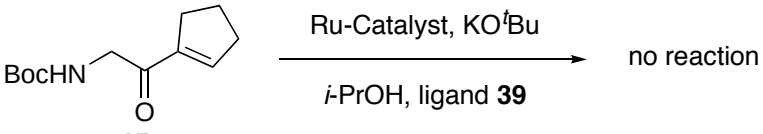
Scheme 2.15 Synthesis of the diphosphonite ligand **39**.



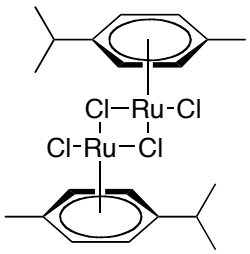
Starting from commercially available 9,9-dimethylxanthene, 4,5-(bis[bis-diethyl-amino]phosphonito)-9,9-dimethylxanthene **38** was obtained by a one-pot two-step synthesis as a yellow solid. Compound **38** was then heated at reflux with (*S*)-BINOL in toluene for 2 days and ligand **39** was obtained.[252] The (*S*,*S*)-ligand was synthesized from (*S*)-BINOL because it had been reported that the (*S*,*S*)-ligand **39** yielded alcohol products with the desired configuration.[251] The ligand with the (*R*,*R*)-configuration could also be synthesized by the same method from (*R*)-BINOL.[252]

Asymmetric transfer hydrogenation was attempted (Scheme 2.16). The catalyst $[RuCl_2(p\text{-cymene})]_2$ and the (*S*,*S*)-ligand **39** were dissolved in anhydrous *i*-PrOH and heated. KO^{t}Bu base was added, followed by α,β-unsaturated ketone **15**. Unfortunately, no reaction was observed after 7 days, and nearly all of the starting material was recovered. The reaction still did not work after changing the amount of catalyst and ligand, or after raising the reaction temperature to 60 °C. The transfer hydrogenation yielded poor results in two classes of ketones: 1) α,β-unsaturated ketones which are slowly converted, but with poor ee %, and 2) saturated ketones containing a heteroatom at the α- or β-position (Prof. Manfred T. Reetz, personal communication). Ketone **15** has both a heteroatom at the β-position and the α,β-unsaturated structure, so it is understandable that the reaction did not work. To date, we have not found an asymmetric reduction method that works for α,β-unsaturated ketone **15**.

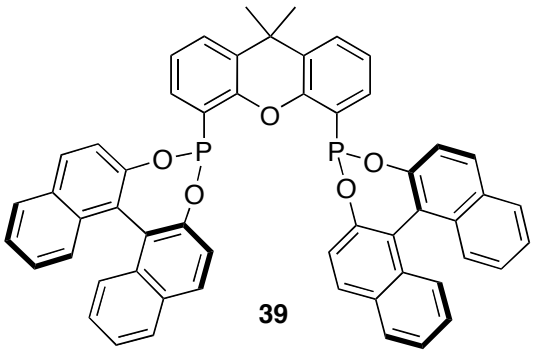
Scheme 2.16 Asymmetric transfer hydrogenation of α,β-unsaturated ketone **15**.



Ru-Catalyst:



Ligand:



Resolution strategies

The second feasible stage for introducing a chiral center was the coupling of a chiral acid to alcohol **19**. There are some requirements for a chiral acid to be a good candidate in this reaction: 1) it must have at least one chiral center; 2) it must be easy to couple, and easy

to cleave; 3) the resulting diastereomers must be separable by flash chromatography because the scale of the reaction would be very large at this step; 4) it must not be too expensive.

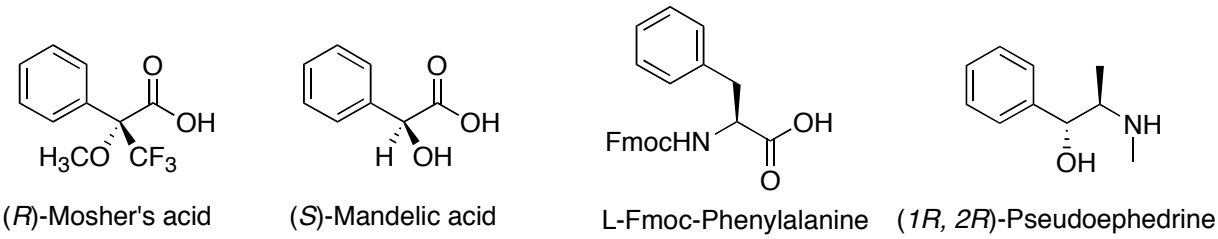
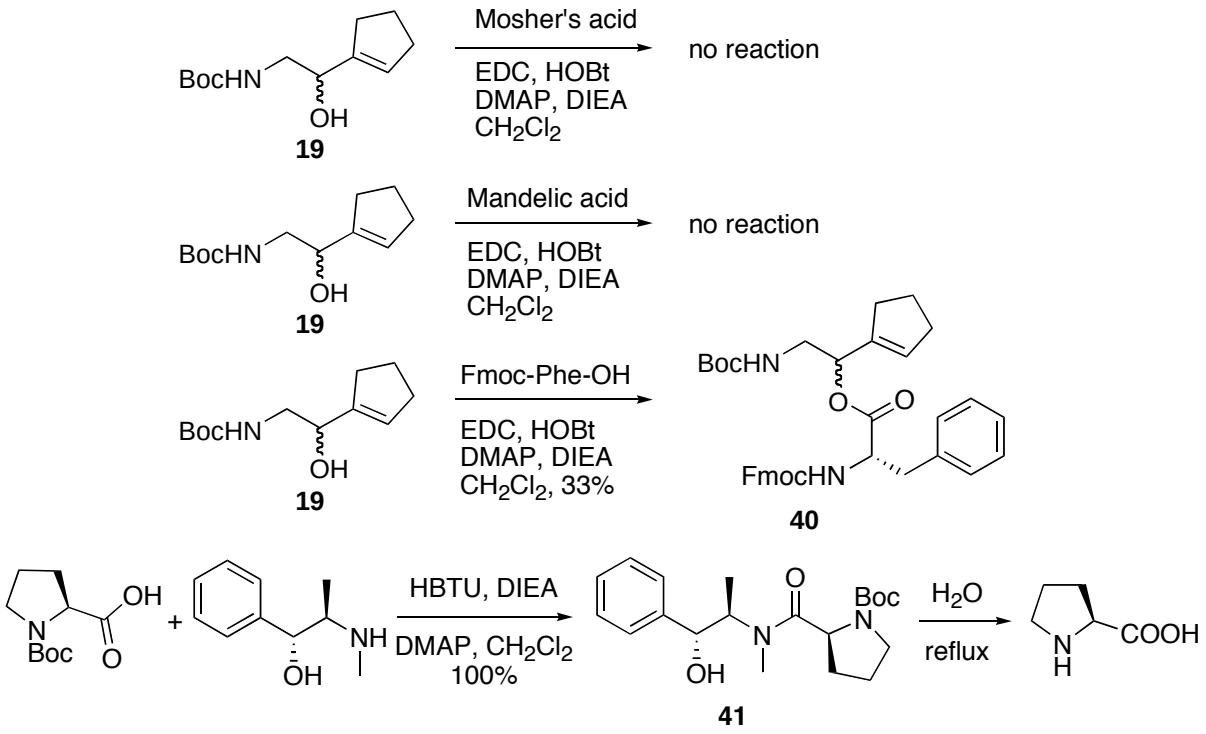


Figure 2.9 Reagents used to make diastereomers from enantiomers.

The best candidate was Mosher's acid, but it is very expensive and not good for large-scale reactions (Figure 2.9). Mosher's acid was coupled first using EDC. No reaction was observed and nearly all of the starting material was recovered (Scheme 2.17). Mandelic acid is much cheaper, but it has a secondary hydroxyl group that can slow the coupling reaction with alcohol **19**, which is also a secondary alcohol. The coupling of (*S*)-mandelic acid with alcohol **19** was performed with HBTU or EDC as the coupling reagent, but still no reaction was observed. An attempt to make (*S*)-Mandelic acid into an acid chloride to increase its reactivity also failed. The reaction between the secondary alcohol **19** and the bulky Mosher's acid or mandelic acid was not favored due to steric hindrance.

Most of the natural amino acids have a chiral center at the C^{α}-position, so they are candidates for this reaction. Fmoc-Phe-OH was used as a chiral acid to couple with alcohol **19** using HBTU (Scheme 2.17). The yield was low (about 33 %). Moreover, the resulting diastereomers showed only one spot on TLC and no separation at all on a column.

Scheme 2.17 Attempts to make diastereomers from the racemic mixture.



Two carboxylic acid intermediates were feasible candidates to make diastereomers from the enantiomers – Boc-protected acid **22** and Fmoc-protected acid **23**. A chiral amine or a chiral alcohol could be introduced to make a pair of diastereomers and the diastereomers could then be separated by normal phase HPLC or chiral HPLC because the scale of the reaction is not very large at this stage. A model reaction with Boc-Pro-OH and (1*R*,2*R*)-pseudoephedrine showed that it was very difficult to cleave the chiral base without affecting the Boc-protecting group. Moreover, 4(*R*)-Hyp had two chiral centers and tripeptide **37** was a pair of diastereomers, so column or HPLC separation of the diastereomers would be performed when tripeptide **37** was synthesized.

Co-crystallization with a chiral base

It would have been feasible to separate the two enantiomers of Boc-protected acid **22** or Fmoc-protected acid **23** by co-crystallization with a chiral base to form diastereomeric salts, which might be separable due to their different physical properties.[254] There are many chiral bases that could be used to make the complex with the acid enantiomers, such as pseudoephedrine and brucine. Based on the availability in our lab, pseudoephedrine and brucine were used as chiral bases to make the chiral crystal complex. Boc-protected acid **22** was used as the source of the acid enantiomers because the Fmoc-protected species often caused solubility issues in many solvents during the crystallization process, and precipitated instead of crystallizing. In most solvent systems, either an oil or a precipitate was obtained (Table 2.1). However, several small pieces of irregular-shaped colorless crystal-like solid were obtained when using MeOH and H_2O as the solvent. Crystals also appeared at 4 °C when MeOH was less than 25 %, but the recovery was always very low (less than 5 %) in each solvent system. The crystals obtained were too thin for X-ray crystal structure determination, and larger crystals were not obtained.

Table 2.1 Co-crystallization results of Boc-protected acid **22** and brucine.

Solvent	Temperature	Result
MeOH	rt	No Precipitate
Acetone	rt	Oil
$CHCl_3$	rt	Oil
MeOH-Ether	rt	Oil
EtOAc-Hexanes	rt	Precipitate
EtOAc-Hexanes	4 °C	Precipitate
10 % MeOH-H_2O	rt	Precipitate
10 % MeOH-H_2O	4 °C	Crystal
25 % MeOH-H_2O	4 °C	Crystal
40 % MeOH-H_2O	4 °C	No Precipitate

The co-crystallization method worked, but there were still some problems affecting the application of this method to this project. 1) The percentage recovery was too low. Only a

couple of milligrams of crystals were obtained from 50 mg of the enantiomers. This method would require a large amount of Boc-protected acid **22**. 2) The crystallization process was very slow, and modification of the conditions was even slower. It was very difficult to predict which solvent system would work. 3) There was no way to predict which enantiomer would crystallize. There was a 50 % chance that the crystal would not be the desired enantiomer. 4) Even after obtaining crystals, it would still be very difficult to determine the absolute stereochemistry because the X-ray structure only provided the relative stereochemistry, and the hydrogen atoms could not be seen by ordinary X-ray crystallography. So the co-crystallization method was abandoned.

HPLC separation of the diastereomers

The last feasible stage to introduce the chiral center was via tripeptide isostere benzyl ester **37** or tripeptide isostere **26**. After coupling with hydroxyproline, the resulting tripeptide isostere was a pair of diastereomers instead of enantiomers. Although the resulting diastereomers still could not be separated by flash chromatography, they were separable by HPLC or chiral HPLC. I chose tripeptide isostere benzyl ester **37** for HPLC separation because 1) the normal phase column usually had a greater number of plates than the reverse phase column of the same length; 2) less polar diastereomers were usually easier to separate than more polar diastereomers.

Tripeptide isostere benzyl ester **37** was purified by flash chromatography with 25 % EtOAc in hexanes to remove highly polar impurities first. An analytical method was developed on a 5 μm normal phase silica column. The two diastereomers were separable with an *i*-PrOH-hexanes solvent system. The isocratic condition with 4 % *i*-PrOH in hexanes provided the best resolution. Although 5 % *i*-PrOH in hexanes also provided good resolution,

the two peaks were too close, and this could have caused difficulty during preparative separations. The system with 3 % *i*-PrOH in hexanes caused some solubility and peak tailing problems. EtOAc was added to solve the solubility problem, but this caused reduced resolution.

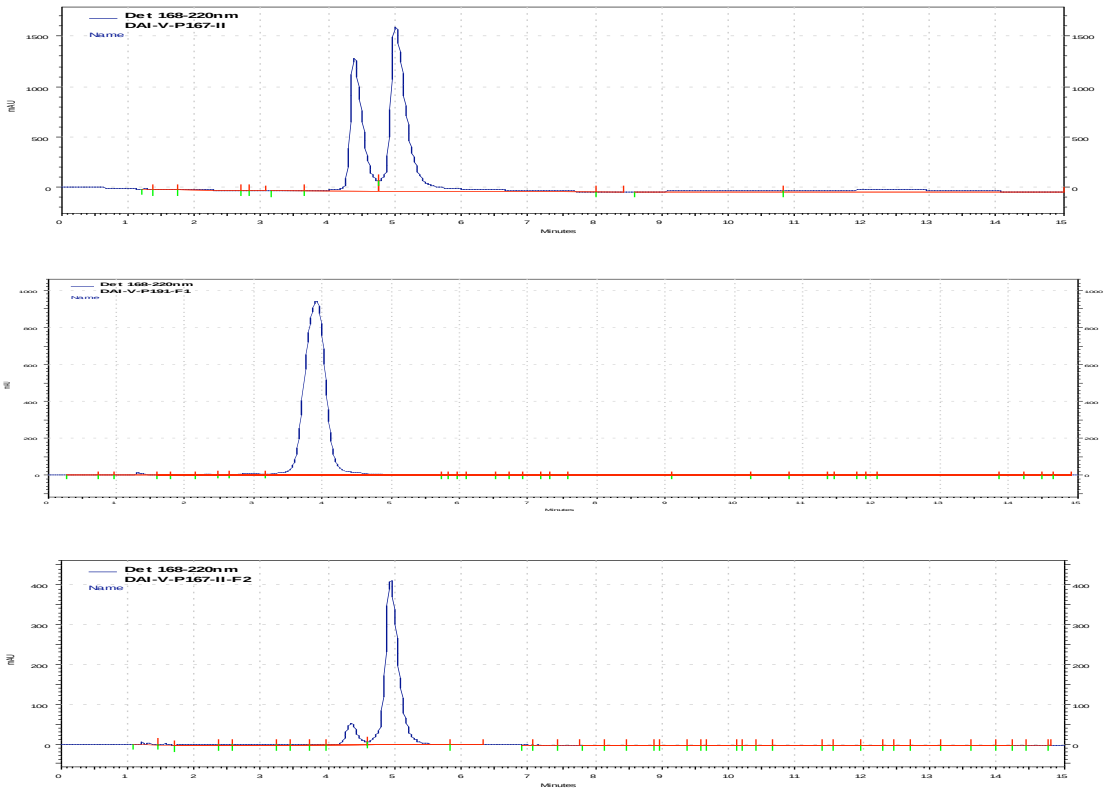


Figure 2.10 HPLC separation of tripeptide isostere benzyl ester **37** by normal phase HPLC. Top: mixture of two diastereomers before separation; middle: the first diastereomer after separation; bottom: the second diastereomer. The analytical chromatograms were obtained on a Beckman P/ACE with a Varian MetaSil 5 μm Si column with 5 % *i*-PrOH in hexanes under isocratic conditions.

The two diastereomers of tripeptide isostere benzyl ester **37** were separated on a 10 μm normal phase preparative column. The resolution was not as good as on the analytical column. The particle size difference between the analytical column and the preparative column might have been the reason for the resolution difference, because small particle sizes usually gave better separation.

After separation, the first fraction was very pure, while the second fraction still contained a little of the first diastereomer, even though the fraction was collected at the maximum of the peak (Figure 2.10). The solubility problem relating to tripeptide isostere benzyl ester **37** in the *i*-PrOH-hexanes system was responsible for the small amount of impurity. A new solvent system that could dissolve tripeptide isostere benzyl ester **37** better without affecting the separation might help to achieve better resolution. The ratio of the two diastereomers in the second fraction was about 1 to 22, as determined by HPLC and NMR. The peptide purity was considered pure enough for peptide synthesis. Although the absolute stereochemistry of the two diastereomers was still unknown, it could be determined after incorporation into the peptide sequence. Both diastereomers were incorporated in the next step of the synthesis.

2.1.7 Deprotection of the benzyl group

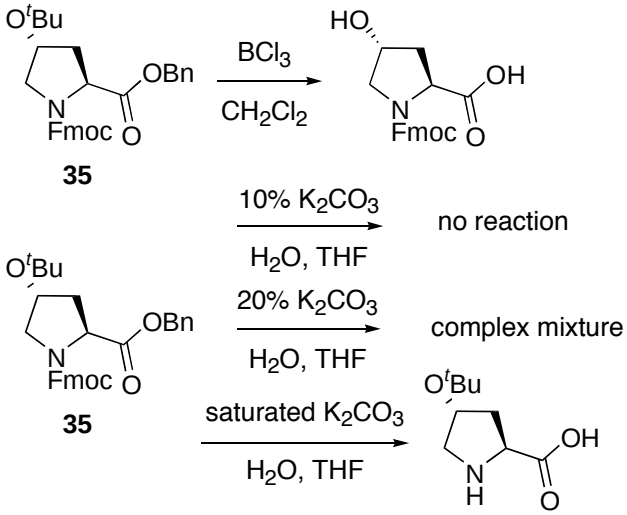
The last synthetic step was the removal of the benzyl protecting group. The most common method to remove the benzyl protection was hydrogenation. But high-pressure hydrogenation conditions required would also cause the Fmoc protecting group to be removed. Raines and co-workers reported a hydrogenation reaction in 50:50 AcOH:MeOH at 1 atm, and the Fmoc-group was not affected.[152] This method worked when hydrogenating the tripeptide benzyl ester Fmoc-Gly-Pro-Hyp-OBn **11** to give the tripeptide acid Fmoc-Gly-Pro-Hyp-OH **1**. But when hydrogenating tripeptide isostere benzyl ester **37**, the carbon-carbon double bond was also partially reduced, as confirmed by NMR and LC-MS. A new deprotection method was needed to remove the benzyl group without affecting either the double bond or the Fmoc-protection.

In tripeptide isostere benzyl ester **37**, there were four functional groups that could react during the deprotection reaction. The Fmoc-group was not stable in strong base; the *tert*-butyl group was not stable in strong acid; the double bond and benzyl groups were not stable during hydrogenation, but the double bond was less reactive than the benzyl protecting group. Considering the reactivity of all four functional groups, three methods might work for tripeptide isostere benzyl ester **37** – weak acid, weak base or less reactive hydrogenation conditions.

Hydrolysis with weak acids or bases

It was reported that the weak Lewis acids, BCl_3 and $AlCl_3$, can be used to cleave the benzyl protecting groups.[255,256] Because solid $AlCl_3$ was moisture sensitive and hard to handle, 1.0 M BCl_3 in CH_2Cl_2 was used in a model reaction first. The benzyl group was cleaved within 2.5 h, while the *tert*-butyl group was also cleaved (Scheme 2.18). This method could not be used to deprotect the benzyl group in tripeptide isostere benzyl ester **37**.

Scheme 2.18 Deprotection of the benzyl group with acids or bases.



Huffman reported a mild hydrolysis condition using a K_2CO_3 solution to deprotect the benzyl group.[257] K_2CO_3 solution (10 % aq.) had been used in our group to open a 7-

membered lactone ring without cleaving the Fmoc-group (Song Zhao, unpublished results). Hydrolysis with weak bases was attempted in a model reaction. Results showed that no reaction was observed in 10 % K_2CO_3 aqueous solution, and starting material was recovered (Scheme 2.18). Using saturated K_2CO_3 solution, both the benzyl and the Fmoc-protecting group were cleaved. When using a 20 % K_2CO_3 solution, the product was complex. Both protected and unprotected compounds with or without the benzyl group or the Fmoc group were found in the crude product mixture. From the TLC, it seemed that the Fmoc-protecting group was cleaved at a slightly faster rate than the benzyl protecting group.

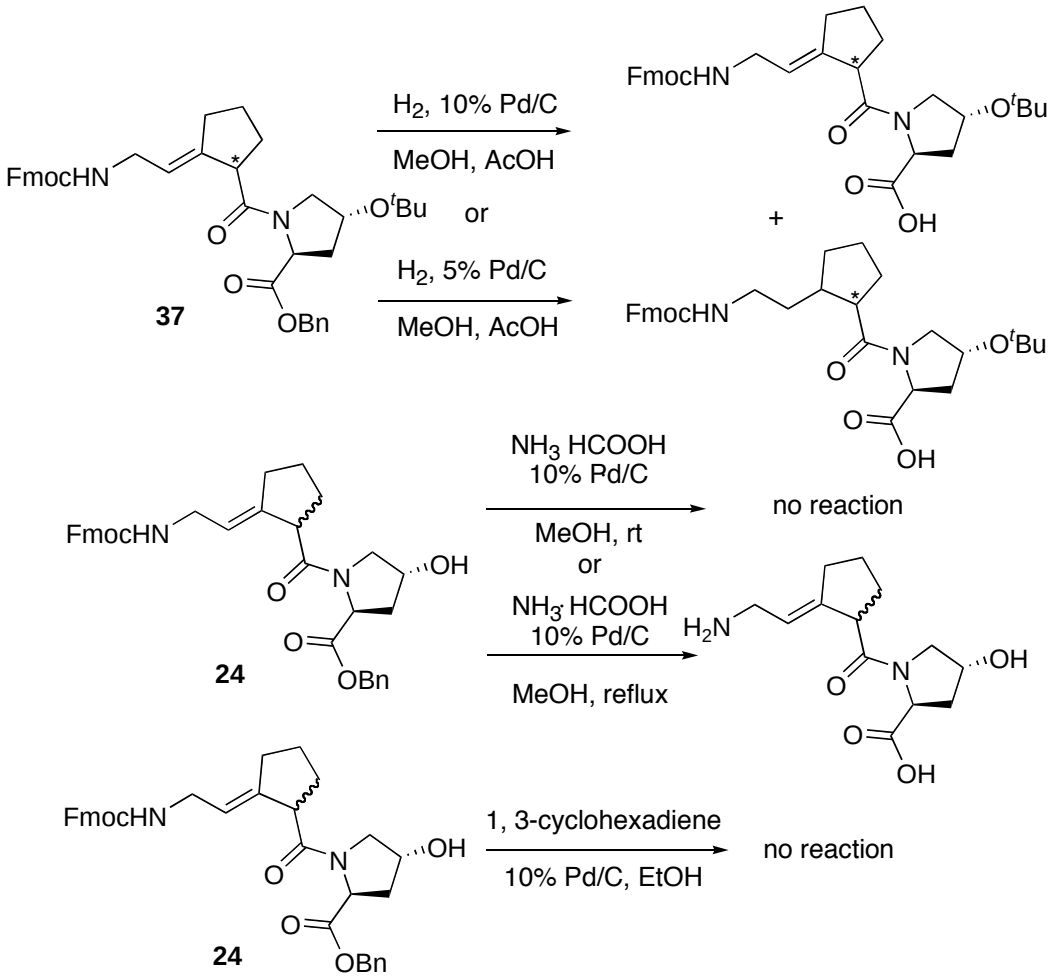
Mild hydrogenation

Mild hydrogenation conditions can be achieved by two methods: a less reactive catalyst or a milder hydrogen source. During the synthesis of Ser-Ψ[(*Z*)CH=C]-Pro isostere, Na/NH_3 was used to cleave the benzyl protecting group without affecting the alkene bond.[234] But the highly basic condition would cleave the Fmoc-protecting group. The less reactive catalyst $Pd(OH)_2/C$ was also tried to remove the benzyl group, but it also cleaved the Fmoc group (Dr. Bailing Xu, Personal communication). The hydrogenation reaction at atmosphere pressure with 10 % Pd/C did not affect the Fmoc-protection,[152] but it partially reduced the alkene (Scheme 2.19). With 5 % Pd/C was used as the catalyst, the reaction was finished within 65 min but the alkene bond was still partially reduced.

Ammonium formate has been used as a mild hydrogen source to replace hydrogen in many hydrogenation reactions.[258-261] A model reaction using ammonium formate as the hydrogen source on 10 % Pd/C was performed in MeOH (Scheme 2.19). No reaction was observed at rt, when the reaction mixture was heated at reflux, both the Fmoc and benzyl

protecting groups were removed. At high temperatures, ammonium formate would release ammonia, which cleaved the Fmoc group.

Scheme 2.19 Mild hydrogenation reactions.



1,4-Cyclohexadiene was another reagent used as a mild hydrogen source.[262] The hydrogenation reaction was performed with 10 % Pd/C as the catalyst in anhydrous EtOH. 1,3-Cyclohexadiene was tried first as the hydrogen source because it was available in our lab. Unfortunately, the reduction did not work at either rt or 50-60 °C (Scheme 2.19). 1,4-Cyclohexadiene was then tried. There was still no reaction at rt, but after heating the reaction mixture to 90-95 °C, the reaction was completed within 45 min (Scheme 2.20). LC-MS, ^{1}H NMR and HSQC showed that the desired tripeptide isostere **26** was obtained as the only

product in high yield (over 90 %). The hydrogenation conditions attempted are listed in Table 2.2.

Scheme 2.20 Synthesis of tripeptide isostere **26**.

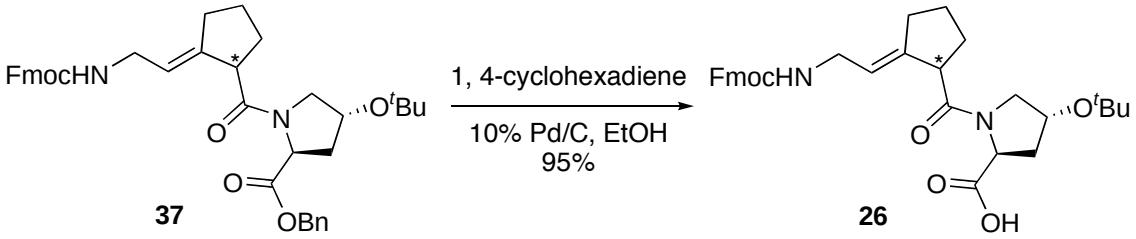


Table 2.2 Methods to remove the benzyl protecting group of a carboxylic acid. [a]

Method	H-source	Catalyst	Temp.	Fmoc	Bn	*t*-Bu	C=C
BCl_3	n.a.	n.a.	rt	–	+	+	n.a.
10 % K_2CO_3	n.a.	n.a.	rt	–	–	–	n.a.
20 % K_2CO_3	n.a.	n.a.	rt	+	+	–	n.a.
Satur. K_2CO_3	n.a.	n.a.	rt	+	+	–	n.a.
Mild Hydrogenation	NH_3·HCOOH	10 % Pd/C	rt	–	–	–	–
	NH_3·HCOOH	10 % Pd/C	80 °C	+	+	–	n.d.
	H_2	10 % Pd/C	rt	–	+	–	+
	H_2	5 % Pd/C	rt	–	+	–	+
	1,3-cyclohexadiene	10 % Pd/C	rt	–	–	–	–
	1,3-cyclohexadiene	10 % Pd/C	60 °C	–	–	–	–
	1,4-cyclohexadiene	10 % Pd/C	rt	–	–	–	–
	1,4-cyclohexadiene	10 % Pd/C	90 °C	–	+	–	–

a) In which "+" means reacted, "–" means no reaction, "n.a." means the reagent is not used or the substrate does no contain the functional group, "n.d." means not determined

Despite our difficulties, the benzyl group is a good protecting group for carboxylic acids. It can be removed in both acidic and basic conditions, and can also be removed by mild hydrogenation without affecting the C=C double bond and Cbz or Fmoc protection. Benzyl protection is very useful in peptide synthesis and organic synthesis.

2.2 Design and Synthesis of Pro-Ψ[(E)CH=C]-Pro isostere

2.2.1 Design of Pro-Ψ[(*E*)CH=C]-Pro isostere

The direct backbone hydrogen bonding pattern of collagen is between the N-H of the Gly residues and the C=O of the Xaa residues of another chain.[6] Each Gly N-H···O=C Xaa hydrogen bond is estimated to contribute about –2.0 kcal/mol to the conformational stability of the triple helical structure.[223,263] It is one of the major forces that stabilize the collagen triple helix.

Jenkins *et al.* used an (*E*)-alkene bond and an ester bond to replace the amide bond between the Yaa and Gly residues in a host-guest system.[223] They found that these replacements caused a significant decrease in the stability of the collagen triple helix. The peptide containing the (*E*)-alkene isostere had the lowest stability.[223] Many factors might be responsible for the destabilizing effect of the (*E*)-alkene isostere. 1) The dihedral angles of the (*E*)-alkene isostere were different from the Φ angle of the Gly residue (–72°) and the Ψ angle of the Yaa residue (152° when Yaa is Pro) in the collagen triple helix.[57] 2) The alkene bond has a reduced dipole moment with respect to the amide bond, which is not favorable for the PPII helix formation.[223] 3) The carbonyl group at the Yaa position is an important hydrogen bond acceptor for water mediated hydrogen bonding, and replacing it with the alkene bond will affect the triple helix solvation.[105] 4) It is possible to form an alkene C-H···O=C hydrogen bond, but this hydrogen bonding pattern is much weaker than the N-H···O=C hydrogen bond.[104,223]

The dipeptide isostere Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH **43** was designed to study how much *cis-trans* isomerization contributes to the stability of the collagen triple helix (Figure 2.11). In this isostere, an (*E*)-alkene bond replaced the amide bond between the Pro residue at

the Xaa position and the Pro residue at the Yaa position, which meant one interchain backbone hydrogen bond between Gly N-H and Xaa C=O was lost. The desired Pro-Pro isostere has the L-stereochemistry to represent the all-L natural amino acids. The Gly residue was coupled during solid phase synthesis to reduce the difficulty of organic synthesis. To minimize problems during solid phase peptide synthesis, the tripeptide building block Fmoc-Gly-Pro-Pro-OH **42** consisting of only natural amino acids was synthesized. This Pro-Pro alkene isostere eliminated *cis-trans* isomerization by locking one Pro amide bond in the *trans*-conformation. By comparison with the Pro-Gly alkene isostere, which also lost one interchain backbone hydrogen bond, but did not affect prolyl *cis-trans* isomerization, we could estimate how much *cis-trans* isomerization contributes to the stability of the collagen triple helix.

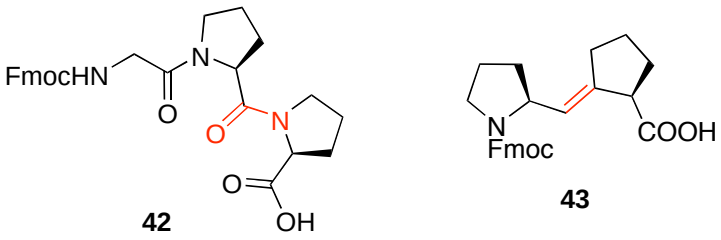
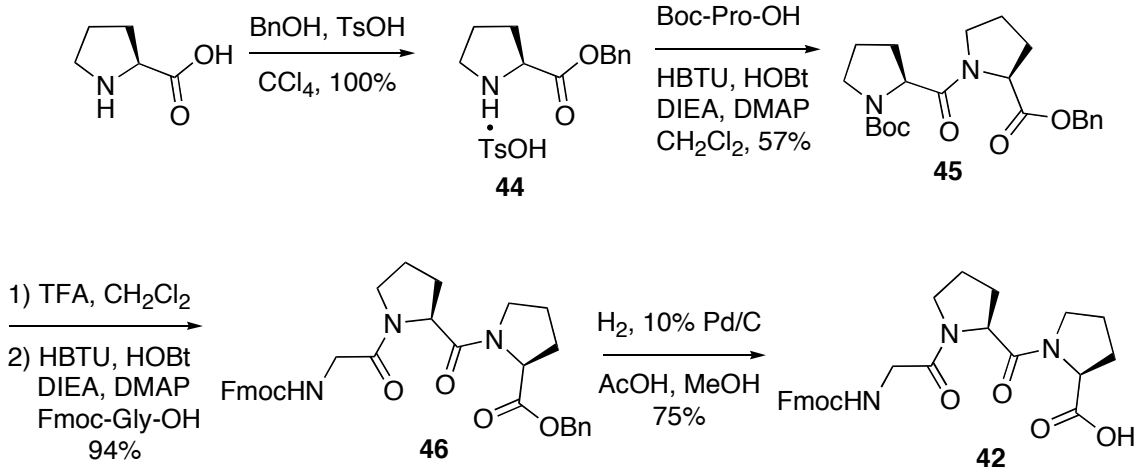


Figure 2.11 Designed dipeptide isostere Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH **43**.

2.2.2 Synthesis of the tripeptide building block Fmoc-Gly-Pro-Pro-OH

Compared to the Hyp containing tripeptide building blocks, tripeptide Fmoc-Gly-Pro-Pro-OH **42** had no hydroxyl side-chain, and it was much easier to synthesize. Either methyl or benzyl could be used to protect the *C*-terminus. Benzyl protection was chosen because in the final deprotection step, the hydrogenation with H_2 on Pd/C was much easier to perform than the $AlCl_3$/DMA deprotection of the methyl group, and the hydrogenation required less reaction time and provided a higher yield.

Scheme 2.21 Synthesis of the tripeptide building block Fmoc-Gly-Pro-Pro-OH **42**.



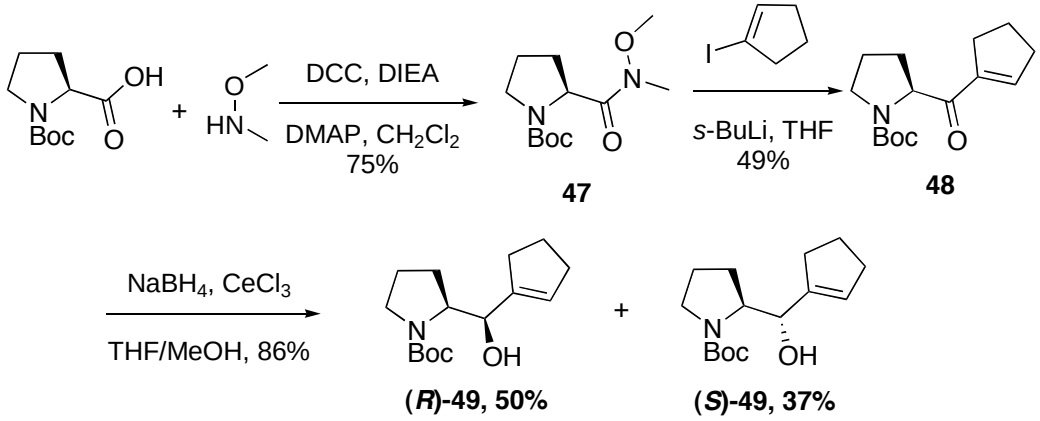
Starting from L-proline, proline benzyl ester **44** was synthesized by Fisher esterification (Scheme 2.21). Unlike hydroxyproline benzyl ester **9**, proline benzyl ester **44** was an oil, and it was very difficult to precipitate from MeOH-Et_2O or EtOAc-hexanes solutions. Excess BnOH was difficult to remove completely. Washing the resulting oil with a lot of Et_2O helped to remove most of the BnOH, but there was still a small amount in the product. Proline benzyl ester **44** was carried to the next step without flash chromatography purification, and the yield from the peptide coupling reaction was lower than for the corresponding reaction with hydroxyproline benzyl ester **9**. After column purification, NMR showed that both the desired dipeptide Boc-Pro-Pro-OBn **45**, and the side product Boc-Pro-OBn were obtained. Dipeptide Boc-Pro-Pro-OBn **45** was the major product. A more pure starting material would improve the yield of this step, and a new solvent system to crystallize proline benzyl ester **44** was necessary. The Boc-protecting group was removed with 25 % TFA in CH_2Cl_2, and tripeptide Fmoc-Gly-Pro-Pro-OBn **46** was obtained by peptide coupling with HBTU as the coupling reagent. The benzyl group was cleaved by hydrogenation with H_2 on 10 % Pd/C, and the crude product was purified by flash chromatography. The

tripeptide building block Fmoc-Gly-Pro-Pro-OH **42** was obtained as a white solid ready for solid phase peptide synthesis.

2.2.3 Synthesis of the dipeptide isostere Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH

The method for the synthesis of the Ser-Ψ[(*E*)CH=C]-Pro isostere has already been reported,[234] and the synthesis of the alkene isostere Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23** was described in Section 2.1.3. The dipeptide isostere Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH **43** had a similar structure to these compounds, so the synthesis method was similar. The synthesis of the Pro-Pro alkene isostere has been reported in the literature.[264]

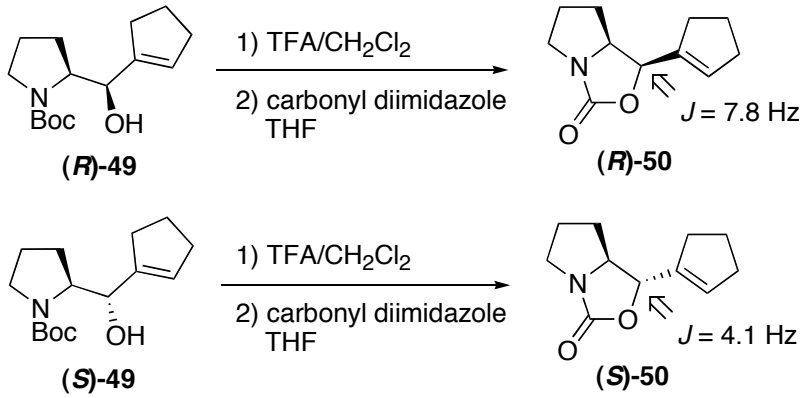
Scheme 2.22 Synthesis of alcohols **(*R*)-49** and **(*S*)-49**.



Starting from commercially available Boc-Pro-OH, the Boc-Pro Weinreb amide **47** was synthesized by DCC coupling with *N,O*-dimethylhydroxylamine (Scheme 2.22). Unlike the Boc-Gly Weinreb amide **14**, the Boc-Pro Weinreb amide **47** did not form crystals in EtOAc-hexanes or MeOH-Et_2O solvent systems. Flash chromatography was required to obtain a pure product. The Boc-Pro Weinreb amide **47** was reacted with cyclopentenyllithium, which was synthesized by mixing iodocyclopentene **13** and *sec*-butyl lithium at –40 °C, and α,β-unsaturated ketone **48** was synthesized. Because the Boc-proline derivative had no N-H proton, the Grignard reagent *i*-PrMgCl was not required in this

reaction. In fact the opposite was observed, and it significantly decreased the yield. The yield of each step was lower than the corresponding step of the Gly-Pro isostere synthesis, because Pro was less reactive than Gly due to steric hindrance. α,β-Unsaturated ketone **48** was reduced by $NaBH_4 \cdot CeCl_3$, and alcohol **49** was obtained. The proline derivative had a chiral α-carbon, and this reaction was stereoselective. The two diastereomers could be separated by flash chromatography, and two stereochemically pure compounds were obtained.

Scheme 2.23 Determination of the absolute stereochemistry of the alcohol **(*R*)-49** and **(*S*)-49**.



The absolute stereochemistry was determined by Mosher's method reported by Williams *et al.*[265] The Boc-protecting group in both diastereomers was cleaved with 25 % TFA in CH_2Cl_2 (Scheme 2.23). The resulting oil was dissolved in anhydrous THF and stirred with *N*,*N*-carbonyl diimidazole. Cyclic carbamates **(*R*)-50** and **(*S*)-50** were obtained, respectively. The coupling constant (*J*) between the C-H proton of the α-carbon and C-H proton adjacent to the hydroxyl group was used to determine the absolute stereochemistry of the two diastereomers **(*R*)-50** and **(*S*)-50**.[234,265,266] The diastereomer with the *syn*-configuration had a larger *J* value, and the one with the *anti*-structure had a smaller *J* value, because the dihedral angle of the two related protons in the *anti*-structure was closer to 90° than the dihedral angle in the *syn*-structure. For compounds **(*R*)-50** and **(*S*)-50**, compound

(*R*)-50 had the *syn*-configuration with a *J* value of 7.8 Hz, while compound **(*S*)-50** had the *anti*-configuration with a *J* value of 4.1 Hz.

The determination of the stereochemistry of the two diastereomers showed that alcohol **(*R*)-49** with the desired stereochemistry was present in a 3:2 ratio with alcohol **(*S*)-49**, which had the undesired stereochemistry. The stereoselectivity of the $NaBH_4 \cdot CeCl_3$ reduction for the Pro-Pro isostere was not as good as that for the Ser-Pro isostere, which had 85:15 selectivity with the desired stereoisomer being dominant.[249] The $NaBH_4 \cdot CeCl_3$ reduction was anti-Felkin-Ahn controlled and the cerium atom acted as a chelating reagent during the reaction.[267-270] For the Ser-Pro isostere, the reaction proceeded through anti-Felkin-Ahn control and the *syn*-product was the major product (Figure 2.12). For the Pro-Pro isostere, the rigid 5-memered-ring of Pro limited the rotation of the transition state and partially exposed the bulky side of the molecule, making it available for hydrogen attack. So the stereoselectivity was not as good as for the Ser-Pro isostere.

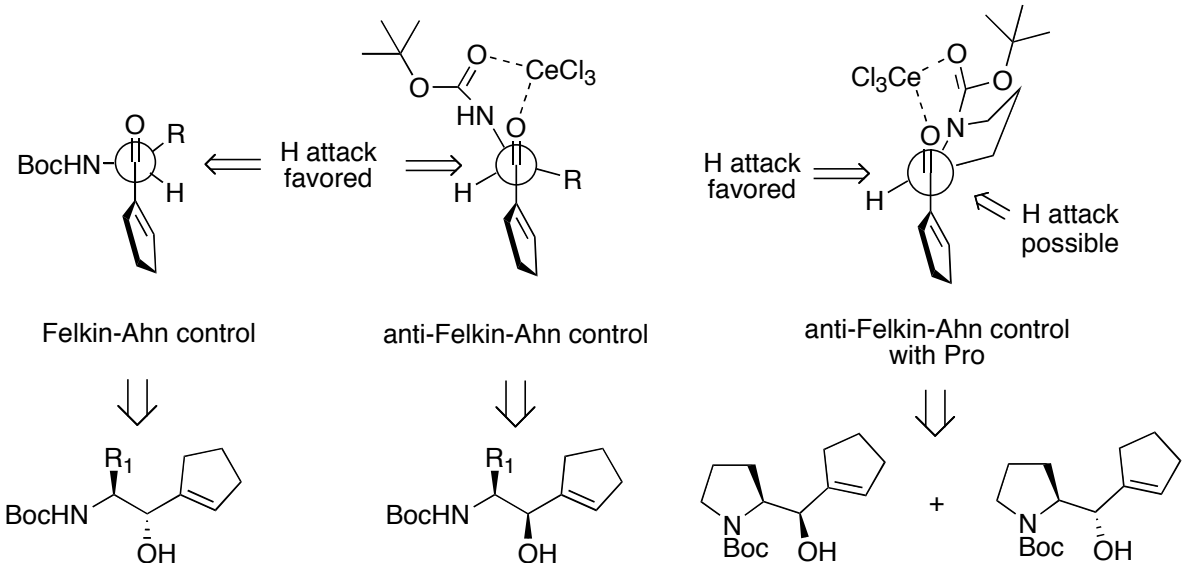
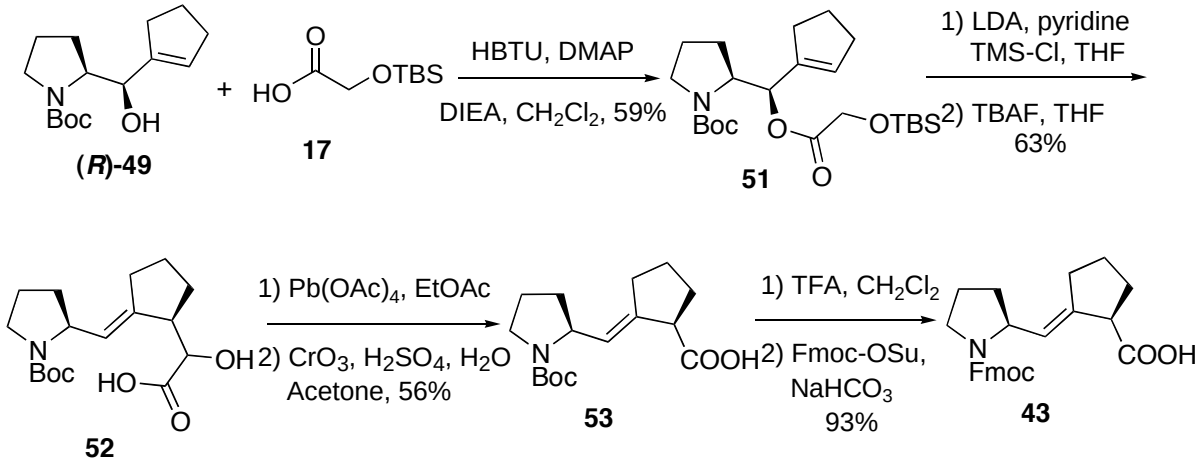


Figure 2.12 Stereoselectivity of the $NaBH_4 \cdot CeCl_3$ reduction.

Both diastereomers were carried to the next step and the synthetic routes were identical. Alcohol **49** was coupled with TBS-protected acid **17** with HBTU as the coupling reagent, and ester **51** was synthesized (Scheme 2.24). The reaction with acid chloride **18** was also attempted, but the yield was always around 25-30 %. The Pro-Pro isostere was more sterically hindered and less reactive, so it was difficult for the hydroxyl group in alcohol **49** to approach the acid chloride group in acid chloride **18** to form a tetrahedral intermediate. α-Hydroxycarboxylic acid **52** was synthesized from ester **51** by the Ireland-Claisen rearrangement, followed by TBS-deprotection with TBAF. The stereochemistry was retained through the chair-form 6-membered ring transition state, and the alkene in the product adopted the (*E*)-configuration. α-Hydroxycarboxylic acid **52** was decarboxylated with $Pb(OAc)_4$ to give the aldehyde, and the resulting aldehyde was oxidized to acid **53** by freshly prepared Jones reagent. Then the Boc group was removed with TFA, and the Fmoc protecting group was added by stirring with Fmoc-OSu in a saturated $NaHCO_3$ solution. The dipeptide isostere Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH **43** was purified by flash chromatography.

Scheme 2.24 Synthesis of the dipeptide isostere Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH **43**.



Comparing the syntheses of the dipeptide isosteres, Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23**, and Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH **43**, the Pro-Pro isostere had stereochemical control,

although the stereoselectivity was not very good. The two diastereomers could be separated easily by flash chromatography, which was even more convenient than for the Ser-Pro isostere. At nearly every step during the synthesis, the Pro-Pro isostere had a lower yield due to steric hindrance of the proline 5-membered ring. The reaction conditions could be optimized to achieve higher yields and better stereoselectivity.

2.3 Synthesis of other tripeptide building blocks for polymerization

To synthesize a polypeptide by polymerization, both the *N*-terminal and *C*-terminal ends of the tripeptide should be free, or readily removed during the polymerization. In this case, Boc-protected amino acids are more convenient to use during the synthesis of tripeptide building blocks because 1) Boc-protection is more stable than Fmoc-protection during solution phase peptide coupling due to the presence of the base DIEA, and coupling usually have higher yields; 2) most of the activated esters are not stable in base, and the Fmoc protecting group is not compatible with the activated ester polymerization; 3) Fmoc-protected amino acids are usually more expensive than the Boc-protected amino acids and the atom economy is not as good with Fmoc-protection.

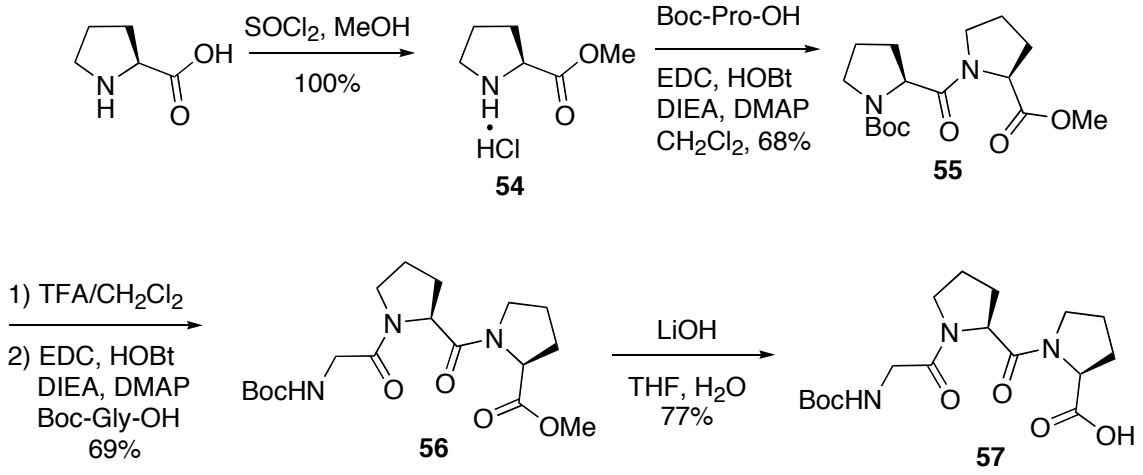
Tripeptides Gly-Pro-Pro and Gly-Pro-Hyp are the most common repeating units in collagen studies. The synthesis of Gly-Pro-Pro and Gly-Pro-Hyp containing polymeric peptides is important for mimicking the polymerization of the conformationally locked (*E*)-alkene Gly-Pro or Pro-Pro isosteres, and provides useful controls for collagen property studies.

2.3.1 Synthesis of the tripeptide building block Boc-Gly-Pro-Pro-OH

Both methyl and benzyl groups could be used to protect the acid because the tripeptide repeat Gly-Pro-Pro had no hydroxyl side-chain. The methyl protecting group was chosen in this synthesis because the basic deprotection would not affect the Boc-protection, and the protection and deprotection of the methyl group was easier to perform than for the benzyl protecting group.

Starting from commercially available H-Pro-OH, proline methyl ester **54** was synthesized by Fisher esterification (Scheme 2.25). The resulting proline methyl ester **54** was dried in vacuo and carried to the next step without further purification. The syntheses of dipeptide **55** and tripeptide **56** were traditional solution phase peptide synthesis. TFA was used to deprotect the Boc-group, and EDC was used as the coupling reagent. Then LiOH was used to remove the methyl group, followed by a work-up with 1 M HCl solution to remove LiOH. The tripeptide building block Boc-Gly-Pro-Pro-OH **57** was obtained as a yellow oil.

Scheme 2.25 Synthesis of the tripeptide building block Boc-Gly-Pro-Pro-OH **57**.



It was very unusual that tripeptide methyl ester **56** could be dissolved in water. During the work-up, about half of the product was dissolved in 1 M HCl solution. More

EtOAc was used to extract the tripeptide back into the organic layer, and both aqueous and organic layers were monitored by TLC during the work-up process.

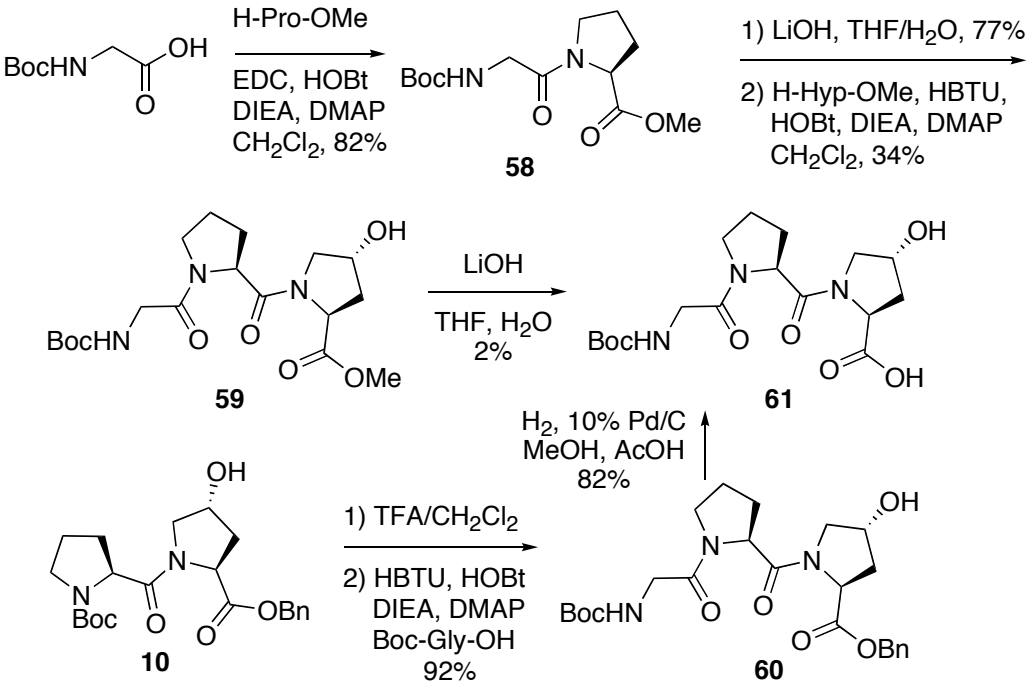
2.3.2 Synthesis of the tripeptide building block Boc-Gly-Pro-Hyp-OH

As described in Section 2.1.2, side-chain unprotected Hyp formed double-coupled products at the hydroxyl group of Hyp during peptide coupling reactions. To minimize the side reactions during peptide synthesis, tripeptide Boc-Gly-Pro-Hyp-OH **61** was synthesized from the *N*-terminus to the *C*-terminus (left to right).

Boc-Gly-OH and H-Pro-OMe **54** were coupled with EDC (Scheme 2.26). Dipeptide Boc-Gly-Pro-OMe **58** was purified by flash chromatography. Then the methyl group was removed with 1 M LiOH solution in THF-H_2O. The resulting dipeptide acid was coupled with hydroxyproline methyl ester **3** using HBTU without further purification, and tripeptide Boc-Gly-Pro-Hyp-OMe **59** was obtained. The yield of this reaction was very low. Double-coupled product was observed, although the amount of the side product was much less than the synthesis of tripeptide Fmoc-Gly-Pro-Hyp-OMe **8**. Tripeptide Boc-Gly-Pro-Hyp-OMe **59** also has good solubility in H_2O, and a significant amount of the product was lost during the work-up. Using TBS-protected hydroxyproline methyl ester could prevent the formation of double-coupled side product, and might improve the yield. The methyl group was removed with 1 M LiOH and the tripeptide building block Boc-Gly-Pro-Hyp-OH **61** was obtained as a clear oil in very low yield (about 10 %). TLC showed that the reaction was finished, but after work-up, very little product was obtained. The high solubility of the product in water might be the reason for the low yield. The same problem occurred when synthesizing another tripeptide Ac-Gly-Pro-Hyp-OH. It seemed that aqueous solvents could not be used in the final step, and the methyl protecting group was not good for the synthesis of the unprotected

Gly-Pro-Hyp building block. Benzyl protection yielded a better result, because the deprotection of benzyl groups did not require aqueous work-up, which could significantly minimize the loss of Boc-Gly-Pro-Hyp-OH **61** in water.

Scheme 2.26 Synthesis of the tripeptide building block Boc-Gly-Pro-Hyp-OH **61**.



The TBS temporary protection of the hydroxyl group could also be used to synthesize tripeptide Boc-Gly-Pro-Hyp-OH **61**. The synthesis of the tripeptide Fmoc-Gly-Pro-Hyp-OMe **8** was described in Section 2.1.2, and replacing Fmoc-Gly-OH with Boc-Gly-OH would yield the desired tripeptide building block.

2.4 Summary

Tripeptides Gly-Pro-Pro and Gly-Pro-Hyp are the most common repeating units used in collagen studies. Fmoc-protected tripeptides Fmoc-Gly-Pro-Pro-OH and Fmoc-Gly-Pro-Hyp-OH were synthesized for solid phase peptide synthesis, while Boc-protected tripeptides Boc-Gly-Pro-Pro-OH and Boc-Gly-Pro-Hyp-OH were synthesized for polymerization. The

Boc-protected tripeptides had good solubility in H_2O, and caused significant loss of the products during aqueous work-up. Fmoc-protected tripeptides did not have this problem.

The conformationally locked alkene isosteres, Gly-*trans*-Pro and Pro-*trans*-Pro, were synthesized. The synthesis of the Gly-*trans*-Pro isostere had no stereochemistry control, and the two diastereomers of the tripeptide isostere Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OBn were separated by normal phase HPLC. Although the stereoselectivity of the asymmetric reduction was not good for the Pro-*trans*-Pro isostere, the resulting diastereomers could be separated by flash chromatography, and the absolute stereochemistry of the two diastereomers was determined by Mosher's method.

During the synthesis, benzyl and methyl groups were used as the protecting groups at the acid end. The methyl group was easy to attach, but always led to problems during deprotection, or during purification. The benzyl group could be removed by Lewis acids, weak bases, or mild hydrogenation without affecting the Fmoc-protection and the C=C double bond. It was the ideal protecting group for peptide synthesis.

Experimental

General Information. Amino acid derivatives, resins, and reagents were purchased. Unless otherwise indicated, all reactions were carried out under N_2 sealed from moisture. Anhydrous THF was obtained by reflux from Na-benzophenone. Anhydrous CH_2Cl_2 was dried by passage through a dry alumina column. Other anhydrous solvents were used directly from sealed bottles, which were stored under Ar. Brine (NaCl), $NaHCO_3$, $Na_2S_2O_3$, and NH_4Cl refer to saturated aqueous solutions unless otherwise noted. Flash chromatography was performed on 32-63 μm or 230-400 mesh silica gel with reagent grade solvents. NMR spectra were obtained at ambient temperature in $CDCl_3$ unless otherwise noted. ^1H and ^{13}C

NMR spectra were obtained at 400 and 100 MHz, respectively, unless otherwise noted. Minor rotamer chemical shifts are shown in parenthesis. Data are reported as follows: chemical shift, multiplicity: singlet (s), doublet (d), triplet (t), quadruplet (q), quintuplet (qt), sextuplet (st), multiplet (m), coupling constants *J* in Hz, and integration. Analytical HPLC was performed on a Beckman PAC/E 168 dual dual-solvent HPLC system equipped with a P/ACE 215 UV detector. The chromatogram was simultaneously monitored at 210 nm and 260 nm. Analytical HPLC was performed on an isocratic condition with 5 % *i*-PrOH in hexanes a Varian MetaSil silica column (5 μm, 4.6 × 50 mm) unless otherwise noted. A gradient from 5 % to 95 % CH_3CN in H_2O in 10 min with a flow rate of 1.0 mL/min on an XBridge C18 reverse phase column (2.5 μm, 4.6 × 50 mm) was used for RP HPLC.

H-Hyp-OMe•HCl, 3. $SOCl_2$ (11.6 mL, 0.161 mol) and MeOH (70 mL) were pre-mixed, and stirred at 0 °C for 30 min. H-Hyp-OH (6.0 g, 0.046 mol) was added. The solution was warmed to rt, and stirred for 24 h. The solvent was removed in vacuo, and the resulting oil was dissolved in a minimal amount of MeOH; Et_2O was added, and a white solid (8.3 g, 99 %) was collected by filtration. ^{1}H NMR (CD_3OD): δ 4.60 (dd, *J* = 7.7, 3.0, 2H), 3.86 (s, 3H), 3.44 (dd, *J* = 12.1, 3.8, 1H), 3.31 (m, 1H), 2.42 (ddt, *J* = 13.6, 7.5, 1.6, 1H), 2.20 (ddd, *J* = 13.8, 10.7, 4.2, 1H). ^{13}C NMR (CD_3OD): δ 170.7, 70.6, 59.4, 55.0, 54.0, 38.6.

Boc-Pro-Hyp-OMe, 4. H-Hyp-OMe **3** (107 mg, 0.589 mmol) was dissolved in DMF (10 mL). EDC (114 mg, 0.595 mmol), HOBt (91 mg, 0.60 mmol), DMAP (7.3 mg, 0.060 mmol) and DIEA (0.40 mL, 2.3 mmol) were added, and the solution was stirred at 0 °C for 10 min. Boc-Pro-OH (128 mg, 0.595 mmol) was added, and the solution was stirred at rt for 6 h. EtOAc (100 mL) was added, and the organic layer was washed with 1 M HCl (2 × 50 mL), H_2O (50 mL), $NaHCO_3$ (2 × 50 mL) and brine (50 mL). After drying over Na_2SO_4,

the solution was concentrated in vacuo and purified by flash chromatography (1 % MeOH in $CHCl_3$). A yellow oil was obtained (116 mg, 57 %). ^{1}H NMR: δ 4.67 (t, *J* = 8.3, 1H), 4.44 (dd, *J* = 7.7, 3.7, 2H), 4.03 (d, *J* = 11.2, 1H), 3.70 (s, 3H), 3.62 (dd, *J* = 11.0, 3.7, 1H), 3.52 (m, 2H), 3.37 (dt, *J* = 10.4, 7.1, 1H), 2.33 (ddt, *J* = 13.2, 8.2, 2.0, 1H), 2.09 (m, 2H), 2.01 (m, 1H), 1.93 (m, 1H), 1.83 (m, 1H), 1.41 (s, 9H). ^{13}C NMR: δ 173.1 (172.8), 171.7 (171.9), 154.9 (154.0), 80.3 (79.8), 70.5 (70.3), 57.5 (58.1), 57.4 (57.9), 55.0 (54.3), 52.3 (52.4), 47.1 (46.7), 37.8 (37.2), 29.4 (30.0), 28.6 (28.5), 24.4 (23.6). Rotamers: 3:1.

Boc-Pro-Hyp(OTBS)-OMe, 6. Boc-Pro-Hyp-OMe **4** (1.23 g, 3.60 mmol) was dissolved in DMF (15 mL), and imidazole (1.47 g, 21.6 mmol) was added. The solution was stirred at rt for 1.5 h. TBS-Cl (1.09 g, 7.25 mmol) was added, and the solution was stirred for another 24 h. EtOAc (200 mL) was added and the organic layer was washed with 1 M HCl (2 × 100 mL), H_2O (100 mL), $NaHCO_3$ (2 × 100 mL) and brine (50 mL). After drying over $MgSO_4$, the solution was concentrated in vacuo and purified by flash chromatography (1 % MeOH in $CHCl_3$). A clear, colorless oil was obtained (1.64 g, 99.6 %). ^{1}H NMR: δ 4.63 (dd, *J* = 8.6, 5.0, 0.5H), 4.59 (d, *J* = 7.6, 0.5H), 4.54 (m, 1H), 4.45 (dd, *J* = 8.2, 3.2, 0.5H), 4.36 (dd, *J* = 8.3, 3.8, 0.5H), 3.76 (dd, *J* = 6.1, 3.3, 0.5H), 3.71 (m, 0.5H), 3.71 (s, 1.5H), 2.70 (s, 1.5H), 3.59 (m, 1H), 3.53 (m, 0.5H), 3.46 (m, 1H), 3.39 (m, 0.5H), 2.14 (m, 2H), 2.00 (m, 3H), 1.83 (m, 1H), 1.43 (s, 4H), 1.39 (s, 5H), 0.87 (s, 5H), 0.86 (s, 4H), 0.07 (s, 3H), 0.06 (s, 3H). ^{13}C NMR: δ 172.9 (173.0), 171.61 (171.58), 154.0 (154.6), 79.7 (79.6), 71.2 (70.7), 58.0 (58.2), 57.7, 55.0 (54.1), 53.4 (52.3), 46.7 (46.9), 38.0 (37.5), 30.0 (29.2), 28.57 (28.62), 25.84 (25.86), 23.6 (24.2), 18.14 (18.07), –4.74 (–4.67). Rotamers: 5:4.

Fmoc-Gly-Pro-Hyp(OTBS)-OMe, 7 and Fmoc-Gly-Pro-Hyp-OMe, 8. Boc-Pro-Hyp(OTBS)-OMe **6** (1.55 g, 3.39 mmol) was dissolved in CH_2Cl_2 (50 mL). TFA (20 mL)

was added, and the solution was stirred at rt for 40 min. The solvent and excess TFA were removed in vacuo. The resulting dark yellow oil was dissolved in CH_2Cl_2 (50 mL). HBTU (1.42 g, 3.75 mmol), HOBt (580 mg, 3.78 mmol), DMAP (42 mg, 0.34 mmol) and DIEA (3.00 mL, 17.2 mmol) were added. The solution was stirred at 0 °C for 20 min, and Fmoc-Gly-OH (1.12 g, 3.76 mmol) was added. The reaction was stirred at rt for 8 h. The solution was concentrated in vacuo and then diluted with EtOAc (250 mL). The organic layer was washed with 1 M HCl (2 × 100 mL), H_2O (100 mL), $NaHCO_3$ (2 × 100 mL) and brine (100 mL). The aqueous layers were extracted with CH_2Cl_2 (4 × 100 mL) and monitored by TLC. The organic layers were combined, dried over Na_2SO_4, concentrated in vacuo and purified by flash chromatography (1 % MeOH in $CHCl_3$). Fmoc-Gly-Pro-Hyp(OTBS)-OMe **7** (461.4 mg, 21.4 %) was obtained as a pale yellow oil, and Fmoc-Gly-Pro-Hyp-OMe **8** (1.29 g, 73.1 %) was obtained as a white solid. Fmoc-Gly-Pro-Hyp(OTBS)-OMe **7:** ^{1}H NMR: δ 7.74 (d, *J* = 7.4, 2H), 7.59 (dd, *J* = 7.4, 2.0, 2H), 7.38 (t, *J* = 7.6, 2H), 7.30 (ddt, *J* = 7.4, 1.8, 1.1, 2H), 5.75 (m, 1H), 4.64 (m, 2H), 4.54 (qt, *J* = 5.3, 1H), 3.43 (m, 2H), 4.21 (d, *J* = 7.4, 1H), 4.09 (dd, *J* = 17.2, 5.1, 1H), 3.96 (dd, *J* = 17.0, 3.2, 1H), 3.76 (dd, *J* = 9.8, 5.6, 1H), 3.72 (s, 3H), 3.66 (dd, *J* = 9.6, 4.6, 1H), 3.60 (m, 1H), 3.50 (m, 1H), 2.18 (m, 3H), 2.04 (m, 3H), 0.89 (s, 9H), 0.09 (s, 3H), 0.08 (s, 3H). ^{13}C NMR: δ 172.8, 170.5, 166.8, 156.3, 144.0, 141.3, 127.7, 127.2 (127.1), 125.3, 120.0, 70.7, 67.2, 58.1, 57.8, 54.4, 52.4, 47.2, 46.2, 43.4, 37.7, 28.3, 25.8, 24.6, 18.0, –4.69 (–4.74). MS calcd. for $C_{34}H_{46}N_3O_7Si$ (MH^+) *m/z* = 636.8, found *m/z* = 636.7. Fmoc-Gly-Pro-Hyp-OMe **8:** ^{1}H NMR: δ 7.73 (d, *J* = 7.5, 2H), 7.59 (d, *J* = 7.4, 2H), 7.37 (t, *J* = 7.4, 2H), 7.28 (tt, *J* = 7.4, 1.2, 2H), 5.91 (t, *J* = 4.4, 1H), 4.65 (m, 2H), 4.43 (m, 2H), 4.35 (d, *J* = 6.8, 2H), 4.20 (t, *J* = 7.0, 1H), 4.06 (dd, *J* = 17.0, 5.5, 1H), 3.92 (m, 1H), 3.69 (s, 3H), 3.60 (m, 2H), 3.43 (dd, *J* = 15.9, 6.8, 1H), 2.63 (br s, 1H), 2.33

(m, 1H), 2.12 (m, 2H), 1.94 (m, 3H). ^{13}C NMR: δ 172.8, 170.7, 167.5, 156.5, 144.0, 141.3, 127.8, 127.1, 125.2, 120.0, 70.5, 67.1, 58.2, 57.8, 55.0, 52.3, 47.2, 46.5, 43.4, 37.6, 28.3, 24.7. MS calcd. for $C_{28}H_{32}N_3O_7$ (MH^+) *m/z* = 522.6, found *m/z* = 522.6.

Fmoc-Gly-Pro-Hyp-OH, 1. Fmoc-Gly-Pro-Hyp-OMe **8** (62.9 mg, 0.120 mmol) was dissolved in CH_2Cl_2 (5 mL). Pre-mixed $AlCl_3$ in *N,N*-dimethylaniline (5:8, 0.22 mL) was added. The resulting blue solution was heated at reflux for 24 h. The reaction was quenched with 1 M HCl (25 mL), and the aqueous layer was extracted with CH_2Cl_2 (4 × 50 mL). The organic layers were combined and dried with Na_2SO_4. The solvent was removed in vacuo, and the resulting dark blue oil was purified by preparative TLC (4 times with 5 % MeOH, 1 % AcOH in $CHCl_3$). A pale yellow solid (54.0 mg, 88.2 %) was obtained. ^{1}H NMR (500 MHz): δ 7.71 (d, *J* = 7.7, 2H), 7.58 (t, *J* = 7.6, 2H). 7.35 (t, *J* = 7.4, 2H), 7.26 (app. t, *J* = 7.2, 2H), 6.17 (m, 1H), 4.60 (m, 1H), 4.40 (m, 2H), 4.34 (m, 1H), 4.29 (m, 1H), 4.17 (t, *J* = 7.0, 1H), 4.03 (m, 1H), 3.86 (d, *J* = 15.4, 1H), 3.78 (d, *J* = 9.0, 1H), 3.53 (m, 2H), 3.37 (m, 1H), 2.33 (m, 1H), 1.98 (m, 4H), 1.86 (m, 1H). ^{13}C NMR: δ 175.8, 171.5, 168.4, 156.8, 143.9, 141.3, 127.8, 127.2, 125.3, 120.0, 70.4, 67.2, 58.4, 55.2, 53.9, 47.2, 46.6, 43.4, 37.2, 28.2, 24.8. MS calcd. for $C_{27}H_{30}N_3O_7$ (MH^+) *m/z* = 508.5, found *m/z* = 508.3.

H-Hyp-OBn•TsOH, 9. H-Hyp-OH (155 mg, 1.18 mmol) was dissolved in a mixture of CCl_4 (1.0 mL) and BnOH (1.0 mL). TsOH (270 mg, 1.42 mmol) was added. The solution was heated at reflux for 24 h, and then cooled to rt. The solvent was removed in vacuo. Et_2O (20 mL) was added, and a white solid precipitated. The white solid (438.3 mg, 94.5 %) was collected by filtration. ^{1}H NMR (CD_3OD): δ 7.71 (d, *J* = 8.1, 2H), 7.39 (m, 5H), 7.23 (dt, *J* = 8.0, 0.5, 2H), 5.32 (d, *J* =11.9, 1H), 5.26 (d, *J* = 12.1, 1H), 4.63 (dd, *J* = 10.6, 7.6, 1H), 4.56 (t, *J* = 3.8, 1H), 3.42 (dd, *J* = 12.2, 3.8, 1H), 3.29 (m, 1H), 2.42 (dddd, *J* = 13.7 7.7, 1.7, 1.4,

1H), 2.37 (s, 3H), 2.16 (ddd, *J* = 13.7, 10.6, 4.0, 1H). ^{13}C NMR (CD_3OD): δ 170.1, 143.5, 141.8, 136.3, 129.81, 129.79, 129.72, 129.67, 127.0, 70.6, 69.4, 59.5, 55.1, 38.6, 21.3. HRMS calcd. for $C_{12}H_{16}NO_3$ (MH^+) *m/z* = 222.1130, found *m/z* = 222.1130.

Boc-Pro-Hyp-OBn, 10. Boc-Pro-OH (110 mg, 0.511 mmol) was dissolved in CH_2Cl_2 (20 mL). HBTU (215 mg, 0.570 mmol), HOBt (93 mg, 0.61 mmol), DMAP (6.6 mg, 0.054 mmol) and DIEA (0.45 mL, 2.6 mmol) were added, and the solution was stirred at 0 °C for 10 min. H-Hyp-OBn **9** (205 mg, 0.520 mmol) was added, and the solution was stirred at rt for 8 h. The solvent was removed in vacuo, and EtOAc (50 mL) was added. The organic layer was washed with 1 M HCl (2 × 25 mL), H_2O (25 mL), $NaHCO_3$ (25 mL) and brine (25 mL). After drying with $MgSO_4$, the solution was concentrated in vacuo, and purified by flash chromatography (2 % MeOH in $CHCl_3$). A pale yellow oil was obtained (217 mg, 100 %). ^{1}H NMR: δ 7.32 (m, 5H), 5.22 (d, *J* = 12.2, 1H), 5.06 (d, *J* = 12.4, 1H), 4.74 (t, *J* = 8.4, 1H), 4.43 (dd, *J* = 7.8, 3.8, 2H), 4.04 (d, *J* = 11.2, 1H), 3.61 (dd, *J* = 11.1, 3.9, 1H), 3.53 (m, 1H), 3.38 (m, 2H), 2.33 (ddt, *J* = 13.4, 8.4, 2.0, 1H), 2.07 (m, 2H), 1.93 (m, 2H), 1.81 (m, 1H), 1.41 (s, 9H). ^{13}C NMR: δ 172.3 (172.0), 171.7 (171.8), 154.9 (154.0), 135.8. 128.6, 128.3, 128.2, 80.3 (79.8), 70.5 (70.3), 66.9 (67.1), 57.63 (58.00), 57.29 (57.58), 55.0 (54.3), 47.1 (46.7), 37.7 (37.1), 29.4 (29.9), 28.6 (28.5), 24.4 (23.6). Rotamers: 4:1.

Fmoc-Gly-Pro-Hyp-OBn, 11. Boc-Pro-Hyp-OBn **10** (217 mg, 0.520 mmol) was dissolved in CH_2Cl_2 (10 mL). TFA (4 mL) was added, and the solution was stirred at rt for 30 min. The solvent and excess TFA were removed in vacuo. The resulting dark yellow oil was dissolved in CH_2Cl_2 (20 mL). HBTU (222 mg, 0.585 mmol), HOBt (88 mg, 0.57 mmol), DMAP (5.3 mg, 0.043 mmol) and DIEA (0.45 mL, 2.6 mmol) were added. The solution was stirred at 0 °C for 20 min, and Fmoc-Gly-OH (154 mg, 0.521 mmol) was added. The reaction was

stirred at rt for 8 h. The solution was concentrated in vacuo, and diluted with EtOAc (100 mL). The organic layer was washed with 1 M HCl (2 × 50 mL), H_2O (50 mL), $NaHCO_3$ (2 × 50 mL) and brine (50 mL). The aqueous layers were extracted with CH_2Cl_2 (4 × 50 mL) and monitored by TLC (5 % MeOH in $CHCl_3$). The organic layers were combined, dried over $MgSO_4$, concentrated in vacuo, and purified by flash chromatography (2 % MeOH in $CHCl_3$). A white solid (257.1 mg, 83.0 %) was obtained. ^{1}H NMR: δ 7.75 (d, *J* = 7.5, 2H), 7.59 (d, *J* = 7.5, 2H, 7.39 (t, *J* = 7.4, 2H), 7.32 (m, 9H), 5.76 (m, 1H), 5.22 (d, *J* = 12.3, 1H), 5.07 (d, *J* = 12.5, 1H), 4.75 (t, *J* = 8.3, 1H), 4.63 (dd, *J* = 8.1, 3.8, 1H), 4.48 (br s, 1H), 4.35 (dd, *J* = 6.6, 3.2, 2H), 4.21 (t, *J* = 7.1, 1H), 4.09 (dd, *J* = 17.3, 5.3, 1H), 4.03 (d, *J* = 11.2, 1H), 3.94 (dd, *J* = 17.2, 3.6, 1H), 3.65 (dd, *J* = 10.9, 4.1, 1H), 3.59 (m, 1H), 3.47 (m, 2H), 2.37 (m, 1H), 2.18 (m, 1H), 2.08 (m, 1H), 1.98 (m, 3H). ^{13}C NMR: δ 172.1, 170.8, 167.3, 156.4, 144.0, 141.4, 135.7, 128.7, 128.4, 128.2, 127.8, 127.2, 125.3, 120.1, 70.6, 67.2, 67.0, 58.0, 57.7, 55.2, 47.2, 46.6, 43.5, 37.7, 28.6, 24.9.

Fmoc-Gly-Pro-Hyp-OH, 1. Fmoc-Gly-Pro-Hyp-OBn **11** (3.05 g, 5.11 mmol) was dissolved in a mixture of MeOH (50 mL) and AcOH (50 mL), and 10 % Pd/C (383 mg) was added. H_2 was introduced with a balloon, and the suspension was stirred at rt for 2 h. The solution was filtered through Celite™, and concentrated in vacuo. The resulting oil was purified by flash chromatography (8 % MeOH, 1 % AcOH in $CHCl_3$), and a white solid was obtained (2.14 g, 82.4 %). ^{1}H NMR (CD_3OD): δ 7.79 (d, *J* = 7.2, 2H), 7.67 (dd, *J* = 7.0, 4.4, 2H). 7.38 (t, *J* =7.4, 2H), 7.30 (t, *J* = 7.6, 2H), 4.71 (dd, 8.8, 3.3, 1H), 4.49 (m, 2H), 4.34 (m, 2H), 4.22 (t, *J* = 7.0, 1H), 4.06 (d, *J* = 17.1, 1H), 3.88 (d, *J* = 17.4, 1H), 3.75 (m, 2H), 3.59 (m, 2H), 2.23 (m, 2H), 2.02 (m, 3H), 1.89 (m, 1H). ^{13}C NMR (CD_3OD): δ 175.4, 172.8, 169.7, 159.0,

145.3, 142.6, 128.8, 128.2, 126.3, 120.9, 71.1, 68.1, 59.8, 59.4, 55.8, 47.6, 44.0, 38.3, 29.1, 25.6.

Cyclopentylidene hydrazone, 12. Cyclopentanone (20 mL, 0.23 mol) and hydrazine monohydrate (45 mL, 0.93 mol) were mixed at rt, and then heated at reflux for 16 h. The solution was cooled to rt, and H_2O (250 mL) was added. The aqueous layer was extracted with CH_2Cl_2 (4 × 100 mL). The organic layers were combined, washed with brine (150 mL), dried over Na_2SO_4, and concentrated in vacuo. A clear, colorless liquid (25.1 g, 100 %) was obtained. ^{1}H NMR: δ 4.81 (br s, 2H), 2.34 (tt, *J* = 7.2, 1.2, 2H), 2.16 (app. tt, *J* = 7.1, 1.1, 2H), 1.84 (qt, *J* = 7.0, 2H), 1.72 (m, 2H). ^{13}C NMR (125 MHz): δ 174.0, 33.1, 29.3, 25.8, 24.8.

1-Iodocyclopentene, 13. I_2 (129 g, 0.508 mol) was dissolved in Et_2O (700 mL). The solution was stirred at 0 °C for 20 min. 1,1,3,3-Tetramethylguanidine (350 mL, 2.79 mol) was dissolved in Et_2O (500 mL), and added dropwise within 3.5 h. The resulting solution was stirred at 0 °C for another 3 h. Cyclopentylidene hydrazone **12** (22.7 g, 0.230 mol) in Et_2O (200 mL) was added dropwise at 0 °C within 3.5 h. The solution was stirred at rt for another 16 h, and then heated at reflux for 2 h. The precipitates were removed by filtration, and the solution was concentrated in vacuo. The resulting solution was heated at reflux at 90 °C for 3 h, and then diluted with Et_2O (700 mL). The organic layer was washed with 2 M HCl (3 × 150 mL), $Na_2S_2O_3$ (3 × 150 mL), $NaHCO_3$ (150 mL) and brine (150 mL), dried over Na_2SO_4, and concentrated in vacuo. A dark purple liquid (28.7 g, 64 %) was obtained without further purification, and stored under Ar at –20 °C. ^{1}H NMR: δ 6.09 (m, 1H), 2.59 (m, 2H), 2.31 (m, 2H), 1.92 (app. q, *J* = 7.6, 2H). ^{13}C NMR: δ 140.1, 92.8, 43.8, 34.1, 24.0.

Boc-Gly-N(OMe)Me, 14. Boc-Gly-OH (5.50 g, 31.3 mmol), *N,O*-dimethyl hydroxylamine hydrochloride (6.22 g, 63.7 mmol) and DCC (7.84 g, 38.0 mmol) were dissolved in CH_2Cl_2 (500 mL), and cooled to 0 °C. DIEA (22.0 mL, 126 mmol) and DMAP (0.10 g, 0.84 mmol) were added and the reaction was stirred for 24 h. The mixture was filtered to remove dicyclohexyl urea and concentrated. The resulting slurry was diluted with EtOAc (500 mL), and washed with NH_4Cl (2 × 100 mL), $NaHCO_3$ (2 × 100 mL), and brine (100 mL). The organic layer was dried over $MgSO_4$ and concentrated. Crystallization with EtOAc-hexanes gave 6.51 g (95 %) colorless needle-like crystals. mp 101-102 °C, lit. mp 100-104 °C. ^{1}H NMR: δ 5.26 (br s, 1H), 4.08 (d, *J* = 4.6, 2H), 3.71 (s, 3H), 3.20 (s, 3H), 1.45 (s, 9H). ^{13}C NMR (125 MHz): δ 170.3, 156.0, 79.7, 61.5, 41.8, 32.5, 28.4. HRMS calcd. for $C_9H_{19}N_2O_4$ (MH^+) *m/z* = 219.1345, found *m/z* = 219.1346.

α,β-Unsaturated ketone, 15. 1-Iodocyclopentene (3.90 g, 20.1 mmol) was dissolved in THF (80 mL) at –40 °C, and *s*-BuLi (1.4 M in cyclohexane, 28.2 mL, 39.5 mmol) was added. The reaction was stirred at –40 °C for 3 h to generate cyclopentenyl lithium. In another flask, Boc-Gly-N(Me)OMe **2** (2.87 g, 13.2 mmol) was dissolved in THF (20 mL). The solution was cooled to –15 °C and *i*-PrMgCl (2.0 M in THF, 6.5 mL, 13 mmol) was added dropwise. After cooling to –78 °C, the cyclopentenyl lithium solution was added to the solution of deprotonated **2** via cannula. The mixture was stirred at –78 °C for 1 h, warmed up slowly to rt and stirred for another 10 h. The reaction was quenched with NH_4Cl (10 mL). The resulting solution was diluted with EtOAc (100 mL), washed with NH_4Cl (2 × 20 mL), $NaHCO_3$ (20 mL), brine (20 mL), dried over $MgSO_4$, and concentrated. Chromatography with 10 % EtOAc in hexanes gave a pale yellow solid (1.77 g. 60 %). ^{1}H NMR: δ 6.78 (s, 1H), 5.36 (br s, 1H), 4.26 (d, *J* = 4.6, 2H), 2.53 (t, *J* = 7.6, 4H), 1.89 (tt, *J* = 7.6, 7.6, 2H),

1.40 (s, 9H). ^{13}C NMR: δ 192.9, 155.8, 144.7, 143.1, 79.7, 47.4, 34.2, 30.6, 28.4, 22.5. HRMS calcd. for $C_{12}H_{20}NO_3$ (MH^+) *m/z* = 226.1443, found *m/z* = 226.1445.

TBS-O-CH_2COO^nBu, 16. Imidazole (15.7 g, 0.231 mol) was dissolved in butyl glycolate (10 mL, 0.077 mol), and stirred at 0 °C for 20 min. TBS-Cl (13.3 g, 0.0882 mol) was added. The resulting slurry was stirred at 0 °C for 20 min, warmed to rt, and stirred for 14 h. After vacuum distillation at 85-90 °C, a clear, colorless liquid (18.3 g, 97 %) was obtained. ^{1}H NMR: δ 4.23 (s, 2H), 4.14 (t, *J* = 6.7, 2H), 1.63 (m, 2H), 1.38 (app. sextet, *J* = 7.5, 2H), 0.92 (m, 12H), 0.10 (s, 6H). ^{13}C NMR: δ 172.0, 64.8, 62.0, 30.8, 25.9, 19.2, 18.6, 13.8, –5.3.

TBS-O-CH_2COOH, 17. TBS-O-CH2COO^{n}Bu **16** (4.0 g, 0.016 mol) was dissolved in THF (25 mL) and stirred at 0 °C for 10 min. A NaOH (653 mg, 0.0163 mol) solution in a mixture of MeOH (2.0 mL) and H_2O (4.0 mL) was added dropwise, and stirred at rt for 12 h. The resulting solution was diluted with H_2O (50 mL), and washed with Et_2O (40 mL). A HCl solution (3.7 M, 5.4 mL) was added, and the aqueous solution was extracted with ether (2 × 40 mL). The ether layers were washed with H_2O (20 mL) and brine (20 mL), dried over $MgSO_4$, and concentrated in vacuo. A white solid (3.07 g, 99 %) was obtained. ^{1}H NMR: δ 4.22 (s, 2H), 0.94 (s, 9H), 0.15 (s, 6H). ^{13}C NMR: δ 61.6, 25.8, 18.3, –5.4.

TBS-O-CH_2COCl, 18. TBS-O-CH_2COOH **17** (214 mg, 1.12 mmol) was dissolved in benzene (5 mL). $(COCl)_2$ (0.20 mL, 2.29 mmol) was added, and the solution was stirred at rt for 30 min. The solution was heated at reflux for another 1.5 h, and the solvent was removed by distillation. The resulting light yellow liquid was used without further purification or characterization.

***rac*-Alcohol, 19.** Ketone **15** (66 mg, 0.29 mmol) was dissolved in THF:MeOH (2.5:1, 7 mL) and cooled to 0 °C. $CeCl_3 \cdot 7H_2O$ (131 mg, 0.352 mmol) was added, followed by $NaBH_4$

(23 mg, 0.60 mmol). After stirring at 0 °C for 2.5 h, the reaction was quenched with NH_4Cl (10 mL). The resulting solution was diluted with EtOAc (50 mL), washed with NH_4Cl (2 × 10 mL), and brine (10 mL), dried on $MgSO_4$, and concentrated. Flash chromatography with 20 % EtOAc in hexanes gave a pale yellow solid (62 mg, 93 %). ^{1}H NMR (500 MHz): δ 5.66 (s, 1H), 4.97 (br s, 1H), 4.30 (br s, 1H), 3.37 (m, 1H), 3.12 (ddd, *J* = 13.0, 6.9, 6.2, 1H), 2.68 (br s, 1H), 2.31 (m, 4H), 1.87 (tt, *J* = 7.3, 7.3, 2H), 1.42 (s, 9H). ^{13}C NMR: δ 156.7, 144.6, 126.4, 79.5, 70.9, 45.3, 32.3, 31.9, 28.4, 23.4. HRMS calcd. for $C_{12}H_{22}NO_3$ (MH^+) *m/z* = 228.1600, found *m/z* = 228.1605.

Ester, 20. Alcohol **19** (3.00 g, 13.2 mmol) was dissolved in CH_2Cl_2 (150 mL), and cooled to 0 °C. HBTU (7.85 g, 19.8 mmol), DMAP (80 mg, 0.66 mmol) and 2-(*tert*-butyldimethylsilyloxy)acetic acid **17** (4.01 g, 21.1 mmol) were added. The mixture was stirred at 0 °C for 10 min. DIEA (9.20 mL, 52.7 mmol) was added and the solution was stirred at rt for 10 h. The solvent was removed in vacuo and the resulting slurry was diluted with EtOAc (200 mL). The organic layer was washed with 0.5 M HCl (2 × 50 mL), $NaHCO_3$ (50 mL), brine (50 mL), dried over $MgSO_4$, and concentrated. Flash chromatography with 10 % EtOAc in hexanes gave a pale yellow solid (4.67 g, 81 %). ^{1}H NMR: δ 5.66 (d, *J* = 1.2, 1H), 5.48 (m, 1H), 4.65 (m, 1H), 4.24 (d, *J* = 1.4, 2H), 3.42 (m, 1H), 3.34 (m, 1H), 2.30 (m, 4H), 1.86 (m, 2H), 1.41 (s, 9H), 0.91 (s, 9H), 0.09 (s, 6H). ^{13}C NMR: δ 171.2, 155.9, 139.9, 128.9, 79.7, 72.8, 61.9, 42.8, 32.5, 32.1, 28.5, 25.9, 23.2, 18.5, –5.3.

α-Hydroxy acid, 21. A solution of LDA (2.0 M in THF, 5.3 mL, 11 mmol) was diluted with THF (20 mL) and stirred at –100 °C. A mixture of Me_3SiCl (3.7 mL, 30 mmol) and pyridine (2.55 mL, 31.5 mmol) in THF (10 mL) cooled to –100 °C was added dropwise to the LDA solution. After 5 min, a solution of ester **20** (1.06 g, 2.64 mmol) in THF (10 mL) cooled to –

100 °C was added dropwise and the reaction was stirred at –100 °C for 25 min, then warmed up slowly to rt over 2.5 h and stirred at rt for 10 h. The mixture was heated to 45 °C for 1.5 h. The reaction was quenched with 1 M HCl (25 mL), and the aqueous layer was extracted with Et_2O (2 × 100 mL). The organic layer was dried over $MgSO_4$, and concentrated to give 1.29 g of a yellow oil. Without further purification, the product was dissolved in THF (10 mL). *n*-Bu_4NF (2.50 g, 7.93 mmol) in THF (10 mL) was added and stirred for 5 min at 0 °C, then at rt for 3 h. The reaction was quenched with 0.5 M HCl (10 mL). The solution was extracted with EtOAc (50 mL), dried over $MgSO_4$ and concentrated. The resulting oil was purified by flash chromatography with 5 % MeOH in $CHCl_3$ and a yellow oil (0.73 g, 97 %) was obtained. ^{1}H NMR (CD_3OD): δ 5.37 (m, 1H), 4.08 (d, *J* = 5.1, 1H) 3.63 (m, 2H), 2.86 (m, 1H), 2.38 (m, 1H), 2.24 (m, 1H), 1.85 (m, 2H), 1.71 (m, 1H), 1.57 (m, 1H), 1.43 (s, 9H). ^{13}C NMR (CD_3OD): δ 177.0, 158.4, 146.2, 120.5, 79.9, 74.7, 40.7, 30.72, 30.66, 28.8, 25.5.

Boc-Gly-Ψ[(*E*)CH=C]-Pro-OH, 22. $Pb(OAc)_4$ (118 mg, 0.266 mmol) in $CHCl_3$ (0.6 mL) was added dropwise to a solution of α-hydroxy acid **21** (73.5 mg, 0.258 mmol) in EtOAc (4.0 mL) at 0 °C. The reaction was stirred for 15 min at 0 °C, then quenched with ethylene glycol (1.5 mL). The solution was diluted with EtOAc (40 mL), washed with H_2O (2 × 10 mL) and brine (10 mL), dried over Na_2SO_4, and concentrated to give a yellow oil (55.2 mg). The crude product was dissolved in acetone (8 mL) and cooled to 0 °C. Freshly prepared Jones reagent (2.7 M CrO_3, 2.7 M H_2SO_4, 0.16 mL, 0.43 mmol) was added dropwise. The reaction was stirred at 0 °C for 30 min, quenched with *i*-PrOH (1 mL), and stirred for another 10 min. The precipitate was removed by filtration, and the solvent was evaporated. The resulting residue was extracted with EtOAc (3 × 10 mL), and the organic layer was washed with H_2O (5 mL) and brine (5 mL), dried over Na_2SO_4, and concentrated.

Flash chromatography with 40 % EtOAc and 0.1 % HOAc in hexanes gave a white solid (34 mg, 53 %). ^{1}H NMR (500 MHz, DMSO-d_6): δ 12.18 (br s, 1H), 6.94 (t, J = 5.4, 1H), 5.38 (m, 1H), 3.5 (m, 2H), 3.16 (m, 1H), 2.29 (m, 1H), 2.21 (m, 1H), 1.80 (m, 3H), 1.54 (m, 1H), 1.36 (s, 9H). ^{13}C NMR (125 MHz, DMSO-d_6): δ 174.8, 155.5, 142.4, 120.2, 77.5, 48.9, 29.6, 28.7, 28.3, 24.5. HRMS calcd. for $C_{13}H_{22}NO_4$ (MH^+) m/z = 256.1549, found m/z = 256.1528.

Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH, 23. Boc-Gly-Ψ[(*E*)CH=C]-Pro-OH **22** (321 mg, 1.26 mmol) was dissolved in CH_2Cl_2 (15 mL) and TFA (5 mL) was added. The solution was stirred at rt for 40 min. The solvent was evaporated and the excess TFA was removed by high vacuum. The resulting dark yellow oil was dissolved in $NaHCO_3$ (36 mL) and stirred at 0 °C for 20 min. Fmoc-*O*-succinimide (637 mg, 1.89 mmol) was dissolved in 1,4-dioxane (8.0 mL) and added dropwise. The solution was stirred at rt for 14 h. The solution was acidified to pH 1 with 1 M HCl. The aqueous solution was extracted with CH_2Cl_2 (4 × 100 mL). The organic layers were combined, washed with water (50 mL), and brine (50 mL). After drying with $MgSO_4$, the organic layer was concentrated in vacuo. Flash chromatography with 5 % MeOH in $CHCl_3$ gave a pale yellow solid (402 mg, 85 %). The (*E*)-configuration was confirmed by 1D NOE spectroscopy by irradiating the alkene C-H peak (Appendix). ^{1}H NMR: δ 7.76 (d, J = 7.5, 2H), 7.59 (d, J = 7.4, 2H), 7.39 (t, J = 7.6, 2H), 7.31 (dt, J = 7.4, 1.1, 2H), 5.56 (br s, 1H), 4.90 (m, 1H), 4.42 (d, J = 10.4, 1H), 4.38 (d, J = 10.4, 1H), 4.20 (t, J = 6.8, 1H), 3.79 (m, 2H), 3.34 (t, J = 6.8, 1H), 2.34 (m, 2H), 1.97 (m, 3H), 1.66 (m, 1H). ^{13}C NMR: δ 179.9, 156.5, 144.3, 144.0, 141.4, 127.8, 127.1, 125.1, 120.1 (2 C, HSQC), 66.8, 49.4, 47.3, 40.5, 30.1, 29.3, 25.1. HRMS calcd. for $C_{23}H_{24}NO_4$ (MH^+) m/z = 378.1705, found m/z = 378.1734.

Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp-OBn, 24. Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23** (133 mg, 0.352 mmol) was dissolved in CH_2Cl_2 (40 mL). HBTU (164 mg, 0.433 mmol), HOBt (66.4 mg, 0.434 mmol), DMAP (4.4 mg, 0.036 mmol) and DIEA (0.30 mL, 1.7 mmol) were added, and the solution was stirred at 0 °C for 10 min. H-Hyp-OBn **9** (167.2 mg, 0.42 mmol) was added, and the solution was stirred at rt for 8 h. The solution was concentrated in vacuo, and diluted with EtOAc (100 mL). The organic layer was washed with 1 M HCl (2 × 50 mL), $NaHCO_3$ (50 mL), and brine (50 mL), dried over Na_2SO_4, and concentrated in vacuo. The resulting oil was purified by flash chromatography (2 % MeOH in $CHCl_3$), and a clear, colorless oil was obtained (114.1 mg, 55.8 %). ^{1}H NMR: δ 7.75 (d, *J* = 7.3, 2H), 7.60 (dd, *J* = 17.9, 7.2, 2H), 7.38 (t, *J* = 7.4, 2H), 7.31 (m, 7H), 5.89 (app. t, *J* = 5.0, 0.5H), 5.35 (m, 0.5H), 5.16 (m, 0.5H), 5.22 (d, *J* = 12.5, 0.5H), 5.21 (d, *J* = 12.3, 0.5H), 5.11 (d, *J* = 12.5, 0.5H), 5.10 (d, *J* = 12.4, 0.5H), 4.76 (t, *J* = 8.1, 0.5H), (t, *J* = 8.2, 0.5H), 4.47 (m, 1H), 4.35 (m, 2H), 4.18 (m, 1H), 3.79 (m, 1H), 3.68 (m, 1.5H), 3.56 (dd, *J* = 11.1, 3.8, 0.5H), 3.42 (d, *J* = 11.1, 0.5H), 3.37 (t, *J* = 7.0, 0.5H), 3.25 (dd, *J* = 11.6, 6.4, 1H), 2.57 (m, 1H), 2.34 (m, 3H), 1.94 (m, 3H), 1.60 (m, 1H). ^{13}C NMR: δ 173.2 (173.5), 172.3 (172.2), 156.8 (156.5), 145.0 (145.6), 144.1 (144.0), 141.4, 135.8 (135.7), 128.62 (128.64), 128.32 (128.35), 128.22 (128.19), 127.8, 127.16 (127.14), 125.2, 120.1, 118.8 (118.6), 70.4 (70.3), 66.9 (66.7), 57.9 (58.1), 55.6 (55.3), 49.2, 47.30 (47.34), 40.5 (40.4), 38.7, 37.6 (37.7), 30.0 (30.2), 29.1 (29.2), 24.6 (24.8). MS calcd. for $C_{35}H_{37}N_2O_6$ (MH^+) *m/z* = 581.27, found *m/z* = 581.28.

Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp-OH, 2. Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp-OBn **24** (114 mg, 0.196 mmol) was dissolved in a mixture of MeOH (3 mL) and AcOH (3 mL), and 10 % Pd/C (15.8 mg) was added. H_2 was introduced with a balloon, and the suspension was stirred at rt for 2 h. The solution was filter through Celite™, and concentrated in vacuo. The

crude ^{1}H NMR and LC-MS showed that the C=C bond was partially reduced. ^{1}H NMR (500 MHz, DMSO-d_6): δ 7.88 (d, *J* = 7.7, 2H), 7.69 (d, *J* = 7.1, 2H), 7.41 (t, *J* = 7.4, 2H), 7.33 (t, *J* = 7.3, 2H), 5.47 (m, 0.5H), 5.36 (m, 0.5H), 5.21 (m, 0.5H), 5.09 (m, 0.5H), 4.34 (app. t, *J* = 3.3, 1H), 4.26 (m, 3H), 4.20 (app t, *J* = 7.0, 1H), 3.66 (dd, *J* = 10.7, 4.4, 0.5H), 3.57 (m, 1.5H), 3.47 (m, 1H), 3.39 (m, 2H), 3.22 (m, 0.5H), 3.10 (m, 0.5H), 2.33 (m, 1H), 2.24 (m, 1H), 2.10 (m, 2H), 1.93 (m, 1H), 1.79 (m, 2H), 1.53 (m, 1H). ^{13}C NMR (125 MHz, DMSO-d_6): δ 173.4 (173.9), 172.1 (171.5), 156.0, 143.9 (144.0), 140.70 (140.72), 139.4 (137.4), 127.6 (128.9), 127.1 (127.3), 125.2 (125.3), 120.1 (121.4), 120.0 (109.8), 68.8 (67.3), 65.4 (67.2), 58.0 (59.7), 55.0 (53.7), 48.6, 48.1 (47.6), 46.7, 37.5, 29.9 (29.8), 28.7 (28.6), 24.0 (24.2). MS calcd. for $C_{35}H_{37}N_2O_6$ (MH^+) *m/z* = 491.22, found *m/z* = 491.31.

Fmoc-Hyp(^{t}Bu)-OBn, 27. Fmoc-Hyp(^{t}Bu)-OH (1.77 g, 4.31 mmol) was dissolved in CH_2Cl_2 (80 mL) at 0 °C. EDC (1.66 g, 8.64 mmol), HOBt (1.33 g, 8.69 mmol), DMAP (24 mg, 0.19 mmol) and DIEA (3.0 mL, 17 mmol) were added. The mixture was stirred at 0 °C for 10 min, and BnOH (0.90 mL, 8.7 mmol) was added. The mixture was stirred at rt for 8 h. The solution was concentrated in vacuo, and EtOAc (150 mL) was added to dilute the mixture. The organic layer was washed with 1 M HCl (2 × 50 mL), $NaHCO_3$ (50 mL), and brine (50 mL), then dried over Na_2SO_4 and concentrated. Flash chromatography with 20 % EtOAc in hexanes gave a clear colorless oil (2.3 g, 100 %). ^{1}H NMR: δ 7.77 (d, *J* = 7.5, 1H), 7.75 (d, *J* = 7.5, 1H), 7.58 (m, 2H), 7.34 (m, 9H), 5.23 (d, *J* = 12.3, 0.5H), 5.17 (d, *J* = 12.3, 0.5H), 5.13 (d, *J* = 12.3, 0.5H), 5.08 (d, *J* = 12.3, 0.5H), 4.54 (dd, *J* = 9.0, 4.6, 0.5 H), 4.48 (dd, *J* = 9.0, 4.6, 0.5 H), 4.37 (m, 1H), 4.28 (m, 2.5H), 4.01 (app. t, *J* = 7.0, 0.5H), 3.81 (dd, *J* = 12.4, 6.2, 0.5H), 3.78 (dd, *J* = 13.3, 6.3, 0.5H), 3.36 (dd, *J* = 10.6, 5.3, 1H), 2.16 (m, 2H), 1.18 (s, 4.5H), 1.16 (s, 4.5H). ^{13}C NMR: δ 177.7 (172.6), 155.1, 144.4 (144.2), 144.1, 141.4,

128.7 (128.6), 128.4 (128.3), 127.8 (127.7), 127.2 (127.1), 125.3 (125.1), 120.12 (120.10), 120.05 (120.03), 74.3, 69.3 (68.4), 67.8 (67.7), 67.1 (67.0), 58.1 (57.9), 53.9 (53.4), 47.3, 38.8 (37.8), 28.4 (28.3). Rotamers: 1:1.

Fmoc-Pro-Hyp(^{t}Bu)-OBn, 28. Fmoc-Hyp(^{t}Bu)-OBn **27** (1.30 g, 2.61 mmol) was dissolved in CH_2Cl_2 (20 mL), and piperidine (10 mL) was added. The solution was stirred at rt for 40 min, and the solvent was removed in vacuo. The resulting slurry was dissolved in CH_2Cl_2 (150 mL). HBTU (1.98 g, 5.22 mmol), HOBt (800 mg, 5.23 mmol), DMAP (26 mg, 0.21 mmol) and DIEA (1.85 mL, 10.6 mmol) were added. The solution was stirred at 0 °C for 20 min, and Fmoc-Pro-OH (1.77 g, 5.23 mmol) was added. The reaction was stirred at rt for 8 h. The solution was concentrated in vacuo, diluted with EtOAc (200 mL), washed with 1 M HCl (2 × 100 mL), H_2O (100 mL), $NaHCO_3$ (2 × 100 mL) and brine (100 mL), dried over Na_2SO_4, and concentrated in vacuo. The resulting yellow oil was purified by flash chromatography (EtOAc:Hexanes = 1:5), and a clear pale yellow oil (858 mg, 55.2 %) was obtained. ^{1}H NMR: δ 7.76 (d, *J* = 7.4, 2H), 7.59 (m, 2H), 7.34 (m, 9H), 5.22 (d, *J* = 12.5, 0.5H), 5.21 (d, *J* = 12.3, 0.5H), 5.09 (d, *J* = 12.3, 0.5H), 5.08 (d, *J* = 12.3, 0.5H), 4.73 (dd, *J* = 8.8, 3.8, 0.5H), 4.52 (m, 1H), 4.44 (m, 0.5H), 4.39 (dd, *J* = 10.8, 4.0, 0.5H), 4.27 (m, 3H), 4.12 (dd, *J* = 14.4, 7.2, 0.5H), 3.76 (m, 1H), 3.64 (m, 1H), 3.53 (m, 1.5H), 3.19 (dd, *J* = 9.5, 6.1, 0.5H), 2.11 (m, 4H), 1.94 (m, 2H), 1.15 (s, 5.5H), 1.08 (s, 3.5H). ^{13}C NMR: δ_172.2 (172.1), 171.0 (170.9), 155.0 (154.4), 144.4 (144.5), 144.1 (144.0), 141.3 (141.4), 128.66 (128.70), 128.45 (128.40), 128.2 (128.3), 127.72 (127.66), 127.2 (127.1), 125.05 (125.01), 119.99 (119.97), 74.16 (74.18), 69.6 (69.5), 67.51 (67.09), 66.98 (67.05), 58.2 (57.7), 57.61 (57.57), 53.0 (52.8), 47.3 (47.4), 46.8 (46.7), 36.43 (36.39), 29.1 (28.4), 28.30 (38.26), 24.3 (23.1). Rotamers: 3:2.

Fmoc-Gly-OSu, 30. Fmoc-Gly-OH (300 mg, 1.01 mmol) and HOSu (117 mg, 1.02 mmol) were dissolved in EtOAc (2.5 mL) and stirred at 0 °C for 10 min. DCC (209 mg, 1.01 mmol) was dissolved in EtOAc (1.5 mL), and added to the mixture of Fmoc-Gly-OH and HOSu dropwise. The solution was stirred at rt for 14 h. The resulting DCU was removed by filtration, and the filtrate was concentrated in vacuo. The resulting oil was dissolved in a minimal amount of EtOAc, and hexanes was added. A white solid was collected by filtration (343 mg, 86.0 %). ^{1}H NMR: δ 7.76 (d, *J* = 7.6, 2H), 7.58 (d, *J* = 7.7, 2H), 7.40 (t, *J* = 7.5, 2H), 7.31 (t, *J* = 7.2, 2H), 5.30 (m, 1H), 4.45 (d, *J* = 7.1, 2H), 4.37 (d, *J* = 5.7, 2H), 4.24 (t, *J* = 7.0, 1H), 2.85 (s, 4H). ^{13}C NMR: δ 168.7, 166.1, 156.1, 143.8, 141.4, 127.9, 127.2, 125.2, 120.1, 67.6, 47.1, 40.8, 25.7. HRMS calcd. for $C_{21}H_{19}N_2O_6$ (MH^+) *m/z* = 395.1243, found *m/z* = 395.1248.

Fmoc-Gly-Pro-OH, 31. Fmoc-Gly-OSu **30** (5.52 g, 14.0 mmol) was dissolved in DMF (60 mL). H-Pro-OH (1.61 g, 14.0 mmol) and DIEA (1.22 mL, 6.99 mmol) were added, and the solution was stirred at rt for 4 h. The solution was diluted with a mixture of EtOAc and 5 % $NaHCO_3$ (1:1 ratio, 50 mL), and the organic layer was removed. The aqueous layer was acidified to pH 1 with 1 M HCl, and extracted with EtOAc (4 × 50 mL). The organic layers were combined, washed with H_2O (2 × 50 mL), dried over Na_2SO_4, concentrated in vacuo, and dissolved in a minimal amount of EtOAc. Hexanes were added and the resulting precipitate was collected and dried in vacuo. A white solid was obtained (5.28 g, 95.6 %). ^{1}H NMR (500MHz): δ 10.37 (br s, 1H), 7.73 (d, *J* = 7.3, 2H), 7.59 (app. dd, *J* = 7.2, 4.0, 2H), 7.36 (t, *J* = 7.4, 2H), 7.27 (dt, *J* = 7.4, 0.9, 2H), 6.08 (t, *J* = 4.9, 1H), 4.55 (dd, *J* = 7.2, 4.8, 1H), 4.32 (dd, *J* = 7.4, 3.3, 2H), 4.19 (t, *J* = 7.2, 1H), 4.09 (dd, *J* = 17.2, 5.4, 1H), 3.95 (dd, *J* = 17.2, 4.2, 1H), 3.55 (m, 1H), 3.44 (dd, *J* = 17.2, 7.4, 1H), 2.11 (m, 2H), 2.06 (m, 1H), 1.96

(m, 1H). ^{13}C NMR (125 MHz): δ 174.5 (174.1), 168.4 (168.0), 156.6 (157.0), 143.90 (143.93), 141.30 (141.28), 127.76 (127.77), 127.1, 125.30 (125.27), 120.0, 67.3 (67.6), 59.3 (58.7), 47.1 (47.0), 46.3, 43.4 (43.3), 28.8 (31.3), 24.6 (22.3). Rotamers: 6:1. HRMS calcd. for $C_{22}H_{23}N_2O_5$ (MH^+) *m/z* = 395.1607, found *m/z* = 395.1584.

Fmoc-Gly-Pro-OSu, 32. Fmoc-Gly-Pro-OH **31** (1.01 g, 2.57 mmol) and HOSu (296 mg, 2.57 mmol) were dissolved in EtOAc (16 mL) and stirred at 0 °C for 10 min. DCC (530 mg, 2.57 mmol) was dissolved in EtOAc (8 mL), and the solution was added to the mixture of Fmoc-Gly-Pro-OH **31** and HOSu dropwise. The solution was stirred at rt for 14 h. The resulting DCU was removed by filtration, and the filtrate was concentrated in vacuo. The resulting oil was dissolved in a minimal amount of EtOAc, and hexanes was added. A white solid was collected by filtration (1.17 g, 93.1 %). ^{1}H NMR: δ 7.76 (d, *J* = 7.6, 2H), 7.61 (d, *J* = 7.3, 2H), 7.39 (t, *J* = 7.4, 2H), 7.31 (t, *J* = 7.4, 2H), 5.87 (m, 1H), 4.83 (dd, *J* = 7.8, 5.0, 1H), 4.36 (d, *J* = 7.7, 2H), 4.22 (t, *J* = 7.0, 1H), 4.11 (m, 1H), 4.00 (dd, *J* = 17.3, 3.5, 1H), 3.62 (m, 1H), 3.52 (m, 1H), 2.82 (s, 4H), 2.33 (m, 2H), 2.14 (m, 2H). ^{13}C NMR: δ 168.8, 167.6, 167.4, 156.3, 144.0, 141.3, 127.8, 127.2, 125.3, 120.0, 67.3, 56.9, 47.2, 46.0, 43.4, 29.4, 25.7, 24.8. HRMS calcd. for $C_{26}H_{26}N_3O_7$ (MH^+) *m/z* = 492.1771, found *m/z* = 492.1798.

Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH, 25. Fmoc-Gly-Pro-OSu **32** (839 mg, 1.71 mmol) was dissolved in DMF (20 mL). H-Hyp(^{t}Bu)-OH (320 mg, 1.71 mmol) and DIEA (0.15 mL, 0.85 mmol) were added, and the solution was stirred at rt for 4 h. The solution was diluted with a mixture of EtOAc and 5 % $NaHCO_3$ (1:1 ratio, 20 mL), and the organic layer was removed. The aqueous layer was acidified to pH 1 with 1 M HCl, and extracted with EtOAc (4 × 50 mL). The organic layers were combined, washed with H_2O (2 × 50 mL), dried over

Na_2SO_4, and concentrated in vacuo. The resulting pale yellow oil was purified by flash chromatography (2 % MeOH, 0.5 % AcOH in $CHCl_3$), and a white solid was obtained (626 mg, 65.1 %). ^{1}H NMR: δ 7.75 (d, *J* = 7.5, 2H), 7.61 (dd, *J* = 7.2, 4.4, 2H), 7.38 (t, *J* = 7.4, 2H), 7.30 (tt, *J* = 7.4, 1.1, 2H), 5.99 (dd, *J* = 5.6, 3.6, 1H), 4.64 (m, 2H), 4.36 (m, 3H), 4.22 (t, *J* = 7.3, 1H), 4.12 (dd, *J* = 17.5, 5.7, 1H), 3.94 (dd, *J* = 17.1, 3.8, 1H), 3.73 (dd, *J* = 9.7, 7.0, 1H), 3.62 (m, 1H), 3.56 (dd, *J* =9.9, 6.4, 1H), 3.47 (m, 1H), 2.27 (m, 1H), 2.15 (m, 3H), 1.99 (m, 2H), 1.18 (s, 9H). ^{13}C NMR: δ 173.5, 172.1, 167.4, 156.5, 144.0, 141.4, 127.8, 127.2, 125.3, 120.1, 74.5, 69.6, 67.3, 58.2, 58.1, 53.3, 47.2, 46.4, 43.3, 35.7, 28.35, 38.33, 24.8. HRMS calcd. for $C_{31}H_{38}N_3O_7$ (MH^+) *m/z* = 564.2710, found *m/z* = 564.2711.

Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OSu, 33. Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23** (289 mg, 0.765 mmol) and HOSu (88.4 mg, 0.765 mmol) were dissolved in EtOAc (10 mL) and stirred at 0 °C for 10 min. DCC (158 mg, 0.765 mmol) was dissolved in EtOAc (5 mL), and the solution was added to the mixture of Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23** and HOSu dropwise. The solution was stirred at rt for 14 h. The resulting DCU was removed by filtration, and the filtrate was concentrated in vacuo. The resulting oil was dissolved in a minimal amount of EtOAc, and hexanes was added. A white solid was obtained by filtration (308 mg, 84.6 %). ^{1}H NMR: δ 7.76 (d, *J* = 7.7, 2H), 7.62 (d, *J* = 7.4, 2H), 7.40 (t, *J* = 7.2, 2H), 7.30 (t, *J* = 7.2, 2H), 5.80 (m, 1H), 5.03 (m, 1H), 4.37 (d, *J* = 7.4, 2H), 4.23 (t, *J* = 7.1, 1H), 3.93 (m, 1H), 3.79 (m, 1H), 3.64 (t, *J* = 6.9,1H), 2.83 (s, 4H), 2.44 (m, 2H), 2.12 (m, 2H), 1.97 (m, 1H), 1.75 (m, 1H). ^{13}C NMR: δ 169.4, 169.3, 156.4, 144.2, 142.9, 141.4, 127.8, 127.1, 125.4, 121.8, 120.1, 66.9, 47.4, 46.7, 40.6, 30.2, 29.2, 25.8, 25.3. HRMS calcd. for $C_{27}H_{27}N_2O_6$ (MH^+) *m/z* = 475.1869, found *m/z* = 475.1846.

Tripeptide isostere, 34. Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OSu **33** (298 mg, 0.626 mmol) was dissolved in DMF (10 mL). H-Hyp(^{t}Bu)-OH (118 mg, 0.626 mmol) and DIEA (55 μL, 0.32 mmol) were added, and the solution was stirred at rt for 4 h. The solution was diluted with a mixture of EtOAc and 5 % $NaHCO_3$ (1:1 ratio, 20 mL), and the organic layer was removed. The aqueous layer was acidified to pH 1 with 1 M HCl, and extracted with EtOAc (4 × 25 mL). The organic layers were combined, washed with H_2O (2 × 25 mL), dried over Na_2SO_4, and concentrated in vacuo. The resulting yellow oil was purified by flash chromatography (5 % MeOH in $CHCl_3$), and a pale yellow solid was obtained (137 mg, 39.9 %). ^{1}H NMR: δ 7.74 (d, *J* = 7.3, 2H), 7.63 (d, *J* = 7.7, 2H), 7.38 (t, *J* = 7.4, 2H), 7.30 (, t, *J* = 7.4, 2H), 6.17 (t, *J* = 4.8, 1H), 4.73 (dd, *J* = 8.8, 5.5, 0.5H), 4.65 (dd, *J* = 8.4, 5.2, 0.5H), 4.35 (m, 1H), 4.25 (m, 2H), 3.84 (m, 0.5H), 3.74 (m, 0.5H), 3.55 (dd, *J* = 10.9, 5.5, 1H), 3.42 (dt, *J* = 7.4, 2.6, 1H), 3.32 (m, 2H), 2.60 (m, 2H), 2.37 (m, 5H), 1.92 (m, 3H), 1.17 (s, 3H), 1.14 (s, 6H). ^{13}C NMR: δ 176.8, 171.3, 168.6, 156.9, 145.7, 144.6, 141.4, 127.7, 127.2, 125.5, 120.0, 74.6, 69.2, 66.7, 54.9, 47.3, 38.8, 36.3, 35.5, 34.3, 29.3, 28.4, 23.2, 20.9. Rotamers: 2:1. MS calcd. for $C_{32}H_{39}N_2O_6$ (MH^+) *m/z* = 547.3, found *m/z* = 547.3.

Fmoc-Hyp(^{t}Bu)-OBn, 35. Fmoc-Hyp(^{t}Bu)-OH (1.77 g, 4.31 mmol) was dissolved in CH_2Cl_2 (80 mL) at 0 °C. EDC (1.66 g, 8.64 mmol), HOBt (1.33 g, 8.69 mmol), DMAP (24 mg, 0.19 mmol) and DIEA (3.0 mL, 17 mmol) were added. The mixture was stirred at 0 °C for 10 min, and BnOH (0.90 mL, 8.7 mmol) was added. The mixture was stirred at rt for 8 h. The solution was concentrated in vacuo, and EtOAc (150 mL) was added to dilute the mixture. The organic layer was washed with 1 M HCl (2 × 50 mL), $NaHCO_3$ (50 mL), and brine (50 mL), then dried over Na_2SO_4 and concentrated. Flash chromatography with 20 % EtOAc in hexanes gave a clear colorless oil (2.3 g, 100 %). ^{1}H NMR: δ 7.77 (d, *J* = 7.5, 1H),

7.75 (d, *J* = 7.5, 1H), 7.58 (m, 2H), 7.34 (m, 9H), 5.23 (d, *J* = 12.3, 0.5H), 5.17 (d, *J* = 12.3, 0.5H), 5.13 (d, *J* = 12.3, 0.5H), 5.08 (d, *J* = 12.3, 0.5H), 4.54 (dd, *J* = 9.0, 4.6, 0.5 H), 4.48 (dd, *J* = 9.0, 4.6, 0.5 H), 4.37 (m, 1H), 4.28 (m, 2.5H), 4.01 (app. t, *J* = 7.0, 0.5H), 3.81 (dd, *J* = 12.4, 6.2, 0.5H), 3.78 (dd, *J* = 13.3, 6.3, 0.5H), 3.36 (dd, *J* = 10.6, 5.3, 1H), 2.16 (m, 2H), 1.18 (s, 4.5H), 1.16 (s, 4.5H). ^{13}C NMR: δ 177.7 (172.6), 155.1, 144.4 (144.2), 144.1, 141.4, 128.7 (128.6), 128.4 (128.3), 127.8 (127.7), 127.2 (127.1), 125.3 (125.1), 120.12 (120.10), 120.05 (120.03), 74.3, 69.3 (68.4), 67.8 (67.7), 67.1 (67.0), 58.1 (57.9), 53.9 (53.4), 47.3, 38.8 (37.8), 28.4 (28.3). Rotamers: 1:1.

H-Hyp(^{t}Bu)-OBn, 36. Fmoc-Hyp(^{t}Bu)-OBn **35** (1.14 g, 2.28 mmol) was dissolved in CH_2Cl_2 (20 mL), and piperidine (5.0 mL) was added. The solution was stirred at rt for 40 min, and concentrated in vacuo. The crude product was purified by flash chromatography (1:2 EtOAc:hexanes, followed by 5 % MeOH in $CHCl_3$), and a clear oil was obtained (501 mg, 79.3 %). ^{1}H NMR: δ 7.34 (m, 5H), 5.14 (d, *J* = 13.3, 1H), 5.09 (d, *J* = 12.2, 1H), 4.26 (tt, *J* = 6.6, 6.6, 1H), 3.54 (dd, *J* = 8.9, 5.9, 1H), 3.29 (dd, *J* = 9.4, 6.8, 1H), 2.51 (dd, *J* = 9.2, 6.0, 1H), 2.07 (m, 2H), 1.13 (s, 9H). ^{13}C NMR: δ 174.0 (175.0), 136.1 (135.9), 128.6 (128.7), 128.29 (128.4), 128.26 (128.3), 73.7 (73.9), 69.8, 66.3 (66.8), 63.2, 59.8 (58.9), 38.3 (39.1), 28.48 (28.49). Rotamers: 5:1.

Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OBn, (*R*)-37 and (*S*)-37. Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-OH **23** (424 mg, 1.12 mmol) and HBTU (686 mg, 1.81 mmol) were dissolved in CH_2Cl_2 (40 mL) and stirred at 0 °C for 10 min, then H-Hyp(^{t}Bu)-OBn **36 (**501 mg, 1.81 mmol) was added and stirred at rt for 8 h. The solvent was removed in vacuo and the resulting slurry was dissolved in EtOAc (200 mL). The organic layer was washed with 0.5 M HCl (2 × 20 mL), $NaHCO_3$ (25 mL), and brine (25 mL), dried over Na_2SO_4, and concentrated. Flash

chromatography with 2 % MeOH in $CHCl_3$ gave a clear colorless oil (416 mg, 68 %). The mixture of Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OBn diastereomers was dissolved in a minimum amount of 4 % *i*-PrOH in hexanes. Analytical HPLC was performed by injecting 20 μL in 5 % *i*-PrOH/hexanes on a 5 μm silica column (100 × 4.6 mm) with isocratic 5 % *i*-PrOH/hexanes at 1.0 mL/min. Preparative HPLC was performed by injecting 1 mL in 5 % *i*-PrOH/hexanes on a 10 μm silica column (100 × 50 mm) with isocratic 4 % *i*-PrOH/hexanes at 40 mL/min. Two diastereomers were obtained as white solids. **Diastereomer (*R*)-37,** (133 mg, 32 % recovery). ^{1}H NMR: δ 7.76 (d, *J* = 7.4, 2H), 7.60 (d, *J* = 7.4, 2H), 7.35 (m, 9H), 5.35 (m, 1H), 5.22 (d, *J* = 12.4, 1H), 5.14 (d, *J* = 12.5, 1H), 4.85 (m, 1H), 4.67 (dd, *J* = 8.4, 6.4, 1H), 4.36 (m, 2H), 4.30 (dd, *J* = 9.7, 5.0, 1H), 4.20 (app. t, *J* = 7.0, 1H), 3.88 (m, 1H), 3.82 (dd, *J* = 10.4, 5.6, 1H), 3.75 (m, 1H), 3.49 (dd, *J* = 10, 3.6, 1H), 3.40 (m, 1H), 2.40 (m, 2H), 2.18 (m, 1H), 2.00 (m, 4H), 1.62 (m, 1H), 1.14 (s, 9H). ^{13}C NMR: δ 173.0, 172.3, 156.3, 145.9, 144.0, 141.3, 135.8, 128.6, 128.3, 128.2, 127.7, 127.1, 125.2, 120.0, 118.3, 74.3, 69.5, 66.9, 66.7, 57.8, 54.6, 48.9, 47.2, 40.4, 37.3, 30.3, 29.3, 28.2, 24.7. HRMS calcd. for $C_{39}H_{45}N_2O_6$ (MH^+) *m/z* = 637.3278, found *m/z* = 637.3281. Anal. retention time 4.95 min. **Diastereomer (*S*)-37** (90 mg, 22 % recovery). ^{1}H NMR: δ 7.76 (d, *J* = 7.0, 2H), 7.60 (d, *J* = 7.4, 2H), 7.39 (m, 4H), 7.31 (m, 5H), 5.43 (m, 1H), 5.19 (d, *J* = 7.6, 1H), 5.14 (d, *J* = 7.6, 1H), 4.78 (m, 1H), 4.66 (dd, *J* = 8.5, 4.5, 1H), 4.35 (m, 3H), 4.21 (m, 1H), 3.89 (dd, *J* = 9.6, 6.8, 2H), 3.68 (m, 1H), 3.38 (dd, *J* = 9.8, 6.0, 2H), 2.40 (m, 2H), 2.12 (m, 2H), 1.95 (m, 3H), 1.63 (m, 1H), 1.15 (s, 9H). ^{13}C NMR: δ 173.0, 172.2, 156.4, 145.7, 144.1, 141.4, 135.8, 128.7, 128.4, 128.2, 127.7, 127.1, 125.2, 120.1, 118.3, 74.3, 69.4, 66.9, 66.7, 57.6, 54.1, 48.8, 47.3, 40.5, 37.0, 30.2, 29.2, 28.3, 24.7. HRMS calcd. for $C_{39}H_{45}N_2O_6$ (MH^+) *m/z* = 637.3278, found *m/z* = 637.3270. Anal. retention time 3.91 min.

Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OH, (*R*)-26 and (*S*)-26. Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OBn **(*S*)-37** (29 mg, 0.046 mmol) was dissolved in anhydrous EtOH (2 mL) and 10 % Pd/C (7.0 mg) was added. 1,4-Cyclohexadiene (50 μL, 0.53 mmol) was added and the solution was heated at reflux at 90 °C for 45 min. After cooling to rt, the solution was filtered through Celite™, and the solvent was removed in vacuo. Flash chromatography with 4 % MeOH and 0.5 % HOAc in $CHCl_3$ gave a white solid (21 mg, 85 %). **Isostere (*S*)-26:** ^{1}H NMR: δ 7.75 (d, *J* = 7.4, 2H), 7.60 (d, *J* = 7.1, 2H), 7.38 (t, *J* = 7.4, 2H), 7.29 (t, *J* =7.5, 2H), 5.66 (br s, 1H), 5.47 (m, 1H), 4.65 (br s, 1H), 4.34 (m, 3H), 4.21 (m, 1H), 3.84 (m, 1H), 3.73 (m, 1H), 3.60 (m, 1H), 3.39 (m, 2H), 2.33 (m, 4H), 1.95 (m, 3H), 1.62 (m, 1H), 1.16 (s, 9H). ^{13}C NMR: δ175.3, 173.8, 156.8, 145.0, 144.2, 141.4, 139.3, 130.1, 127.7, 127.2, 125.3, 120.0, 119.3, 74.5, 69.4, 66.9, 54.6, 48.9, 47.3, 40.5, 36.2, 31.9, 30.6, 29.3, 28.7, 28.3, 24.8. HRMS calcd. for $C_{32}H_{39}N_2O_6$ (MH^+) *m/z* = 547.2808, found *m/z* = 547.2825. **(*R*)-26** was prepared by a similar method to give 15 mg of a white solid (75 %). **Isostere (*R*)-26:** ^{1}H NMR: δ 7.76 (d, *J* = 7.6, 2H), 7.59 (d, *J* = 7.2, 2H), 7.40 (t, *J* = 7.6, 2H), 7.31 (dt, *J* = 7.6, 1.2, 2H), 5.34 (br s, 1H), 5.04 (br s, 1H), 4.69 (m, 1H), 4.36 (m, 2H), 4.21 (m, 1H), 3.87 (m, 1H), 3.75 (m, 1H), 3.50 (m, 1H), 3.42 (m, 1H), 2.64 (m, 1H), 2.40 (m, 3H), 2.14 (m, 1H), 1.98 (m, 3H), 1.63 (m, 1H), 1.18 (s, 9H). ^{13}C NMR: δ 175.5, 173.8, 156.5, 145.3, 144.1, 141.4, 127.8, 127.1, 125.2, 120.1, 119.0, 74.6, 69.3, 66.8, 54.9, 48.9, 47.3, 40.4, 36.4, 30.6, 29.4, 28.4, 28.3, 24.8. HRMS calcd. for $C_{32}H_{39}N_2O_6$ (MH^+) *m/z* = 547.2808, found *m/z* = 547.2834.

4,5-(Bis[bis-diethylamino]phosphonito)-9,9-dimethylxanthene, 38. By the method of Goertz *et al.*[252] 9,9-Dimethyl xanthene (292 mg, 1.39 mmol) and *N,N,N',N'*-tetramethylethylendiamine (0.52 mL, 3.5 mmol) were dissolved in Et_2O (5.0 mL), and stirred

at –40 °C for 20 min. Then *n*-BuLi (1.39 mL, 3.48 mmol) was added dropwise, and the resulting red solution was stirred at rt for 16 h. Chloro(bis-diethylamino)phosphane (0.73 mL, 3.5 mmol) was dissolved in pentane (3.0 mL), and stirred at –78 °C for 10 min. The solution was transferred into the mixture via cannula dropwise at –78 °C. The yellow solution was stirred at rt for 16 h, and the resulting precipitates were removed by filtration. The filtrate was concentrated and dried in vacuo. A yellow solid was obtained (597.8 mg, 76.9 %). ^{1}H NMR: δ 7.27 (m, 3H), 6.96 (m, 3H), 3.00 (m, 16H), 1.52 (m, 6H), 1.01 (app. qt, *J* = 6.9, 18H), 0.93, (t, *J* = 7.0, 6H). ^{13}C NMR: δ 150.41, 150.39, 130.0, 129.8, 127.3, 127.2, 126.1, 125.7, 123.0, 122.9, 122.5, 122.0, 116.7, 116.3, 45.45, 45.34, 45.29, 45.18, 43.9, 43.7, 43.2, 43.1, 43.0, 38.96, 38.93, 38.90, 38.87, 37.74, 37.67, 33.91, 33.90, 33.01, 32.98, 32.4, 32.24, 32.21, 30.8, 15.15, 15.11, 14.79, 14.77, 14.49, 14.47, 14.46, 14.40, 14.37, 14.27, 14.25.

Diphosphonite ligand, 39. 4,5-(Bis[bis-diethylamino]phosphonito)-9,9-dimethyl-xanthene **38** (597.8 mg, 1.07 mmol) and (*S*)-BINOL (612.9 mg, 2.14 mmol) were dissolved in toluene (20 mL), and a catalytic amount of tetrazole was added. The mixture was heated at reflux for 16 h, and the solvent was removed in vacuo. The resulting slurry was dissolved in CH_2Cl_2 (1.0 mL), and CH_3CN (10 mL) was added. A pale yellow solid was collected by filtration, and dried in vacuo (889 mg, 99 %). ^{1}H NMR: δ 7.71 (m, 9H, 7.18 (m, 21H), 1.51 (m, 6H).

Ester, 40. Fmoc-Phe-OH (50.4 mg, 0.130 mmol) was dissolved in CH_2Cl_2 (15 mL). HBTU (49.5 mg, 0.131 mmol), HOBt (19.9 mg, 0.130 mmol), DMAP (0.7 mg, 0.006 mmol), and DIEA (57 μL, 0.33 mmol) were added, and the solution was stirred at 0 °C for 10 min. Alcohol **19** (21 mg, 0.092 mmol) was added, and the solution was stirred at rt for 14 h. The resulting solution was concentrated in vacuo, diluted with EtOAc (50 mL), washed with 1 M

HCl (2 × 25 mL), H_2O (25 mL), $NaHCO_3$ (2 × 25 mL), and brine (25 mL), dried over Na_2SO_4, and purified by flash chromatography (EtOAc:Hexanes = 1:5). A colorless oil was obtained (23.3 mg, 42.3 %). ^{1}H NMR: δ 7.77 (d, *J* = 7.6, 2H), 7.56 (d, *J* = 6.6, 2H), 7.41 (t, *J* = 7.4, 2H), 7.31 (t, *J* = 7.5, 2H), 7.26 (m, 3H), 7.18 (d, *J* = 7.3, 1H), 7.13 (d, *J* = 6.7, 1H), 5.61 (d, *J* = 10.9, 1H), 5.52 (m, 0.5H), 5.44 (m, 0.5H), 5.30 (dd, *J* = 12.0, 8.0, 1H), 4.85 (m, 0.5H), 4.65 (dd, *J* = 15.6, 7.3, 1H), 4.54 (m, 0.5H), 4.44 (m, 1H), 4.33 (m, 1H), 4.21 (t, *J* = 6.9, 1H), 3.48 (m, 0.5H), 3.38 (m, 0.5H), 3.14 (m, 3H), 2.29 (m, 4H), 1.86 (m, 2H), 1.41 (s, 9H). ^{13}C NMR: δ 171.1, 155.9, 143.9, 141.4, 139.4, 135.9, 135.8, 129.5, 129.4, 128.84, 128.76, 127.9, 127.4, 127.2, 127.2, 125.3, 125.2, 120.1, 79.6, 74.1, 74.0, 67.2, 55.3, 55.1, 47.2, 42.9, 42.7, 38.5, 38.2, 32.5, 32.1, 32.0, 28.5, 23.2.

Amide, 41. Boc-Pro-OH (150 mg, 0.696 mmol) was dissolved in CH_2Cl_2 (25 mL). HBTU (398 mg, 1.05 mmol), DMAP (5.5 mg, 0.045 mmol), and DIEA (0.49 mL, 2.8 mmol) were added, and the solution was stirred at 0 °C for 10 min. (*1R,2R*)-(–)-Pseudoephedrine (173 mg, 1.05 mmol) was added, and the solution was stirred at rt for 14 h. The resulting solution was concentrated in vacuo, diluted with EtOAc (50 mL), washed with 1 M HCl (2 × 25 mL), H_2O (25 mL), $NaHCO_3$ (2 × 25 mL) and brine (25 mL), dried over Na_2SO_4, and purified by flash chromatography (3 % MeOH in $CHCl_3$). A colorless oil was obtained (260 mg, 100 %). ^{1}H NMR: δ 7.31 (m, 5H), 4.67 (t, *J* = 7.6, 0.5H), 4.54 (m, 2H), 4.40 (m, 0.5H), 4.02 (m, 0.5H), 3.62 (m, 0.5H), 3.54 (dt, *J* = 10.4, 6.8, 0.5H), 3.40 (dt, *J* = 10.3, 6.8, 1H), 2.97 (s, 2H), 2.78 (s, 1H), 2.10 (m, 1H), 2.00 (m, 1.5H), 1.79 (m, 1.5H), 1.42 (s, 6H), 1.40 (s, 3H), 1.26 (d, *J* = 7.0, 1H), 1.01 (d, *J* = 5.8, 2H). ^{13}C NMR: δ 174.5 (175.1), 154.9 (153.8), 142.7 (141.8), 128.3 (127.6), 128.2 (127.4), 126.9 (125.8), 80.1 (79.7), 75.7 (76.1), 57.4, 56.3, 47.0 (46.6), 29.3 (30.1), 28.5 (28.3), 24.8, 23.1, 14.38 (14.44). Rotamers: 7:4:1.

H-Pro-OBn•TsOH, 44. H-Pro-OH (3.22 g, 28.0 mmol) was dissolved in a mixture of CCl_4 (25 mL) and BnOH (20 mL). TsOH (6.39 g, 33.6 mmol) was added. The solution was heated at reflux for 24 h, and cooled to rt. The solution was concentrated in vacuo, and the remaining BnOH was removed by Et_2O wash. After drying under high vacuum, a colorless oil (9.4 g, 100 %) was obtained. ^{1}H NMR: δ 7.74 (d, *J* = 8.0, 2H), 7.31 (m, 5H), 7.14 (d, *J* = 7.9, 2H), 5.16 (d, *J* = 12.2, 1H), 5.08 (d, *J* = 12.2, 1H), 4.56 (m, 1H), 3.48 (m, 2H), 2.34 (s, 3H), 1.97 (m, 4H). ^{13}C NMR: δ 169.0, 141.3, 140.7, 134.7, 129.1, 128.7, 128.5, 126.1, 68.3, 59.6, 46.5, 28.8, 23.6, 21.5. HRMS calcd. for $C_{12}H_{16}NO_2$ (MH^+) *m/z* = 206.1181, found *m/z* = 206.1192.

Boc-Pro-Pro-OBn, 45. Boc-Pro-OH (6.62 g, 30.8 mmol) was dissolved in CH_2Cl_2 (200 mL). HBTU (11.66 g, 30.7 mmol), HOBt (4.29 g, 30.8 mmol), DMAP (171.2 mg, 1.4 mmol), and DIEA (19.5 mL, 0.11 mol) were added, and the solution was stirred at 0 °C for 10 min. H-Pro-OBn **44** (9.4 g, 28 mmol) was added, and the solution was stirred at rt for 14 h. The resulting solution was concentrated in vacuo, diluted with EtOAc (250 mL), washed with 1 M HCl (2 × 75 mL), H_2O (75 mL), $NaHCO_3$ (2 × 75 mL), and brine (75 mL), dried over Na_2SO_4, and purified by flash chromatography (33 % EtOAc in hexanes). A colorless oil was obtained (6.40 g, 56.8 %). ^{1}H NMR: δ 7.33 (m, 5H), 5.22 (d, *J* = 12.2, 0.5H), 5.21 (d, *J* = 12.4, 0.5H), 5.06 (d, *J* = 12.2, 0.5H), 5.04 (d, *J* = 12.3, 0.5H), 4.64 (m, 1H), 4.49 (dd, *J* = 8.4, 2.9, 0.5H), 4.37 (dd, *J* = 8.4, 4.2, 0.5H), 3.76 (ddd, *J* = 9.2, 7.3, 7.3, 0.5H), 3.58 (m, 2.5H), 3.42 (m, 1H), 2.01 (m, 7H), 1.80 (m, 1H), 1.45 (s, 5H), 1.38 (s, 4H). ^{13}C NMR: δ 172.3 (172.0), 171.2 (171.7), 154.7 (153.8), 135.75 (135.67), 128.61 (128.59), 128.30 (128.36), 128.22 (128.26), 79.6 (79.5), 67.0 (66.8), 58.9, 57.8 (57.7), 46.9 (46.7), 46.5 (46.6), 29.1 (30.0), 28.8 (28.9), 28.6 (28.4), 25.0 (25.1), 24.1 (23.6). Rotamers: 5:4.

Fmoc-Gly-Pro-Pro-OBn, 46. Boc-Pro-Pro-OBn **45** (6.40 g, 15.9 mmol) was dissolved in CH_2Cl_2 (100 mL), and TFA (30 mL) was added. The solution was stirred at rt for 30 min, and the solvent was removed in vacuo. The resulting slurry was dissolved in CH_2Cl_2 (200 mL), and HBTU (7.35 g, 19.4 mmol), HOBt (2.92 g, 19.1 mmol), DMAP (97 mg, 0.79 mmol) and DIEA (11.1 mL, 63.6 mmol) were added, and stirred at 0 °C for 20 min. Fmoc-Gly-OH (5.50 g, 18.5 mmol) was added, and the solution was stirred at rt for 14 h. The resulting solution was concentrated in vacuo, diluted with EtOAc (250 mL), washed with 1 M HCl (2 × 100 mL), H_2O (100 mL), $NaHCO_3$ (2 × 100 mL), and brine (100 mL), dried over Na_2SO_4, and purified by flash chromatography (2 % MeOH in $CHCl_3$). A white solid (8.67 g, 93.8 %) was obtained. ^{1}H NMR: δ 7.75 (d, *J* = 7.4, 2H), 7.59 (dd, *J* = 7.4, 3.0, 2H), 7.39 (t, *J* = 7.3, 2H), 7.33 (m, 7H), 5.75 (t, *J* = 4.0, 1H), 5.21 (d, *J* = 12.3, 1H), 5.06 (d, *J* = 12.2, 1H), 4.66 (ddd, *J* = 15.4, 8.0, 3.6, 2H), 4.35 (m, 2H), 4.21 (t, *J* = 7.3, 1H), 4.10 (dd, *J* = 17.3, 5.4, 1H), 3.97 (dd, *J* = 17.3, 3.6, 1H), 3.80 (m, 1H), 3.59 (m, 2H), 3.44 (m, 1H), 2.16 (m, 3H), 1.98 (m, 5H). ^{13}C NMR: δ 172.0, 170.2, 166.9, 156.3, 144.0, 141.4, 135.7, 128.6, 128.4, 128.2, 127.8, 127.2, 125.3, 120.0, 67.2, 67.0, 59.0, 58.2, 47.2, 46.8, 46.3, 43.4, 28.9, 28.2, 25.0, 24.6.

Fmoc-Gly-Pro-Pro-OH, 42. Fmoc-Gly-Pro-Pro-OBn **46** (8.56 g, 14.7 mmol) was dissolved in a mixture of MeOH (75 mL) and AcOH (75 mL), and 10 % Pd/C (520 mg) was added. H_2 was introduced with a balloon, and the suspension was stirred at rt for 2 h. The solution was filter through Celite™, and concentrated in vacuo. The resulting oil was purified by flash chromatography (5 % MeOH in $CHCl_3$), and a white solid was obtained (5.42 g, 75.0 %). ^{1}H NMR: δ 7.74 (d, *J* = 7.4, 2H), 7.60 (t, *J* = 6.6, 2H), 7.37 (t, *J* = 7.4, 2H), 7.28 (app. t, *J* =7.4, 2H), 5.95 (m, 1H), 4.62 (m, 1H), 4.56 (app. dd, *J* = 8.2, 3.2, 1H), 4.35 (d, *J* = 7.4, 2H), 4.20

(t, *J* = 7.1, 1H), 4.09 (dd, *J* = 17.1, 6.1, 1H), 3.90 (dd, *J* = 17.0, 3.0, 1H), 3.76 (dd, *J* =16.2, 7.2, 1H), 3.57 (m, 2H), 3.42 (m, 1H), 2.15 (m, 2H), 2.01 (m, 5H), 1.90 (m, 1H). ^{13}C NMR: δ 174.0, 171.1, 167.4, 156.5, 143.9 (143.8), 141.2, 127.6, 127.0, 125.2 (125.1), 120.0, 67.0, 59.1, 58.1, 47.0, 46.9, 46.3, 43.2, 38.6, 28.5, 28.1, 24.8 (24.6). HRMS calcd. for $C_{27}H_{30}N_3O_6$ (MH^+) *m/z* = 492.2135, found *m/z* = 492.2130.

Boc-Pro-N(OMe)Me, 47. Boc-Pro-OH (6.41 g, 29.8 mmol), *N*,*O*-dimethyl hydroxylamine hydrochloride (5.82 g, 59.6 mmol), HOBt (5.47 g, 35.7 mmol), and DCC (7.41 g, 35.9 mmol) were dissolved in CH_2Cl_2 (300 mL) and cooled to 0 °C. DIEA (20.8 mL, 119 mmol) and DMAP (182 mg, 1.48 mmol) were added, and the reaction was stirred for 24 h. The mixture was filtered to remove dicyclohexyl urea and concentrated. The resulting slurry was diluted with EtOAc (300 mL), and washed with NH_4Cl (2 × 100 mL), $NaHCO_3$ (2 × 100 mL), and brine (100 mL). The organic layer was dried over $MgSO_4$, concentrated in vacuo, and purified by flash chromatography (20 % EtOAc in hexanes). A pale yellow solid was obtained (5.75 g, 74.8 %). mp 46-47 °C. ^{1}H NMR: δ 4.64 (dd, *J* = 8.3, 2.9, 0.5H), 4.55 (dd. *J* = 8.3, 3.7, 0.5H), 3.72 (s, 1.5H), 3.66 (s, 1.5H), 3.51 (m, 1H), 3.35 (m, 1H), 3.14 (s, 3H), 2.14 (m, 1H), 1.80 (m, 3H), 1.40 (s, 4.5H), 1.36 (s, 4.5H). ^{13}C NMR: δ 173.9 (173.3), 153.9 (154.5), 79.5 (79.4), 61.3 (61.4), 56.8 (56.5), 46.9 (46.6), 32.4 (32.3), 30.5 (29.6), 28.4 (28.5), 23.4 (24.1). HRMS calcd. for $C_{12}H_{23}N_2O_4$ (MH^+) *m/z* = 259.1658, found *m/z* = 259.1655. Rotamers: 1:1. Anal. HPLC: retention time 8.95 min (> 97 % purity).

α,β-Unsaturated ketone, 48. 1-Iodocyclopentene **13** (4.35 g, 22.4 mmol) was dissolved in THF (80 mL) at –40 °C, and *s*-BuLi (1.4 M in cyclohexane, 32.0 mL, 44.8 mmol) was added. The reaction was stirred at –40 °C for 3 h to generate cyclopentenyl lithium. In another flask, Boc-Pro-N(Me)OMe **47** (3.84 g, 14.9 mmol) was dissolved in THF (30 mL). The

solution was cooled to –78 °C, and the cyclopentenyl lithium solution was added via cannula. The mixture was stirred at –78 °C for 1 h, warmed up slowly to rt, and stirred for another 10 h. The reaction was quenched with NH_4Cl (10 mL). The resulting solution was diluted with EtOAc (100 mL), washed with NH_4Cl (2 × 20 mL), $NaHCO_3$ (20 mL), brine (20 mL), dried over $MgSO_4$, and concentrated. Chromatography with 10 % EtOAc in hexanes gave a pale yellow solid (1.91 g. 48.7 %). mp 45-49 °C. ^{1}H NMR: δ 6.82 (m, 0.5H), 6.76 (m, 0.5H), 4.95 (dd, *J* = 9.0, 3.6, 0.5H), 4.77 (dd, *J* = 8.8, 4.4, 0.5H), 3.53 (m, 1.5H), 3.41 (m, 0.5H), 2.58 (m, 4H), 2.18 (m, 1H), 1.89 (m, 5H), 1.44 (s, 4H), 1.33 (s, 5H). ^{13}C NMR: δ 197.6 (197.1), 154.0 (154.6), 144.12 (143.52), 143.34 (143.47), 79.7 (79.6), 62.0 (61.5), 46.7 (46.9), 34.32 (34.35), 31.3 (31.1), 30.3, 28.4 (28.6), 23.8 (24.3), 22.63 (22.55). HRMS calcd. for $C_{15}H_{24}NO_3$ (MH^+) *m/z* = 266.1756, found *m/z* = 266.1759. Anal. HPLC: retention time 2.93 min (> 99 % purity). Rotamers: 5:4.

Alcohol 49: Ketone **48** (1.20 g, 4.52 mmol) was dissolved in THF:MeOH (2.5:1, 65 mL) and cooled to 0 °C. $CeCl_3 \cdot 7H_2O$ (4.21 g, 11.3 mmol) was added, followed by $NaBH_4$ (860 mg, 22.7 mmol). After stirring at 0 °C for 2.5 h, the reaction was quenched with NH_4Cl (50 mL). The resulting solution was diluted with EtOAc (150 mL), washed with NH_4Cl (2 × 50 mL), and brine (50 mL), dried over $MgSO_4$, concentrated in vacuo, and purified by flash chromatography (20 % EtOAc in hexanes). Both **(*R*)-49** (596.1 mg, 49.4 %) and **(*S*)-49** (445.7 mg, 36.9 %) were obtained as pale yellow solids. **Diastereomer (*R*)-49:** $[\alpha]_D^{25}$ = –46.3° (*c* 0.36, $CHCl_3$). ^{1}H NMR: δ 5.62 (m, 1H), 4.49 (m, 1H), 4.10 (m, 1H), 3.46 (m, 1H), 3.20 (m, 1H), 2.31 (m, 4H), 1.86 (m, 6H), 1.48 (s, 9H). ^{13}C NMR: δ 165.4, 144.4, 127.3, 80.1, 73.9, 62.0, 48.1, 32.2, 32.1, 28.6, 27.5, 24.3, 24.0. HRMS calcd. for $C_{15}H_{26}NO_3$ (MH^+) *m/z* = 268.1913, found *m/z* = 268.1908. Anal. HPLC: retention time 2.03 min (> 98 %

purity). **Diastereomer (*S*)-49:** $[\alpha]_D^{25}$ = –62.6° (*c* 0.26, $CHCl_3$). ^{1}H NMR: δ 5.61 (m, 1H), 5.21 (br s, 1H), 4.15 (m, 1H), 3.98 (m, 1H), 3.42 (m, 1H), 3.33 (m, 1H), 2.52 (m, 1H), 2.31 (m, 2H), 2.21 (m, 1H), 1.83 (m, 5H), 1.61 (m, 1H), 1.48 (s, 9H). ^{13}C NMR: δ 158.2, 145.0, 128.6, 80.7, 76.1, 61.3, 47.5, 32.1, 30.1, 28.6, 28.5, 24.0, 23.4. HRMS calcd. for $C_{15}H_{26}NO_3$ (MH^+) *m/z* = 268.1913, found *m/z* = 268.1927. Anal. HPLC: retention time 1.90 min (> 98 % purity).

Cyclic carbamates, (*R*)-50 and (*S*)-50. Alcohol **(*S*)-49** (27 mg, 0.099 mmol) was dissolved in CH_2Cl_2 (3.0 mL), and TFA (1.0) mL was added. The solution was stirred at rt for 30 min, and the solvent was removed in vacuo. The resulting oil was dissolved in THF (3.0 mL), and carbonyl diimidazole (0.33 mL, 0.099 mmol) was added. The solution was stirred at rt for 24 h, concentrated in vacuo, diluted with EtOAc (50 mL), washed with 1 M HCl (2 × 10 mL), $NaHCO_3$ (2 × 10 mL), and brine (10 mL), dried over Na_2SO_4, and purified by flash chromatography (20 % EtOAc in hexanes). A colorless oil was obtained (17.8 mg, 92.6 %). **Diastereomer (*S*)-50:** $[\alpha]_D^{25}$ = –82.4° (*c* 0.64, $CHCl_3$).^{1}H NMR: δ 5.80 (m, 1H), 4.88 (d, *J* = 4.1, 1H), 4.18 (dd, *J* = 9.2, 4.4, 1H), 4.15 (dd, *J* = 9.2, 4.6, 1H), 3.66 (m, 1H), 2.38 (m, 4H), 2.06 (m, 2H), 1.94 (m, 3H), 1.53 (app. dt, *J* = 9.4, 3.5, 1H). ^{13}C NMR: δ 161.3, 141.2, 129.5, 79.6, 63.7, 45.9, 32.5, 30.9, 30.6, 25.9, 23.4. HRMS calcd. for $C_{11}H_{16}NO_2$ (MH^+) *m/z* = 194.1181, found *m/z* = 194.1186. Anal. HPLC: retention time 5.47 min (> 98 % purity). **(*R*)-50** was prepared by a similar method. **Diastereomer (*R*)-50:** ^{1}H NMR: δ 5.82 (dtt, *J* = 2.0, 2.0, 2.0, 1H), 5.22 (d, *J* = 7.8, 1H), 3.91 (ddd, *J* = 10.3, 7.7, 5.8, 1H), 3.67 (dt, *J* = 11.5, 7.9, 1H), 3.17 (ddd, *J* = 11.3, 9.5, 4.0, 1H), 2.37 (m, 2H), 2.27 (m, 2H), 1.94 (m, 3H), 1.70 (dddd, *J* = 12.6, 7.5, 5.5, 1.8, 1H), 1.55 (m, 1H), 1.42 (dddd, *J* = 12.4, 10.6, 10.6, 8.9, 1H). ^{13}C NMR: δ 166.9, 138.1, 127.8, 75.7, 63.1, 46.1, 32.8, 32.3, 25.8, 25.2, 23.6. HRMS calcd. for

$C_{11}H_{16}NO_2$ (MH^+) *m/z* = 194.1181, found *m/z* = 194.1187. Anal. HPLC: retention time 5.42 min (> 98 % purity).

Ester, 51. Alcohol **(*R*)-49** (263 mg, 0.984 mmol) was dissolved in CH_2Cl_2 (20 mL) and cooled to 0 °C. HBTU (559 mg, 1.47 mmol), DMAP (5.9 mg, 0.048 mmol), and 2-(*tert*-butyldimethylsilyloxy)acetic acid **17** (385 mg, 2.02 mmol) were added. The mixture was stirred at 0 °C for 10 min. DIEA (0.70 mL, 4.0 mmol) was added, and the solution was stirred at rt for 10 h. The solvent was removed in vacuo, and the resulting slurry was diluted with EtOAc (50 mL). The organic layer was washed with 1 M HCl (2 × 25 mL), $NaHCO_3$ (25 mL), brine (25 mL), dried over $MgSO_4$, and concentrated. Flash chromatography with 10 % EtOAc in hexanes gave a pale yellow solid (254 mg, 58.8 %). **Diastereomer (*R*)-51:** $[\alpha]_D^{25}$ = –67.5° (*c* 0.55, $CHCl_3$). ^{1}H NMR: δ 5.93 (br s, 0.5H), 5.90 (br s, 0.5H), 5.51 (m, 1H), 4.25 (d, *J* = 16.6, 1H), 4.19 (d, *J* = 16.8, 1H), 4.10 (m, 0.5H), 3.94 (m, 0.5H), 3.46 (m, 0.5H), 3.32 (m, 0.5H), 3.24 (m, 1H), 2.29 (m, 4H), 1.87 (m, 6H), 1.48 (s, 5H), 1.43 (s, 4H), 0.90 (s, 9H), 0.09 (s, 6H). ^{13}C NMR: δ 170.8 (170.7), 154.3 (154.6), 140.8 (140.9), 126.3 (126.8), 79.9 (79.4), 74.4 (73.9), 61.9 (62.0), 58.5 (58.3), 46.9 (47.0), 33.1 (32.9), 32.3, 28.6, 26.2 (25.6), 25.9, 23.9 (24.5), 23.3 (23.4), 18.5, –5.3 (–5.4). HRMS calcd. for $C_{23}H_{42}NO_5Si$ (MH^+) *m/z* = 440.2832, found *m/z* = 440.2805. Anal. HPLC: retention time 1.53 min (> 96 % purity). Rotamers: 5:4. **Diastereomer (*S*)-51:** ^{1}H NMR: δ 5.71 (br s, 0.5H), 5.33 (m, 1H), 5.53 (m, 0.5H), 4.20 (m, 2H), 4.11 (m, 0.5H), 3.99 (m, 0.5H), 3.43 (m, 0.5H), 3.33 (m, 0.5H), 3.23 (m, 1H), 2.29 (m, 4H), 1.81 (m, 6H), 1.47 (s, 4.5H), 1.43 (s, 4.5H), 0.89 (s, 9H), 0.07 (s, 6H). ^{13}C NMR: δ 170.7, 154.7, 140.4 (140.2), 129.6 (128.6), 79.9 (79.3), 73.7 (72.9), 61.84 (61.75), 58.6 (58.0), 46.6, 33.3 (32.2), 32.3, 28.6, 27.2, 25.9, 23.5 (24.8), 23.4 (23.2), 18.5, –5.36. Rotamers: 1:1.

α-Hydroxy acid, 52. A solution of LDA (2.0 M in THF, 1.2 mL, 2.3 mmol) was diluted with THF (8.0 mL) and stirred at –100 °C for 30 min. A mixture of Me_3SiCl (0.80 mL, 6.4 mmol) and pyridine (0.56 mL, 6.9 mmol) in THF (4.0 mL), cooled to –100 °C, was added dropwise to the LDA solution. After 5 min, a solution of ester **51** (254 g, 0.581 mmol) in THF (4.0 mL), cooled to –100 °C was added dropwise and the reaction was stirred at –100 °C for 25 min, then warmed slowly over 2.5 h to rt, and stirred for 10 h. Then the mixture was heated to 45 °C for 1.5 h. The reaction was quenched with 1 M HCl (10 mL), and the aqueous layer was extracted with Et_2O (2 × 100 mL). The organic layer was dried over $MgSO_4$ and concentrated to give 271 mg of a yellow oil. Without further purification, the product was dissolved in THF (10 mL). *n*-Bu_4NF (547 mg, 1.73 mmol) in THF (4.0 mL) was added and stirred for 5 min at 0 °C, then at rt for 3 h. The reaction was quenched with 0.5 M HCl (10 mL). The solution was extracted with EtOAc (50 mL), dried over $MgSO_4$ and concentrated. The resulting oil was purified by flash chromatography with 5 % MeOH in $CHCl_3$ and a yellow oil (118 mg, 62.9 %) was obtained. **Diastereomer (*R*)-52:** ^{1}H NMR (CD_3OD): δ 5.30 (d, *J* = 8.5, 1H), 4.33 (m, 1H), 4.07 (m, 1H), 3.34 (t, *J* = 6.9, 1H), 3.23 (m, 0.5H), 3.12 (m, 0.5H), 2.86 (m, 1H), 2.54 (br s, 1H), 2.17 (m, 1H), 2.04 (m, 1H), 1.84 (m, 4H), 1.69 (m, 3H), 1.56 (m, 1H), 1.41 (s, 9H). **Diastereomer (*S*)-52:** ^{1}H NMR (DMSO-d_6): δ 5.35 (m, 1H), 5.02 (m, 1H), 4.21 (m, 1H), 3.73 (m, 1H), 3.26 (m, 3H), 2.24 (m, 1H), 2.00 (m, 2H), 1.75 (m, 6H), 1.52 (m, 1H), 1.37 (s, 9H).

Boc-Pro-Ψ[(*E*)CH=C]-Pro-OH, 53. $Pb(OAc)_4$ (162 mg, 0.365 mmol) in $CHCl_3$ (2.0 mL) was added dropwise to a solution of α-hydroxy acid **52** (118 mg, 0.365 mmol) in EtOAc (5.0 mL) at 0 °C. The reaction was stirred for 15 min at 0 °C, then quenched with ethylene glycol (1.5 mL). The solution was diluted with EtOAc (40 mL), washed with H_2O (2 ×

10 mL), brine (10 mL), dried over Na_2SO_4, and concentrated to give a yellow oil. The crude product was dissolved in acetone (10 mL) and cooled to 0 °C. Freshly prepared Jones reagent (2.7 M CrO_3, 2.7 M H_2SO_4, 0.27 mL, 0.73 mmol) was added dropwise. The reaction was stirred at 0 °C for 30 min, quenched with *i*-PrOH (1 mL), and stirred for another 10 min. The precipitate was removed by filtration, and the solvent was evaporated. The resulting residue was extracted with EtOAc (3 × 10 mL), and the organic layer was washed with H_2O (5 mL) and brine (5 mL), dried over Na_2SO_4, and concentrated. Flash chromatography with 5 % MeOH in $CHCl_3$ gave a white solid (59.6 mg, 55.6 %). **Diastereomer (*R*)-53:** $[\alpha]_D^{25}$ = – 62.5° (*c* 0.20, $CHCl_3$). ^{1}H NMR: δ 5.44 (d, *J* = 7.5, 1H), 4.28 (m, 1H), 3.34 (m, 3H), 2.52 (m, 1H), 2.24 (m, 1H), 1.90 (m, 6H), 1.61 (m, 2H), 1.40 (s, 9H). ^{13}C NMR: δ 180.1, 154.8, 139.7, 126.3, 79.2, 56.7, 49.4, 46.5, 33.0, 30.1, 28.9, 28.7, 25.3, 23.7. HRMS calcd. for $C_{16}H_{26}NO_4$ (MH^+) *m/z* = 296.1862, found *m/z* = 296.1857. Anal. HPLC: retention time 5.95 min (RP HPLC, > 98 % purity). **Diastereomer (*S*)-53:** ^{1}H NMR (CD_3OD): δ 5.44 (ddt, *J* = 8.9, 2.4, 2.4, 1H), 4.33 (m, 1H), 3.37 (t, *J* = 6.8, 2H), 3.24 (app. t, *J* = 7.2, 1H), 2.51 (m, 1H), 3.32 (m, 1H), 2.11 (m, 1H), 1.89 (m, 3H), 1.67 (m, 2H), 1.40 (s, 9H). ^{13}C NMR (CD_3OD): δ 177.6, 156.6, 142.2, 126.1, 80.8, 58.1, 50.5, 47.5, 33.9, 30.9, 29.8, 28.7, 26.0, 24.6. MS calcd. for $C_{16}H_{26}NO_4$ (MH^+) *m/z* = 296.2, found *m/z* = 296.2.

Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH, 43. Boc-Pro-Ψ[(*E*)CH=C]-Pro-OH **53** (72.1 mg, 0.245 mmol) was dissolved in CH_2Cl_2 (3.0 mL) and TFA (1.0 mL) was added. The solution was stirred at rt for 40 min. The solvent was evaporated and the excess TFA was removed under high vacuum. The resulting dark yellow oil was dissolved in $NaHCO_3$ (8.0 mL) and stirred at 0 °C for 20 min. Fmoc-OSu (124 mg, 0.366 mmol) was dissolved in 1,4-dioxane (3.0 mL) and added dropwise. The solution was stirred at rt for 14 h. The solution was

acidified to pH 1 with 1 M HCl. The aqueous solution was extracted with CH_2Cl_2 (4 × 50 mL). The organic layers were combined, washed with water (50 mL), and brine (50 mL). After drying with $MgSO_4$, the organic layer was concentrated in vacuo. Flash chromatography with 5 % MeOH in $CHCl_3$ gave a pale yellow solid (94.8 mg, 93.0 %). $[\alpha]_D^{25}$ = –40.1° (*c* 0.83, $CHCl_3$). ^{1}H NMR: δ 7.76 (d, *J* = 7.5, 2H), 7.61 (d, *J* = 7.2, 2H), 7.39 (t, *J* = 7.4, 2H), 7.30 (t, *J* = 7.0, 2H), 5.53 (m, 1H), 4.39 (m, 3H), 4.23 (m, 1H), 3.48 (m, 2H), 3.40 (m, 1H), 2.73 (m, 1H), 2.32 (m, 2H), 1.96 (m, 6H), 1.69 (m, 1H). ^{13}C NMR: δ 179.8 (180.1), 155.0, 144.3, 142.1 (140.8), 141.5, 127.8, 126.2 (124.9), 127.1, 125.3, 120.1, 67.1 (67.2), 57.2 (56.8), 49.4, 47.5, 46.5 (47.0), 32.1 (33.3), 30.1, 29.1, 25.3, 24.4 (23.6). HRMS calcd. for $C_{26}H_{28}NO_4$ (MH^+) *m/z* = 418.2018, found *m/z* = 418.1987. Anal. HPLC: retention time 7.92 min (RP HPLC, > 95 % purity).

H-Pro-OMe•HCl, 54. $SOCl_2$ (4.0 mL, 55 mmol) and MeOH (50 mL) were mixed and stirred at 0 °C for 30 min. H-Pro-OH (2.0 g, 17 mmol) was added, and the mixture was stirred at rt for 48 h. The solvent was removed in vacuo. A light green solid (2.3 g, 99 %) was obtained. ^{1}H NMR (CD_3OD): δ 4.44 (t, *J* = 7.8, 1H), 3.86 (s, 3H), 3.39 (m, 2H), 2.43 (m, 1H), 2.11 (m, 3H). ^{13}C NMR (CD_3OD): δ 170.5, 60.7, 54.0, 47.2, 29.3, 24.5.

Boc-Pro-Pro-OMe, 55. H-Pro-OMe **54** (2.0 g, 17 mmol) was dissolved in DMF (100 mL). EDC (6.66 g, 34.7 mmol), HOBt (5.34 g, 39.5 mmol), DMAP (303 mg, 2.48 mmol) and DIEA (30.0 mL, 172 mmol) were added, and the solution was stirred at 0 °C for 10 min. Boc-Pro-OH (5.61 g, 26.1 mmol) was added, and the mixture was stirred at rt for 6 h. The resulting solution was diluted with EtOAc (250 mL), washed with 1 M HCl (2 × 100 mL), H_2O (100 mL), $NaHCO_3$ (2 × 100 mL), and brine (100 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (25 % EtOAc in hexanes) to give a yellow

oil (3.48 g, 67.7 %). ^{1}H NMR (500 MHz): δ 4.59 (dd, *J* = 8.6, 4.2, 0.5H), 4.55 (dd, *J* = 8.5, 4.2, 0.5H), 4.50 (dd, *J* = 8.5, 3.0, 0.5H), 4.39 (dd, *J* = 8.4, 4.0, 0.5H), 3.77 (dt, *J* = 9.3, 7.3, 0.5H), 3.71 (s, 1.5H), 3.70 (s, 1.5H), 3.59 (m, 2.5H), 3.46 (dt, *J* = 10.2, 7.9, 0.5H), 3.39 (dt, *J* = 10.4, 7.4, 0.5H), 2.17 (m, 2H), 2.02 (m, 5H), 1.85 (m, 1H), 1.45 (s, 5H), 1.39 (s, 4H). ^{13}C NMR (125 MHz): δ 172.9 (172.7), 171.2 (171.6), 154.6 (153.7), 79.42 (79.39), 58.7, 57.71 (57.76), 52.1 (52.1), 46.9 (46.7), 46.5, 29.1 (30.0), 28.7 (28.8), 28.5 (28.4), 25.0 (25.1), 24.1 (23.6). Rotamers: 3:2.

Boc-Gly-Pro-Pro-OMe, 56. Boc-Pro-Pro-OMe **55** (3.48 g, 10.7 mmol) was dissolved in CH_2Cl_2 (50 mL), and TFA (10 mL) was added. The solution was stirred at rt for 40 min, and the solvent was removed in vacuo. The resulting dark yellow oil was dissolved in DMF (100 mL), and HOBt (3.38 g, 22.1 mmol), EDC (4.18 g, 21.8 mmol), DMAP (144 mg, 1.18 mmol) and DIEA (18.7 mL, 107 mmol) were added. The solution was stirred at 0 °C for 20 min, and Boc-Gly-OH (3.78 g, 22.0 mmol) was added. The mixture was stirred at 0 °C for 8 h, then diluted with EtOAc (250 mL), washed with 1 M HCl (2 × 100 mL), H_2O (100 mL), $NaHCO_3$ (2 × 100 mL), and brine (100 mL). The aqueous layers were extracted with EtOAc (4 × 150 mL), and monitored by TLC (50 % EtOAc in hexanes). The organic layers were combined, dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (50 % EtOAc in hexanes). A white solid was obtained (2.9 g, 69 %). ^{1}H NMR: δ 5.36 (m, 1H), 4.69 (dd, *J* = 8.2, 3.7, 1H), 4.55 (dd, *J* = 8.7, 4.5, 1H), 4.03 (dd, *J* = 17.4, 5.6, 1H), 3.89 (d, *J* = 3.8, 1H), 3.85 (m, 1H), 3.71 (s, 3H), 3.62 (m, 2H), 3.47 (dt, *J* = 9.8, 7.0, 1H), 2.20 (m, 3H), 2.06 (m, 3H), 1.97 (m, 2H), 1.43 (s, 9H). ^{13}C NMR: δ 172.8, 170.4, 167.4, 155.9, 79.6, 58.8, 58.1, 52.3, 46.8, 46.3, 43.1, 28.9, 28.4, 28.3, 25.1, 24.7. Rotamers: 8:1.

Boc-Gly-Pro-Pro-OH, 57. Boc-Gly-Pro-Pro-OMe **56** (405 mg, 1.06 mmol) was dissolved in THF (20 mL), and LiOH solution (1.0 M, 11 mL) was added. The mixture was stirred at rt for 1 h, monitored by TLC, and quenched with 1 M HCl (50 mL). The aqueous solution was extracted with CH_2Cl_2 (4 × 100 mL). The organic layers were combined, dried over Na_2SO_4, and removed in vacuo. A pale yellow oil was obtained (298 mg, 81 %). ^{1}H NMR: δ 5.40 (m, 1H), 4.65 (dd, *J* = 8.2, 3.9, 1H), 4.59 (dd, *J* = 8.0, 3.6, 1H), 4.03 (m, 1H), 3.83 (m, 2H), 3.60 (m, 2H), 3.28 (dd, *J* = 16.0, 6.4, 1H), 2.18 (m, 3H), 2.02 (m, 4H), 1.90 (m, 1H), 1.42 (s, 9H). ^{13}C NMR: δ 174.8, 171.4, 168.1, 156.2, 79.8, 59.2, 58.1, 47.1, 46.4, 42.9, 28.5, 28.4, 28.2, 24.9, 24.7.

Boc-Gly-Pro-OMe, 58. H-Pro-OMe **54** (0.0558 mol) was dissolved in CH_2Cl_2 (350 mL). EDC (12.8 g, 0.0670 mol), HOBt (10.3 g, 0.0671 mol), DMAP (342 mg, 2.80 mmol) and DIEA (40.0 mL, 0.229 mol) were added, and the solution was stirred at 0 °C for 10 min. Boc-Gly-OH (9.83 g, 0.0561 mol) was added, and the mixture stirred at rt for 8 h. The solution was concentrated in vacuo, diluted with EtOAc (300 mL), washed with 1 M HCl (2 × 100 mL), H_2O (100 mL), $NaHCO_3$ (2 × 100 mL) and brine (100 mL), dried over Na_2SO_4, and purified by flash chromatography (25 % EtOAc in hexanes). A pale yellow solid was obtained (13.0 g, 81.4 %). ^{1}H NMR: δ 5.45 (m, 1H), 4.52 (d, *J* = 6.6, 1H), 3.95 (m, 2H), 3.73 (s, 3H), 3.59 (m, 1H), 3.48 (m, 1H), 2.20 (m, 2H), 2.04 (m, 2H), 1.44 (s, 9H). ^{13}C NMR: δ 172.4 (171.9), 167.4 (167.6), 155.8, 79.5, 58.8 (58.4), 52.3 (52.7), 45.8 (46.6), 42.9 (42.8), 29.0 (31.3), 28.3, 24.6 (22.2). Rotamers: 5:1.

Boc-Gly-Pro-Hyp-OMe, 59. Boc-Gly-Pro-OMe **58** (116 mg, 0.405 mmol) was dissolved in THF (6 mL), and 1 M LiOH (4.0 mL aq., 4.34 mmol) was added dropwise. The solution was stirred at rt for 2 h, acidified to pH 1 with 1 M HCl, and extracted with CH_2Cl_2 (6 × 25 mL).

The resulting organic layers were combined, dried with $MgSO_4$, and concentrated in vacuo. A clear oil (86.8 mg, 78.5 %) was obtained. ^{1}H NMR: δ 8.83 (br s, 1H), 5.61 (m, 1H), 4.50 (m, 1H), 3.99 (d, *J* = 17.2, 1H), 3.86 (d, *J* = 17.2, 1H), 3.54 (m, 1H), 3.44 (m, 1H), 2.11 (m, 2H), 1.98 (m, 1H), 1.86 (m, 1H), 1.38 (s, 9H). Boc-Gly-Pro-OH (138 mg, 0.506 mmol) was dissolved CH_2Cl_2 (20 mL). HBTU (231 mg, 0.610 mmol), HOBt (93.2 mg, 0.609 mmol), DMAP (5.9 mg, 0.048 mmol) and DIEA (0.35 mL, 2.0 mmol) were added, and the solution was stirred at 0 °C for 10 min. H-Hyp-OMe **3** (91.9 mg, 0.506 mmol) was added, and the mixture was stirred at rt for 8 h. The solution was concentrated in vacuo, diluted with EtOAc (50 mL), washed with 1 M HCl (2 × 25 mL), H_2O (25 mL), $NaHCO_3$ (25 mL) and brine (25 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (2 % MeOH in $CHCl_3$). A yellow oil was obtained (68.6 mg, 33.9 %). ^{1}H NMR: δ 5.59 (m, 1H), 5.41 (m, 1H), 4.65 (app. t, *J* = 8.3, 1H), 4.56 (dd, *J* = 8.2, 4.2, 1H), 4.22 (d, *J* = 11.3, 1H), 4.03 (m, 1H), 3.91 (app. d, *J* = 15.1, 1H), 3.82 (m, 1H), 3.74 (s, 3H), 3.59 (m, 1H), 3.49 (m, 1H), 2.46 (m, 1H), 2.29 (m, 1H), 2.15 (m, 2H), 2.02 (m, 2H), 1.46 (s, 5H), 1.43 (s, 4H). ^{13}C NMR: δ 172.2 (171.2), 170.3 (169.8), 167.7, 156.00 (155.96), 80.1 (79.8), 73.67 (73.73), 60.6 (59.0), 57.7 (57.6), 57.55 (57.49), 52.5, 46.3 (47.1), 42.9 (42.8), 34.3 (34.4), 28.5 (28.6), 28.37 (28.43), 25.0. MS calcd. for $C_{18}H_{30}N_3O_7$ (MH^+) *m/z* = 400.5, found *m/z* = 400.6. Rotamers: 3:2.

Boc-Gly-Pro-Hyp-OBn, 60. Boc-Pro-Hyp-OBn **10** (6.37 g, 15.2 mmol) was dissolved in CH_2Cl_2 (60 mL). TFA (20 mL) was added, and the solution was stirred at rt for 30 min. The solvent and excess TFA were removed in vacuo. The resulting dark yellow oil was dissolved in CH_2Cl_2 (200 mL). HBTU (9.08 g, 22.8 mmol), HOBt (3.50 g, 22.8 mmol), DMAP (93.6 mg, 0.766 mmol) and DIEA (13.3 mL, 76.2 mmol) were added. The solution was

stirred at 0 °C for 20 min, and Boc-Gly-OH (4.01 g, 22.9 mmol) was added. The reaction was stirred at rt for 8 h. The solution was concentrated in vacuo, and diluted with EtOAc (300 mL). The organic layer was washed with 1 M HCl (2 × 150 mL), H_2O (150 mL), $NaHCO_3$ (2 × 150 mL) and brine (100 mL). The aqueous layers were extracted with CH_2Cl_2 (4 × 150 mL) and monitored by TLC (5 % MeOH in $CHCl_3$). The organic layers were combined, dried over $MgSO_4$, concentrated in vacuo, and purified by flash chromatography (2 % MeOH in $CHCl_3$). A white solid (6.65 g, 91.9 %) was obtained. ^{1}H NMR (500 MHz): δ 7.33 (m, 5H), 5.62 (m, 0.5H), 5.47 (dt, *J* = 4.8, 4.8, 1H), 5.38 (m, 0.5H), 5.24 (d, *J* = 12.1, 0.5H), 5.21 (d, *J* = 12.1, 0.5H), 5.08 (d, *J* = 12.2, 0.5H), 5.06 (d, *J* = 12.1, 0.5H), 4.70 (dt, *J* = 8.2, 3.8, 1H), 4.64 (dd, *J* = 8.4, 4.2, 0.5H), 4.55 (dd, *J* = 8.4, 3.8, 0.5H), 4.49 (m, 1H), 4.00 (m, 1.5H), 3.86 (m, 2H), 3.65 (dd, *J* = 10.6, 4.2, 0.5H), 3.56 (m, 1H), 3.45 (m, 1H), 2.46 (m, 1H), 2.35 (m, 1H), 2.10 (m, 2H), 1.95 (m, 2H), 1.45 (s, 4.5H), 1.43 (s, 4.5H). ^{13}C NMR (125 MHz): δ 172.0, 171.3 (170.8), 167.8 (166.0), 155.95 (156.02), 135.6 (135.4), 128.58 (128.62), 128.32 (128.42), 128.15 (128.20), 79.7 (80.1), 70.4, 66.9 (67.1), 59.0 (60.6), 58.0 (57.8), 57.7 (55.0), 46.4 (46.3), 42.9 (43.0), 38.7 (37.5), 34.2 (34.3), 28.39 (28.36), 24.7 (24.9). Rotamers: 1:1.

Boc-Gly-Pro-Hyp-OH, 61. Boc-Gly-Pro-Hyp-OBn **60** (417 mg, 0.878 mmol) was dissolved in MeOH (20 mL), and 10 % Pd/C (40 mg) was added. H_2 was introduced with a balloon, and the suspension was stirred at rt for 3.5 h. The solution was filtered through Celite™, and concentrated in vacuo. A clear, colorless oil (278 mg, 82 %) was obtained. ^{1}H NMR (500 MHz, CD_3OD): δ 5.41 (m, 1H), 4.72 (dd, *J* = 8.7, 3.8, 0.5H), 4.67 (dd, *J* = 8.3, 3.9, 0.5H), 4.50 (m, 1H), 4.10 (d, *J* = 11.9, 0.5H), 3.98 (app. t, *J* = 16.2, 0.5H), 3.89 (m, 0.5H), 3.84 (m, 0.5H), 3.81 (m, 1H), 3.77 (m, 2H), 3.59 (m, 2H), 2.69 (m, 0.5H), 2.61 (app. t, *J* = 5.6, 0.5H),

2.24 (m, 2H), 2.07 (m, 1H), 2.03 (m, 1H), 1.89 (m, 1H), 1.45 (s, 4H), 1.44 (s, 5H). ^{1}H NMR (125 MHz, CD_3OD): δ 175.2 (174.6), 173.1 (172.8), 169.90 (169.86), 158.3 (158.5), 80.5 (80.6), 71.2 (71.1), 61.7 (60.1), 59.7 (59.5), 59.3 (59.2), 55.7 (53.5), 47.6 (47.5, 46.3), 43.6 (43.3), 37.8 (35.4), 29.0 (29.2), 28.7, 25.5 (25.7). Rotamers: 3:1:1.

Chapter 3 Design, synthesis and characterization of collagen-inspired polypeptides

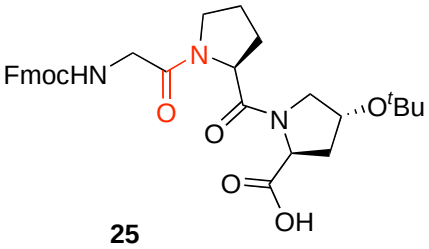
*3.1 Design of peptides containing the Gly-*trans*-Pro isostere*

Synthetic collagen-like polypeptides have been used widely to elucidate the three-dimensional structure and the conformational stability of the collagen triple helix.[4] Although natural collagen can be either homotrimeric or heterotrimeric, the homotrimeric, self-associating peptides with the sequence of $(Gly\text{-}Xaa\text{-}Yaa)_n$ are often used as models of synthetic collagen polypeptides because they are easier to synthesize by various organic or peptide synthetic methods. Peptides with the sequence of $(Gly\text{-}Pro\text{-}Pro)_n$ and $(Gly\text{-}Pro\text{-}Hyp)_n$ are the most widely used synthetic polypeptides in collagen studies because they can self-associate into triple helical structures.[116,215] Studies on collagen peptides with homogeneous repeating units showed that except for the peptides with 4(*R*)-Flp at the Yaa position, the collagen-like polypeptide $(Gly\text{-}Pro\text{-}Hyp)_n$ has the highest T_m value.[10,45,138]

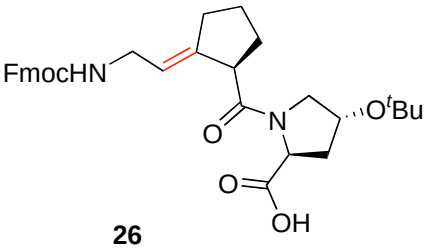
It is useful to predict the stability of the triple helix before the actual synthesis of the polypeptides. The host-guest system was used to study the effect of a single amino acid residue on the overall structural stability. O'Neil and DeGrado,[123] and Smith *et al.*[124] applied this system to the study of the α-helix and β-sheet respectively. Bachinger and Davis first suggested the parameters that allowed the prediction of the triple helical propensity of each amino acid in collagen.[271] Brodsky and co-workers applied the host-guest system to collagen stability studies.[44,120,125,129,205] In their design, a guest tripeptide of interest is flanked by several host triplets, which are used to increase the thermal stability of the triple-helical structure into the ambient temperature region. Usually, three or four repeats of Gly-Pro-Hyp

are added to each end of the guest sequence. Studies showed that Pro at the Xaa, and Hyp at the Yaa positions yielded the most stable triple helical structure.[44]

Previous studies showed that the thermal stability of the collagen-like polypeptides with free acid or amine ends is pH-dependent, with a higher T_m value under acidic or basic conditions, and a lower T_m value under neutral conditions.[272] These results show that a charge on the backbone will affect the stability of the collagen triple helix. The host-guest system that Brodsky designed has the sequence, Ac-(Gly-Pro-Hyp)$_3$-Gly-Xaa-Yaa-(Gly-Pro-Hyp)$_4$-Gly-Gly-NH$_2$, with no charge on the backbone. The thermal stability of the resulting triple helix is pH-independent if the side-chain of the guest triplet is not charged.[126]



25



26

Ac-(Gly-Pro-Hyp)$_3$-Gly-Pro-Hyp-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH$_2$
62

Ac-(Gly-Pro-Hyp)$_3$-Gly-Ψ[(*E*)CH=C]-L-Pro-Hyp-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH$_2$
63

Ac-(Gly-Pro-Hyp)$_3$-Gly-Ψ[(*E*)CH=C]-D-Pro-Hyp-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH$_2$
64

Figure 3.1 Design of the collagen-like host-guest polypeptides.

Based on these previous studies and the model reactions we had performed, we designed our collagen-like polypeptides. Peptide **62** has Gly-Pro-Hyp as the guest triplet and acts as control peptide for our stability studies (Figure 3.1). Peptide **63** has the alkene isostere Gly-Ψ[(*E*)CH=C]-L-Pro-Hyp, which has the natural Pro-like stereochemistry, as the guest peptide. Peptide **64** has the alkene isostere Gly-Ψ[(*E*)CH=C]-D-Pro-Hyp, which has the

unnatural Pro-like stereochemistry, as the guest peptide. Although the two diastereomers of the tripeptide isostere **26** were separated, the absolute stereochemistry of the diastereomers was not determined. Both diastereomers were incorporated into the sequence; the peptide with the isostere Gly-Ψ[(*E*)CH=C]-D-Pro-Hyp would be unable to form a triple helical structure based on previous studies with D-amino acids.[43,273] Moreover, we also put one Tyr residue at the amine end for the quantitative measurement of the peptide concentration with UV, which provides a more accurate assessment of the concentration than the measurement of the weights of the dry peptides.[274,275]

*3.2 Synthesis of peptides containing the Gly-*trans*-Pro isostere*

Before putting the synthetic tripeptide isostere Gly-Ψ[(*E*)CH=C]-Pro-Hyp into the sequence, model reactions were performed with only the Gly-Pro-Hyp isostere to find the best tripeptide building block and reaction conditions.

Rink amide MBHA resin was used as the matrix for solid phase peptide synthesis because it provides an amide group at the *C*-terminus after cleavage from the resin. A model reaction with side-chain unprotected tripeptide Fmoc-Gly-Pro-Hyp-OH **1** was performed. The loading was measured after coupling the first amino acid residue. (Applied Biosystems. Determination of the Amino Acid Substitution Level via an Fmoc Assay; Technical Note 123485 Rev 2, accessed Jun. 19th, 2008) The loading percentage was usually around 86 % to 93 % for Rink amide MBHA resin, which was much higher than for the 2-chlorotrityl chloride resin (about 35 %-60 %). The peptide with the sequence Ac-(Gly-Pro-Hyp)$_7$-Gly-Gly-NH_2 was synthesized by solid phase peptide synthesis and characterized by LC-MS. Because the LC-MS has a 1500 Da mass limit and the molecular weight of the synthesized peptide is greater than the detectable mass upper-limit, the molecular ion peak MH^+ was not

available. The doubly charged peak $MH_2^{2+}/2$ or the triply charged peak $MH_3^{3+}/3$ was used to identify the polypeptide. From the LC-MS chromatogram, in addition to the desired $MH_2^{2+}/2$ peak, a $[M+45]^{2+}/2$ peak was also observed. This indicated that during the final acetylation step, the acetyl group was attached to both the amine end and the unprotected hydroxyl group of one of the Hyp residues. From the integration of the LC-MS chromatogram, the $[M+45]^{2+}/2$ peak was even larger than the $MH_2^{2+}/2$ peak, and the crude product was a very complicated mixture, which was not worth purifying. The side-chain unprotected tripeptide Fmoc-Gly-Pro-Hyp-OH **1** was not a good tripeptide building block for solid phase peptide synthesis of the desired host-guest polypeptides **62** and **63**.

The side-chain protected tripeptide building block **25** was synthesized. The model reaction with tripeptide **25** (Figure 3.1) was performed to synthesize the peptide Ac-(Gly-Pro-Hyp)$_7$-Gly-Gly-NH_2, and the peptide synthesis was monitored by the Kaiser test.[276] No $[M+45]^{2+}/2$ peak was observed in the LC-MS chromatogram of the crude peptide mixture, which indicated that the acetylation of the free hydroxyl group was successfully avoided. But for most of the model reactions, the crude polypeptides were still very impure. Some factors significantly affected the quality of the synthesized polypeptides, and these are summarized as follows:

1) **The purity of the tripeptide building block**. Ottl *et al.* synthesized the tripeptide building block **25** without column purification, and the final product was obtained by precipitation from a MeOH-Et_2O solvent system.[239] But we found that the quality of tripeptide **25** from precipitation was not good enough for solid phase peptide synthesis. Therefore, tripeptide **25** we prepared was purified by flash chromatography followed by precipitation from MeOH-Et_2O. The resulting solid tripeptide **25** was

more pure, and provided a cleaner crude polypeptide from solid phase peptide synthesis.

2) **The quality of the reagents used in SPPS, especially DIEA.** The quality of the reagents is very important for solid phase synthesis. Usually the coupling reagents, such as HBTU and HOBt, do not affect the quality of the product if they are properly stored, although the use of old HATU has caused failures of solid phase peptide synthesis in our group (Dr. Tao Liu, unpublished results). However, the base DIEA can be problematic in solid phase peptide synthesis. We found that fresh DIEA, which is usually very light yellow in color, can provide a very pure crude peptide from solid phase synthesis. Once the DIEA aged, which was yellow or even darker in color, the crude peptides became very impure, and sometimes even no desired peptide was obtained (Dr. Song Zhao and Ana Y. Mercedes-Camacho, unpublished results). It is important to test the purity of the base by NMR before solid phase synthesis.

3) **Reaction time and temperature.** Solid phase peptide synthesis is a fast reaction and usually finishes within 30 min. The synthesis of a peptide with nine amino acid residues using 10 min coupling time for each residue gave a crude peptide that was very pure and did not need HLPC purification (Ana Y. Mercedes-Camacho, unpublished results). The longer the reaction time runs, the more chance that the peptide synthesis will include side reactions. Unfortunately, for the synthesis of the collagen-like peptides, except for the first three amino acid residues, the coupling reactions do not finish within 30 min. The coupling of tripeptide building block required at least 1-1.5 h monitored by HPLC. For the coupling of the last two or three tripeptide units, the Kaiser test showed that a small amount of resin still turned blue

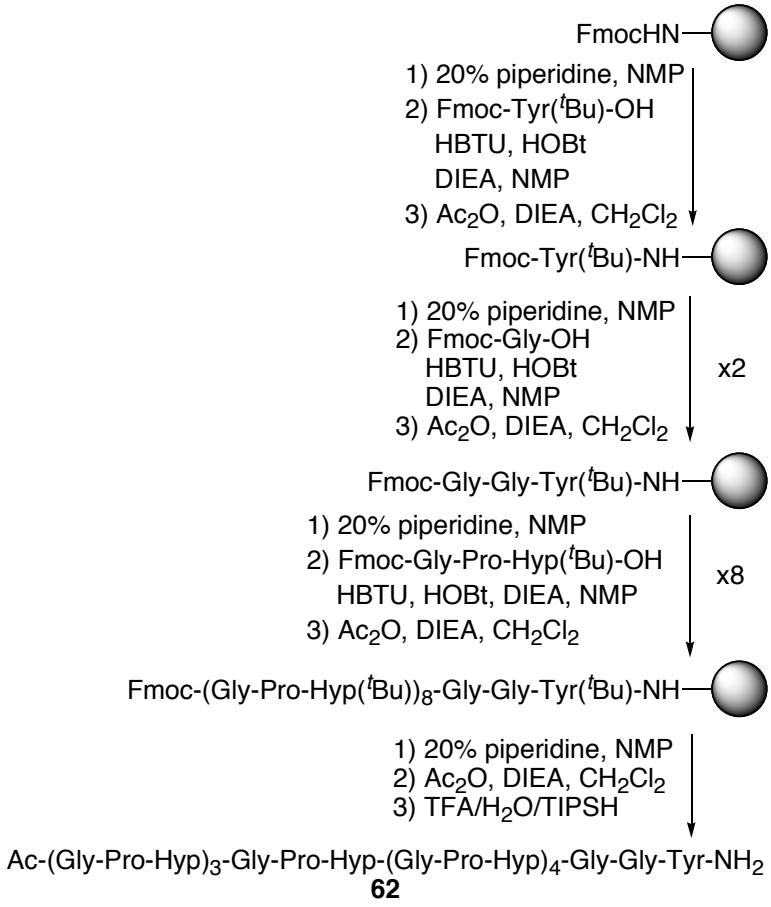
even after 3-4 h, which meant that the reaction was not finished. For this reason, a capping reaction with 10 % Ac_2O and 10 % DIEA in CH_2Cl_2 was performed after each coupling. This creates greater differences in peptide length to simplify HPLC purification. Increasing the reaction temperature can increase the coupling reaction rate, but it also increases the reaction rate of side reactions. Moreover, it caused some problems during the capping process, and the top of the reaction tube popped off due to the built-up pressure. Consequently, the reaction temperature could not be very high.

4) **Solvent.** For solution phase peptide synthesis, CH_2Cl_2 is the most common solvent used because it is easy to remove. While for solid phase synthesis, CH_2Cl_2 causes solubility problems with HBTU and HATU, and it is not as good for fast peptide synthesis. DMF is widely used in solid phase peptide synthesis, but HBTU still dissolves slowly in DMF. NMP is another possible solvent for solid phase synthesis, and HBTU dissolves better in NMP than in DMF. It is still a matter of debate as to which solvent is better. For the synthesis of peptides containing Ser-Pro isosteres, NMP was a better solvent system than DMF; (Dr. Song Zhao, personal communication) for the synthesis of short peptides, DMF provided a cleaner product. (Ana Y. Mercedes-Camacho, personal communication) For collagen-like polypeptide synthesis, no significant difference between DMF and NMP as the solvent was observed, although it seems the coupling reaction was a little faster in DMF.

Based on the model reactions, the best conditions for solid phase peptide synthesis are shown in Scheme 3.1. The whole solid phase peptide synthesis was performed at 30 °C. Rink amide MBHA resin and the solvent NMP were used. Standard Fmoc/HBTU conditions were

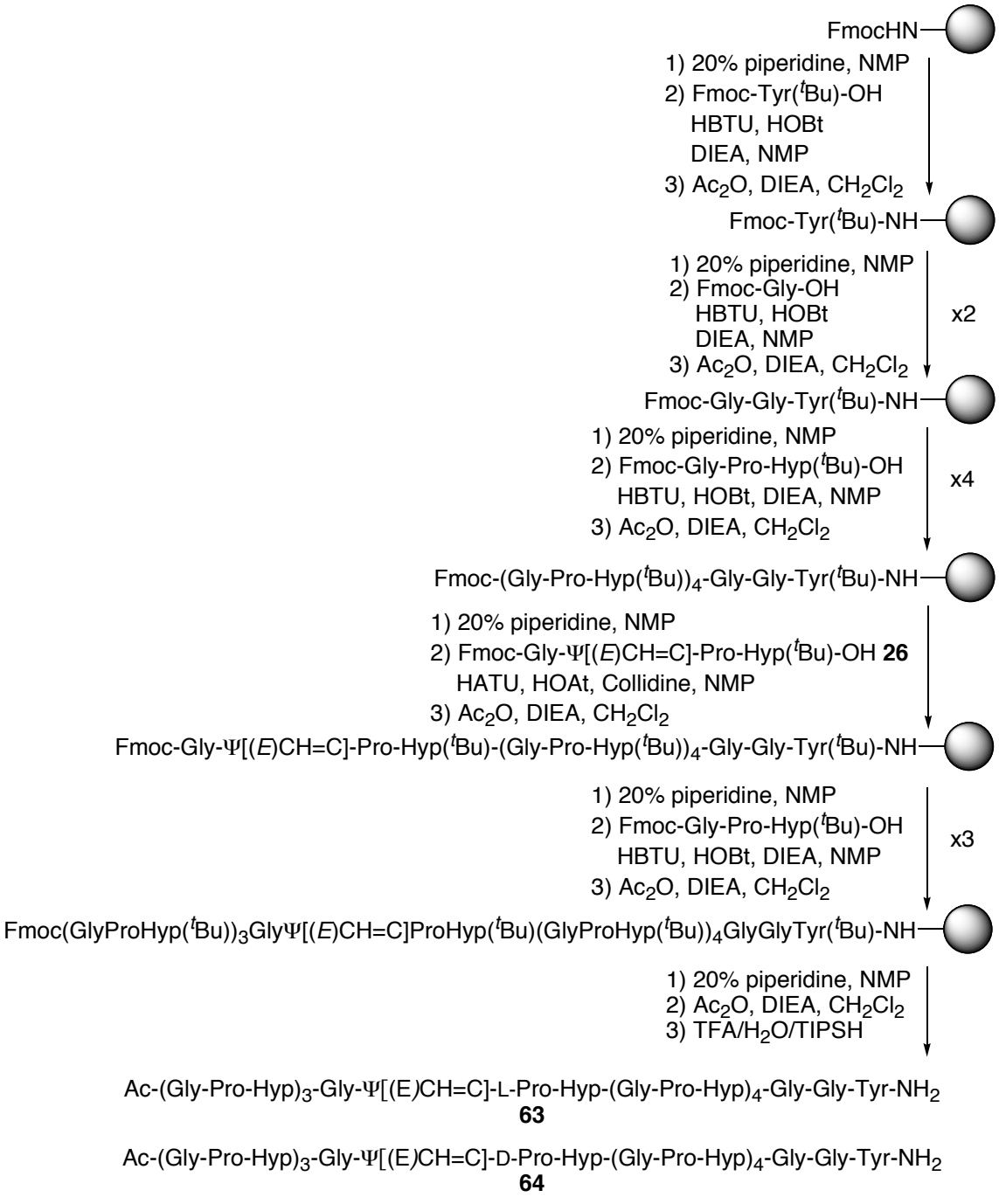
used, with 30 min coupling for single amino acids, and 2 h coupling for tripeptide building blocks.

Scheme 3.1 Solid phase peptide synthesis of control peptide **62**.



The synthesis of peptides **63** and **64**, which contain the conformationally locked Gly-Pro isostere, is similar to that of control peptide **62**. During the coupling of the tripeptide alkene isostere, Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OH L- or D-**26**, HATU and HOAt were used as the coupling reagents, and collidine was used as the base to prevent the double-bond migration during solid phase synthesis (Scheme 3.2).[224] Because tripeptide isostere **26** is difficult to synthesize and is very precious, exactly 1.0 equivalent of the tripeptide isostere **26** was used and the coupling reaction was performed only once at 35 °C for 2.5-3 h. Both peptides **63** and **64** were obtained as white solids, and ready for purification.

Scheme 3.2 Solid phase peptide synthesis of the Gly-Pro isostere containing peptide **63** and **64**.



*3.3 Purification of peptides containing the Gly-*trans*-Pro isostere*

The crude peptides were purified by reverse phase HPLC. Usually, all the impurities should be visible at 210 nm, so the chromatogram at 210 nm was used to obtain better separation. Since TFA cannot be used in the solvents for LC-MS, and formic acid has a cut-off wavelength of 250 nm, the chromatogram at 260 nm was obtained to compare with the LC-MS results to identify the peak of interest.

A slow gradient with two solvents was used to achieve a good resolution: solvent A was 0.1 % TFA in H_2O, solvent B was 0.1 % TFA in CH_3CN. Solvents without TFA were also tried, but significant tailing of the peaks was observed. Solvents with 1 % formic acid were used in the LC-MS analysis, and were also tried in the analytical HPLC system to identify the peak order in the chromatograms.

Two reverse phase columns were tried in the analytical HPLC separation: a Waters XBridge C18 reverse phase column (2.5 μm, 4.6 × 50 mm) and a VYDAC Protein C4 214TP54 reverse phase column (5 μm, 4.6 × 250 mm). A relatively fast gradient from 10 % to 90 % solvent B over 20 min was tried to see which column gave the best separation. Although the separation on the VYDAC Protein C4 column was not very good at this gradient, the desired the peak could be identified. The desired peak was not found in the chromatogram obtained with the Waters XBridge C18 column. There are two possible reasons for this result: 1) the synthetic collagen-like peptide is a very polar 27-mer, and it elutes too fast; 2) the synthetic collagen-like peptide has a very strong interaction with the C-18 stationary phase and it elutes very slowly. From my experience with helix-turn-helix peptides (29-mer), the latter reason is more probable.

A Protein C4 reverse phase column was chosen to separate the crude peptides. The gradient from 10 % to 90 % CH_3CN over 20 min was still too fast, and the desired peak eluted at about 8 min. Methods with a slower gradient were tried, and the one that gave the best separation is shown in Table 3.1.

Table 3.1 Method for HPLC separation of the collagen-like peptides **62**, **63** and **64**.

Time (min)	0	20	22	30	32	35
% B [a]	5	25	60	60	5	5

a) Solvent A is 0.1 % TFA in H_2O; solvent B is 0.1 % TFA in CH_3CN.

The solvent B gradient ran from 5 % to 25 % over 20 min, and the desired peak eluted at about 15 min. Although this condition still did not give very good peak resolution, it gave the best separation of all the gradients. Heating the sample may help the separation of the collagen-like peptides. (Mathew Shoulders, personal communication) The sample was heated to 60 °C followed by an immediately injection onto the column, but the peak resolution appeared unchanged. Heating the column to approximately 70 °C was also tried on the LC-MS system, but still no significant improvement to the separation was observed.

Before preparative HPLC separation, LC-MS with the same VYDAC Protein C4 column was performed to confirm the identity of the peaks. The preparative HPLC separation was performed with the same method at a flow rate of 15 mL/min. The chromatogram was monitored at 210 nm, and fractions were collected manually. The purity of the peptides was confirmed by analytical HPLC.

*3.4 Characterization of peptides containing the Gly-*trans*-Pro isostere*

3.4.1 NMR

Two ^{1}H NMR experiments were performed to identify the peptides. All three purified peptides were dissolved in 0.5 mL DMSO-d_6 and the ^{1}H NMR experiments were then performed. The spectra showed the presence of all proton peaks including the N-H proton. There are many peaks with a chemical shift above 5.0 ppm, and it was very difficult to determine which peak represents the alkene C-H proton. A proton exchange ^{1}H NMR experiment was performed. A few drops of D_2O were added to the NMR tube, and the solution was mixed well. The NMR tube was left at rt for about 2 h to allow complete deuterium exchange with the active N-H protons. The exchange ^{1}H NMR spectra showed that only three peaks have chemical shifts above 5.0 ppm. A peak at 7.0 ppm and another peak at 6.65 ppm are associated with the benzene ring of the Tyr residue. For peptide **63**, there is an additional peak at 5.09 ppm, while there is a peak at 5.26 ppm for peptide **64**. These two peaks were tentatively assigned as the alkene C-H in the two peptides respectively. The ^{1}H NMR results confirmed the presence of the alkene H-C=C bond, so this bond does not migrate into the 5-membered ring. The integration was not very good because the amount of peptides used in this study was very small (< 1 mg), and it was difficult to get a good baseline for accurate integration.

3.4.2 LC-MS and MALDI-TOF

LC-MS experiments were performed to prove the identities of the purified peptides. The gradient was from 5 % to 95 % solvent B (1 % formic acid in CH_3CN) over 35 min with

a flow rate of 0.2 mL/min. The peaks of interests were eluted at 13-15 min. LC-MS results are shown in Table 3.2.

Table 3.2 Mass of the collagen-like peptides **62**, **63** and **64** from LC-MS. [a]

	MH^+		$MH_2^{2+}/2$		$MH_3^{3+}/3$	
	Calcd.[b]	Exp.	Calcd.[a]	Exp.	Calcd.[b]	Exp.
62	2475.1	ND[c]	1238.1	1238.0	825.7	825.9
63	2458.1	ND[c]	1229.6	1229.8	820.0	820.0
64	2458.1	ND[c]	1229.6	1229.7	820.0	820.1

a) LC-MS was performed with a RP C-18 column (100 × 2.1 mm), electrospray ionization (positive ion mode), and a triple-quadrupole mass separator; b) The calculated mass was the calculated highest peak, not the average MW, c) ND = not detectable. The MH^+ peak was above the detectable upper-limit of the instrument.

The molecular mass peaks were not available because the mass detector has a 1500 Da detection limit. The doubly charged ($MH_2^{2+}/2$) and triply charged ($MH_3^{3+}/3$) peaks were used to identify the desired product peaks. The doubly charged and triply charged ions appeared at the same retention time in the chromatogram. Theses two ions have mass values that are very close to the calculated values, which indicated that the purified peptides have the correct mass.

The peptides were submitted for MALDI-TOF analysis (VBI core lab, Virginia Tech) so that the molecular mass could be determined. The peptide was dissolved in water, and 50 μL of each sample were taken. Ziptip C18 pipette tips were used to enrich the peptides. α-Cyano-4-hydroxycinnamic acid (CHCA) in 50 % CH_3CN was used as the matrix. To obtain a more accurate mass, a standard peptide (ACTH 18-39, MW = 2464.1989) with some BSA digest was added to act as the internal standard. The results are shown in Table 3.3. Unfortunately, no molecular ion MH^+ peak was observed in the MALDI-TOF spectra, the major peaks in the desired region have masses close to those calculated for the molecule plus sodium ion ($[M+Na]^+$, MW = M+23) or potassium ion ($[M+K]^+$, MW = M+39).

An ion-exchange experiment was performed with the peptides **63** and **64** to prove their identities. A few drops of the peptide solutions were added to 1 N LiCl solution and left at rt for about 30 min to exchange the Na^+ and K^+ ions with the Li^+ ion. Then a MALDI-TOF experiment was performed with 2,5-dihydroxybenzoic acid (DHBA) as the matrix to study the ion-exchange result. The spectrum showed that, before the ion exchange, there was a peak at 2480 Da, which was thought to be the $[M+Na]^+$ peak. After the ion-exchange experiment, the peak at 2480 Da became smaller, and a new peak at 2463 Da appeared and became the highest peak in the spectrum. This result confirmed that the peak at 2480 Da is the $[M+Na]^+$ peak (M+23), and the peak at 2463 Da is the $[M+Li]^+$ peak (M+6). Then, the molecular weight of the peptides **63** and **64** can be deduced to be 2457 Da, and the molecular weight of peptide **62** is 2474 Da, which matched the calculated values.

From the ^1H NMR, LC-MS and MALDI-TOF results, the identities of the peptides were confirmed. All three synthesized peptides had the right mass, and the Gly-Pro isostere containing peptides **63** and **64** had the alkene double bond in the sequence. The alkene bond was at the correct position, and did not shift into the ring during solid phase peptide synthesis.

Table 3.3 Mass of the collagen-like peptides **62**, **63** and **64** from MALDI-TOF.

	Internal standard [a]		$[M+Na]^+$		$[M+K]^+$	
	Calc. [b]	Exp.	Calc. [b]	Exp.	Calc. [b]	Exp.
62	2465.20	2465.86	2497.12	2497.76	2513.12	2513.78
63	2465.20	2465.11	2480.13	2480.06	2496.13	2496.03
64	2465.20	2464.90	2480.13	2479.85	2496.13	2495.82

a) The standard peptide (ACTH 18-39, MW = 2464.1989) with some BSA digest was added as the internal standard. b) The calculated mass was the calculated highest peak, not the average molecular weight.

3.4.3 CD analysis

Principle of circular dichroism

Natural light can be polarized electrically or magnetically to plane-polarized light. Plane-polarized light can be assigned to two components with equal magnitude: the counter-clockwise rotating component (left-handed, L), and the clockwise rotating component (right-handed, R). When plane-polarized light passes through the a sample, if the L and R components are not absorbed or are absorbed to equal extents, the recombination of L and R components will regenerate the radiation in the original plane. [277-279] However, if L and R components are absorbed differently, the resulting radiation will have elliptical polarization. Circular dichroism (CD) refers to the differential absorption of two circular-polarized components of the plane-polarized light.[277-279] Elliptical polarization can be observed when a chromophore is optically active for one of the following reasons: 1) intrinsically chiral, or optically active due to the torsion angle of the C-S-S-C chain; 2) covalently linked to a chiral center; 3) in an 3-dimensional asymmetric environment.[277-279] CD signals only appear when the differential absorption of radiation occurs, so the spectral bands are associated with the distinct structural features of a molecule.[277]

Many structural features in proteins can cause CD signals. Peptide bonds (< 240 nm), aromatic side-chains (260-320 nm), disulphide bonds (~260 nm, weak), even non-protein cofactors (in a wide wavelength range), and ligands, which may not be intrinsically chiral, but can acquire their chirality when bound in an asymmetric environment, are possible chromophores that can be observed in a CD spectrum.[277] The CD spectra in the far-UV region (< 250 nm), which can be attributed to the amide bonds in proteins, are associated with various types of secondary structure. In particular, the CD spectrum between 240 nm

and 200 nm, which provides useful structural information, is widely used to estimate the protein conformation.[280]

Circular dichroism is an important tool to characterize the structure of short peptides. Collagen-like polypeptides have a characteristic maximum at 225 nm and an intense minimum around 200 nm in the CD spectrum.[122] The heat-induced unfolding process from the collagen triple helix to the unfolded state is highly cooperative and occurs within a very narrow temperature range. The temperature at the midpoint of this transition T_m is used to describe the stability of the collagen triple helix.[40]

Concentration determination

All three peptides were dissolved in 10 mM PBS buffer (pH 7.0, 0.15 M NaCl, 10 mM phosphate buffer). Concentrated guanidine hydrochloride solution (6 M) was used to denature the peptides to produce the completely unfolded state and achieve an accurate UV measurement. The concentration can be determined by Beer's law:

$$A = \varepsilon bc \quad (1)$$

Where A is the absorbance at 280 nm; ε is the molar extinction coefficient, which is 1490 $M^{-1} \cdot cm^{-1}$ for Tyr at 280 nm;[281,282] b is the concentration; c is the length of the cell, which is 1 cm for this measurement. The calculated concentrations of the three peptides are shown in Table 3.4.

Table 3.4 Concentration of the collagen-like peptides **62**, **63** and **64** solutions.

	Peptide **62**	Peptide **63**	Peptide **64**
Concentration (mg/mL)	0.22	0.40	0.60

Two-state model for data processing

A two-state model has been used in protein studies to describe the transition between the all-native and the all-denatured state.[283,284] The folding process of the collagen-like

peptide has a triple helix state and an unfolded state, and the folding process is highly cooperative. The values of T_m can be determined by fitting the data to the two-state model.

The molar ellipticity of the triple helical conformation was found to be independent of temperature, while the ellipticities of the unfolded forms show a linear temperature dependence in all solvent systems.[122,285] Using the "all-or-none" model, the equilibrium constant K can be shown as:

$$K = \frac{[H_{3n-2}]}{[C]^3} = \frac{F}{3c_0^2(1-F)^3} \quad (2)$$

Where F is the fraction of chains or tripeptide units in the triple helical state, which means $F = 3[H]/[C_0]$, in which [H] is the concentration of triple helical state, $[C_0]$ is the initial concentration of coil state, and c_0 is the total concentration of the peptide.[168]

The Gibbs free energy can be shown to be:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = -RT\ln K \quad (3)$$

Where ΔG^0 is the standard Gibbs free energy, ΔH^0 is the standard enthalpy, and ΔS^0 is the standard entropy. When $F = 0.5$, which means that half of the tripeptide units are in the triple helical structure, the temperature T is T_m. Then the equation becomes:

$$T_m = \frac{\Delta H^0}{\Delta S^0 + R\ln(0.75c_0^2)} \quad (4)$$

It is clear that the melting point T_m is concentration dependent, and can be calculated from Eq. 4. By combining Eq. 3 and Eq. 4, the equilibrium can also be shown to be:

$$K = \exp\left[\frac{\Delta H^0}{RT}\left(\frac{T}{T_m} - 1\right) - \ln(0.75c_0^2)\right] \quad (5)$$

The temperature dependence of ΔH^0 is very small and the specific heat of the coil to triple helix transition is close to 0.[134] The relationship between the fraction conversion F and the measured molar ellipticity [Θ] can be shown to be:

$$F = \frac{[\Theta] - [\Theta]_c}{[\Theta]_h - [\Theta]_c} \tag{6}$$

$$K = \frac{F}{1 - F} \tag{7}$$

In which $[\Theta]_h$ and $[\Theta]_h$ are the ellipticities of the peptide in the complete triple helical conformation or in the unfolded conformation respectively. Eq. 7 is the simplified form of Eq. 2 during the denaturing process. The linear temperature dependencies of the observed ellipticities in the triple helix state $[\Theta]_h$ and the coil state $[\Theta]_c$ can be described as:

$$\begin{aligned} [\Theta]_h &= [\Theta]_{h,T_m} + a(T - T_m) \\ [\Theta]_c &= [\Theta]_{c,T_m} + b(T - T_m) \end{aligned} \tag{8}$$

In which the theoretical ellipticities of the triple helix and the unfolded state at the melting point $[\Theta]_{h,\,Tm}$ and $[\Theta]_{c,\,Tm}$ are reference values.[40,122,168] The obtained ellipticities were fitted to the two state model with Sigma-Plot v10.0 by combining Eq. 2, 5, 6, 7 and 8, and fitting the fraction conversion F to the equilibrium constant K. The initial slopes a, b, and the Y-axis intercepts were calculated with Excel from the first 6, and the last 5 data points from the CD spectra respectively.

Results and discussion

Purified peptides were dissolved in PBS buffer solution (pH 7.0, 10 mm), and incubated at 4-5 °C for 72 h. The CD spectra were obtained in ~3 °C increments with a 5 min equilibration time, and the ellipticity at 226 nm was monitored. The unfolding process of each peptide was measured twice, and the data obtained were processed separately and

averaged to obtain the T_m value. The cell was checked before each measurement because air bubbles in the cell significantly decreased the ellipticity.

In the full-range spectra, all three peptides showed a maximum at around 225 nm, and a minimum around 210 nm (Figure 3.2). The results confirmed the formation of the PPII helixes for all three peptides. Since the lamp was very old, and it had a limited detection range, the spectrum below 210 nm exceeded the negative detection limit and was very noisy, and could not be used to reliably identify the structure of the peptides. The maximum at 225 nm in the spectra of peptides **63** and **64** was less intense than for control peptide **62**, while the minimum around 205 nm was also less intense, which indicated that the peptides with the alkene isosteres were less helical than the all-natural control peptide. The value of the molar ellipticity was not accurate because: 1) the calibration of the CD spectropolarimeter is not good, the CD spectropolarimeter was calibrated before use and the CD spectrum of the standard showed that the wavelength calibration was very accurate, but the ellipticity value was about 3.5 times greater than it should be, and we could not adjust it to acquire a more accurate value, and 2) the concentration of the peptides was not very accurate. Using UV to measure the concentration of the peptides was more accurate than using the weight to estimate the concentration, because the collagen-like peptides contained a lot of water in the molecule which was very difficult to remove. The accurate UV absorbance range is around 0.1-0.7. But in our measurements, the UV absorbance was around 0.015-0.03, which was too small to be very accurate even with multiple measurements (> 5 times). To obtain the melting point T_m, we needed to measure the *change* in the ellipticity, but the absolute value of the ellipticity does not affect the value of T_m. The general slope of the curve confirmed that peptide **63** had more PPII helical character, while peptide **64** had less PPII helical character.

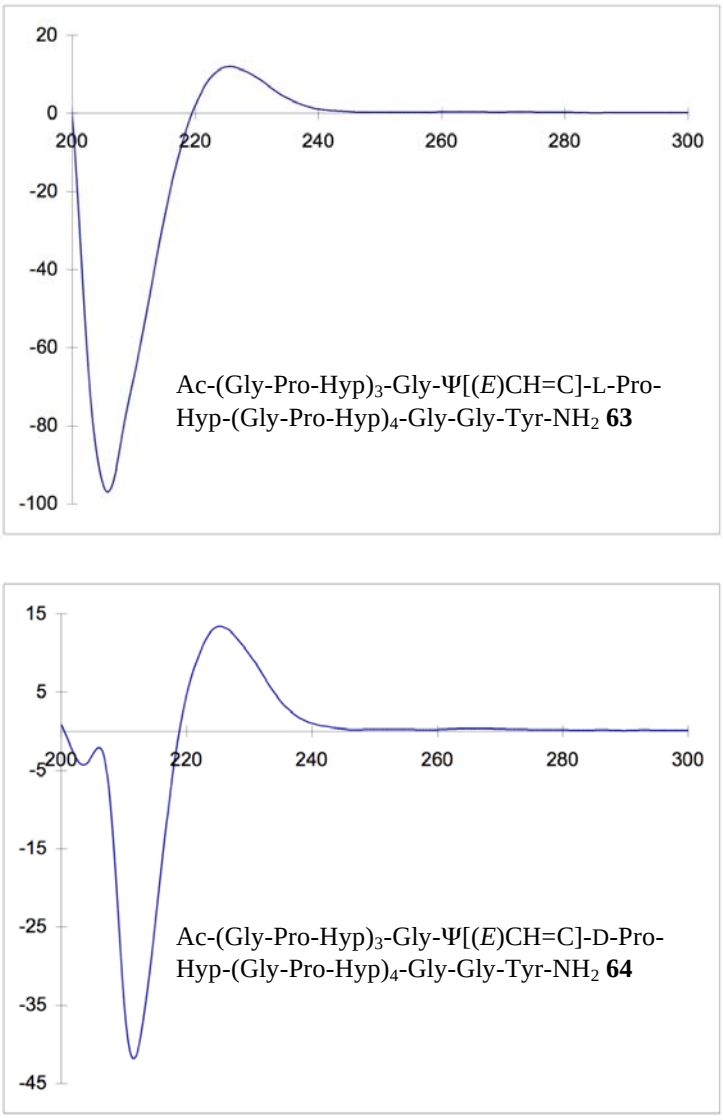


Figure 3.2 Full-range CD spectra of the peptides **63** (top, 0.40 mg/mL) and **64** (bottom, 0.60 mg/mL) from 300 nm to 190 nm.

To measure the ellipticities of the peptides at different temperatures, the CD spectra were obtained by scanning from 250 nm to 215 nm in 0.5 nm increments and averaged over three consecutive scans. For each peptide, the ellipticity at 225 nm decreased with the increase of temperature (Figure 3.3). At high temperature (over 65 °C), the maximum at 225 nm was very low (close to zero). With the increase of temperature, especially at high temperatures, the maximum moved to lower wavelengths. But the wavelength change was

very small and it was difficult to determine a trend due to a lack of data. The minimum also decreased with the increase of temperature, and the spectra below 220 nm became very noisy at high temperatures, which indicated the disassociation of the triple helix or PPII helical structure.

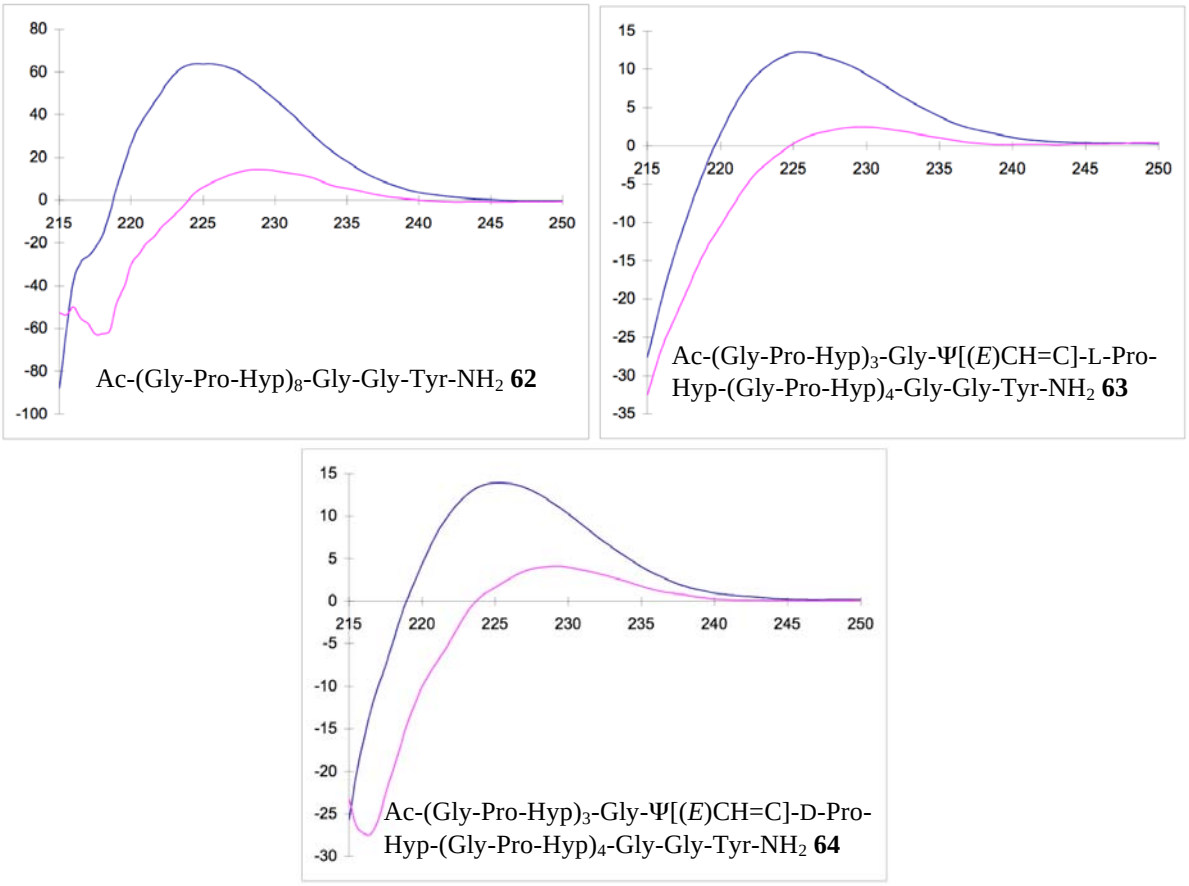


Figure 3.3 CD spectra of peptides **62**, **63** and **64**. Upper-left: control peptide 62 (0.22 mg/mL), upper-right: peptide 63 (0.40 mg/mL), bottom: peptide 64 (0.60 mg/mL), blue line: CD spectra at 3-4 °C, purple line: CD spectra at 67-68 °C.

The ellipticity at 226 nm was recorded and fitted to the two-state model. Control peptide **62** had a T_m value of 50.0 °C, which was close to the literature T_m value of 47.3 °C for Ac-(Gly-Pro-Hyp)$_8$-Gly-Gly-NH$_2$.[120] Peptide **63** formed a stable triple helix, but it had a lower T_m value of 28.3 °C. Peptide **64** showed a linear decrease of the ellipticity with

increasing temperature, which meant it did not form a triple helix. We therefore assumed that peptide **63** contained the Gly-L-Pro isostere, and peptide **64** contains the Gly-D-Pro isostere, because the D-amino acid could not form a triple helix in the host-guest system.[43,273]

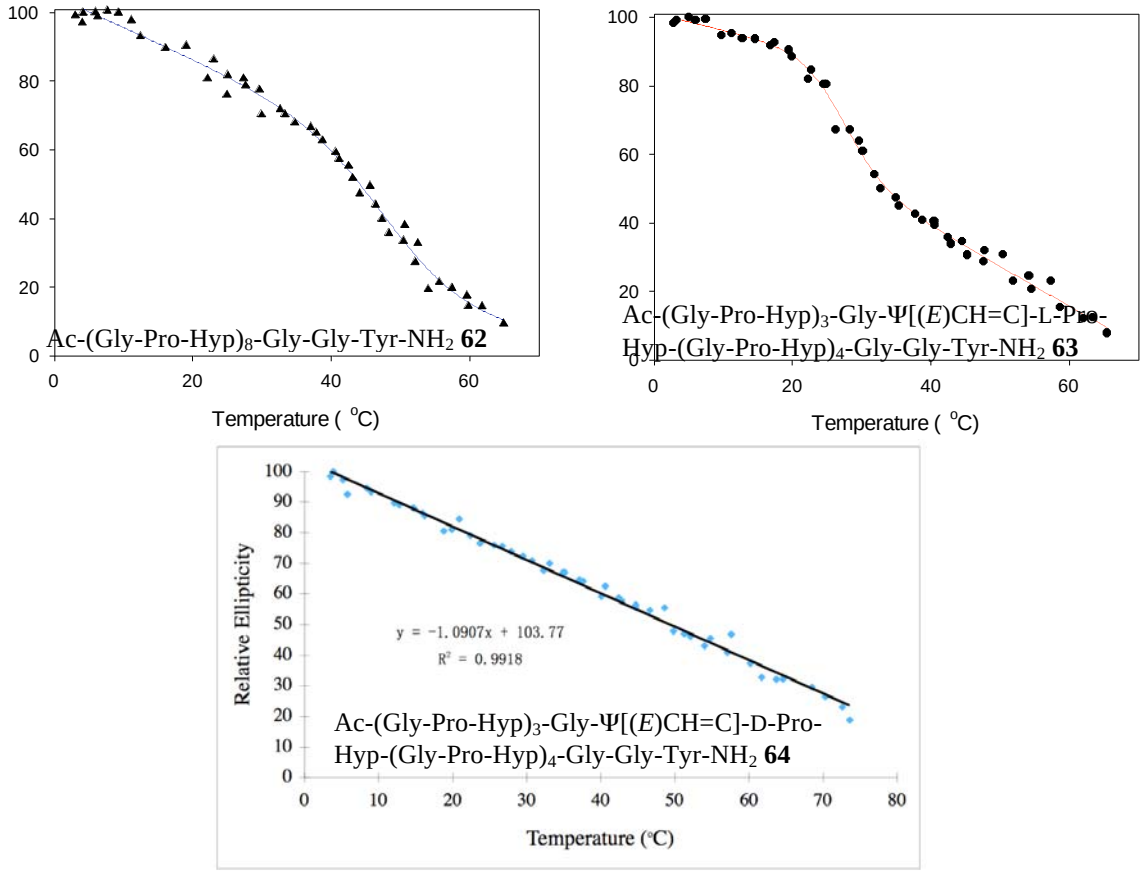


Figure 3.4 Relative ellipticity-temperature dependence of peptides **62**, **63** and **64**. The observed ellipticity at 4 °C was set to 100 % helical and zero observed ellipticity was set to 0 % helical. Upper-left: control peptide **62**; upper-right: peptide **63**; bottom: peptide **64**.

The ellipticity-temperature plot of control peptide **62** was divided into four regions, a near flat region below 10 °C, a linear regression region from 10 °C to 30 °C, a highly cooperative disassociation region from 35 °C to 55 °C, and another linear regression region above 60 °C (Figure 3.4). The ellipticity at 226 nm appeared not to change when the temperature was below 10 °C, which agreed with the fact that the ellipticity of the triple helix was temperature independent.[122,285] From 10 °C to about 30 °C, the ellipticity at 226 nm

showed a linear decrease with increasing temperature, which indicated that a small number of the triple helixes unfolded, and the unfolded peptide spectrum showed a linear temperature dependence.[286] The relatively low concentration was also responsible for the appearance of this region.[168] After the triple helix completely melted, there was still a small maximum around 225 nm, which showed a linear decrease with increasing temperature. This means that after the triple helixes completely dissociated to the coil state, some single strand peptide chains were in a PPII helical structure, and the PPII helix has a linear temperature dependence.[287,288]

Compared with the two linear temperature-dependent regions in the CD spectra of peptides **62** and **63**, the slope was steeper at high temperature, which meant more PPII helix-to-coil transitions occurred in the unfolded state. It seems that the triple helix did not unfold directly to the coiled state. With the increase of temperature, the three chains of the collagen triple helix unfolded and formed three separated single stands, but these still had the PPII helical structure. The PPII helixes unfolded to give complete coils upon further heating.

From Figure 3.4, the unfolding process of peptide **62** started with the triple helix beginning to unfold above 10 °C, a process that was slow. When the temperature was close to the melting point, the triple helix quickly unfolded to single strands within a relatively narrow temperature range. After the triple helix had completely unfolded, many single strands still had the PPII helix conformation, and their helicities decreased with increasing temperature.

Peptide **63**, which is assumed to have the Gly-L-Pro isostere in the sequence, showed a similar ellipticity-temperature correlation to control peptide **62** (Figure 3.4). The plot of the unfolding process of peptide **63** also had four regions: a near flat region below 10 °C, a linear

regression region from 10 °C to about 23 °C, a cooperative disassociating region from 23 °C to 40 °C, and another linear regression region above 45 °C. The relative molar ellipticity of peptide **63** was much lower than that of control peptide **62** in both the triple helical state and the unfolded state, which meant that the helicity of peptide **63** was smaller than the helicity of control peptide **62**, both in the triple helical and the single strand PPII helical state. At very high temperatures (over 60 °C), the molar ellipticities of the two peptides were very close, which meant that the two peptides were close to being completely random coils. Of the two linear temperature-dependent regions, the high temperature region had a steeper slope, which indicates a higher fraction of the peptides were in the single strand PPII helical state.

Peptide **64** showed a linear decrease in ellipticity at 226 nm with increasing temperature, which indicated that no triple helixes were formed due to the lack of a cooperative region. The maximum around 225 nm and the minimum at about 200 nm indicated the presence of the PPII helix, and the whole linear regression region represented the transition from the PPII helix to random coil. It was noteworthy that peptide **64** and the unfolded state of peptide **63** had similar slopes for the ellipticity with respect to temperature. This indicated that the two peptides had similar PPII helical structures in the unfolded state; therefore they showed similar temperature-dependent behavior. The difference in the molar ellipticity might be due to the inaccuracy of the measurements of the peptide concentrations.

The fraction conversion F was calculated from the equilibrium constant K with Eq. 7, and the equilibrium constant K was calculated with the calculated values of ΔH^0 and T_m with Eq. 5. It is obvious that the triple helixes unfolded to coils within a narrow temperature range (Figure 3.5).

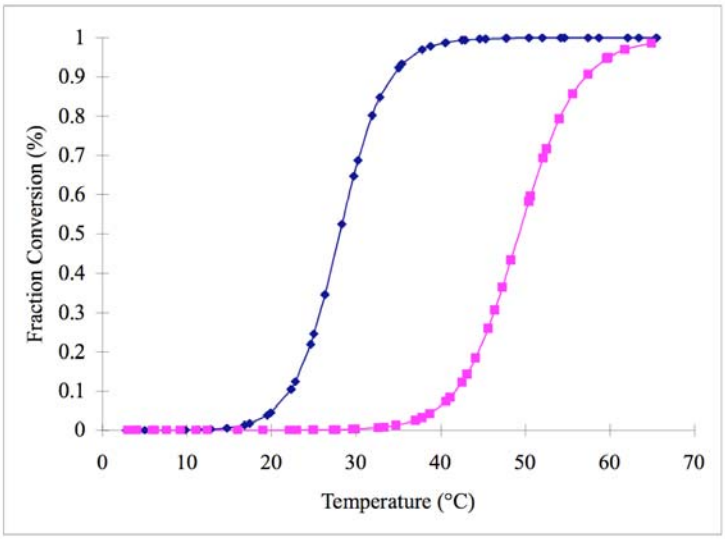


Figure 3.5 Fraction conversion-temperature dependence for the unfolding process of the peptides **62** and **63**. Purple line: control peptide **62**, blue line: peptide **63**.

3.4.4 Discussion

The stability of a host-guest peptide does not always match the stability of the homogeneous peptide containing exclusively the guest triplet. The Gly-Pro-Hyp guest tripeptide had the highest melting point of all the guest triplets in the host-guest system Ac-(Gly-Pro-Hyp)$_3$-Gly-Xaa-Yaa-(Gly-Pro-Hyp)$_4$-Gly-Gly-NH$_2$.[45] Moreover, the Pro-Pro-Gly guest triplet in this system had a higher T_m value (45.5 °C) than the Gly-Pro-Flp triplet (43.7 °C). While the homogeneous peptide (Pro-Flp-Gly)$_{10}$ formed a superstable collagen-like triple helix with a T_m value of 87 °C,[118] which was much higher than the T_m value of the homogeneous peptide (Pro-Pro-Gly)$_{10}$ (32.6 °C).[45] Persikov *et al.* believed that the Gly-Pro-Hyp host environment had a more stabilizing effect on Gly-Pro-Hyp and Gly-Pro-Pro due to the hydration network that was established between the host system and the guest triplet.[45] The Flp at the Yaa position stabilized the triple helix by pre-organizing the prolyl amide bond to the *trans*-conformation in the unfolded state, and helped the formation of the PPII helix entropically.[15,118] But no such preferential PPII helical structure was observed for Gly-

Pro-Flp in the host-guest system, and all ellipticities and temperature-dependent slopes were similar.[45]

Table 3.5 Calculated thermodynamic data of the collagen-like peptides **62** and **63**.

	T_m (°C)	ΔH^0 (kcal/mol)	ΔS^0 (cal/mol·K)
Peptide **62**	50.0	34.5	69
Peptide **63**	28.3	59.1	162

The calculated enthalpy was obtained by fitting the fraction conversion *F* to the equilibrium constant *K* with the combination of Eq. 2, 5, 6, 7 and 8 (Table 3.5). Although the value was not accurate, and varied with different fitting models, it provided a rough description of the thermodynamics of the triple helix-to-coil transition. Control peptide **62** had lower enthalpic and entropic barriers than the Gly-L-Pro isostere containing peptide **63** for the transition from the triple helix to the unfolded state.

It is noticeable that even after the disassociation of the collagen triple helix, peptide **62**, **63**, and **64** still showed a small maximum around 225 nm in the CD spectra, which indicated the presence of the PPII helix. These results indicated that the Gly-Pro-Hyp host peptides could form a single-strand PPII helix even if there was no collagen triple helix formation.

3.4.5 Summary

The all-natural control peptide **62** and the Gly-Pro alkenyl peptides **63** and **64** were successfully synthesized and purified by HPLC. Although the molecular mass ions were not observed in the MALDI-TOF spectra, the ion exchange experiment and LC-MS results confirmed the identities of the synthetic peptides, and ^{1}H NMR confirmed the presence of the intact alkene isostere.

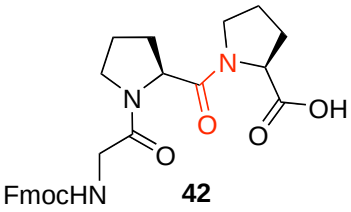
The melting point studies with CD showed that control peptide **62** formed a stable triple helix with a T_m value of 50.0 °C, which was very close to the literature value of 47.3 °C. The peptide **63** formed a stable triple helix with a T_m value of 28.3 °C, and it is assumed to contain the Gly-L-Pro isostere. The peptide **64** showed a linear decrease of the ellipticity with the increase of temperature, which indicates that no triple helix formed; it is therefore assumed that peptide **64** contains the Gly-D-Pro isostere. These results will be discussed all together at the end of this chapter.

*3.5 Design of the peptide containing the Pro-*trans*-Pro isostere*

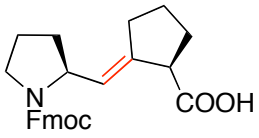
The Gly-Pro alkenyl peptide **63** showed an unexpected decrease in the collagen triple helix stability. The replacement of one amide bond between Gly-Pro with an (*E*)-alkene bond in the middle of a host-guest system, in which the one amide bond preceding the Pro residue at the Xaa position was locked in the *trans*-conformation, and the interchain backbone hydrogen bonds remain intact, showed a –21.7 °C decrease in T_m of the resulting collagen triple helix, compared to its control peptide Ac-(Gly-Pro-Hyp)$_8$-Gly-Gly-Tyr-NH_2 **62**. While the replacement of one Pro-Gly amide bond with a trans alkene in a host–guest peptide, in which one interchain backbone N–H···O=C hydrogen bond was removed with no improvement of *cis–trans* isomerization, resulted in a –57.5 °C decrease in the melting temperature (T_m) compared to its control peptide (Pro-Pro-Gly)$_{10}$.[223] These results lead to a question: how much does *cis-trans* isomerization affect the stability of the collagen triple helix?

A collagen-like polypeptide containing the conformationally locked Pro-*trans*-Pro alkene isostere was designed to study the influence of the amide bond preceding the Pro residue at the Xaa position, and to study how much *cis-trans* isomerization affects the

stability of the collagen triple helix (Figure 3.6). The host-guest peptide H-(Pro-Pro-Gly)$_4$-Pro-Ψ[(*E*)CH=C]-Pro-Gly-(Pro-Pro-Gly)$_5$-OH **66** and the control peptide H-(Pro-Pro-Gly)$_{10}$-OH **65** were designed. This host-guest peptide has the same host system as the one that Jenkins *et al*. used to study the triple helix stability of the Pro-Gly alkenyl peptide and ester peptide,[223] so it provided a better comparison.



42



43

Ac-(Gly-Pro-Hyp)$_3$-Gly-Pro-Pro-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH$_2$
65

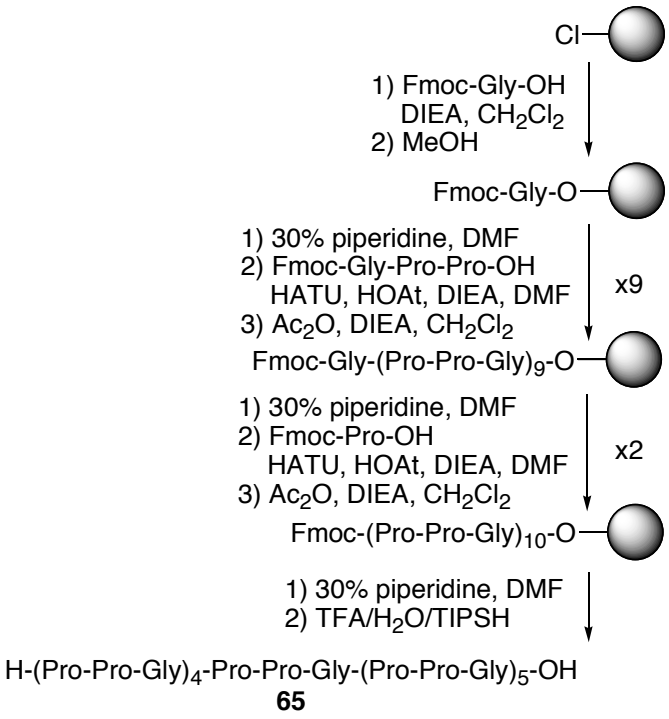
Ac-(Gly-Pro-Hyp)$_3$-Gly-Pro-Ψ[(*E*)CH=C]-L-Pro-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH$_2$
66

Figure 3.6 Design of the collagen-like host-guest polypeptides with the Pro-Pro isostere.

*3.6 Synthesis of the peptide containing the Pro-*trans*-Pro isostere*

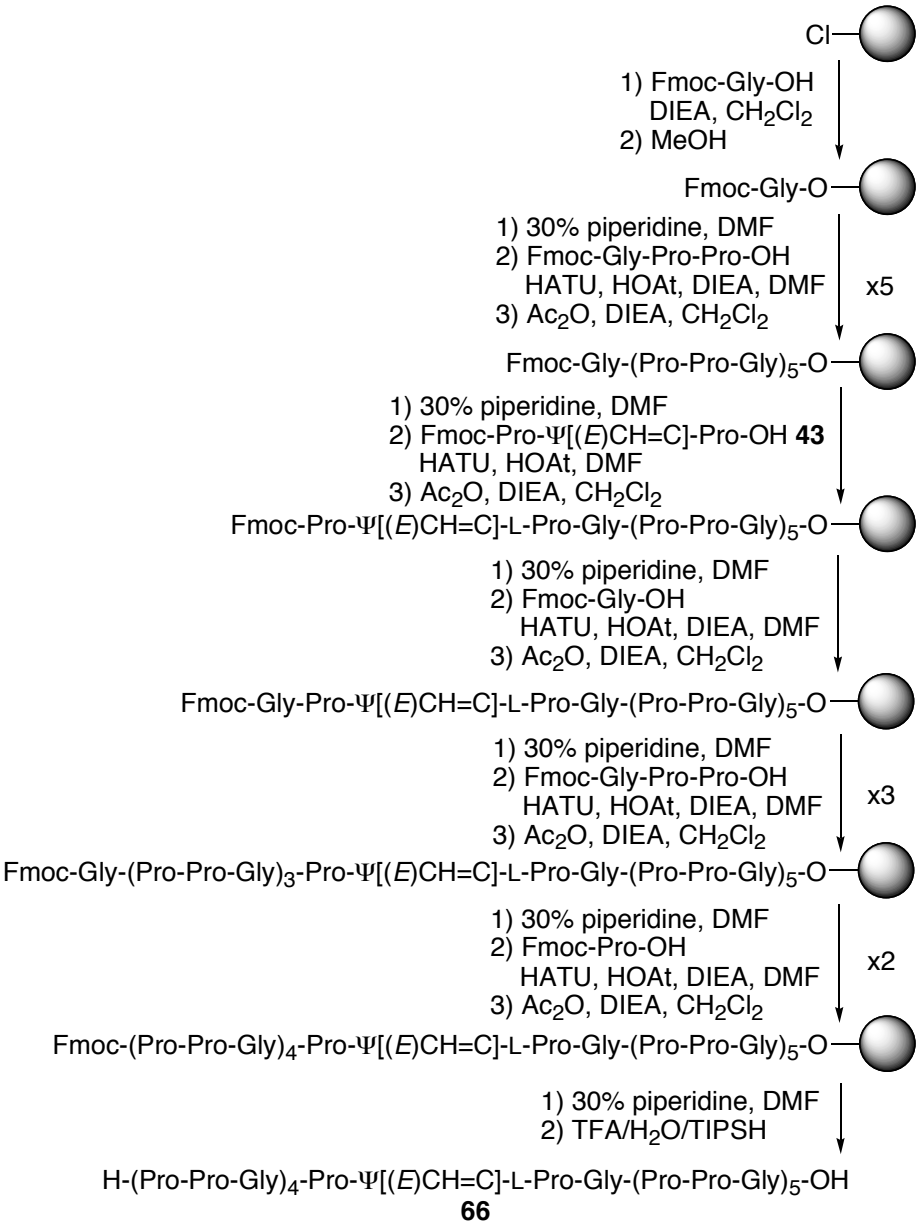
Control peptide **65** was synthesized by solid phase peptide synthesis on a 2-chlorotrityl chloride resin (Scheme 3.3). After the first amino acid was coupled, the loading was measured to obtain an accurate yield. A model synthesis using HATU-HOAt as the coupling reagent showed that the crude peptides were more pure, and the yield was higher. So HATU and HOAt were used as coupling reagents, and DMF was used as the solvent to synthesize peptides **65** and **66**.

Scheme 3.3 Solid phase peptide synthesis of control peptide **65**.



Peptide **66**, which contains the Pro-*trans*-Pro alkene isostere, was synthesized by a similar method used in the synthesis of peptide **65**. In a model reaction with collidine as the base for coupling alkene isostere **43**, an LC-MS chromatogram showed two peaks with the desired mass, which indicated that there was still partial double bond migration into the ring, even with the weak base collidine. So the dipeptide alkene isostere **43** was coupled with HATU-HOAt as the coupling reagents with no base present (Scheme 3.4), and monitored by reverse phase HPLC on a C18 column. No double bond migration was observed. There was only one peak with the desired mass in the LC-MS chromatogram, and the alkene proton was confirmed by NMR. Crude peptide **66** was obtained as a white solid.

Scheme 3.4 Solid phase peptide synthesis of the Gly-Pro isostere containing peptide **66**.



*3.7 Purification of the peptide containing the Pro-*trans*-Pro isostere*

The crude peptides were purified by reverse phase HPLC with the same procedure used for peptides **62, 63** and **64**. The same slow gradient (Table 3.1) from 5 % to 25 % for solvent B over 20 min gave the best separation, and the desired peak eluted at about 15 min.

The identity of the product was confirmed by LC-MS with the same analytical VYDAC protein C4 column. The preparative HPLC separation was performed on a VYDAC protein C4 reverse phase column, and fractions were collected manually. The purity of the peptides was confirmed by analytical HPLC and LC-MS. The pure peptides were obtained as white solids, and stored at –20 °C under N_2.

*3.8 Characterization of the peptide containing the Pro-*trans*-Pro isostere*

3.8.1 NMR

Rt and variable temperature ^{1}H NMR experiments were performed to identify the peptides. Peptide **66** was dissolved in 0.5 mL DMSO-d_6, and the ^{1}H NMR experiment was performed at rt. The spectra showed the presence of all proton peaks including the N-H protons. A proton exchange ^{1}H NMR experiment was performed. Two drops of D_2O were added to the NMR tube, and the solution was mixed and left standing for 1.5 h at rt to allow the complete deuterium exchange with N-H protons. The exchange ^{1}H NMR spectra showed two peaks with chemical shifts between 5.0 to 6.0 ppm. Both peaks at 5.3 and 5.5 ppm belong to the alkene proton. A variable temperature NMR experiment showed that these two peaks became closer with the increase of temperature, and nearly merged as one at 90 °C. These results indicated that the two alkene peaks were due to the conformation difference of the alkene proton in difference chains, possibly rotamers of the peptide.

3.8.2 LC-MS

LC-MS experiments were performed to prove the identities of the purified peptides. A gradient from 5 % to 95 % solvent B over 35 min was used, and peaks of interests were eluted around 15 min. The results are shown in Table 3.6.

The molecular mass peaks were not available because of the1500 Da cut-off of the detector. The doubly charged ($MH_2^{2+}/2$) and triply charged ($MH_3^{3+}/3$) peaks were used to identify the desired product peaks. The doubly charged and triply charged ions were from the same peak in the HPLC chromatogram. Theses two ions had mass values that were very close to the calculated values, which confirmed the identity of the purified peptides.

Table 3.6 Mass of the collagen-like peptides **65** and **66** from LC-MS. [a]

	MH^+		$MH_2^{2+}/2$		$MH_3^{3+}/3$	
	Calcd.[b]	Exp.	Calcd.[a]	Exp.	Calcd.[b]	Exp.
65	2530.28	ND[c]	1266.14	1266.3	844.43	844.6
66	2513.29	ND[c]	1257.64	1257.8	838.76	838.8

a) LC-MS was performed with a RP C-18 column (100 × 2.1 mm), electrospray ionization (positive ion mode), and a triple-quadrupole mass separator; b) The calculated mass is the calculated highest peak, not the average MW, c) ND = not determined. The MW is above the detectable upper-limit of the instrument.

3.8.3 CD analysis

The alkenyl peptide **66** removed an interchain backbone hydrogen bond that was present in the Pro-Pro-Gly host system, so we expected that peptide **66** would have a better or similar T_m value as the alkenyl peptide Raines and coworker studied. The melting point might not be measurable in ordinary PBS buffer.

To study the melting point of some very stable peptides and proteins, trimethylamino *N*-oxide (TMAO) was used.[128,223,289] TMAO is a natural osmolyte, which can thermodynamically stabilize folded proteins against the denaturing process by reducing the degree of backbone solvation.[289-292] Beck *et al.* used a high concentration of TMAO in

phosphate-buffered silane (PBS) to study the stability of some very unstable collagen peptides.[128] They found that the T_m value was linearly dependent on the concentration of TMAO.[128] Raines and coworkers used the same method to measure the melting temperature of ester and alkenyl collagen peptides by extrapolating the data to 0 M TMAO, and also observed the linear dependence of T_m with TMAO concentration.[130]

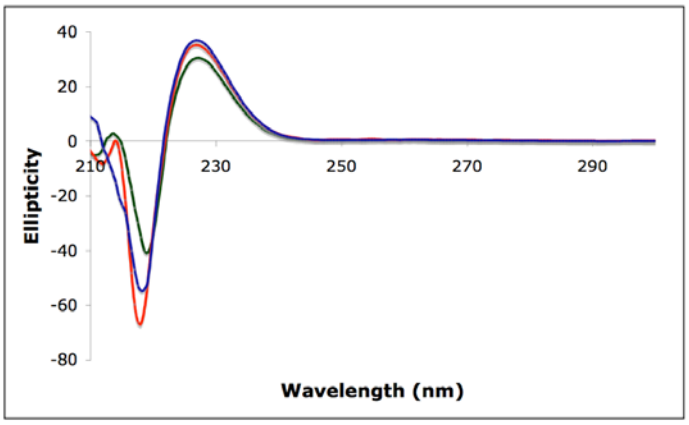


Figure 3.7 Full range CD spectra of control peptide **65** (1.3 mg/mL) in PBS buffer (9.6 mM, pH 7.4) at 3 °C. Green: in 0 M TMAO; red: in 0.5 M TMAO; blue: in 1.0 M TMAO.

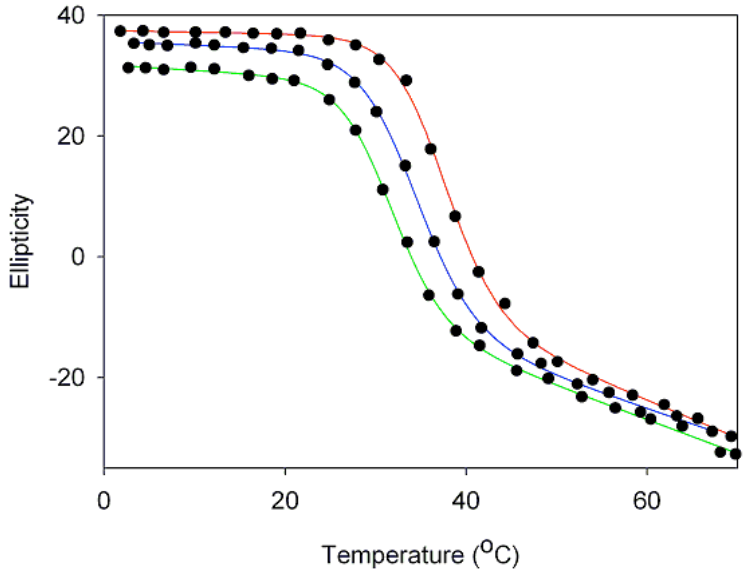


Figure 3.8 Molar ellipticity-temperature dependence of control peptide **65** in different concentrations of TMAO. All solutions were in 9.6 mM (pH 7.4) PBS buffer. Peptide concentration was 1.3 mg/mL. Red: 1.0 M TMAO; blue: 0.5 M TMAO; green: 0 M TMAO.

To obtain a better comparison, we used exactly same conditions that Raines used. Both control peptide **65** and the Pro-Pro alkenyl peptide **66** were dissolved in 9.6 mM PBS buffer (pH 7.4) in the presence of different concentrations of TMAO. The concentrations of peptide solutions were determined by the weight of the peptides. Because the concentrations of the peptide solutions were very close, and could be normalized in different concentrations of TMAO, the relative ellipticity of each peptide could be compared on the same graph. For control peptide **65**, the full range CD spectrum showed a maximum around 225 nm, and a minimum around 215 nm, which indicated the formation of the PPII helix. Peptide **65** showed a cooperative denaturing process with increasing temperature, which indicated the formation of the collagen triple helix. When the concentration of TMAO was less than 1.0 M, the ellipticity of the peptide increased with increasing TMAO concentration (Figure 3.7). After fitting the data to a two-state model, the T_m value also increased with the increase of TMAO concentration (Figure 3.8). However, unlike the T_m value, the increase of absolute ellipticity was not linear over the increase of TMAO concentration. This result indicated that molar ellipticity did not directly correspond to the T_m value.

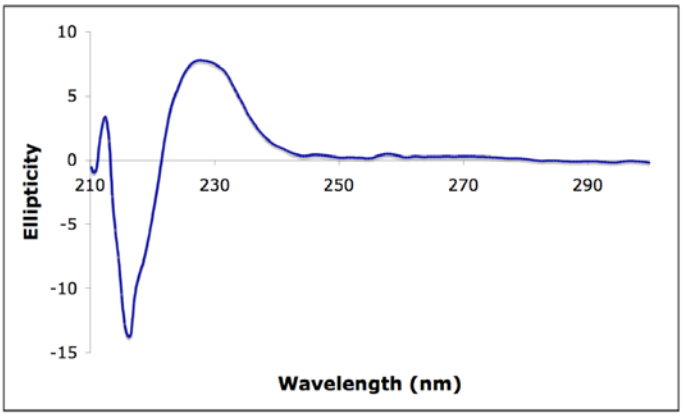


Figure 3.9 Full range CD spectra of control peptide **65** (1.3 mg/mL) in PBS buffer (9.6 mM, pH 7.4) in 3.0 M TMAO at 3 °C.

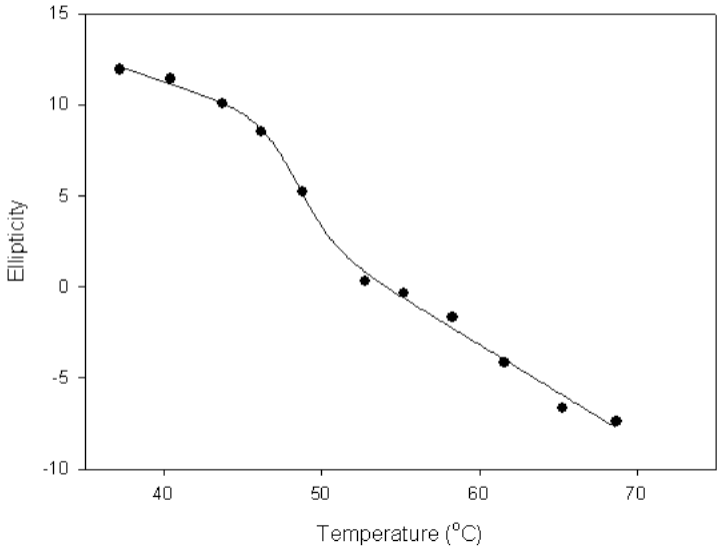


Figure 3.10 Molar ellipticity-temperature dependence of control peptide **65** in 3.0 M TMAO solutions.

For control peptide **65** in 3.0 M TMAO, the result was very interesting. The absolute ellipticity was much lower than in 1.0 M TMAO (Figure 3.9). But after fitting the data to the two-state model, the T_m value was right on the line (Figure 3.10). The slope and intercept of the T_m–[TMAO] plot were very close with or without the data at 3.0 M TMAO. To obtain a more accurate intercept, the data at 3.0 M TMAO was included.

The T_m values at each TMAO concentration were determined by fitting the data to a two state model, and the results are shown in Table 3.7. The melting point of control peptide **65** at 0 M TMAO was obtained by extrapolating the T_m value to 0 M TMAO. The T_m value was determined to be 31.6 °C at 0 M TMAO (Figure 3.11), which was very close to the measured T_m value of 31.5 °C in 9.6 mM PBS buffer with no TMAO present. Compared to the literature value, this result was close to the T_m value determined in the same concentration of PBS buffer (32.8 °C),[130] and the value determined in 50 mM AcOH

(31.4 °C).[141] The agreement of the extrapolation result and direct measurement indicated that this method worked in this host-guest peptide system.

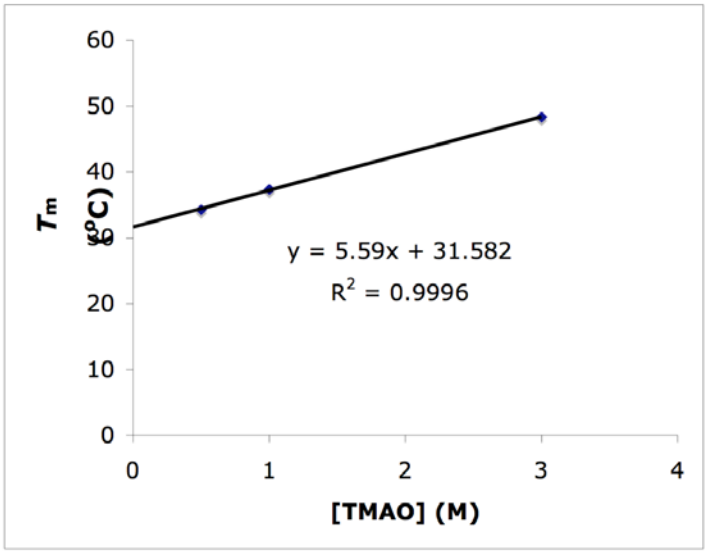


Figure 3.11 T_m–[TMAO] dependence of peptide **65**.

Table 3.7 T_m values of control peptide **65** at different concentrations of TMAO.

[TMAO] (M)	0	0.5	1.0	3.0
T_m (°C)	31.5	34.2	37.3	48.3

For the Pro-Pro alkenyl peptide **66**, the full range CD spectrum showed a maximum around 225 nm, and a minimum around 215 nm when the TMAO concentration was higher than 2.5 M, which indicated the formation of the PPII helix. When [TMAO] ≥ 2.5 M, peptide **66** showed a cooperative denaturing process with the increase of temperature, which indicated the formation of the collagen triple helix. For peptide **66**, the molar ellipticity also increased with the increase of the TMAO concentration (Figure 3.12). After fitting the data to a two-state model, the T_m value also increased with the increase of TMAO concentration (Figure 3.13). The molar ellipticity – TMAO concentration dependence of peptide **66** was not linear either.

The T_m values at each TMAO concentration were also determined by fitting the data to a two state model, and the results are shown in Table 3.8. The melting point of the Pro-Pro alkenyl peptide **66** also showed linear dependence on the concentration of TMAO when [TMAO] ≤ 3.5 M. The T_m value in 4.0 M TMAO was also measured. The molar ellipticity was higher than at lower concentrations of TMAO, but the T_m value was close to the one in 3.5 M TMAO. We believed that the T_m value of a peptide could not increase with the increase of TMAO concentration linearly at high concentrations. When the TMAO concentration reached a certain level, the melting point would be stable with [TMAO]. This might be the reason why the TMAO concentrations were never higher than 4.0 M in peptide stability studies.[128,223]

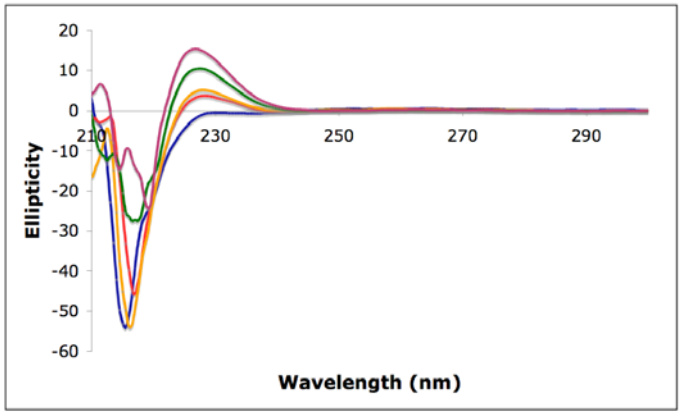


Figure 3.12 Full range CD spectra of peptide **66** (0.99 mg/mL in 3.25 M TMAO, 1.3 mg/mL for the rest) in PBS buffer (9.6 mM, pH 7.4) at 3 °C. Blue: in 2.5 M TMAO; red: in 3.0 M TMAO; yellow: in 3.25 M TMAO; green: in 3.5 M TMAO; purple: in 4.0 M TMAO.

After extrapolating the data to 0 M TMAO, the melting point of alkenyl peptide **66** was determined to be –22.0 °C (Figure 3.14). This number was much lower than the Gly-Pro alkenyl peptide **64** (T_m = 28.3 °C), and a little higher than the Pro-Gly alkenyl peptide that Jenkins *et al.* synthesized (T_m = –24.7 °C).[223] These results will be discussed later.

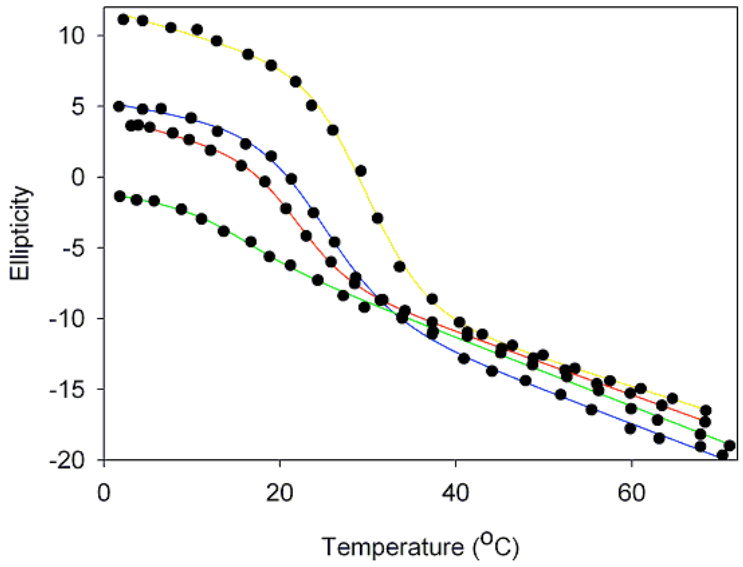


Figure 3.13 Molar ellipticity-temperature dependence of peptide **66** in different concentrations of TMAO. All solutions were in 9.6 mM (pH 7.4) PBS buffer. Peptide concentration was showed in parenthesis. Yellow: 3.5 M TMAO (1.3 mg/mL); blue: 3.25 M TMAO (0.99 mg/mL); red: 3.0 M TMAO (1.3 mg/mL); green: 2.5 M TMAO (1.3 mg/mL).

Table 3.8 T_m values of alkenyl peptide **66** in different concentration of TMAO.

[TMAO] (M)	2.5	3.0	3.2	3.5	4.0
T_m (°C)	14.4	22.1	24.6	29.3	28.7

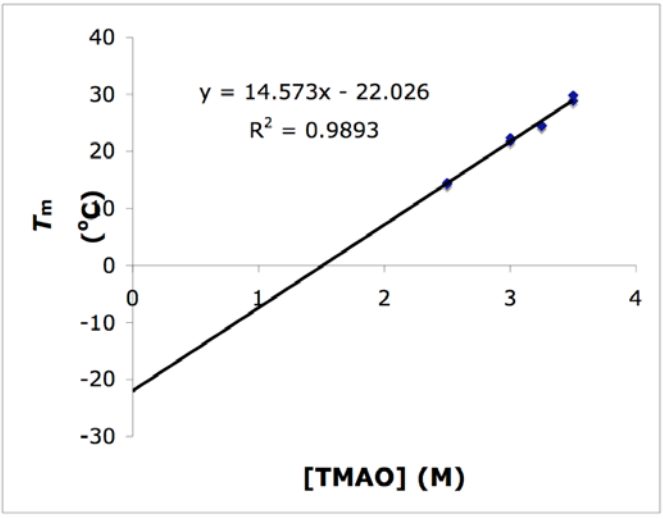


Figure 3.14 T_m–[TMAO] dependence of peptide **66**.

3.8.4 Discussion

The control peptide H-(Pro-Pro-Gly)$_{10}$-OH **65** and the Pro-Pro alkenyl peptide H-(Pro-Pro-Gly)$_4$-Pro-Ψ[(*E*)CH=C]-Pro-Gly-(Pro-Pro-Gly)$_5$-OH **66** showed similar CD spectra with or without TMAO. They both had maxima around 225 nm, which indicated the formation of the PPII helix, and minima around 215 nm. The minimum was not reliable because it exceeded the detection limit of the old lamp. With the increase of the TMAO concentration, the molar ellipticity usually increased. However, in 3.0 M TMAO, the molar ellipticity of control peptide **65** showed an unexpected decrease, while the T_m value still showed a linear increase with the increase of TMAO concentration. I think this can be explained by hydration of collagen-like peptides. The collagen triple helix has a very high degree of hydration, while the osmolyte reagent, TMAO, stabilizes proteins and peptides by reducing the degree of backbone solvation.[289-292] For some very stable collagen-like peptides, they already had strong hydration networks to stabilize the triple helix. High concentrations of TMAO would affect the hydration network, and caused the decrease of molar ellipticity. But during the denaturing process, the osmolyte effect still worked, and the T_m value was linear over the TMAO concentration in a certain range.

Unlike the Gly-Pro-Hyp based host-guest peptide system, the Pro-Pro-Gly based host-guest system was not as stable. The molar ellipticity was lower for the Pro-Pro-Gly host-guest peptide of the same length. During the denaturing process, when the triple helix dissociated, the maximum around 225 nm also disappeared, which indicated that there were no single-strand PPII helixes formed. These results indicated that monomeric Pro-Pro-Gly sequences were not stable in a PPII helix conformation, while monomeric Pro-Hyp-Gly peptides probably did form the PPII helixes.

For control peptide **65**, the T_m value showed linear dependence over the TMAO concentration until 3.0 M, which was the highest measured TMAO concentration (Figure 3.10). For the Pro-Pro alkenyl peptide **66**, the T_m value showed linear dependence over the TMAO concentration until 3.5 M. At 4.0 M TMAO, the T_m value was close to that at 3.5 M TMAO. As stated before, we believed this is due to the saturation of TMAO. Usually, osmolyte effect studies used 0 to 1 M TMAO. In our system, 4.0 M TMAO obviously already exceeded the linear range, so it was not included in the plot.

The denaturing process of control peptide **65** in 2.0 M TMAO was also measured. The melting region (sharp decrease region) was about 50-70 °C, when the molar ellipticity at 225 nm was negative in that temperature range, which indicated that no PPII helix was formed. It is very unlikely that the collagen triple helix forms without forming the PPII helixes, so the result was not reliable. One possible explanation is the competition between the hydration network and the osmolyte effect. The osmolyte effect can decrease the peptide backbone hydration degree, and may cause a decrease of molar ellipticity. At certain TMAO concentrations, the osmolyte effect and hydration both have a large influence on the peptide molar ellipticity. At lower temperature, the osmolyte effect is more important, and causes low molar ellipticity numbers. At higher temperature, the hydration network is more important, and causes a sharper decreasing curve in the molar ellipticity, and affects the determination of the melting point. We already had enough data points to determine the T_m value without TMAO, so the experiment in 2.0 M TMAO was not repeated.

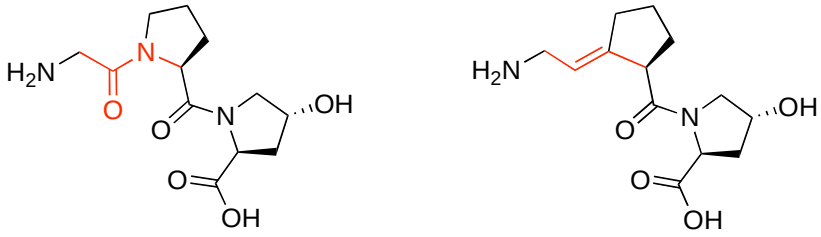
3.8.5 Summary

The all-natural control peptide **65** and the Pro-Pro alkenyl peptides **66** were successfully synthesized and purified by HPLC. LC-MS results confirmed the identities of

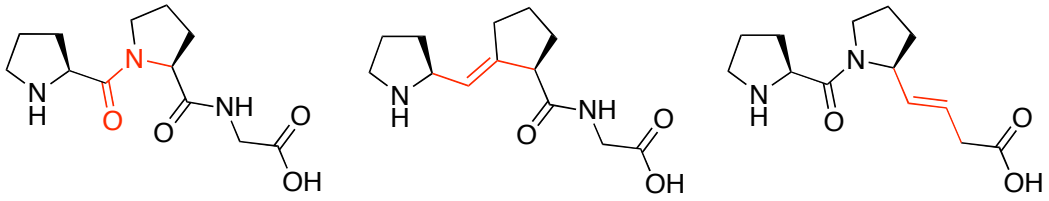
the synthesized peptides. ^{1}H NMR and variable temperature ^{1}H NMR confirmed the presence of the intact alkene isostere.

The melting point studies in PBS buffer in the presence of different concentrations of TMAO were performed. After extrapolation to 0 M TMAO, control peptide **65** showed a T_m value of 31.6 °C, which was very close to the T_m value of 31.5 °C measured in 0 M TMAO. The melting point of the Pro-Pro alkenyl peptide **66** was measured by the same method, and the T_m value was determined to be –22.0 °C after extrapolating to 0 M TMAO.

3.9 Discussion and conclusion



Control: Ac-(Gly-Pro-Hyp)$_3$-Gly-Pro-Hyp-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH$_2$ **62**
Gly-Pro Isostere: Ac-(Gly-Pro-Hyp)$_3$-Gly-Ψ[(*E*)CH=C]-L-Pro-Hyp-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH$_2$ **63**



Control: H-(Pro-Pro-Gly)$_4$-Pro-Pro-Gly-(Pro-Pro-Gly)$_5$-OH **65**
Pro-Pro Isostere: H-(Pro-Pro-Gly)$_4$-Pro-Ψ[(*E*)CH=CH]-Pro-Gly-(Pro-Pro-Gly)$_5$-OH **66**
Pro-Gly Isostere: H-(Pro-Pro-Gly)$_4$-Pro-Pro-Ψ[(*E*)CH=CH]-Gly-(Pro-Pro-Gly)$_5$-OH **67**

Figure 3.15 Structure of alkene isosteres, control and alkenyl peptides.

All of the alkene isosteres that have been studied are shown in Figure 3.15. The alkenyl peptides containing Gly-Pro and Pro-Pro alkene isosteres have been reported in this chapter. The Pro-Gly alkene isostere was studied by Raines and coworkers.[223] Both Gly-Pro and Pro-Pro alkene isosteres eliminated *cis-trans* isomerization by locking one prolyl amide bond in the *trans*-conformation with an (*E*)-alkene bond (Table 3.9). The Pro-Pro alkene

isostere also removed the C=O of the Xaa residue, and one triple helix interchain backbone hydrogen bond was broken. The Pro-Gly alkene isostere removed the N-H of Gly, so the interchain backbone hydrogen bond was also broken. The Pro-Gly alkene isostere also has an (*E*)-alkene bond, but the secondary amide bond preceding the Gly residue is already 99.9 % in the *trans*-conformation.[159] So there was no improvement of *cis*-*trans* isomerization for Pro-Gly alkene isostere.

From the structural analysis, we expected that the Gly-Pro alkenyl peptide would have the highest stability, then the Pro-Pro alkenyl peptide, and the Pro-Gly alkenyl peptide would have the lowest stability. Experimental results showed that the order of melting point differential from its control (ΔT_m) was just as expected (Table 3.9). The Gly-Pro alkenyl peptide showed a ΔT_m of –21.2 °C, the Pro-Pro alkenyl peptide showed a ΔT_m of –53.6 °C, and the Pro-Gly alkenyl peptide showed a ΔT_m of –57.5 °C compared to each of their control peptides, respectively.

Combining the experimental results and structural analysis, the stabilizing effects of each factor can be estimated. The Pro-Pro and Pro-Gly alkenyl peptides both broke an interchain backbone hydrogen bond, while the Pro-Pro alkene isostere locked one prolyl amide bond to the *trans*-conformation, and showed a 3.9 °C improvement of the ΔT_m value ($\Delta\Delta T_m$). This number is close to the ΔT_m (3.5 °C) of replacing a Pro residue with a Hyp in the middle of the host-guest peptide, $(Pro\text{-}Pro\text{-}Gly)_9$,[293] in which *cis*-*trans* isomerization was improved by the *exo*-puckering of Hyp. We believe this 3.9 °C improvement of $\Delta\Delta T_m$ can be attributed to the elimination of *cis*-*trans* isomerization by locking the prolyl amide bond at the Xaa position to the *trans*-conformation. This result indicated that the elimination of *cis*-

trans isomerization had a small contribution to the overall collagen triple helix stability, but it was not as significant as expected.

Table 3.9 Melting point comparisons and structural analysis of alkenyl peptides.

Gly-Pro-Hyp HG System [a]	T_m (°C)	ΔT_m (°C)	Interchain H-Bond [c]	*cis-trans* isomerization [c]
Control Peptide **62**	50.0	–	No	No
Gly-Pro isostere **63**	28.3	–21.7	No	Yes
Pro-Pro-Gly HG System [b]	T_m (°C)	ΔT_m (°C)	Interchain H-Bond	*cis-trans* isomerization
Control peptide **65**	31.6, 32.8[223]	–	No	No
Pro-Pro isostere **66**	–22.0	–53.6	Yes	Yes
Pro-Gly isostere **67**	–24.7	–57.5	Yes	No

(a) In 10 mM PBS buffer (pH 7.0); (b) in 9.6 mM PBS buffer (pH 7.4) with various concentration of TMAO, T_m value was obtained by extrapolation to 0 M TMAO; c) "yes" means change to its control; "no" means no change to its control.

The Gly-Pro and Pro-Pro alkenyl peptides both locked a prolyl amide bond in the *trans*-conformation, while the Gly-Pro alkene isostere did not break any interchain backbone hydrogen bond to disrupt the triple helix. The –31.9 °C $\Delta\Delta T_m$ between the Pro-Pro and the Gly-Pro alkenyl peptides could be attributed to the missing interchain backbone hydrogen bond. The special stabilizing effect of Hyp surely has a stabilizing effect on the resulting collagen triple helix, so this estimation is not accurate.

The Gly-Pro alkene isostere did not break any interchain backbone hydrogen bond, and eliminated *cis-trans* isomerization. The resulting peptide showed a ΔT_m of –21.7 °C compared to its control peptide. This indicates that some intrinsic properties of the amide bond have a huge influence on the stability of the collagen triple helix. Some factors that may be responsible for this stability decrease are discussed below.

1. Dipole moment. The dipole moment μ for the *N*-methyl acetamide is about 3.5-3.6 D, and the dipole moment μ for the methyl amide bond is 1.7 D; but for the alkene C=C bond, the dipole moment is much smaller (close to 0).[294] Favorable dipole-dipole interactions are very

important for the formation of the α-helix.[295,296] They may also be important for the formation of the PPII helix, too. The three chains of the collagen triple helix are parallel.[6,297] From the X-ray crystal structures, the direct interchain backbone hydrogen bonds point in the same direction along the axis of the collagen triple helix.[105] The dipole moment may also play an important role in directing the propagation of folding triple helixes.

2. Water-mediated hydrogen bonds and hydration network. The carbonyl groups of the Gly and Yaa residues are solvent exposed, and they do not participate in any direct peptide hydrogen bonds within the triple helix,[57] but they are possible hydrogen bond acceptors for water-mediated hydrogen bonds. The carbonyl group at the Yaa position is more hydrated than the carbonyl group of Gly residues. The crystal structures show that the carbonyl groups of the Gly residues are single-hydrated, and the carbonyl groups of the Yaa residues are double-hydrated.[53,105,298] The replacement of any amide bond with an alkene bond will break these water-mediated hydrogen bonds, and cause a decrease of the triple helix stability. The collagen triple helix has a high degree of hydration, and the hydration network is also important to the stability of the collagen triple helix. The replacement of the amide bond with a *trans*-alkene bond will also affect the hydration network, and decrease the stability of the collagen triple helix.[45] It is possible to form a C-H···O=C hydrogen bond,[223,299] but the bond strength is very weak, and it has a limited contribution to the stability of the collagen triple helix.

3. Unfavorable torsion angles. In the crystal structure of the collagen-like peptide (Pro-Hyp-Gly)$_{10}$, the average dihedral angle Φ of the Pro residues is –69.8°, and the average dihedral angle Ψ of the Gly residues is 174.9° (Figure 3.16).[298] The amide bond has partial double bond character,[154] but it is less rigid than a C=C double bond. The alkene bond in the

isosteres may make it more difficult for the resulting peptides to adopt favorable Φ, Ψ torsion angles.[223]

4. n→π* interactions. The non-bonding orbital of the oxygen atom of one amide bond and the π* orbital of the adjacent carbonyl group may have n→π* interactions, and stabilize the collagen triple helix.[300] We hypothesize that the n→π* interaction may be as important as the interchain backbone hydrogen bonding in stabilizing the collagen triple helix. During the collagen folding process, it helps directing the chains to the desired position, lowers the energy barrier of *cis-trans* isomerization, and preferentially stabilizes the amide bond preceding prolyl residues to the *trans*-conformation.[300] Replacing one amide bond with an alkene will cause the loss of six pairs of n→π* interactions in the collagen triple helix, and causes significant destabilization.

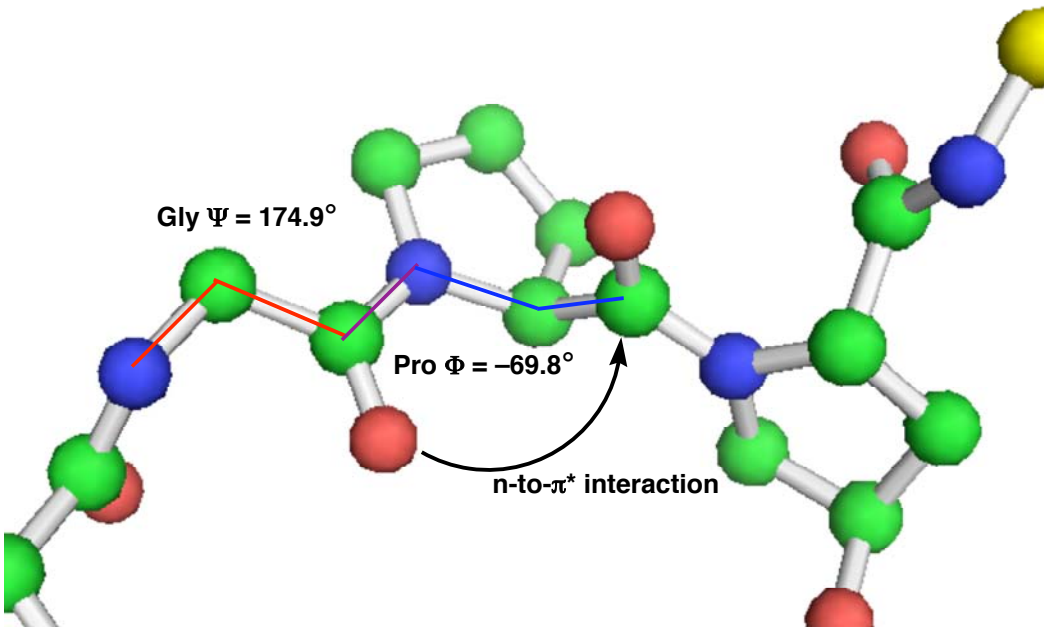


Figure 3.16 Torsion angles and possible n→π* interactions in the collagen-like peptide (Pro-Hyp-Gly)$_{10}$. (Generated from PDB file 1V7H[301] with MacPyMol 1.0)

Jenkins *et al.* also studied the peptide with an ester bond replacing one Pro-Gly amide bond.[223] This ester isostere also removed one interchain backbone hydrogen bond due to the removal of a Gly N-H. The ester group is not planar, which may alter the conformation of the resulting triple helix. The T_m value of the ester peptide was determined to be 10.7 °C by extrapolation to 0 M TMAO, which was –22.1 °C lower than the control peptide (Pro-Pro-Gly)$_{10}$.[223] This result confirmed that the backbone interchain hydrogen bond is a major factor for the stability of the collagen triple helix. Compared to the –57.5 °C ΔT_m value of the Pro-Gly alkenyl peptide, the importance of the carbonyl group was also proven.

In summary, replacing any amide with an (*E*)-alkene bond caused a significant stability decrease in the resulting collagen triple helix. We confirmed that the backbone interchain hydrogen bond between the Gly N-H and the Xaa C=O the adjacent chain is one of the major forces in stabilizing the collagen triple helix. *Cis-trans* isomerization has a limited contribution to the triple helix stability. The intrinsic properties of the amide bond also have a huge influence on the stability of the collagen triple helix.

3.10 Synthesis of collagen-like polypeptide by polymerization

Polymerization from tripeptide monomers is another method to synthesize collagen-like polypeptides. Compared to solid phase peptide synthesis, polymerization is easier to handle and yields much longer polypeptides. Although the polymers obtained by polymerization methods had high polydispersity, this method yielded polymers with long strands (10-20 nm) which still adopted triple helical structures.[302] Kishimoto *et al.* used the tripeptide Pro-Hyp-Gly as the monomer and synthesized a polypeptide with a molecular weight of over 10,000 Da.[303]

Two methods are widely used to synthesize polypeptides from tripeptide monomers. One method is the direct polymerization of the tripeptide monomer with both free amine and acid termini.[302-304] This method needs a weak base, usually tertiary amines, and a coupling reagent such as DCC, EDC or HBTU. Another method is via the activated ester method. In this method, an activated ester monomer is synthesized first, and the activated ester monomer is then polymerized with one equivalent of tertiary amine in DMF, DMSO, or NMP.[305] The most common activated esters are *p*-nitrophenyl ester (-ONP)[306,307] and pentachlorophenyl ester (-OPCP).[308] These methods are faster and cleaner than the direct polymerization method, but it takes more effort to synthesize the activated ester monomer. When these two methods are compared, direct polymerization of a tripeptide monomer with both free amine and acid terminals appears to be easier, and the required reagents are more readily available. Therefore, the free amino acid coupling method was used first.

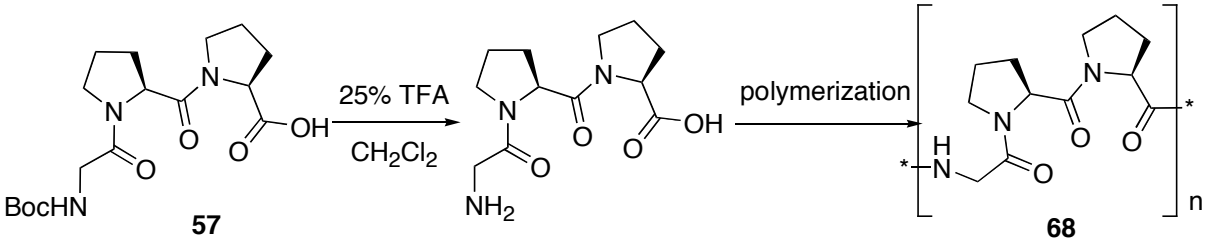
3.10.1 Free amino acid coupling

Tripeptide Gly-Pro-Pro was used as the model monomer for polymerization to avoid any problems that may be caused by the unprotected hydroxyl group of Hyp. The Boc-protected tripeptide Boc-Gly-Pro-Pro-OH **57** had already been synthesized. The Boc-protecting group was removed by 25 % TFA in CH_2Cl_2 and the excess TFA was removed under high vacuum. The free tripeptide monomer H-Gly-Pro-Pro-OH was obtained as dark yellow oil, and used without further purification (Scheme 3.5).

Several combinations of coupling reagents and solvents were tried in the synthesis the poly-(Gly-Pro-Pro) polymer (Table 3.10). First, HBTU was used as the coupling reagent and DIEA as the base. HOBt was added to minimize the racemization of the amino acid residue with a chiral α-carbon. The reaction was carried out in DMF at 60 °C for three days. White

solids precipitated as the reaction proceeded. The resulting white solid could not be dissolved in DMF, H_2O, acetone, CH_2Cl_2 or chloroform, while it could be dissolved in NMP or TFA. By Gel Permeation Chromatography (GPC) analysis, the molecular weight of this solid was less than 900 Da, which indicated that it was not the desired polypeptide. The ^{1}H NMR spectrum showed that it was possibly HBTU derivatives, as the spectrum was very similar to the standard spectrum of pure HBTU. The result indicated that the HBTU and HOBt might also undergo some polymerization-like reactions under these conditions.

Scheme 3.5 Polymerization with free amino acid coupling.



57 **68**

The polymerization was then performed in NMP, with HBTU-HOBt as the coupling reagents and DIEA as the base at 60 °C for three days. This time no precipitate was found, and only a brown clear liquid was observed. The GPC analysis showed a clear polymer peak, but the mass range of the synthesized polymer could not be obtained. To obtain an accurate molecular weight range for a polymer, the signals from all three detectors are required. The synthesized polymer only showed a peak in two of the detectors, which was inadequate for the calculation of the mass range. A ^{1}H NMR spectrum was also obtained, and it indicated that the desired $(Gly\text{-}Pro\text{-}Pro)_n$ polypeptide had been synthesized. From the MALDI-TOF analysis, the molecular weight was not as high as expected (about 2000-3000 Da). This might be because the polymerization process needed a very high concentration of the monomer, after several initial couplings, the concentration became much lower, thus preventing the formation of high molecular weight polymers. Moreover, it was very difficult to completely

remove the HOBt. Dialysis with a membrane could remove most of the HOBt, but it also caused a great loss of the product when the molecular weight was not high.

Table 3.10 Combinations of coupling reagents and solvents.

	1	2	3	4	5
Solvent	DMF	NMP	THF	HMPA	THF
Reagent	HBTU	HBTU	EDC	DCC	DCC

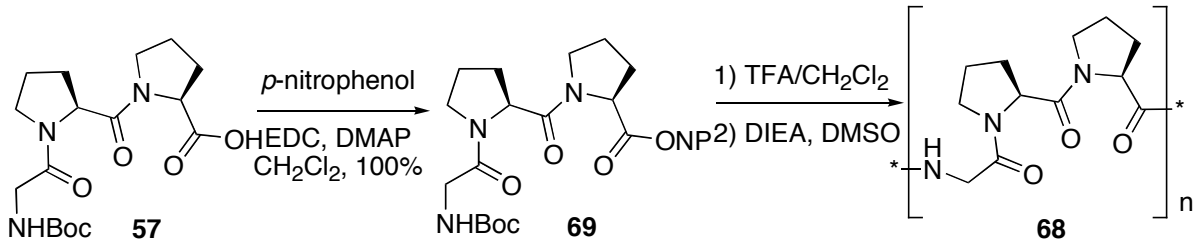
The polymerization was also performed in HMPA with DCC, HOBt and Et_3N at rt for 10 days,[304] in THF with EDC, HOBt and DIEA at rt for 7 days, and in THF with DCC, HOBt and DIEA. (Dr. Yinnian Lin, Personal communication) The same purification problem remained. By using DCC, DCU was removed by filtration, and Et_3N or DIEA by rotary evaporation. Most of the HOBt could be removed by acidic work-up followed by filtration to remove the precipitates, but still some of the HOBt remained mixed with the desired polymer. Additionally, the molecular weight range was not high enough (less than 2000 Da).

The free amino acid coupling method yielded the desired polypeptide, but again the molecular weight was not high enough. The concentration of the monomer was very important for the formation of polymers with high molecular weight. Purification was also a problem because traditional flash chromatography could not be used to purify the polar polypeptides. Once polypeptides with high molecular weights were synthesized, a dialysis membrane with a higher cut-off MW could be used, which might help with the removal of low MW impurities. It has also been reported that the polymerization of the Pro-Hyp-Gly monomer in aqueous solutions yielded a polymer with a high molecular weight.[303] More model reactions are needed to optimize the polymerization conditions.

3.10.2 Activated ester

The activated ester method to synthesize the model polypeptide $(Gly\text{-}Pro\text{-}Pro)_n$ was also tried. Due to the availability of the reagents, *p*-nitrophenyl ester (-ONP) was used as the activated ester. Starting from the Boc-protected tripeptide Boc-Gly-Pro-Pro-OH **57**, *p*-nitrophenol was coupled with EDC (Scheme 3.6). No base was used because the activated ester was not stable in base. The Boc-group was removed, and the resulting TFA salt was polymerized without purification with equivalent Et_3N in DMSO. A brown glue-like slurry was obtained, and the polymer was precipitated in Et_2O.

Scheme 3.6 Polymerization with activated ester.



Two types of polymers were obtained. A colorless film-like polymer was obtained when Et_2O was added, but it disappeared during the membrane dialysis (cut-off MW = 3000 Da) in water and became a protein-like white solid. A brown, sand-like solid was also obtained when Et_2O was added, but it did not disappear during the dialysis process. The sand-like polymer did not dissolve in water, MeOH, EtOH or DMF, and it had limited solubility in the presence of TFA. From MALDI-TOF, the highest observed molecular weight was 4872 Da, with a repeating unit of 251 Da, which meant that the resulting polypeptide had at least 19 repeating units.

Both methods gave polymer-like polypeptides, but the molecular weights of the synthesized polypeptides were not high enough. More model reactions are needed to

optimize the polymerization conditions before polymerizing the tripeptide with alkene isosteres.

3.11 Summary

The Gly-Pro alkenyl peptides **63**, **64** and their control peptide Ac-(Gly-Pro-Hyp)$_8$-Gly-Gly-Tyr-NH_2 **62** were synthesized and purified. All three peptides showed a maximum around 225 nm and a minimum close to 200 nm, which indicated the formation of the PPII helixes. Control peptide **62** with Gly-Pro-Hyp as the guest triplet formed a triple helix with a T_m value of 50.0 °C. Peptide **63** with Gly-Ψ[(*E*)CH=C]-L-Pro-Hyp as the guest triplet also formed a stable triple helix, but the T_m value was 28.3 °C, 21.7 °C lower than the control. Peptide **64** with Gly-Ψ[(*E*)CH=C]-D-Pro-Hyp as the guest triplet showed a linear decrease in the ellipticity at 226 nm with the increase of temperature, which indicated that no triple helix was formed.

The Pro-Pro alkenyl peptide **66** and its control peptide H-(Pro-Pro-Gly)$_{10}$-OH **65** were also synthesized and purified. The T_m value of control peptide **65** was determined to be 31.6 °C by extrapolation to 0 M TMAO in PBS buffer, which was very close to the value of 31.5 °C measured in 0 M TMAO. The Pro-Pro alkenyl peptide **66** began to show a maximum around 225 nm when the concentration of TMAO was higher than 2.5 M. After extrapolation to 0 M TMAO, the T_m value of the Pro-Pro alkenyl peptide **66** was determined to be –22.0 °C. These results indicated that replacing any amide with an (*E*)-alkene bond caused a significant stability decrease in the resulting collagen triple helix. The backbone interchain hydrogen bond between the Gly N-H and the Xaa C=O the adjacent chain is one of the major forces in stabilizing the collagen triple helix. *Cis-trans* isomerization has a limited

contribution to the triple helix stability. The intrinsic properties of the amide bond also have a huge influence on the stability of the collagen triple helix.

The model polypeptide $(Gly\text{-}Pro\text{-}Pro)_n$ was synthesized by two methods. The free amino acid coupling method yielded a polymer with a molecular weight around 2000-3000 Da, but the impurities were very difficult to remove when the MW of the polymer was not high. The activated ester method was faster and cleaner, and yielded a polypeptide with a molecular weight around 5000 Da. More model reactions are needed to optimize the polymerization conditions.

Experimental

General Information. Amino acid derivatives, resins, and reagents were purchased. Unless otherwise indicated, all reactions were carried out under N_2 sealed from moisture. Anhydrous THF was obtained by refluxing from Na-benzophenone. Anhydrous CH_2Cl_2 was dried by passing through a dry alumina column. Other anhydrous solvents were used directly from sealed bottles, which were stored under Ar. Brine (NaCl), $NaHCO_3$, $Na_2S_2O_3$, and NH_4Cl refer to saturated aqueous solutions unless otherwise noted. Flash chromatography was performed on 32-63 μm or 230-400 mesh silica gel with reagent grade solvents. NMR spectra were obtained at ambient temperature in $CDCl_3$ unless otherwise noted. ^1H and ^{13}C NMR spectra were obtained at 400 and 100 MHz, respectively, unless otherwise noted. Minor rotamer chemical shifts are shown in parenthesis.

HPLC purification: Analytical HPLC was performed on a VYDAC Protein C4 reverse phase column (5 μm, 4.6 × 250 mm) on a Beckman P/ACE 168 dual-solvent HPLC system equipped with a P/ACE 215 UV detector. A binary solvent system (A: 0.1 % TFA in H_2O, B: 0.1 % TFA in CH_3CN) with a flow rate of 1.0 mL/min was used. The chromatogram was

simultaneously monitored at 210 nm and 260 nm. Preparative HPLC was performed on VYDAC Protein C4 reverse phase column (10 μm, 22 × 250 mm) with a Varian PrepStar 218 dual-solvent HPLC system with a ProStar 320 UV detector. A binary solvent system (A: 0.1 % TFA in H_2O, B: 0.1 % TFA in CH_3CN) with a flow rate of 15 mL/min was used. LC-MS experiments were performed on an Agilent 1100 HPLC system with a Thermo Finnigan TSQ triple quad mass detector on a reverse phase C18 column (2.1 × 100 mm). A binary solvent system (A: 1 % formic acid in H_2O, B: 1 % formic acid in CH_3CN) with a flow rate of 0.2 mL/min was used.

UV measurement: The concentration of the peptide solutions was determined by UV absorbance at 280 nm with a Shimadzu UV 1601 spectrometer. The peptide solutions (100 μL) were measured with a volumetric automatic pipette and diluted to 1 mL in the volumetric flask with 6 M guanidine hydrochloride solution.

Solid phase peptide synthesis: Ac-(Gly-Pro-Hyp)$_8$-Gly-Gly-Tyr-NH_2 **62**, Ac-(Gly-Pro-Hyp)$_3$-Gly-Ψ[(*E*)CH=C]-L-Pro-Hyp-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH_2 **63**, Ac-(Gly-Pro-Hyp)$_3$-Gly-Ψ[(*E*)CH=C]-D-Pro-Hyp-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH_2 **64**. Rink amide MBHA resin (30 mg, 0.66 mmol/g) was placed in a 5 mL polypropylene tube and swelled in CH_2Cl_2 (2 mL) for 1 h, then 20 % piperidine in NMP (2 mL) was added, and the mixture was shaken for 40 min to remove the Fmoc group. To couple single amino acids (Gly and Tyr), HBTU (22 mg, 0.059 mmol), HOBt (9.1 mg, 0.059 mmol), Fmoc-amino acid (0.059 mmol) and DIEA (0.021 mL, 0.12 mmol) were dissolved in NMP (2 mL), added to the resin, and shaken at 30 °C for 30 min, and the coupling was repeated. Fmoc-Gly-Pro-Hyp(^{t}Bu)-OH **25** (34 mg, 0.059 mmol), HBTU (22 mg, 0.059 mmol), HOBt (9.1 mg, 0.059 mmol), and DIEA (0.021 mL, 0.12 mmol) were dissolved in NMP (2 mL) and the mixture was shaken at 30 °C

for 1 h, and the coupling was repeated. Alkene isostere **26** (11 mg, 0.021 mmol), HATU (23 mg, 0.059 mmol), HOAt (8.1 mg, 0.059 mmol), and 2,4,6-collidine (0.016 mL, 0.12 mmol), were dissolved in NMP (2 mL), and the mixture was shaken at 30 °C for 2 h. After each coupling, the peptide was capped with 10 % Ac_2O and 10 % DIEA in CH_2Cl_2 (2 mL) for 20 min. The Fmoc group was removed with 20 % piperidine in NMP for 10 min and 20 min, respectively. The final Fmoc group was removed with 20 % piperidine in NMP, and the peptide was capped with 10 % Ac_2O and 10 % DIEA in CH_2Cl_2 (2 mL) for 20 min. The peptide was cleaved from the resin by treatment with 2 % H_2O, 3 % triisopropyl silane in TFA (5 mL) for 3.5 h. After removing the resin by filtration and concentrating the solution in vacuo, Et_2O (10 mL) was added, and the white solid that precipitated was collected. Reverse phase analytical HPLC was performed by injecting 20 μL in MeOH on a 5 μm protein C4 column (250 × 4.6 mm), and semi-preparative HPLC by injecting 1.0 mL in MeOH on a 10 μm protein C4 column (250 × 22 mm), using solvents A: 0.1 % TFA in H_2O, and B: 0.1 % TFA in CH_3CN, with UV detection at 210 nm.

CD analysis. The concentrations of peptides in PBS (10 mM Na_xPO_4, 150 mM NaCl, pH 7.0) were determined by the UV absorption of the Tyr residue at 280 nm. The concentration of peptide **62** was 0.087 mM, **63** was 0.16 mM, and **64** was 0.24 mM. The peptide solutions were incubated at 4-5 °C for ≥ 72 h. The CD spectra were obtained with a spectropolarimeter in 0.5 nm increments, 1 nm bandwidth, and 0.2 cm pathlength at a scan speed of 100 nm/min. The spectra were averaged over three consecutive scans, and blank buffer scans were subtracted from the baseline. Full scan spectra from 190 to 300 nm were obtained at 4 °C. The solutions were heated from 3 °C to 70 °C in 3 °C increments with a 5 min equilibration time at each temperature before measurement. The temperature was measured in the cell with

a temperature probe. The ellipticity at 226 nm was monitored at each temperature and averaged over three consecutive measurements. The unfolding process of each peptide was measured twice, and the data obtained were processed separately and averaged to obtain the T_m value. The data were fit to the following equations using SigmaPlot v.10.0.

$$K = \exp\left[\frac{\Delta H^0}{RT}\left(\frac{T}{T_m} - 1\right) - \ln(0.75c_0^2)\right]$$

$$F = \frac{[\Theta] - [\Theta]_c}{[\Theta]_h - [\Theta]_c}$$

$$K = \frac{F}{1 - F}$$

$$[\Theta]_h = [\Theta]_{h,T_m} + a(T - T_m)$$

$$[\Theta]_c = [\Theta]_{c,T_m} + b(T - T_m)$$

Solid phase peptide synthesis: H-(Pro-Pro-Gly)$_{10}$-OH **65**, H-(Pro-Pro-Gly)$_4$-Pro-Ψ[(*E*)CH=C]-L-Pro-Gly-(Pro-Pro-Gly)$_5$-OH **66** were synthesized by solid phase peptide synthesis. 2-Chlorotrityl chloride resin (400 mg, 1.2 mmol/g) was placed in a 5 mL polypropylene tube, swelled in CH_2Cl_2 (2 mL) for 1.5 hr, capped with MeOH (2 mL), and was dried in vacuo for 14 h. The resin (10 mg) was dried in vacuo for 16 h, stirred in and stirred in 30 % piperidine in DMF (0.5 mL) for 30 min, then diluted to 20.0 mL with EtOH. UV analysis at 300 nm (ε_{300} = 6566 M^{-1} cm^{-1}) showed that the actual loading was 0.67 mmol/g (56 %). 2-Chlorotrityl chloride resin was separated into two polypropylene tubes (110 mg, 0.67 mmol/g). To couple single amino acids (Gly and Pro), HATU (85 mg, 0.22 mmol), HOAt (30 mg, 0.22 mmol), Fmoc-amino acid (0.22 mmol) and DIEA (0.075 mL, 0.43 mmol) were dissolved in DMF (2 mL), added to the resin, and shaken at 30 °C for 30 min, and the coupling was repeated. Fmoc-Gly-Pro-Pro-OH (74 mg, 0.15 mmol), HATU (57 mg, 0.15 mmol), HOAt (20 mg, 0.15 mmol), and DIEA (0.050 mL, 0.29 mmol) were dissolved in DMF (2 mL) and the mixture was shaken at 30 °C for 2 h.

Fmoc-Pro-Ψ[(*E*)CH=C]-L-Pro-OH (33 mg, 0.079 mmol), HATU (57 mg, 0.15 mmol), and HOAt (20 mg, 0.15 mmol) were dissolved in DMF (2 mL), and the mixture was shaken at 30 °C for 2.5 h, monitoring with reverse phase HPLC. After each coupling, the peptide was capped with 10 % Ac_2O and 10 % DIEA in CH_2Cl_2 (2.5 mL) for 15 min. The Fmoc group was removed with 30 % piperidine in DMF for 20 min and 15 min, respectively. The peptide was cleaved from the resin by treatment with 2 % H_2O, 3 % triisopropyl silane in TFA (5 mL) for 3.5 h. After removing the resin by filtration and concentrating the solution in vacuo, Et_2O (10 mL) was added, and the white solid that precipitated was collected. Reverse phase analytical HPLC was performed on a 5 μm protein C4 column (250 × 4.6 mm), semi-preparative HPLC on a 10 μm protein C4 column (250 × 22 mm), using solvents (A) 0.1 % TFA in H_2O, and (B) 0.1 % TFA in CH_3CN, with UV detection at 210 nm.

CD analysis. The concentrations of peptides in PBS (20.4 mg KCl, 800 mg NaCl, 19.9 mg KH_2PO_4, 116 mg Na_2HPO_4, in 100 mL H_2O, pH 7.4) were determined by the weight of the peptides. The concentration of control peptide **65** was 1.26 mg/mL, and the concentration of peptide **66** was 1.32 mg/mL. The peptide solutions were incubated at 5 °C for ≥ 48 h. The CD spectra were obtained with a spectropolarimeter in 0.5 nm increments, 1 nm bandwidth, and 0.2 cm pathlength at a scan speed of 100 nm/min. The spectra were averaged over four consecutive scans. The solutions were heated from 2 °C to 70 °C in 3 °C increments with a 10 min equilibration time at each temperature before measurement. The temperature was measured in the cell with a temperature probe. The ellipticity at 227 nm was monitored at each temperature and averaged over four consecutive measurements. The unfolding process of each peptide was measured twice, and the data obtained were processed separately and

averaged to obtain the T_m value. The data were fit to the following equations using SigmaPlot v.10.0.[309,310]

$$\Theta = \frac{(a_n + b_n T) + (a_d + b_d T)\exp(\frac{-\Delta G_U^0(T)}{RT})}{1 + \exp(\frac{-\Delta G_U^0(T)}{RT})}$$

$$\Delta G_U^0(T) = \Delta H^0(T_m)(1 - \frac{T}{T_m}) - \Delta C_p^0[(T_m - T) + T\ln(\frac{T}{T_m})]$$

$$\Delta C_p^0 = 0$$

Boc-Gly-Pro-Pro-ONP, 69. Boc-Gly-Pro-Pro-OH **56** (1.73 g, 6.43 mmol) was dissolved in CH_2Cl_2 (100 mL). EDC (1.88 g, 9.83 mmol), DMAP (62.2 mg, 0.509 mmol) were added. The solution was stirred at 0 °C for 20 min and *p*-nitrophenol (1.36 g, 9.74 mmol) was added. The mixture was stirred at rt for 12 h, and the solvent was removed in vacuo. The resulting oil was dissolved in EtOAc (150 mL), washed with 1 M HCl (2 × 100 mL), dried over $MgSO_4$, concentrated in vacuo, and purified by flash chromatography (1 % MeOH in $CHCl_3$). A brown oil (2.19 g, 98.3 %) was obtained. ^{1}H NMR: δ 8.24 (d, *J* = 9.2, 2H), 7.27 (d, *J* = 9.2, 2H), 5.34 (m, 1H), 4.72 (dd, *J* = 8.2, 4.4, 2H), 4.06 (dd, *J* = 17.2, 5.6, 1H), 3.89 (m, 2H), 3.68 (m, 2H), 3.50 (dt, *J* = 9.6, 6.7, 1H), 2.38 (m, 1H), 2.17 (m, 5H), 2.01 (m, 2H), 1.43 (s, 9H). ^{13}C NMR: δ 171.3, 169.8, 168.1, 156.2, 155.2, 145.6, 125.3, 122.3, 80.3, 59.4, 58.4, 47.1, 46.5, 43.0, 28.9, 28.35, 28.30, 25.2, 24.8. Impurity (*p*-nitrophenol): ^{1}H NMR: δ 8.09 (d, *J* = 9.0, 2H), 6.89 (d, *J* = 9.4, 2H). ^{13}C NMR: δ 163.2, 140.8, 126.2, 115.7.

Poly-(Gly-Pro-Pro)$_n$, 68. Boc-Gly-Pro-Pro-ONP **69** (124 mg, 0.252 mmol) was dissolved in CH_2Cl_2 (10 mL), and TFA (3 mL) was added. The solution was stirred at rt for 40 min, and the solvent was removed in vacuo. The resulting dark brown oil was dissolved in DMSO (0.2 mL), and Et_3N (36 μL, 0.26 mmol) was added. The solution was stirred at rt for 2 days to yield a dark brown glue-like liquid. Solvent and resulting *p*-nitrophenol were removed by

washing with Et_2O. The resulting brown solid was dissolved in water and dialyzed against water 6 times for 24 h. A light brown solid (19.5 mg) was obtained. MALDI-TOF analysis: highest observed MW: 4872 Da, repeating unit: 251 Da.

Chapter 4 Structure and stability of the α-helix

4.1 Structure of the α-Helix

The α-helix is one of the most important and abundant secondary structures in proteins and about 30 % of amino acid residues are in the α-helical structures.[311] It is believed that the α-helix is essential for the initiation of protein folding.[284] Unlike the β-sheet, the α-helix is mainly stabilized by internal hydrogen bonding and hydrophobic interactions along the main-chain,[312] and they can function as autonomous folding units in proteins.[313] The α-helix can occur as left-handed and right-handed. For L-amino acids, the right-handed structure is more stable by about 2.0 kcal/residue than the left-handed one due to the steric effects.[314,315] For the same reason, the left-handed structure is more stable for D-amino acids.

Three conformational and geometric parameters and two Ramachandran angles are used to characterize the α-helix: *p* - pitch of the helix (Å), *n* - the number of residues per turn, *r* - the radius of the helix (Å) and conformational angles Φ and Ψ. The ideal α-helix has 3.6 residues per turn and a rise of 1.5 Å per residue (5.4 Å per turn) with a radius of 2.3 Å (Figure 4.1).[312,316] Therefore, side-chains spaced (*i*, *i* + 3), (*i*, *i* + 4) and (*i*, *i* + 7) are close in space, which are the possible side-chain-side-chain interaction pairs. Side-chains spaced (*i*, *i* + 2), (*i*, *i* + 5) and (*i*, *i* + 6) are on opposite sides of the helix, and no side-chain-side-chain interaction will occur.[317] There is only one backbone hydrogen bond for each residue from the C=O of the *i* residue to the N-H of the *i* + 4 residue toward the *C*-terminus.[154] The dihedral angles Φ and Ψ of different residues do not have the same values, but a range with

mean values of –62° and –41° respectively. It is believed that the peptide sequence and hydrogen bonding to water are responsible for the variation.[311]

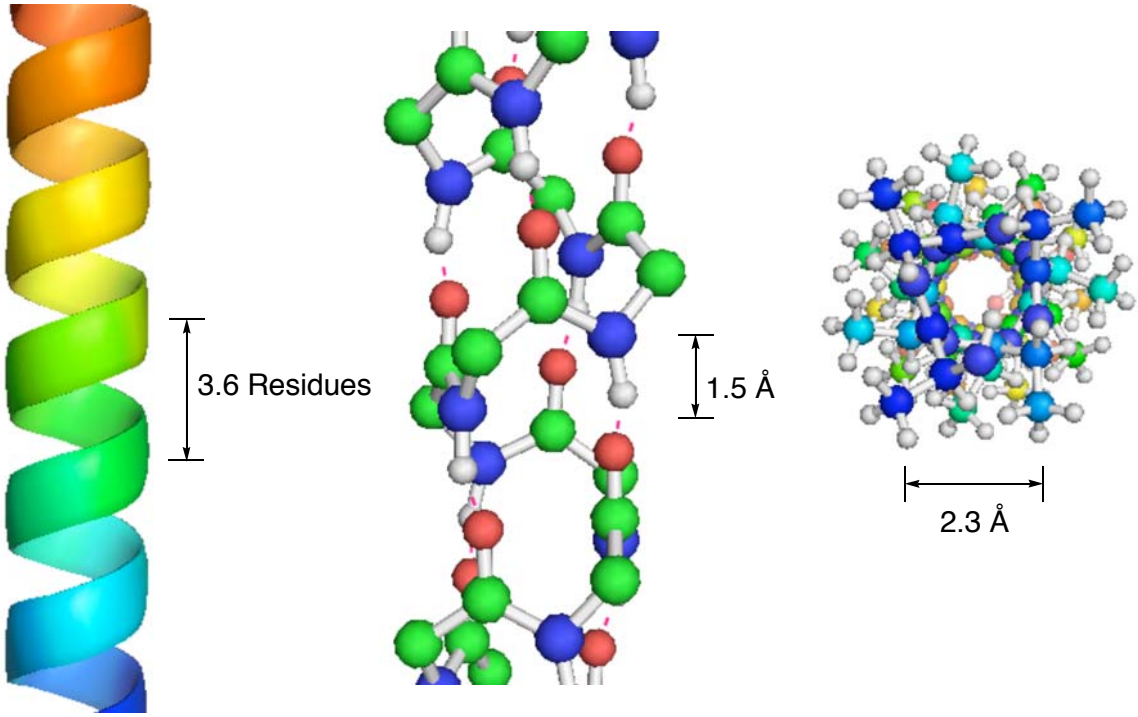


Figure 4.1 Idealized diagram of the main chain in an α-helix. Model was built with MacPyMOL 1.0.

All the side-chains point outside of the α-helix and are exposed to the solvent.[154] The surface of the α-helix can be either hydrophilic or hydrophobic, depending on the property of the side-chains.[318] In globular proteins, an α-helix can exhibit two "faces" – an α-helix containing predominantly hydrophobic amino acids orients toward the interior of the protein in the hydrophobic core, and an α-helix containing predominantly polar amino acids orients toward the hydrophilic surface of the protein. Amphiphilicity, the property of hydrophobic interactions, is also responsible for the helix bending.[312,319]

The length of the α-helix varies in different proteins. Most of the α-helixes are short, with 5-14 residues. [320] In all-β proteins (proteins exclusively composed of β-sheets), no

helixes are longer than 11 residues and 75 % of the helixes are shorter than 2 turns; while in all-α proteins (proteins composed of the α-helixes), only 12 % of the helixes have 2 turns or less and 20 % of the helixes are longer than 17 residues.[311]

4.2 Amino acid propensity

It was noticed a long time ago that different amino acids showed different propensities for natural α-helixes.[321,322] Alanine and leucine have very high propensities, while glycine has a low propensity to be in the middle of an α-helix.[320] Proline does not have a free N-H for hydrogen bonding, and a Pro residue in the middle of a sequence will break the backbone hydrogen bonding pattern of an α-helix. So it is highly disfavored in the middle of the α-helix. Statistical studies showed that proline residues are often observed at the end of helixes and regarded as helix breakers.[323-325] The Pro residue can be in the middle of an α-helix in some proteins, and still allow continuation of the helix.[311] But a "kinked" structure with a curvature from 40 to 100 Å is observed.[311]

A host-guest system was used to determine the stabilizing energy of a single amino acid residue for the overall α-helical structure.[326,327] O'Neil and Degrado used a host-guest system in α-helical two-stranded coiled-coils to study the helix-forming tendencies of natural amino acids.[123] They determined that the helical preferences of the natural amino acids are:[123]

A > R > K > L > M > W > F > S > Q > E > C > I > Y > D > V > T > N > H > G > P

Part of the order was confirmed by NMR and CD studies of the helix formed by the host-guest α-helical peptides with the sequence of $YSE_4X_3E_4K_4$.[328] Myers *et al.* reported that the helix propensities are similar in short peptides as in proteins, with the exception of some

polar amino acids.[329,330] Pace and Scholtz averaged and normalized the propensities from different proteins and peptides.[331]

4.3 Models for the α-helix

Many models have been built to describe the structure of the α-helix. Two mathematical models are used most often: Lifson-Roig and Zimm-Bragg theory.

4.3.1 Lifson-Roig Theory

The Lifson-Roig theory was proposed in 1961.[332] In this theory, only the residues with α-carbons on both sides of a peptide unit (CONH group) can be counted. For example, if a peptide has the structure of NH_3^+-ABCDE-COO^-, the number of residues of this peptide is three, and only B, C and D are regarded as residues. Thus, Lifson-Riog theory predicts the modification of the first amino acid will have no effect on the whole helix properties, which has been experimentally disproven.[333]

In Lifson-Roig theory, all residues in a polypeptide chain only have two states: helix or coil. Residues with α-helical conformational dihedral angles Φ and Ψ are defined as the helix state (h-state) and all the others are defined as the coil state (c-state). Each residue has a statistical weighting determined by both its own state and the states of the adjacent residues on both sides (a triplet). The weightings are defined as: u = coil state, v = helix state adjacent to one or two coil states, w = helix state with two helix states on both sides. If setting u-weightings to 1, the Lifson-Roig definition is shown in a matrix (m), and the partition function (Z) for a polypeptide with N residues can be generated from this matrix.

$$Z = \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix} m^N \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Since capping effects are not included in the Lifson-Roig theory, some modifications were made to incorporate the new findings.[334] In the modified Lifson-Roig theory, the number of residues of peptide NH_3^+-ABCDE-COO^- were counted as five. The *N*-capping residue is in the coil state with a weighting of *n* and the adjacent residue is in the helix state, while the *C*-capping residue is also in the coil state with a weighting of *c*. The modified definition is shown in a new matrix (Figure 4.2).[334]

$m =$

	$\bar{h}h$	$\bar{h}c$	$\bar{c}h$	$\bar{c}c$
$h\bar{h}$	w	v	0	0
$h\bar{c}$	0	0	1	1
$c\bar{h}$	v	v	0	0
$c\bar{c}$	0	0	1	1

$l =$

	$\bar{h}h$	$\bar{h}c$	$\bar{c}h$	$\bar{c}c$
$h\bar{h}$	w	v	0	0
$h\bar{c}$	0	0	$\sqrt{nc}$	c
$c\bar{h}$	v	v	0	0
$c\bar{c}$	0	0	n	1

Figure 4.2 Matrix for Lifson-Roig definition (Left) and modified Lifson-Roig definition (Right).[334]

Stapley *et al.* introduced *p* (*i*, *i* + 4 interaction parameter) and *q* (*i*, *i* + 3 interaction parameter) into the modified Lifson-Roig theory to include the side-chain-side-chain interactions.[335] These modifications included some proven stabilizing factors for the α-helix, and made the Lifson-Roig model more close to the real α-helical peptides. Of course, these modifications also made mathematical calculations more complex.

4.3.2 Zimm-Bragg Theory

The Zimm-Bragg theory was proposed by Zimm and Bragg in 1959.[336] Unlike the Lifson-Roig theory, the units in the Zimm-Bragg theory are peptide groups (CONH), and

they are classified by whether the N-H groups are in hydrogen bonds. If the N-H group participates in hydrogen bonding, the unit is given a code of 1; if not in a hydrogen bond, the code is 0. The difference between the Zimm-Bragg theory and the Lifson-Roig theory is shown in Figure 4.3.

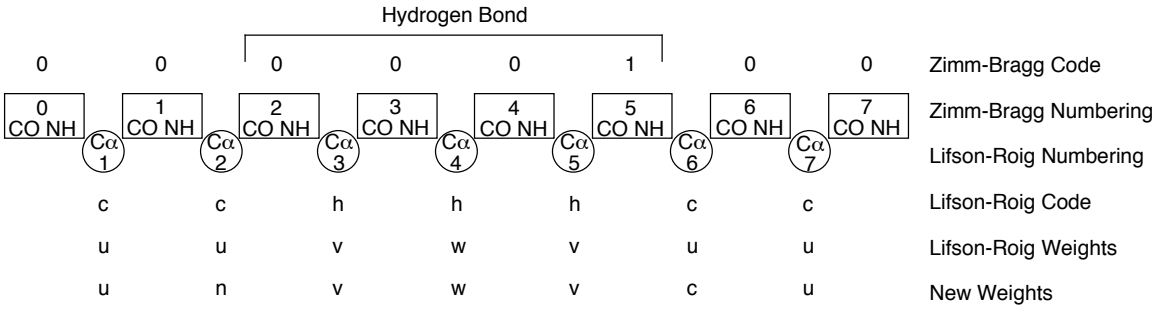


Figure 4.3 Codes, numberings and statistical weights for the Zimm-Bragg, Lifson-Roig and modified Lifson-Roig models in a partially helical peptide.[334] (Reprinted with permission from Doig, A. J. *et al.*; *Biochemistry* **1994**, Copyright 1994, American Chemical Society.)

Since the Zimm-Bragg theory cannot treat the effect of a single residue substitution, it is not convenient when dealing with heteropolypeptides, such as host-guest studies.

4.4 Stability of the α-helix

Backbone intra-chain hydrogen bonding between the C=O of the *i* residue and the N-H of the *i* + 4 residue was initially believed to be the major force stabilizing the α-helical structure.[154] The stabilizing energy of each backbone hydrogen bond is still under debate. Pauling *et al.* estimated that each N-H···O=C hydrogen bond contributed about 8 kcal/mol to the stability of the α-helix.[154] Schellman estimated that an intrapeptide hydrogen bond was about 1.5 kcal/mol favored over a peptide-H_2O hydrogen bond enthalpically.[337] Calorimetry studies showed that the enthalpy of helix formation was favored by about 1 kcal/mol per hydrogen bond for polyalanine[338-340] and some helixes with other amino acid residues.[341] Myers and Pace reported that the estimated stabilization energy became 2.2 kcal/mol per

hydrogen bond if corrected for differences in side-chain hydrophobicity and conformational entropy.[342] But a hydrogen bond in the interior of a protein contributed more than 1-2 kcal/mol to the stability of the proteins. Both theoretical and experimental studies showed that the enthalpy cost of breaking an internal hydrogen bond was at least 5-6 kcal/mol.[343-347] The enthalpy of an intrapeptide hydrogen bond buried in the hydrophobic protein interior could be up to 12 kcal/mol per hydrogen bond.[348]

Backbone hydrogen bonding is not the only factor affecting the stability of the α-helix. Other backbone interactions, such as hydrophobic packing, dipole-dipole interaction, and *N*- and *C*-terminal capping effects are also significant. In addition, side-chain-side-chain interactions, such as steric forces through the conformation of side-chains, electrostatic, inductive and dispersive forces, hydrophobic interactions, salt bridges, hydrogen bonding between side-chains of residues, and interactions of charged residues with the macrodipole of the helix are responsible for the stability of the α-helix.[312,320]

4.4.1 The Helix Dipole and Capping Effects

The α-helix dipole plays an important role in modulating the properties of proteins. Many folding patterns in proteins can be explained by the helix dipole.[349] The α-helix dipole originates from the dipole moment of each peptide unit due to the partial double bond character of the amide bond.

The direction of this single unit dipole moment is parallel to the C=O and N-H bonds with a value of 3.5 D.[294] The dipole of a helix of n residues is approximately n × 3.5 D. The *N*-terminus of a helix is the positive end, and the *C*-terminus is the negative end.[349] The helix dipole can be quite well estimated by placing a half positively charged unit near the *N*-terminus and a half negatively charged unit near the *C*-terminus. Because of the repeating

structure of helixes, all dipole charges are cancelled except for the *N*-terminal and *C*-terminal charges. Negatively charged residues have a tendency to appear near *N*-termini, while positively charged residues tend to appear near *C*-termini in order to counteract the helix dipole.[349] This phenomenon is called the helix capping effect.

The conventional nomenclature for helixes and their flanking residues are given as follows:[350]

$$\cdots\text{– N'' – N' – Ncap – N1 – N2 – N3 – } \cdots \text{ – C3 – C2 – C1 – Ccap – C' – C'' –}\cdots$$

The Ncap (or Ccap) is defined as the interface residue "half-in and half-out" of the helix.[351] C1, C2 and C3 are helical residues based on Pauling's model, and the α-helix has $i \rightarrow i + 4$ hydrogen bonds between every N-H and C=O group along the main chain. Thus, the first four N-H groups and the last four C=O groups lack this hydrogen bonding because of the absence of partners. New alternative hydrogen bond patterns are found in the initial and final turns of the helix and are called helix caps.[352] The amino acid residue at the *N*-cap or *C*-cap position is the first (or last) residue that does not have the characteristic dihedral angles of the α-helix. For the sp^3 hybridized atoms, the preference for the dihedral angles are +60° (*gauche*$^-$), –60° (*gauche*$^+$) and 180° (*trans*).[353-355] When Thr and Ser residues are at the *N*-cap position, the side-chains adopt the *gauche*$^-$ conformation to form a hydrogen bond with the N3 N-H with dihedral angle $\Psi = 164 \pm 8°$. When Asp or Asn is the *N*-cap, the side-chain adopts the *gauche*$^-$ conformation to form a hydrogen bond with the N3 N-H with $\Psi = 172 \pm 10°$, or the side-chain adopts the *trans* conformation to form a hydrogen bond with both the N2 and N3 N-H with dihedral angle $\Psi = 107 \pm 19°$. For the other residues at the *N*-cap position, the side-chain favors the *gauche*$^+$ conformation, and dihedral angle Ψ is unrestricted.[356]

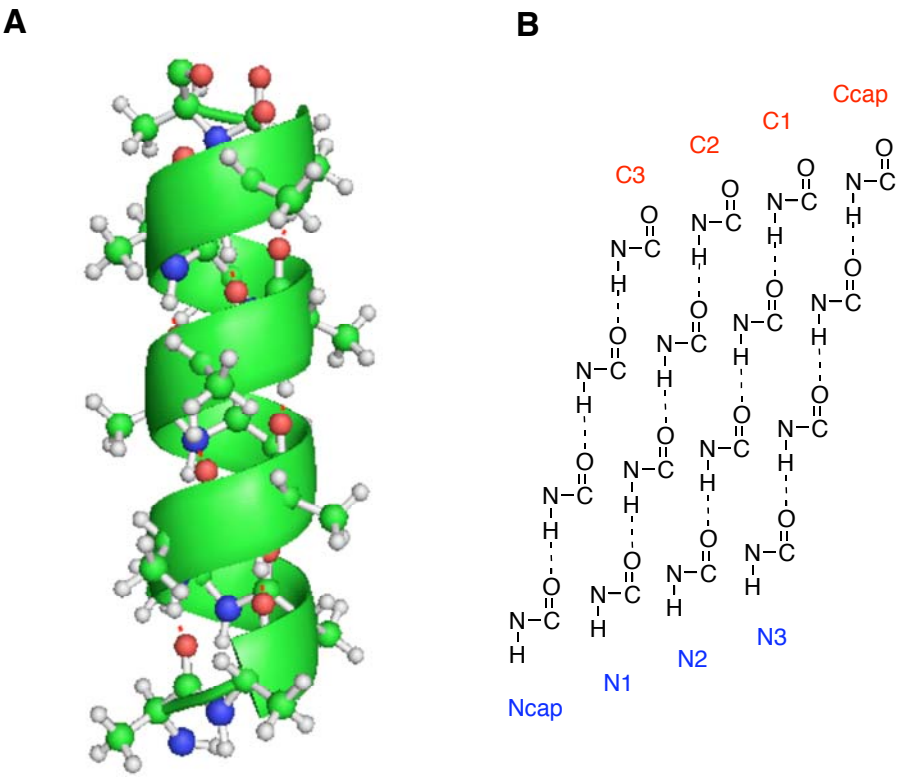


Figure 4.4 Two views of the α-helix. A: Backbone of the α-helix; B: The first four N-H groups and the last four C=O groups cannot make intrahelical hydrogen bonds.[350] (Reprinted from Aurora *et al.*, *Protein Sci.* **1998**, Copyright 1998, with permission from Prof. George D. Rose, Courtesy of Protein Science.)

For traditional intrahelical interactions, capping interactions can be put into three categories: short-range, mid-range, and long-range.[350] If the partners of the hydrogen bonds and/or hydrophobic interactions are within three or four amino acid residues from the *N*-cap/*C*-cap, it is called a short-range interaction. If the partners of the hydrogen bonds and/or hydrophobic interactions are more than seven residues from the *N*-cap/*C*-cap, it is a long-range interaction. All the others 3 to 7 residues apart are mid-range interactions. Most of the capping interactions are short-range and mid-range interactions.[350]

The helix capping theory was first proposed by Presta and Rose in 1988. They believed that the *C*-cap and *N*-cap have different capping patterns, although hydrogen bonds are still

the major force for capping effects.[352] Later studies showed that both hydrogen bonds and hydrophobic interactions are involved in helix capping.

Hydrogen bonds are the major force in helix capping, and polar side-chains are usually involved in hydrogen bonding. Negatively charged residues, such as Asp and Glu are favored at the *N*-terminus; while positively charged residues, such as Lys and His are favored at the *C*-terminus because of the helix dipole.[357] The C=O groups at the *C*-terminus are mainly satisfied by the backbone N-H groups from the turn following the helix while the N-H groups at the *N*-terminus are mainly satisfied by hydrogen bonding acceptors in side-chains.[352] The residues with longer side-chains, such as Glu, Gln, Arg, Lys and His, form hydrogen bonds with residues distal ($> \pm 4$) in the sequence while residues with shorter side-chains, such as Asp, Asn, Ser, and Thr, form hydrogen bonds with residues close ($< \pm 4$) in the sequence.[358]

There are two possibilities for intra-helical helix capping: side-chain to backbone capping, and backbone to backbone capping. Backbone to backbone capping is preferred more at the *C*-terminus and side-chain to backbone capping is preferred more at the *N*-terminus. About 66 % of the helixes have at least one short- or mid-range backbone to backbone capping interaction at the *C*-terminus, while 31 % of the helixes contain at least one short- or mid-range side-chain to backbone capping interaction at the *N*-terminus. The residue chirality is believed to be the major factor in this preference. For natural L-amino acids, the Cα → Cβ vector is oriented toward the *N*-terminus in the helix, and the helix C → O vector points to the *C*-terminus. To make a backbone to backbone hydrogen bond at the *N*-terminus, a "U-turn" is required which would greatly increase the energy and is unlikely to happen. Since side-chain donors rarely follow the helix immediately, side-chain to backbone capping is rare at the *C*-terminus.[350]

Although the hydrophobic interaction was recognized later than hydrogen bonds,[359] it is very common in the α-helixes that at least one residue at either the *N*- or the *C*-terminus participates in hydrophobic interactions. Hydrophobic capping is caused by two sequentially close hydrophobic residues. Usually, one residue is at the first or last turn and the other is just outside of the helix. Hydrophobic capping does not show special preference between the *N*- and *C*-termini. About 81 % of the helixes contain at least one short- or mid-range hydrophobic interaction at the *N*-terminus and the percentage is 87 % at the *C*-terminus.[350] But specific positional preferences were found as follows:

N-terminus: N3 ~ N4 > N2 > Ncap ~ N1

C-terminus: C3 > C2 > Ccap > C1 > C4

Capping effects are more complicated than described above, and classical intrahelical hydrogen bonds are not always effective, especially when the initial or final four amino acid residues are polar. The solvent will shield the polar groups and inhibit intermolecular interactions. Thus, special conformational arrangements at the termini are required to provide the intrahelical hydrogen bond partners in this case. Richardson and Richardson found that some amino acid residues are especially preferred at the ends of the helixes. Gly represents about 34 % at the *C*-cap position and Asn is preferred at the *N*-cap position.[351] Further studies show that Met has a high occurrence as the first residue at the *N*-terminus of helixes while Lys is very common as the last residue at the *C*-terminus.[360]

Ser, Lys and Leu have a higher preference at the *C*-terminus (C1-C3),[361] while Asn, Asp, Ser, Thr and Cys are more preferred at the *N*-cap.[362] Forood *et al.* found a trend in capping abilities:[363]

At the *N*-terminus: Asn > Asp > Ser > Glu > Gln > Ala

At the *C*-terminus: Arg > Lys > Ala

The energy difference between Asn and Gln at the *N*-terminus is about 2 kcal/mol.[334]

4.4.2 Side-Chain-Side-Chain Interactions

Side-chain-side-chain interactions were first reported by Marqusee and Baldwin in 1987 by studying the effects of Gln/Lys ion pairs on the helical stability of a 17-residue monomeric peptide containing 11 Ala residues.[364] After that, more and more side-chain-side-chain interaction patterns have been found, and the effect of side-chains on helical stability was studied thoroughly in the past 20 years. In 1995, Stapley *et al.* made modifications of the Lifson-Roig theory and made it applicable for side-chain- side-chain effects.[335]

Side-chain-side-chain interactions refer to non-covalent interactions. There are two categories: classical and non-classical side-chain-side-chain interactions. Classical side-chain-side-chain interactions include ion pairs (salt bridges), hydrogen bonds, and hydrophobic interactions. Non-classical side-chain-side-chain interactions include non-classical cation-π interactions and C-H···O hydrogen bonds.[365,366]

Side-chain-side-chain interactions are very complicated, and it is very difficult to confirm what parameters really affect the helix stability. For example, the interactions between Arg or Lys and aromatic residues were usually attributed to cation-π interactions. But Andrew *et al.* showed that sometimes the helix-stabilizing effects were caused by hydrophobic interactions between the CH_2 groups of Lys or Arg and the aromatic rings, and these will give nearly the same stabilizing energy (–0.10 to –0.18 kcal/mol) when the spacing is (i, i + 4) despite the difference in polarity between Lys or Arg and the aromatic residues.[367]

Side-chain-side-chain interactions also participate in helix capping, but the strengths are not identical at the termini as in the middle of the helix. For example, side-chain ion-pairs

were observed in helix capping. In some positions, Lys and Arg functioned as hydrophobic residues due to their long, flexible side-chains, to avoid the charged terminal groups.[350]

4.4.3 Ion Pairs

An ion pair is a pair of oppositely charged residues, such as Asp or Glu with Arg, Lys or His. If the centroids of the side-chain charged-groups are close (within 4.0 Å), it is also called a salt bridge.[357] Hydrogen-bonded ion pairs between side-chains are very common in proteins, and electrostatic ion pairs have an important influence on the stability of the proteins both inter- and intra-molecularly. For example, the Asp70-His31 ion pair contributes 3-5 kcal/mol to the folding of T4 lysozyme.[368] In the α-helixes, charged residues at the *C*- and *N*-termini are important in helix capping effects. In the middle of the α-helixes, side-chain ion pairs tend to appear at adjacent positions (i, $i \pm 1$) or on the same face (i, $i \pm 3$ or $i \pm 4$). Studies show that this type of ion pair also stabilizes helical structures.[368-370]

Several factors affect the strength of the ion pairs: environment, the nature of the residues, spacing and orientation. Most ion-pairs are pH dependent. Because of the pK_a of the functional groups of the side-chains, different ion pairs have different pH ranges. For example, the charged ion pair of Gln-Asp is stronger at pH 7 than the ion dipole interaction at pH 2,[371] while the Glu-His ion pair prefers pH 5.5.[372]

Lyu *et al.* studied the side-chain contributions to the stability of the α-helix and found that the side-chain effects are not identical when the residues are different.[327] It is plausible that the strengths of ion pairs are different when the ion pair patterns are different. The strengths of some ion pairs have been determined. For example, the strength of the (i, $i + 4$) Glu-Lys ion pair was –0.5 kcal/mol.[373] The general trend in helicity is given as follows:[372]

Glu-Lys > Asp-Lys > Glu-His

The ion pairs may occur at (*i*, *i* + 1, 2, 3 or 4), but only (*i*, *i* + 3) and (*i*, *i* + 4) ion pairs are reported to have helix-stabilizing ability.[357] That is because the α-helix has 3.6 residues per turns. The ion pairs are spatially close when the side-chains are spaced (*i*, *i* + 3) or (*i*, *i* + 4). Ion pairs (*i*, *i* + 5) are also reported but no helix-stabilizing effect was observed.[372] Usually this type of ion pair was used as a reference in the peptide study. Ion pairs (*i*, *i* + 7) and (*i*, *i* + 8) are also possible at some specific positions, such as a Glu 2^{-}···Arg 10^{+} ion pair.[374] It was reported that a majority of long-range ion pairs will destabilize the helical structures.[375]

Table 4.1 Stabilizing energy of the side-chain ion pairs.[372,376-382]

i, *i* + 3		*i*, *i* + 4	
Ion pair	ΔΔG (kcal/mol)	Ion pair	ΔΔG (kcal/mol)
Asp-Lys	–0.12	Asp-Lys	–0.24
Lys-Asp	–0.4	Lys-Asp	–0.58
Asp-His	–0.63	Asp-His	–0.63
His-Asp	–0.53	His-Asp	–2.38
Asp-Arg	–0.8	Asp-Arg	n. a.
Glu-His	–0.23	Glu-His	–0.10
His-Glu	–0.45	His-Glu	–0.54
Glu-Lys	–0.38	Glu-Lys	–0.44
Lys-Glu	–0.38	Lys-Glu	–0.46
		pSer-Asp	–0.45
		pSer-Lys	–1.6

The orientations of the ion pairs are also responsible for the stability of the helix (Table 4.1). That is, if the acidic (A) and basic (B) residues in the sequence are not identical, the effects of the ion pairs are not identical either; even when the spacing is the same. By studying the stabilization effects of the ion pairs Asp-Arg, Glu-Arg, Glu-Lys and Asp-Lys, the order of the ion-pair strength:[383]

$$(i, i + 4)\ \mathrm{AB} > (i, i + 4)\ \mathrm{BA} \sim (i, i + 3)\ \mathrm{AB} > (i, i + 3)\ \mathrm{BA}$$

It is believed that intrinsic features of the sequence are responsible for this pattern, and this pattern is independent of the chemical types of the residues.

Some amino acid derivatives were also used in the sequence to increase the stability of the α-helix. Phosphoserine (pSer) has two negative charges, which can form ion-pairs with positively charged side-chains.[381] Errington and Doig reported that the pSer-Lys ion pair contributed about 1.6 kcal/mol to the stability of the α-helix when the spacing was (*i*, *i* + 4), which was believed to be the strongest side-chain-side-chain interaction that had been measured.[381]

The cooperative salt bridge triplet was reported to have a stabilizing effect in barnase with spatially close residues: Asp8-Arg110-Asp12.[384] The introduction of the Arg-Glu-Arg cooperative salt bridge to the helix surface of the GCN4 leucine zipper resulted a stabilizing energy of 1.7 kcal/mol.[385] Later, studies showed that the bridged triplet can stabilize the α-helical structure in the synthetic peptide when the spacing is (*i*, *i* + 3) or (*i*, *i* + 4).[366] The stabilizing energy of the Arg-Glu-Arg and Glu-Lys-Glu salt bridge triplets are given in Table 4.2.

Table 4.2 Stabilizing energy of the cooperative salt bridge triplet.[366,386]

ΔG (kcal/mol)	(*i*, *i*+3), (*i*, *i*+3)	(*i*, *i*+3), (*i*, *i*+4)	(*i*, *i*+4), (*i*, *i*+3)	(*i*, *i*+4), (*i*, *i*+4)
Arg-Glu-Arg	–1.0	–0.3	–0.1	–1.5
Glu-Lys-Glu				–1.02

4.4.4 Hydrogen Bonds

Side-chain-related hydrogen bonds (H-bonds) include side-chain-main-chain H-bonds,[387] side-chain-side-chain H-bonds, and side-chain-solvent (usually water) H-bonds.[388] Here, hydrogen bond refers to the classical non-covalent side-chain hydrogen bond. That is, N-H or

O-H groups will act as hydrogen donors, while O or N atoms will be the acceptors. Like the backbone hydrogen bonds, side-chain hydrogen bonds greatly stabilize the helical structures.

Hydrogen bond pairs may be neutral or charged. Since most proton donors and acceptors can be acidic or basic in different surroundings, the strength of a specific hydrogen bond is pH-dependent. The (i, $i + 3$) or (i, $i + 4$) His-Asp side-chain interaction shows different properties at different pH values. At pH 2, a strong side-chain-side-chain hydrogen bond is formed between neutral Asp and positively charged His. The hydrogen bonding pattern can be observed as His^{+}-Asp with both (i, $i + 3$) and (i, $i + 4$) spacing. At pH 5.5, His and Asp are both charged and a salt bridge forms. At pH 8.5, His is neutral and Asp is negatively charged. The charged hydrogen bond can be seen as His-Asp^{-} only at (i, $i + 3$) spacing.[378] Studies show that ion-pairs and charged-dipole hydrogen bonds cannot be completely separated because of the nearly identical composition.[383]

The spacing of the hydrogen bond pairs is usually (i, $i + 3$) and (i, $i + 4$), and the energy difference between these two positions is very small.[372] No significant hydrogen-bonding interaction was found for (i, $i + 1$) or (i, $i + 2$) spacing.[379] The (i, $i + 5$) spacing is usually used in reference peptides, and it is believed to have no hydrogen bond interaction. The order of helix stabilization ability for spacing is:[379]

$$(i, i + 4) > (i, i + 3) >> (i, i + 2) \sim (i, i + 1)$$

The bridged hydrogen bond is also reported to have helix-stabilizing ability. The Asp8-Ser77-Arg17 triplet in the λ repressor forms a salt bridge and a hydrogen bond, which stabilize the structure by 1.5 and 0.8 kcal/mol, respectively.[380] The Glu^{0}-Lys-Glu^{0} bridged hydrogen bond contributes 0.47 kcal/mol to the stability of the synthetic α-helix.[386]

The orientations of the hydrogen bonds show different effects in different pairs. The (*i*, *i* + 4) Gln/Asn pair forms a hydrogen bond with a free energy of –0.4 to –0.7 kcal/mol when the orientation is Gln-Asn, while the interaction energy is nearly zero for the reverse orientation.[389] The Asp/His pair only shows an interaction when the orientation is His-Asp.[378] But the helix-stabilizing ability for the Glu/Lys pair is independent of the side-chain orientation.[379] Water molecules function as hydrogen bonding bridges when the residue pairs are spaced (*i*, *i* + 3, 4 or 5) and stabilize helical strutures.[390]

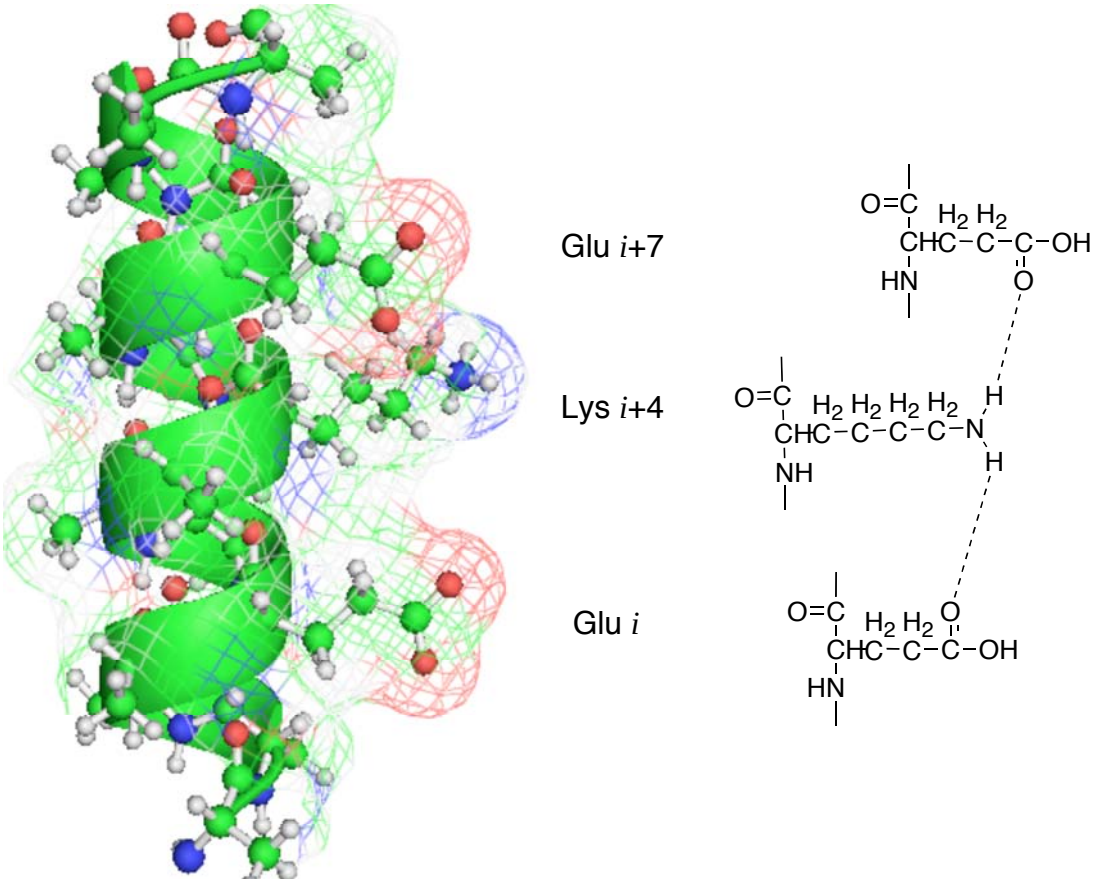


Figure 4.5 (*i*, *i* + 3 or *i* + 4, *i* + 7) hydrogen bond in the α-helix. Model was built with MacPyMOL 1.0.

4.4.5 Hydrophobic Interactions

Hydrophobic interactions are thought to be the major force for protein folding when non-polar surfaces are in contact.[391] Since forming a helix-stabilizing hydrophobic interaction will

lead to the loss of side-chain conformational entropy when two non-polar side-chains are fixed in specific conformations, the contribution of the hydrophobic interactions to the stability in an isolated helix was not found until 1994. Padmanabhan and Baldwin found hydrophobic interactions between Tyr and Leu or Val when the spacing was (i, i + 4).[392]

The mechanism of hydrophobic interactions is still under debate. It is believed that electrostatic fluctuations, changes in water structure, bridging microscopic bubbles, and a drying transition induced by hydrophobic surfaces can cause hydrophobic interactions.[393] It is also reported that removing water from some backbone hydrogen bonds may cause hydrophobic packing and greatly improve the stability of the protein.[394]

Hydrophobic interactions have more restraints than ion pairs because of the requirements of conformation. All possible hydrophobic interaction pairs are from a small group of amino acid residues, including Ala, Cys, His, Ile, Leu, Met, Phe, Trp, Tyr, and Val.[395] One study showed that any hydrophobic helix-stabilizing interaction between non-polar residues spaced (i, i + 4) strongly depends on the limited flexibility of at least one residue. If the side-chain conformations are very flexible and cannot be fixed to a single conformation, there would be a great loss of conformational entropy. Then this pair will have very little or no helix-stabilizing effect. This may explain why straight side-chain pairs are less important in helix-stabilization than a pair containing a straight side-chain and a β-branched side-chain or an aromatic side-chain, such as Tyr-Leu.[396]

Unlike ion pairs, hydrophobic interactions do not show significant differences when the hydrophobicities of the side-chains are different (Table 4.3). For example, the (i, i + 4) hydrophobic interactions between Leu and Val or Ile are similar even though the hydrophobicity of Ile is greater than Val (Ile-Leu: –0.35 kcal/mol; Val-Leu: –0.31kcal/mol

when the spacing is *i*, *i* + 4).[395,396] But if the side-chains contain aromatic rings, such as Phe, Tyr, Trp and His, the interactions are much greater[395] (Leu-Trp: –1.0 kcal/mol when the spacing is *i*, *i* + 4).[397]

Hydrophobic interactions are more position specific than other classical side-chain interactions. Hydrophobic pairs have great helix-stabilizing or destabilizing ability when the spacing is (*i*, *i* + 4). When the spacing is (*i*, *i* + 3), the interactions are much weaker (about 2.7 kJ/mol difference in Leu-Leu pair).[392,398] No hydrophobic effect is observed when the spacing is (*i*, *i* + 2) because of the structure of the α-helix.[395]

Table 4.3 Stabilizing energy of the side-chain hydrophobic interactions.

i, *i* + 3		*i*, *i* + 4	
Amino acid pair	ΔΔG (kcal/mol)	Amino acid pair	ΔΔG (kcal/mol)
Ile-Leu [a]	–0.26	Ile-Leu	–0.39
Leu-Ile [a]	–0.43	Leu-Ile	–0.47
Leu-Leu [a]	–0.35	Leu-Leu	–0.40
Leu-Tyr [b]	–0.44	Leu-Tyr	–0.65
Tyr-Leu [b]	–0.02	Tyr-Leu	–0.44
Leu-Val [a]	–0.21	Leu-Val	–0.19
Val-Leu [a]	–0.49	Val-Leu	–0.54
Phe-Leu [a]	–0.15	Phe-Leu	–0.10
Tyr-Val [b]	–0.13	Tyr-Val	–0.31
Val-Val [a]	–0.17	Val-Val	–0.28

a) Free energy derived from a protein dataset;[395] b) experimental data.[399]

Some side chain aromatic-aromatic interactions are also regarded as hydrophobic interactions when the side chains are uncharged, such as the (*i*, *i* + 4) Phe-His interaction. Charged His^+ is usually orthogonal to the Phe aromatic ring (edge-to-face).[400] When His is uncharged, it acts more as a π-system, and forms a π-π interaction in a parallel manner.[401,402] This aromatic-aromatic interaction has a weak stabilizing effect to the α-helix.[400]

Like other side-chain-side-chain interactions, aromatic-aromatic interactions are also sequence-dependent. For example, the Phe-Phe interaction contributes –0.27 kcal/mol to the

stability of a 18-residue α-helix in the middle of the sequence (Phe7-Phe11), while the interaction is much greater (–0.8 kcal/mol) at the *C*-terminus (Phe13-Phe17).[365]

4.4.6 Cation-π Interactions

Non-classical side-chain-side-chain interactions mainly include cation-π interactions, non-covalent C-H···O hydrogen bonds, and other unconventional hydrogen bonds such as C-H···π, and O-H···π hydrogen bonds.[366] Although the free energies of these interactions are relatively weak, they still have cumulative effects on the stability of helical structures and protein folding.

The cation-π interaction is the interaction between a charged or partially charged group perpendicular to the plane of an aromatic system. The cation-π interaction was first reported in the gas phase by Deakyne and Meot-Ner.[403,404] They showed the possibility of X-Hδ^+···π interactions in specific peptide conformations.[335] Perutz *et al.* studied the interactions of ammonium groups with aromatic rings in human hemoglobin with drugs.[405] Burley and Petsko found that positively charged or δ^+ amino groups of Lys, Arg, Asp, Glu and His showed great preference to locate within 6 Å of the ring centroids of Phe, Tyr and Trp side-chains in proteins.[406]

Most of the cation-π interactions in the α-helixes are amino-aromatic interactions. Studies show that the cation-π interactions are not as weak as expected. The cation-π interactions can be as strong as ion pairs,[407] and some cation-π interactions may be even stronger than salt bridges.[408] Even the simplest cation-π interaction between an ammonium cation and a phenyl group could contributes 0.4 kcal/mol, nearly as much as that of salt bridges, to the helix stability.[409]

Cation-π interactions can occur between an aromatic residue, such as Phe, Tyr or Trp, and a cationic amino acid, such as Lys, Arg or His.[409] The contributions to the stability of the α-helixes are different if the pairs are different. Of the aromatic residues, the order of stabilization is: Trp > Tyr > Phe.[410] The reason for this trend can be explained by the electron density of the six-membered ring. Trp has greater electron density in the benzene ring than Tyr and Phe, so the cation-π interaction will be greater.[411] Arg is preferred as the cationic amino acid when forming cation-π interactions.[412] This can be explained by the difference in the geometries of the interactions. Arg prefers to interact with an aromatic ring in a stacked geometry, in which van der Waals force may also contribute to the stability of the helical structure and reduce the desolvation energy for interacting with an aromatic ring. Lys has no van der Waals force to counteract the desolvation energy, so Lys is less preferred in cation-π interactions.[401]

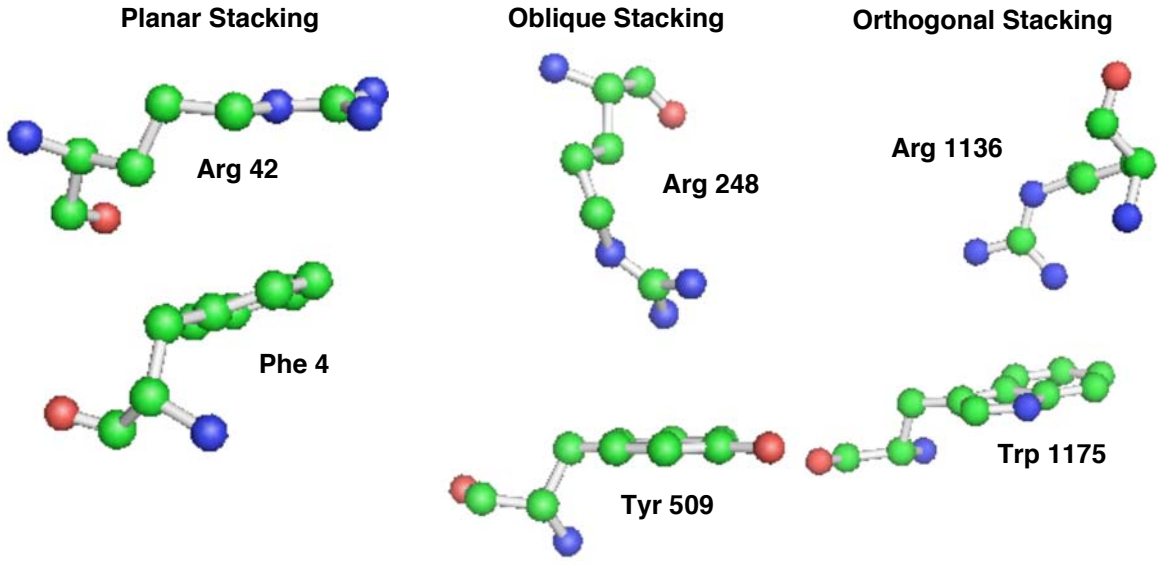


Figure 4.6 Cation-π interaction arrangements. Left: planar stacking (4PTI),[413] middle: oblique stacking (1YCS),[414] right: orthogonal stacking (1REK).[415] All models were generated from Protein Data Bank with MacPymol 1.0.

Three categories of interaction arrangements were assigned based on the interplanar angle (α): planar ($0° < \alpha < 30°$), oblique ($30° < \alpha < 60°$), and orthogonal ($60° < \alpha < 90°$) (Figure 4.6).[416] Arg prefers planar stacking, in which the guanidinium group is directly over the center of the aromatic ring, and the two planes are nearly parallel. The Arg-π interaction contributes more to the stability of helical structures, because this arrangement does not affect the side-chains ability to form hydrogen bonds elsewhere.[417] Charged His^+ prefers orthogonal stacking, although this interaction is also categorized into edge-to-face hydrogen bonding.[400]

The sequence order of the cation-π interactions is also very important to the stability of helical structures. The solvent-exposed Trp→Arg (i, $i + 4$) interaction contributes –0.4 kcal/mol to the helix stability, while the reverse order Arg→Trp (i, $i + 4$) interaction has little contribution to the helix stability.[418]

4.4.7 C-H···O Hydrogen Bonds and Other Non-Classical Hydrogen Bonds

C-H···O hydrogen bond and other non-classical hydrogen bonds were proposed a long time ago.[419] But since they are so weak, their effects were always neglected. In 1982, Taylor and Kennard provided strong crystallographic evidence for the existence of a C-H···O hydrogen bond.[420] In 1995, this type of hydrogen bond was found to stabilize the secondary structure of proteins (helixes and parallel or anti-parallel β-strands).[421]

Unlike classical hydrogen bonds, in which N-H or O-H groups act as proton donors, the C-H group is the donor in C-H···O hydrogen bond (Figure 4.7). Since the C-H group is not a good donor, the strength of the C-H···O hydrogen bond is considered to be much weaker than that of the ordinary hydrogen bonds. The free energy of O-H···O hydrogen bonds is greater

than 10 kcal/mol,[422] but only 0.6-1.5 kcal/mol for C-H···O hydrogen bonds.[423] But due to the abundance of C-H groups in proteins, the occurrence of this type of hydrogen bond is very high and the total contributions to the stability of proteins are still considerable.[422] Side-chain π-system acceptors also interact with C-H donors to form hydrogen bonds,[424] but the strength is even weaker. The free energy of a C-H···π hydrogen bond is 0.5-1.0 kcal/mol, the weakest hydrogen bond of all.[422]

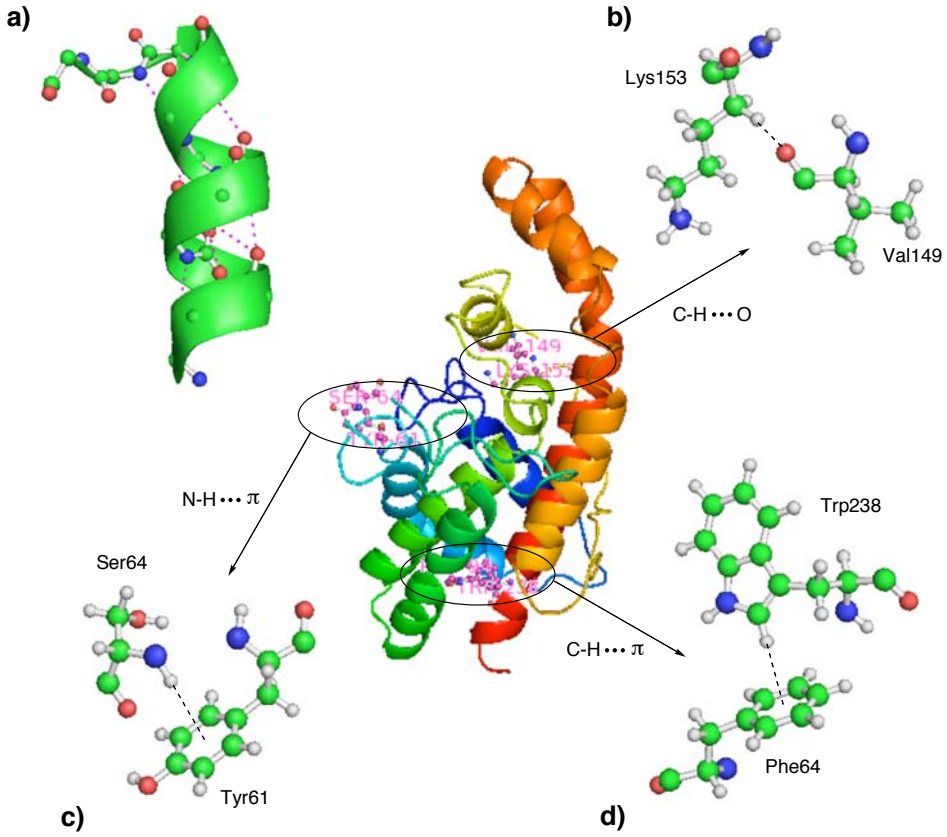


Figure 4.7 Structure of some classical and non-classical hydrogen bonds: (a) classical N-H···O hydrogen bond in the α-helixes (main-chain atoms of the helix 193–204); (b) non-classical C^{β}-H···O hydrogen bond between Lys153 and Val149; (c) non-classical N-H···π hydrogen bond between Ser64 backbone N-H and aromatic ring of Tyr61; (d) non-classical $C^{\delta 1}$-H···π hydrogen bond between Trp238 and aromatic ring of Phe64.[422] (Reprinted from Weiss *et al.*, *Trends Biochem. Sci.* **2001**, Copyright 2001, with permission from Elsevier.)[425]

When hard/soft-acid/base (HSAB) theory was applied to non-classical hydrogen bonds, N-H and O-H groups are regarded as hard acids while C-H groups are soft acids. Hard-acids

have greater affinity to hard-bases – N or O acceptors; while soft-acids prefer soft-bases. The π system is regarded as soft here. This may explain why π hydrogen bond acceptors have a greater preference to interact with C-H groups than O-H, N-H or S-H groups.[422]

Non-classical hydrogen bonds also stabilize the α-helical structure. Proline tends to form C-H···O hydrogen bonds to stabilize the helical structure because of the lack of N-H groups (Figure 4.8).[426] The Pro C^{δ} protons are usually involved to form two C^{δ}-H···O hydrogen bonds with the C=O groups in the preceding turn (i-3 and i-4 positions). It is also possible that the C^{δ} hydrogen atoms form just one C^{δ}-H···O with one C=O group (type 3 or 4) or H_2O molecules, Alternatively one C^{δ}-H donor can combine with two acceptors simultaneously to form a three-centered bond (type 5) and leave the second hydrogen atom free.[426]

The orientation of a C-H···O hydrogen bond is also responsible for the strength of the interaction. The C-H···O hydrogen bond between Glu and Phe contributed about –0.5 kcal/mol to the stability of a helix when the orientation is Glu→Phe (N→C) with the spacing (*i*, *i* + 4), while no stabilizing effect was observed when the orientation was Phe→Glu with the same spacing.[427]

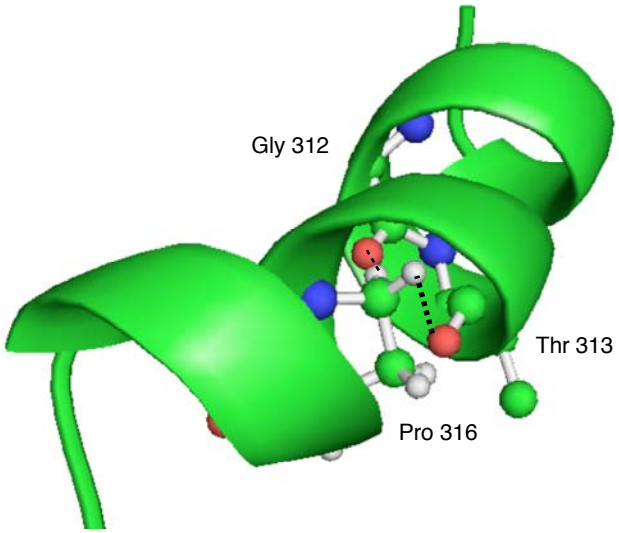


Figure 4.8 Structure of proline C-H···O hydrogen bond – type 4/3.[426] (Reprinted from Chakrabarti *et al.*, *J. Mol. Biol.* **1998**, Copyright 1998, with permission from Elsevier.)

If the O-acceptors are in different functional groups, the strengths of C-H···O hydrogen bonds will be different. The binding energy of C^{α}-H to a water molecule is in the range of –1.9 to –2.5 kcal/mol depending on the polarity of the amino acid,[428] while C-H···O=C hydrogen bonds only contribute –0.5 kcal/mol to the stability of the helical structure in the Glu→Phe pair.[427] In the α-helixes, C-H···O=P hydrogen bonds also stabilize helical structures in the absence of strong hydrogen bonding patterns.[429]

4.5 De Novo *design of the α-helix*

Proteins are the basis of life, and it is believed that the structure of a protein determines the function of this protein. How proteins fold and function has always been mysterious to scientists due to the involvement of various interactions. To study the subtle interactions in proteins, *de novo* design was adopted to evaluate protein folding and determine the desired interactions and functions.[430] Alanine-based peptides are used in most studies because of two factors: 1) the side-chain interactions are minimal thus the structural and thermodynamic changes are mainly caused by the peptide backbones;[431,432] 2) alanine can stabilize the helical structure intrinsically.[433,434] Studies show that four types of interactions are responsible for protein structure, stability, and folding: hydrophobic forces, ionic interactions, van der Waals forces, and hydrogen bonds.[366]

Proteins are mainly composed of 20 natural L-α-amino acid residues and these amino acid residues form stable secondary structures, such as α-helix, β-sheet and β-turn. The structure of the α-helix was first proposed by Pauling *et al.* in 1951,[154] which is regarded as a landmark in the study of proteins. As one of the most abundant and important secondary structures in nature, the α-helix has been important for biologists and chemists in the past

fifty years. After 1980, new techniques, such as NMR and CD, were introduced into the study of proteins that made it possible to reveal the roles of "weak interactions" in the stability and folding of the α-helixes. It is known that backbone hydrogen bonds are the major forces to stabilize the helical structure.[154] In addition to backbone hydrogen bonds, several factors are also responsible for the stability of the α-helix. These include helix nucleation, propagation constants, helix capping effects and side-chain-side-chain interactions.[366,435] These interactions have different contributions to the stability of the α-helix. It was reported that in an alanine based peptide with 15-20 residues, the free energy barrier for nucleation is about 4 kcal/mol, while the free energy for helix propagation is only 0.4 kcal/mol.[436] The details of peptide folding are revealed with mimic peptides by *de novo* design.

4.6 Summary

The α-helix is one of the most important and abundant secondary structures in proteins. The peptide chain consisting of only natural L-α-amino acids forms a right-handed α-helix with 3.6 residues per turn and a pitch of 1.5 Å per residue. All the side-chains point outside of the α-helix and are exposed to the solvent.

The backbone intra-helical hydrogen bonds between *i* and *i* + 4 residues, and various side-chain-side-chain interactions are responsible for the stability of the α-helix. These enable our ability to modify the sequence of a peptide to design a more stable α-helix. Some strong interactions, such as *N*- and *C*-terminal capping effects, salt bridges, and hydrogen bonding between side-chains of spatially close residues, are widely used in *de novo* design to evaluate protein folding and determine the desired interactions and functions.

Chapter 5 Design of the helix-turn-helix peptides

5.1 Introduction to the helix-turn-helix motif

A tri-helical DNA-binding domain was reported in phage λ transcription regulators, Cro and the cI repressor, and lacI, the lactose operon repressor in 1982.[437-441] The second and the third helixes of this tri-helical domain form a α-helix-turn-α-helix motif, which is critical for the interaction with DNA.[442] This domain was called a helix-turn-helix (HTH) domain, and the α-helix-turn-α-helix structure was called a helix-turn-helix motif.

The HTH motif is a tertiary structure, which can be found in many DNA-binding proteins from prokaryotes to eukaryotes, especially in the cro repressor protein (cro), catabolite gene activator protein (CAP), and λ repressor.[443] The HTH motif has at least 20 residues, including an α-helix, a turn of three or four amino acids, and a second α-helix. The two helixes pack together at an angle of 90° and form a "criss-cross" shape,[444] and the amino acid residues of the turn form a "pocket". The "core" of this turn is hydrophobic, and the hydrophobic interactions are the major force between these two helixes.[445]

The HTH-turn is quite different from the common β-turn and γ-turn. The β-turn has exactly 4 amino acid residues and there is a hydrogen bond between the C=O group of the first residue and the N-H group of the fourth residue.[444] The γ-turn has three amino acid residues and hydrogen-bonding between the C=O group of the first residue and the N-H group of the third residue to form a seven-membered ring.[446,447] The turn in the HTH motif is tighter, and has three or four amino acid residues and lacks the hydrogen bond.[445] The Ramachandran conformational angles Φ and Ψ are also different.[448]

Table 5.1 Dihedral angles Φ and Ψ of the amino acid residues in the β-turn, γ-turn and HTH-turn.[446,447,449-453]

	Φ_i	Ψ_i	Φ_{i+1}	Ψ_{i+1}	Φ_{i+2}	Ψ_{i+2}
HTH-turn Cro [a]	–87	+14	+60	+27	–105	–172
HTH-turn Oct-1 [b]	–109	+22	+52	–3	–98	+164
β-Turn, Type I			–60	–30	–90	0
β-Turn, Type II			–60	+120	+80	0
β-Turn, Type III			–60	–30	–60	–120
β-Turn, Type VIa			–60	120	–90	0
β-Turn, Type VIb			–120	120	–60	0
γ-Turn, inverse			–70 to –85	60 to 70		
γ-Turn, classic			70 to 85	–60 to –70		

a) From the X-ray crystal structure of phage 434 Cro/O_R1 complex;[454] b) from the X-ray crystal structure of an Oct-1 POU domain-octamer DNA complex;[455]

The HTH motif is not an independently stable domain.[456] It needs other secondary elements to form a stable, globular structure. In the tri-helical engrailed homeodomain, the first helix immediately preceding the HTH motif covers the hydrophobic face of the HTH motif to stabilize the HTH motif.[457] In the tetra-helical DNA-binding domains, such as λ and 434 repressors,[458,459] 434 Cro[460] and the POU-specific domain,[455] the first and fourth helixes immediately preceding and following the HTH motifs also have a stabilizing effect on the HTH motif.

The two α-helixes in the HTH motif are responsible for recognition and binding to DNA, but the two helixes are not functionally identical. Usually, the helix at the *C*-terminus is oriented in a position that is close to the major groove of DNA, and contributes more to the DNA recognition process.[461,462] This helix is called a recognition helix and it carries the main amino acid residues that bind to DNA and recognize a specific sequence of nucleotides through a series of hydrogen bonds and various Van der Waals interactions with the exposed bases. The other α-helix stabilizes the structure and assists the interactions between protein and DNA, but does not significantly contribute to the recognition.[439,443] In many different

HTH proteins the recognition helix has various orientations in the major grooves of DNA. The orientation of the helix in eukaryotes is much less variable.[312]

5.2 De Novo *Design of HTH Peptides*

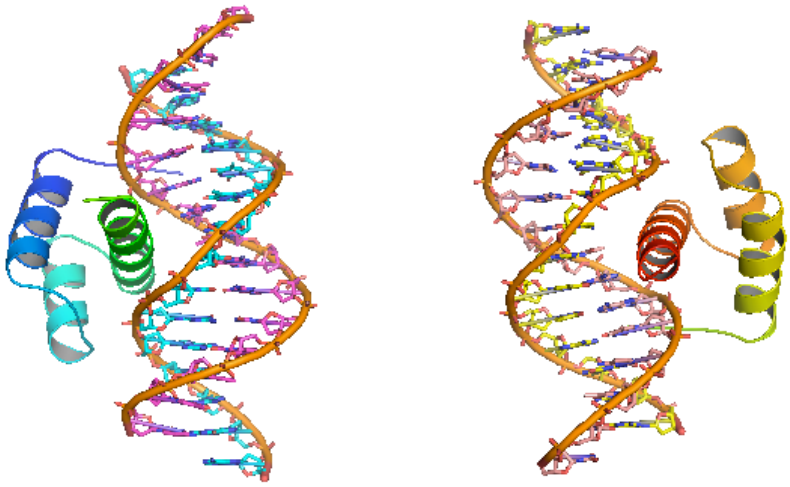


Figure 5.1 Antp-DNA binding complex. The crystal structure was obtained from the Protein Data Bank (9ANT),[463] and processed with MacPyMOL v1.0.

Antennapedia (Antp) is a protein that controls the development of *Drosophila*.[464,465] The Antennapedia homeodomain consists of 60 amino acid residues, and has a very high α-helical content.[466] The structure of the Antennapedia homeodomain was determined by NMR in aqueous solution[467,468] and by X-ray crystallography (Figure 5.1).[463] The NMR studies showed that the Antennapedia homeodomain consists of three well-defined α-helixes, which are residues 10-21, 28-38 and 42-52, and a relatively flexible short helix residues 53-59. Helixes II and III are connected by a tripeptide that forms a turn, and together they constitute a helix-turn-helix motif (residues 30-50).[468] The X-ray crystal structure mostly agreed with the NMR results, but only three α-helixes were observed: helix I from residue 10 to 22, helix II from residue 28 to 38, helix III from residue 42 to 58.[463] The recognition helix penetrates into the DNA major groove, and forms direct, or water-mediated, intermolecular hydrogen

bonds for specific protein-DNA contacts. Helix III, and perhaps helix IV as well, act as the recognition helixes for the Antennapedia homeodomain, and makes contact with the surface of the major groove of the DNA.[463,467,469]

The native Antp HTH peptide contains two α-helixes (helix II from T27 to H36; helix III from T41 to K55) and one short hydrophobic turn (A37-L38-S39-L40). Stepwise modifications of the Antennapedia HTH peptide was designed to improve the helicity (F. A. Etzkorn, unpublished results), and new HTH peptides were synthesized (University of Virginia, Biomolecular Research Facility). The DNA-binding residues, identified in the co-crystal structure of Antp/DNA, and the hydrophobic core residues were retained as much as possible (F. A. Etzkorn, unpublished results). The sequences of the native and modified peptides are shown in Figure 5.2.

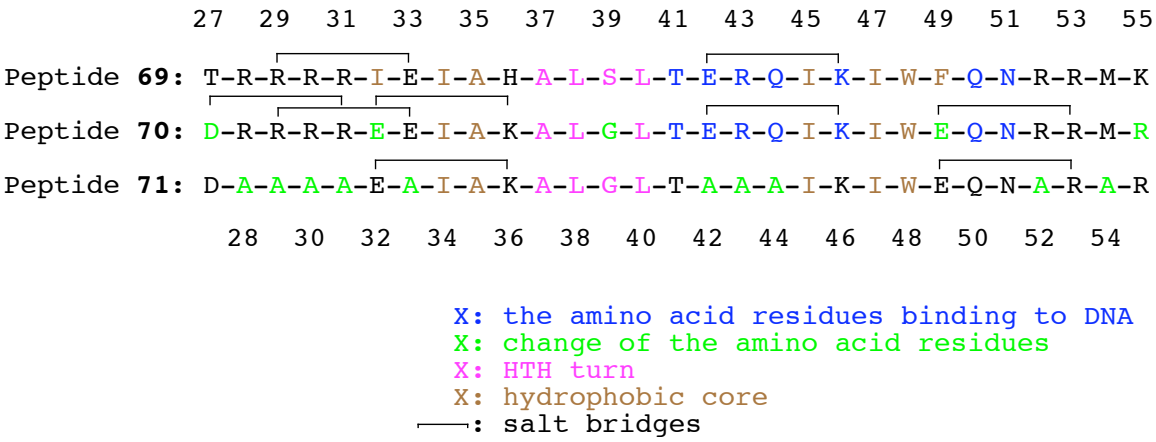


Figure 5.2 Stepwise changes to Antennapedia HTH peptide (27–55). (F. A. Etzkorn, unpublished results)

Compared to the native Antennapedia peptide (27–55), the Cys residue at residue 39 was replaced with a Ser residue in peptide **69** to prevent oxidative dimerization.[470] Peptides **70** and **71** have a negatively charged Asp (D) residue at the partially positively charged *N*-terminus, and a positively charged Arg (R) residue at the partially negatively charged *C*-terminus to cap the helix dipole. Additionally, in peptide **70**, more ion pairs are introduced to

stabilize the helix. In addition to the original Arg29-Glu33 and Glu42-Lys46 ion pairs, peptide **70** contains three more ion pairs spaced (*i*, *i* + 4) to stabilize the two helixes: Asp27-Arg31, Glu32-Lys36, and Glu49-Arg53. The two helixes in peptide **70** are expected to have greater stability with these oppositely charged (*i*, *i* + 4) ion pairs, even though this meant removing a few DNA-binding (E42, R43, Q44) and hydrophobic core residues (I32, F49), but these are thought to be less important for the DNA recognition process.

Different amino acid residues have different contributions to the helicity of peptides. Circular dichroism studies indicated that the extent of the helicity of short peptides has the order:[327]

Ala > Leu > Met > Gln > Ile > Val > Ser > Thr > Asn > Gly

Short alanine-based peptides tend to be very helical.[434] In peptide **71**, more Ala residues were included to increase the helicity.

5.3 Purification of HTH Peptides

Peptides can be purified efficiently on a reverse phase column with a binary solvent system of H_2O and CH_3CN in the presence of 0.1 % TFA. Two solvent systems were tried to purify the HTH peptides. The first solvent system was a pH 7.0 ammonium acetate buffer. Ammonium acetate does not dissolve well in CH_3CN, so water was added to increase the solubility. This HPLC solvent system consisted of two solvents: solvent A was 20 % CH_3CN in H_2O with 10 mM ammonium acetate buffer (pH 7.21); solvent B was 20 % H_2O in CH_3CN with 10 mM ammonium acetate buffer (pH 6.88). A fast gradient from 10 % to 95 % solvent B over 25 min was tried on a VYDAC protein C4 reverse phase column, but the separation was poor. Another solvent system, in which solvent A was 0.1 % TFA in H_2O, and solvent B was 0.1 % TFA in CH_3CN, was tried with the same gradient. The separation

was much better, so the binary solvent system of H_2O and CH_3CN with 0.1 % TFA was used as the mobile phase to purify the peptides.

Three reverse phase columns were tried to obtain a better separation: Varian Polaris C18 reverse phase column, VYDAC protein C8 reverse phase column, and VYDAC protein C4 reverse phase column. On the Varian Polaris C18 reverse phase column, the desired peak did not elute during the gradient, and an additional 5 min at 95 % solvent B was required to elute the desired peak. VYDAC protein C8 and protein C4 reverse phase columns both gave good resolution. The resolution was a little better on the C8 column with a longer retention time. Since the C8 column has no corresponding preparative column, the VYDAC protein C4 reverse phase semi-preparative column was used to purify the HTH peptides.

Due to the difference in polarity and impurity in different peptides, no one method was good for all peptides. Different gradients were used to achieve the best resolution, and the methods are given in the Experimental section.

5.4 CD analysis of HTH peptides

Circular dichroism (CD) can be used to measure the helicity of synthesized peptides. In folded proteins, different secondary structures have different characteristic CD spectra. The random coil has a maximum at 212 nm corresponding to a $\pi \rightarrow \pi^*$ transition and a minimum at 195 nm assigned to an $n \rightarrow \pi^*$ transition. The β-sheet has a minimum at 218 nm related to a $\pi \rightarrow \pi^*$ transition and a maximum at 196 nm corresponding to an $n \rightarrow \pi^*$ transition. The α-helix exciton coupling of the $\pi \rightarrow \pi^*$ transitions leads to $(\pi \rightarrow \pi^*)_{perpendicular}$ maximum at 192 nm and $(\pi \rightarrow \pi^*)_{parallel}$ minimum at 209 nm, it has

another minimum at 222 nm, which is red shifted, related to an n→π* transition.[471,472] The molar ellipticities are shown in Table 5.7.

Table 5.2 CD peaks of different secondary structures in the folded proteins.[471,472]

Structure	Maximum	Transition	Peak (nm)	Molar ellipticity (deg·cm^2·dmol^{-1})
α-Helix	Positive	π→π*	190-195	60,000 to 80,000
α-Helix	Negative	π→π*	208	–36,000 ± 3,000
α-Helix	Negative	n→π*	222	–36,000 ± 3,000
β-Sheet	Positive	π→π*	195-200	30,000 to 50,000
β-Sheet	Negative	n→π*	215-220	–10,000 to –20,000
Random Coil	Negative	π→π*	ca. 200	–20,000
Random Coil	Positive	n→π*	220	Small

The helical content of a peptide is proportional to its mean residue ellipticity measured at a wavelength of 222 nm. The fraction of helix (f_H) is defined as the ratio of the average number of backbone hydrogen bonds divided by the total number of backbone hydrogen bonds available in a peptide with *n* residues.[473] The measured value of f_H can be calculated by equation 1:[372]

$$f_H = \frac{[\theta]_{obs} - [\theta]_C}{[\theta]_H - [\theta]_C} \quad (1)$$

Where $[\theta]_{obs}$ is the observed mean residue ellipticity with units of deg·cm^2·dmol^{-1}; $[\theta]_C$ is the mean residue ellipticity for the random coil state, and $[\theta]_H$ is the mean residue ellipticity for a completely helical state. The values of $[\theta]_C$ and $[\theta]_H$ at 0 ºC can be described as:[372]

$$[\theta]_C = +640 \text{ deg·cm}^2\text{·dmol}^{-1}$$

$$[\theta]_H = -42500(1 - 3/n) \text{ deg·cm}^2\text{·dmol}^{-1}$$

Where *n* is the number of residues; 3 means three C=O groups that are not H-bonded in the peptide. The molar ellipticity $[\theta]$ is temperature dependent, and therefore all experiments must be performed at the same temperature, usually 0 ºC to maximize helicity.[473]

$$\theta_H = (\theta_H(0) + (\partial\theta_H / \partial T)T)(1 - x/n) \quad (2)$$

Two methods can be used to investigate the effect of salt bridges without the synthesis of new control peptides: varying pH and ionic strength. For example, Glu and Lys are both charged at pH 7.0 and the ion pair can be shown as Glu^{-}-Lys^{+}, while at pH 2.5, Glu is in the uncharged form and the pair will be Glu^{0}-Lys^{+}.[372] The contribution to the helicity of the Glu-Lys pair is expected to be much smaller at pH 2.5 than at pH 7.0. If the residues of an ion pair are both in their charged forms, the strength of the salt bridges diminish with increasing salt concentration.[372] The ellipticities of synthesized peptides at different NaCl concentrations were measured to show the effect of salt bridges.

5.4.1 CD results

The molar ellipticity is temperature-dependent.[473] Usually, the molar ellipticity at rt is lower than the molar ellipticity at 1.5 ºC. Because the circulating water bath was not very efficient, the internal temperature in the cell could not reach 0 ºC. From the literature, the acceptable temperature range was 0-4 ºC,[372,383,418] so CD data was obtained at 1.5 ºC. The absolute molar ellipticity and percentage helicity were not available due to lamp and calibration problems.

The peptides were dissolved in 10 mM phosphate buffer, and the stock concentration of peptides **69**, **70** and **71** was determined by single Trp UV absorbance in 6 M guanidine hydrochloride at 280 nm (ε_{280} = 5690 M^{-1} cm^{-1}).[474] The solutions for CD analysis were diluted from the stock solution with 10 mM phosphate buffer (pH 7.0).

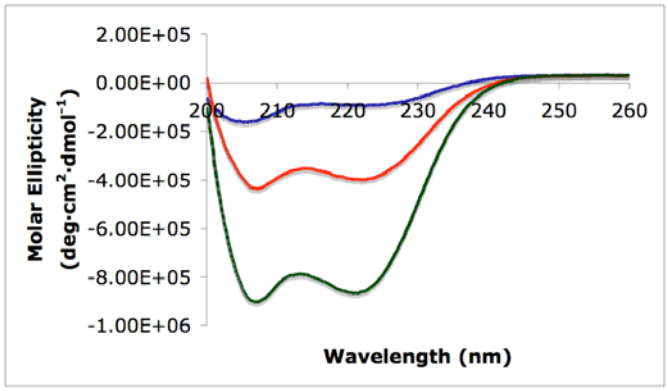


Figure 5.3 Full range CD spectrum of peptides **69**, **70**, and **71**. The peptide concentrations were around 20 μM in 10 mM phosphate buffer. Blue line: the Antp native peptide **69**, red line: peptide **70**, green line: peptide **71**.

The full range CD spectra of peptides **69**, **70** and **71** were obtained at a similar concentration for better comparison (Figure 5.3). All three peptides showed a minimum around 222 nm, and another minimum around 209 nm, indicating the presence of the α-helixes. The molar ellipticity of the native Antp peptide **69** was –93129 deg·cm^2·dmol^{-1} at 222 nm. The molar ellipticity of the modified Antp peptide **70**, which had three more salt bridges and *N*- and *C*-capping effects, was –398924 deg·cm^2·dmol^{-1} at 222 nm, 4.28 times greater than the native Antp peptide **69**. The molar ellipticity of peptide **71**, which had ten more Ala residues and *N*- and *C*-capping effects, was –863374 deg·cm^2·dmol^{-1} at 222 nm, 9.27 times greater than the native peptide.

The peptide concentration dependence experiments were performed (Figure 5.4). The full range CD spectra of peptides **69**, **70** and **71** were obtained at different peptide concentrations, and the molar ellipticities at 222 nm were recorded. The molar ellipticity of all three peptides increased with increasing peptide concentrations. The concentration dependence of the native Antp peptide **69** was significant, while the concentration

dependence of the modified Antp peptides **70** and **71** was not as significant. The concentration dependence was more significant when the molar ellipticity was low.

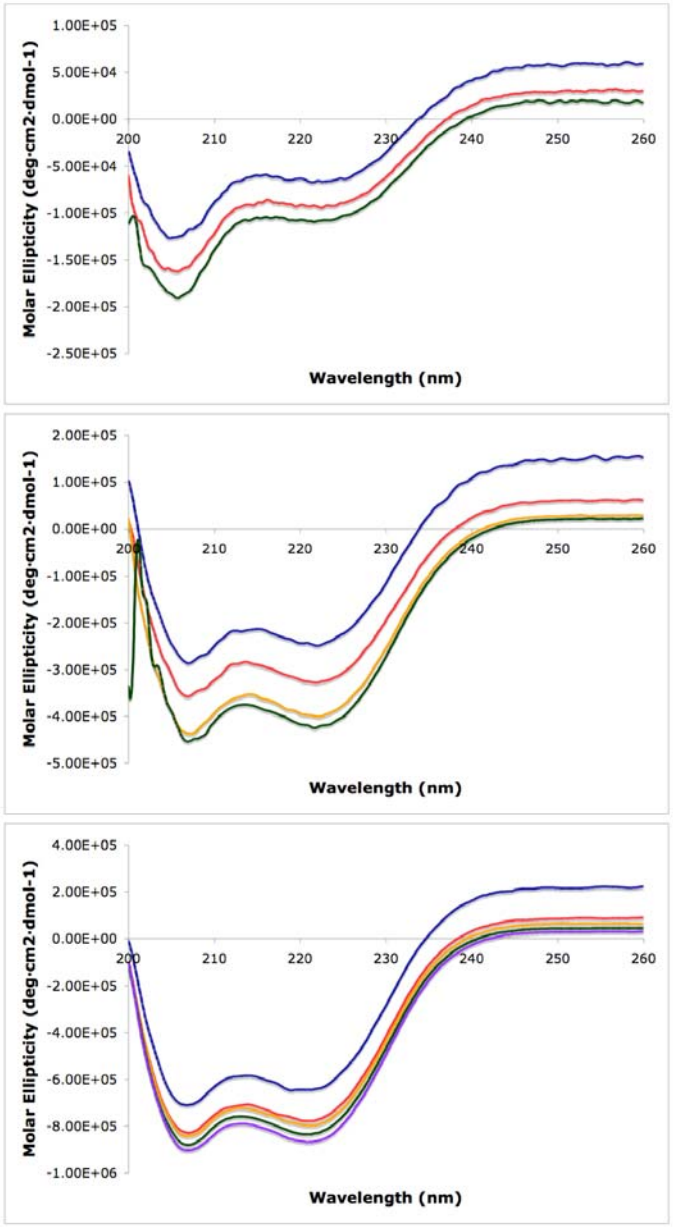


Figure 5.4 Molar ellipticity-concentration dependence of peptides **69**, **70** and **71**. Top: the native Antp peptide **69**. Blue line: 10 μM; red line: 20 μM; green line: 30 μM. Middle: peptide **70**. Blue line: 3.8 μM; red line: 9.8 μM; yellow line: 19.6 μM; green line: 29.5 μM. Bottom: peptide **71**. Blue line: 2.8 μM; red line: 7 μM; yellow line: 9.8 μM; green line: 14 μM; purple line: 21 μM.

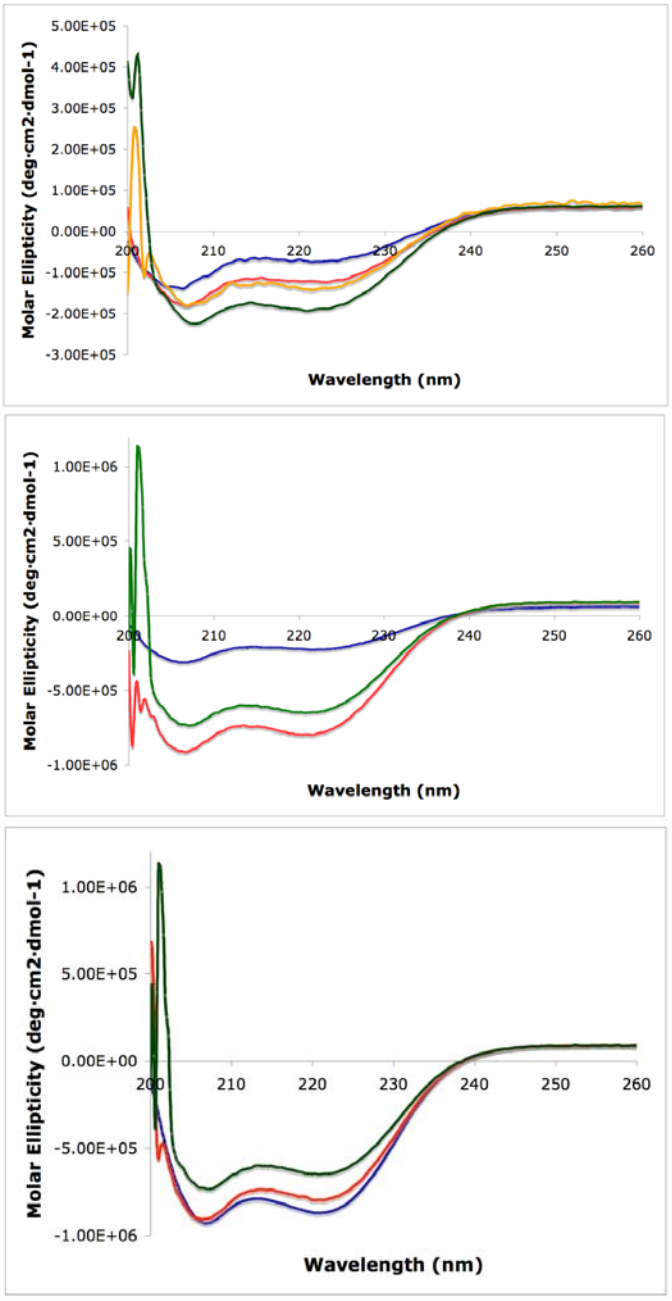


Figure 5.5 Salt effect of peptides **69**, **70** and **71**. Top: the native Antp peptide **69** (10 μM). Blue line: 10 mM NaCl; red line: 0.5 M NaCl; yellow line: 1 M NaCl; green line: 2.5 M NaCl. Middle: peptide **70** (9.8 μM). Blue line: 10 mM NaCl; red line: 1 M NaCl; green line: 2.5 M NaCl. Bottom: peptide **71** (7.0 μM). Blue line: 10 mM NaCl; red line: 1 M NaCl; green line: 2.5 M NaCl.

The experiments of salt effects were also performed. The full range CD spectra of peptides **69**, **70** and **71** in 10 mM, 0.5 mM, 1 M, and 2.5 M NaCl solution were obtained,

respectively (Figure 5.5). High concentrations of salt can interact with charged side-chains, and break the side-chain-side-chain ion pairs. We expected that the molar ellipticity would decrease with the increase of salt concentration. The results showed that only the molar ellipticity of peptide **71** decreased with increasing salt concentration. Peptide **69** and **70** showed an increase of the molar ellipticity when salt concentration increased, which opposed to our hypothesis. These results will be discussed in Section 5.4.2.

5.4.2 Discussion

The native Antp peptide **69** has two ion pair, Arg29-Glu33 and Glu42-Lys46, and no *N*- or *C*-capping effects. It has only two Ala residues in the sequence, which is not enough for Ala intrinsic stabilizing effect. Peptides **70** and **71** have a negatively charged Asp residue at the partially positively charged *N*-terminus, and a positively charged Arg residue at the partially negatively charged *C*-terminus to cap the helix dipole. In peptide **70**, three more ion pairs spaced (i, $i + 4$), Asp27-Arg31, Glu32-Lys36, and Glu49-Arg53, were introduced to stabilize the helix. In peptide **71**, 10 more Ala residues were introduced, and the high content of Ala residues should stabilize the α-helixes intrinsically.

From these results, the molar ellipticities of modified peptides **70** and **71** were higher than the native Antp peptide **69**. The molar ellipticity of peptide **70** was 4.3 times greater than the native Antp peptide **69**, which meant the additional three salt bridges and the *N*- and *C*-terminal capping effects made a significant contribution to the stability of the α-helix. The molar ellipticity of the alanine-based peptide **71** was 9.3 times greater than the native Antp peptide **69**, which agreed with the hypothesis that Ala could intrinsically stabilize the α-helix.[434]

The molar ellipticity of all three peptides increased with the increase of peptide concentration. For the native Antp peptide **69**, this concentration dependence was significant. For peptides **70** and **71**, this concentration dependence was also observed, but not as significant. The peptide concentration had more influence on the molar ellipticity when the concentration was low.

The salt effect of the α-helical content was very complicated. High concentrations of salt could screen the side-chain ion-pairs, and break the salt bridges. But the stability of the resulting helix did not always decrease as expected.[364,434] Even in similar peptides, the salt effect is still very difficult to predict and explain. For example, the (*i*, *i* + 4) and (*i*, *i* + 3) Glu^{-} $\cdots Lys^{+}$ ion pairs in a short Ala-based peptide showed a decrease in the molar ellipticity with an increase in the NaCl concentration, while the (*i*, *i* + 4) and (*i*, *i* + 3) $Lys^{+}\cdots Glu^{-}$ ion pairs in the same peptide showed a slight increase with an increase in the NaCl concentration.[364]

The molar ellipticities of the alanine-based peptide **71** decreased with an increase in the salt (NaCl) concentration, which agrees with the hypothesis. This means that both the helical propensity of Ala, which was believed to be the major stabilizing force, and the two salt bridges in the sequence, contribute stabilizing effects to the stability of the helical structure of the alanine-based peptide **71**. When the salt concentration in the peptide solution increased, the salt bridges broke and the helicity of the peptide decreased. But in the native peptide **69** and the modified peptide **70**, the molar ellipticities increased with an increase in the salt concentration, which opposes the hypothesis. The salt bridges are responsible for the stability of the helical structures in the native peptide **69** and the modified peptide **70**, but they may not be the major force in stabilizing the helixes. Some other interactions, such as hydrophobic interactions and dipole-dipole interactions, increase the stability of the α-helix

with the increase in the salt concentration, and are more important to the stability of the α-helical structure in peptides **69** and **70**.[434] The helix capping effects are also responsible for the stability of the helixes in the modified peptide **70**, but their contribution to the stability is difficult to determine.

5.5 Other HTH peptides

Two control peptides for HTH-turn mimics were also purified and characterized. Peptide **72** (Sequence T-R-R-R-R-I-E-I-A-H-A-F-G-A-T-E-R-Q-I-K-I-W-F-Q-N-R-R-M-K), which replaced the Ser residue in the turn with a Gly residue, was nearly not helical. The full range CD spectrum showed a very small minimum around 222 nm, and a small minimum around 205 nm (Figure 5.6). The molar ellipticity at 222 nm was only –3269 deg·cm^2·dmol^{-1} at 24 μM peptide concentration, which indicated that the peptide helicity decreased significantly after the Ser → Gly mutation. The helicity of peptide **72** increased with increasing peptide concentrations, but the molar ellipticity values were always very small.

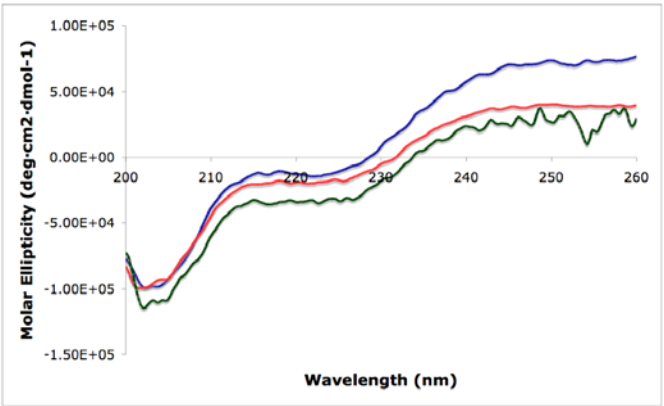


Figure 5.6 Molar ellipticity-concentration dependence of peptide **72**. Blue line: 8.0 μM; red line: 16 μM; green line: 24 μM.

Peptide **73** (Sequence T-R-R-R-R-I-E-I-A-H-A-chA-G-A-T-E-R-Q-I-K-I-W-F-Q-N-R-R-M-K) has a cyclohexylalanine (chA) residue at the position of Ser in the turn. The full

range CD spectrum showed a minimum at 222 nm and another minimum around 209 nm, indicating the presence of the α-helix (Figure 5.7). The molar ellipticity at 222 nm was –152684 deg·cm^2·dmol^{-1} at 21 μM peptide concentration, which was a litter higher than the native Antp peptide **69**, much higher than another control peptide **72**. The concentration dependence experiment showed that the helicity of peptide **73** was independent on the peptide concentrations. The salt effect experiment showed that the helicity increased significantly with the increase of salt concentration, which also opposed to the hypothesis.

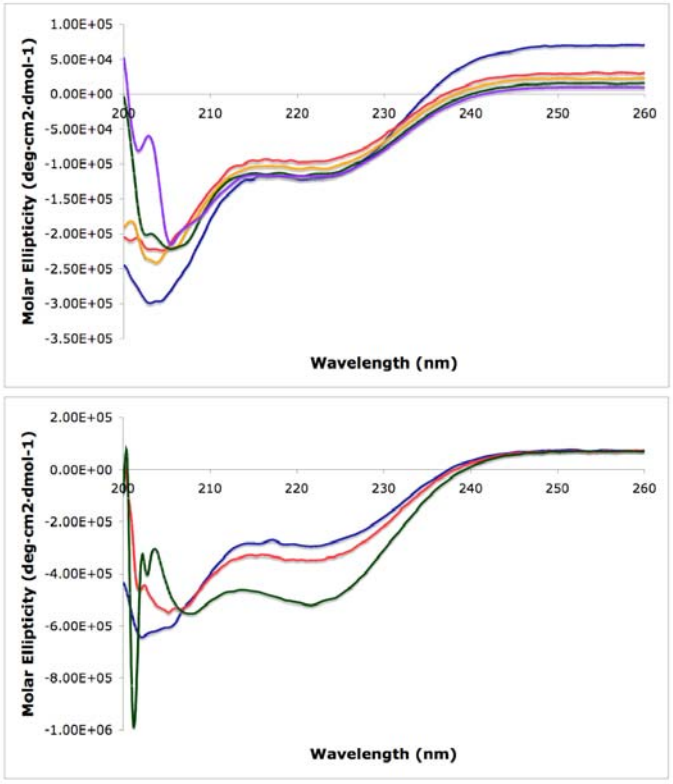


Figure 5.7 Molar ellipticity-concentration dependence and salt effect of control peptide **73**. Up: concentration dependence. Blue line: 8.4 μM; red line: 21 μM; yellow line: 30 μM; green line: 42 μM; purple line: 63 μM. Bottom: salt effect. Blue line: 10 mM NaCl; red line: 0.5 M NaCl; green line: 2.5 M NaCl.

5.6 Summary

The helix-turn-helix motif is an important tertiary structure in DNA-binding proteins. It has at least 20 residues, including an α-helix, a turn of three or four amino acids, and a second α-helix. The two α-helixes in the HTH motif are responsible for the recognition and binding to DNA. In many different HTH proteins the recognition helix has various orientations in the major grooves of DNA.

Stepwise modifications of the Antennapedia HTH peptide (27-55) were performed to improve the helicity and stability. Designed HTH peptides were synthesized at Biomolecular Research Facility, University of Virginia. The peptides were purified with HLPC on a protein C4 column, and CD analysis was performed. The peptide **70** with more side-chain ion-pairs was over 4 times more helical than the native Antp peptide **69**, while the Ala-based peptide **71** was over 9 times more helical than the native peptide. Due to the problem with the old lamp, the absolute molar ellipticity was not available.

The molar ellipticity of all three peptides increased with the increase of peptide concentration. For the native Antp peptide **69**, this concentration dependence was significant. For peptides **70** and **71**, this concentration dependence was observed, but not very significant.

The salt effect on the HTH peptides was also studied. The molar ellipticity of the Ala-based peptide **71** decreased with the increase of NaCl concentration. The native Antp peptide **69** and the more ion-pair containing peptide **70** showed unusual increases in the molar ellipticity with the increase of salt concentration, which indicated that ion-pairs may not be the major factor in stabilizing these HTH peptides.

Experimental

HPLC purification: Both the analytical and preparative HPLC were performed on a Varian PrepStar 218 dual-solvent HPLC system with a ProStar 320 UV detector. The chromatograms were monitored at 254 nm in analytical HPLC, and at 210 nm in preparative HPLC. Analytical HPLC was performed on a VYDAC Protein C4 reverse phase analytical column (5 μm, 4.6 × 250 mm). A binary solvent system (A: 0.1 % TFA in H_2O, B: 0.1 % TFA in CH_3CN) with a flow rate of 2.0 mL/min was used. Preparative HPLC was performed on VYDAC Protein C4 reverse phase semi-preparative column (10 μm, 22 × 250 mm). A binary solvent system (A: 0.1 % TFA in H_2O, B: 0.1 % TFA in CH_3CN) with a flow rate of 20 mL/min was used. The native Antp peptide **69** was purified with a gradient from 10 % to 70 % solvent B in 40 min, and the retention time was 38.12 min. The modified Antp peptide **70** was purified with a gradient from 20 % to 60 % solvent B in 40 min, and the retention time was 22.79 min. The modified Antp peptide **71** was purified with a gradient from 10 % to 55 % solvent B in 30 min, and the retention time was 31.06 min. Peptide **72** was purified with a gradient from 10 % to 45 % solvent B in 35 min, and the retention time was 22.83 min. Peptide **73** was purified with a gradient from 10 % to 95 % solvent B in 40 min, and the retention time was 21.45 min.

UV measurement: The concentration of the peptide solutions was determined by UV absorbance at 280 nm with a Shimadzu UV 1601 spectrometer. The peptide solutions (100 μL) were measured with a volumetric automatic pipette and diluted to 1 mL in the volumetric flask with 6 M guanidine hydrochloride solution. The concentrations of peptides in phosphate buffer (10 mM Na_xPO_4, pH 7.0) were determined by the UV absorption of the Trp residue at 280 nm (ε_{280} = 5690 M^{-1} cm^{-1}).[474] The stock concentration of the native Antp HTH

peptide **69** was determined to be 400.7 μM. The stock concentration of the modified Antp peptides **70** and **71** was 393.7 μM and 282.2 μM, respectively. The stock concentration of peptides **72** and **73** was determined to be 323.4 μM and 843.6 μM, respectively.

CD analysis. The peptide solutions were incubated at 4 °C for ≥ 72 h. The CD spectra were obtained with a spectropolarimeter in 0.2 nm increments, 1 nm bandwidth, and 1.0 cm pathlength at a scan speed of 50 nm/min. The spectra were averaged over four consecutive scans, and blank buffer scans were subtracted from the baseline. Full scan spectra from 200 to 260 nm were obtained at 1.5 °C after 30 min equilibrium time. The temperature was measured in the cell with a temperature probe. Salt solutions for salt effect studies were 10 mM or 5 M NaCl, diluted to the desired concentrations. The molar ellipticity at 222 nm was recorded for data analysis.

Chapter 6 Design and synthesis of the conformationally locked HTH-turn mimic

6.1 Design of the HTH-Turn Mimic

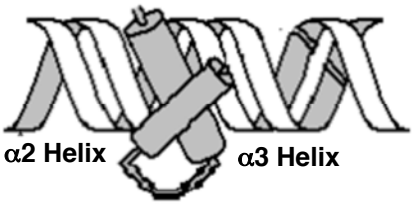
Short peptides are not conformationally stable when separated from the entire protein due to the lack of constraints from other residues.[434,475] Various covalent bonds, such as disulfide bonds, amide bonds and carbon-carbon bonds, can be used to constrain the conformation. Conformationally locked 12-membered-ring HTH-turn mimics were designed to stabilize the tertiary structure of the HTH motif.[448] Introduction of additional hydrophobic contacts are also favored in the core of the HTH motif.[476]

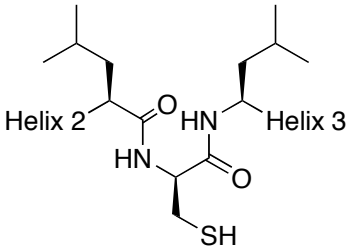
Two principles were followed when designing mimics for the HTH-turn: a) the designed turns should not affect the property of the DNA binding sites; b) the HTH motif with the designed turn should retain the tertiary structure.[477] The designed mimic does not contain a hydrophobic side-chain, but it has a carbon-carbon covalent bond to stabilize the tertiary structure, and to satisfy the steric constraints in the core of the turn. Meanwhile, the 12-membered ring is somewhat flexible to allow the mimic to work even if it cannot exactly match the natural motif.[448]

In the HTH-turn mimic **74**, the side-chains of the first amino acid residue and the third residue were linked by a covalent bond to form a 12-membered ring. A phenyl group was introduced into the turn to increase the hydrophobic interactions between helix 2 and helix 3 of the Antp HTH motif.[448] The Boc-protected HTH-turn mimic was synthesized earlier, but solid phase peptide syntheses with Boc-protected amino acids were not successful, and no desired Antp peptide with the HTH-turn mimic was obtained.[477] So the

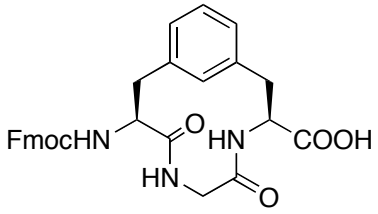
Fmoc protecting group was used to protect the *N*-terminus instead of Boc, this was supposed to make HTH-turn mimic suitable for solid phase peptide synthesis. The Fmoc-protected HTH-turn mimic **74** was successfully synthesized by a method similar to the Boc-protected HTH-turn mimic.[476]

The HTH-turn mimic **75** is the control mimic for the HTH-turn mimic **74**. Like the HTH-turn mimic **74**, the HTH-turn mimic **75** is also linked by a covalent bond to form a 12-membered ring, and was Fmoc-protected at the amine end, but it had an alkene bond in the place of the phenyl group. The HTH-turn mimic **75** was designed to be less hydrophobic than the HTH-turn mimic **74**, which might result in a less stable HTH peptide when the mimic was incorporated into the peptide sequence. The double bond of **75** retained two of the three sp^2 centers in the 12-membered ring of compound **74**, while the position of the double bond did not affect the property of the mimic. The HTH-turn mimic **75** was my target compound.

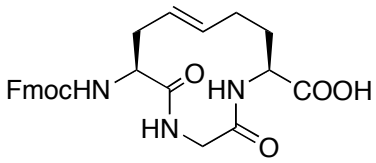




Antp27-55:T-R-R-R-R-I-E-I-A-H-A-L-C-L-T-E-R-Q-I-K-I-W-F-Q-N-R-R-M-K
HTH mimic:T-R-R-R-R-I-E-I-A-H-A---X---T-E-R-Q-I-K-I-W-F-Q-N-R-R-M-K
Control**72**:T-R-R-R-R-I-E-I-A-H-A-L-G-L-T-E-R-Q-I-K-I-W-F-Q-N-R-R-M-K
Control**73**:T-R-R-R-R-I-E-I-A-H-A-L-chA-L-T-E-R-Q-I-K-I-W-F-Q-N-R-R-M-K



74



75

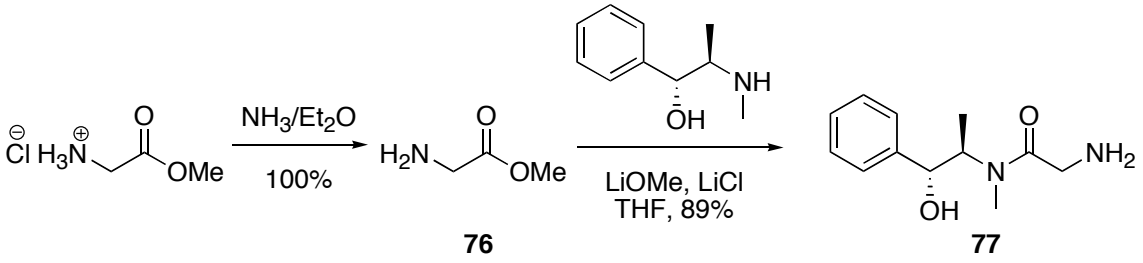
Figure 6.1 Designed conformationally locked HTH-turn mimics.

6.2 Synthesis of the HTH-turn mimic

6.2.1 Asymmetric synthesis of L-allylglycine and L-homoallylglycine

The synthesis of the Boc-protected HTH-turn mimic with a similar structure to the target molecule has been described by Travins.[477] The Fmoc-protected HTH-turn mimic **75** was designed to be compatible with solid phase peptide synthesis methods. The synthesis of L-allylglycine has already been reported by Myers and Gleason.[478] Starting from the commercially available glycine methyl ester hydrochloride, NH_3 gas was bubbled through a solution containing this reagent for 3 h to free the amine group (Scheme 6.1). Glycine methyl ester **76** was then reacted with (*R,R*)-(–)-pseudoephedrine in the presence of the base LiOMe to make (*R,R*)-(–)-pseudoephedrine glycinamide **77**. The pure compound **77** was obtained by crystallization from toluene.

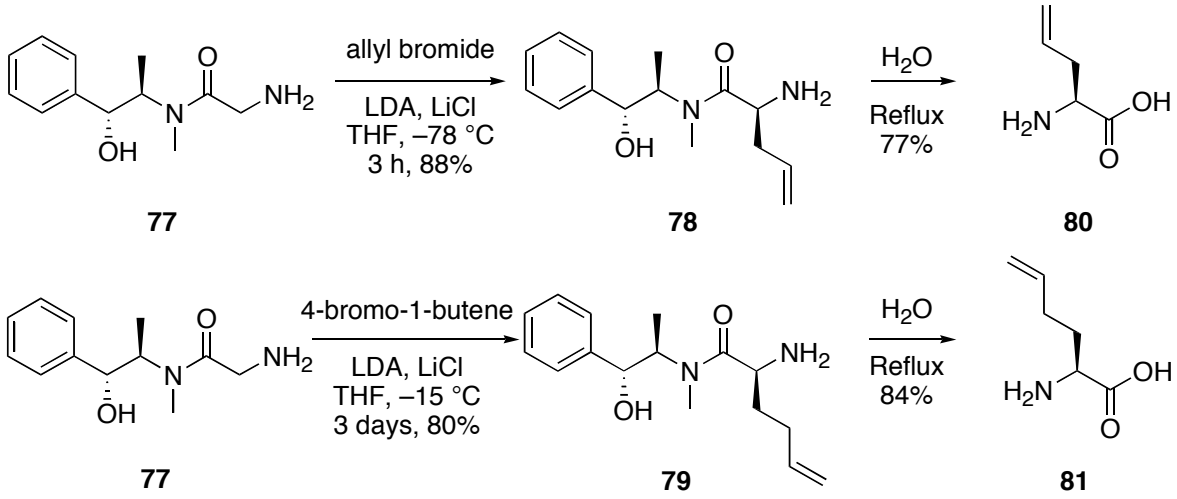
Scheme 6.1 Synthesis of (*R,R*)-(–)-pseudoephedrine glycinamide **77**.



Allylbromide and 4-bromo-1-butene was used to replaced one C-H hydrogen atom of Gly asymmetrically in the presence of LDA (Scheme 6.2). (*R,R*)-(–)-Pseudoephedrine acted as a chiral auxiliary to direct the chirality of product resulting from the replacement of the allyl and homoallyl groups. Pseudoephedrine allylglycinamide **78** and pseudoephedrine homoallylglycinamide **79** were synthesized with a similar method. Allyl bromide was very reactive, and the reaction only needed 2-3 h at –78 °C. 4-Bromo-1-butene was much less reactive, and it required a much longer reaction time. When the reaction time was 3 h, less

than 10 % of the desired product was obtained; the yield increased to about 20-30 % when the reaction time was increased to 16-24 h. When the reaction was carefully controlled at –15 to –25 °C for 3 days, the yield increased to 80 %. A longer reaction time (14 h) was also tried for the synthesis of pseudoephedrine allylglycinamide **78**, but the yield decreased from 60 % to about 30-40 %.

Scheme 6.2 Synthesis of L-allylglycine **80** and L-homoallylglycine **81**.



The chiral auxiliary (*R*,*R*)-(–)-pseudoephedrine was removed by heating pseudoephedrine allylglycinamide **78** and pseudoephedrine homoallylglycinamide **79** at reflux in water for 12-16 h. The resulting L-allylglycine **80** and L-homoallylglycine **81** were obtained as light yellow solids. The products were usually very pure and no flash chromatography or HPLC was needed. Glycine was observed as a minor impurity sometimes, but it could be removed by flash chromatography after the amine or acid end were protected. The stereoselectivity of L-allylglycine **80** was measured by optical rotation. The synthesized L-allylglycine **80** had an $[\alpha]_D$ value of –35.6° at 20 °C in H_2O, which was close to the literature $[\alpha]_D$ value of –37.2° at 23 °C.[478] The stereo-purity of L-homoallylglycine **81** was not confirmed due the lack of a literature value for comparison.

6.2.2 Protecting group manipulation

The protecting groups at the amine and acid ends needed to be chosen to protect L-allylglycine **80** and L-homoallylglycine **81**, respectively. The desired 12-membered ring mimic **75** was Fmoc-protected, so the amine end of L-allylglycine **80** or L-homoallylglycine **81** had be protected with the Fmoc-group. Since Fmoc-protected amino acids were not as stable as Boc-protected amino acids, it was difficult to choose a protecting group for the acid end.[477] In the original synthetic route, a benzyl group was used to protect the carboxyl end. However, if the Fmoc-group was used to protect the amine, the benzyl group would be very difficult to remove because the catalytic hydrogenation procedure would also cleave the Fmoc-group. So new synthetic routes were designed to achieve the desired product.

***tert*-Butyl protection**

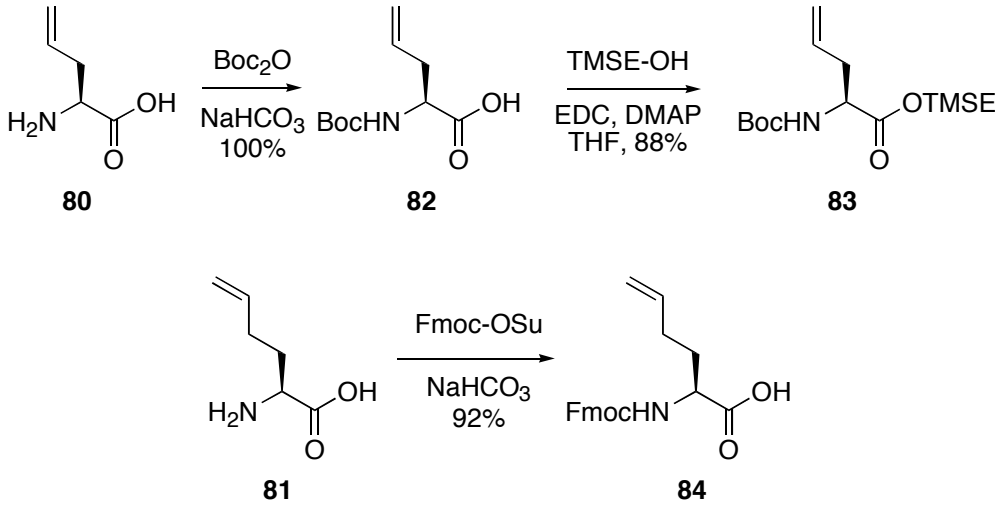
The *tert*-butyl group was used to protect the acid first because it required fewer synthetic steps. Two different conditions were used to couple the *tert*-butyl group to the carboxyl end. The traditional isobutene-H_2SO_4 method[479,480] was tried twice with homoallylglycine **81**, but no desired product was obtained. The alkene bond might not be very stable in highly acidic conditions over a long reaction time (> 7 days), and this generated many side products. A new method was performed with *tert*-butyl acetate and perchloric acid ($HClO_4$).[481] A model reaction with L-leucine was performed first, but still no desired product was obtained. The oxidative nature of the perchloric acid was believed to be responsible for the failure of the reaction. Then concentrated hydrochloric acid with *tert*-butyl acetate was tried, but still no reaction was observed. I believe that this reaction might only work for some specific substrates, such as cysteine. Moreover, perchloric acid could act as an oxidation reagent and oxidize the carbon-carbon double bond in the allyl or homoallyl

glycine. Although the *tert*-butyl group was a perfect protecting group, which was orthogonal to the Fmoc protecting group and the alkene, coupling the *tert*-butyl group to the acid was difficult.

Trimethylsilylethyl (TMSE) protection

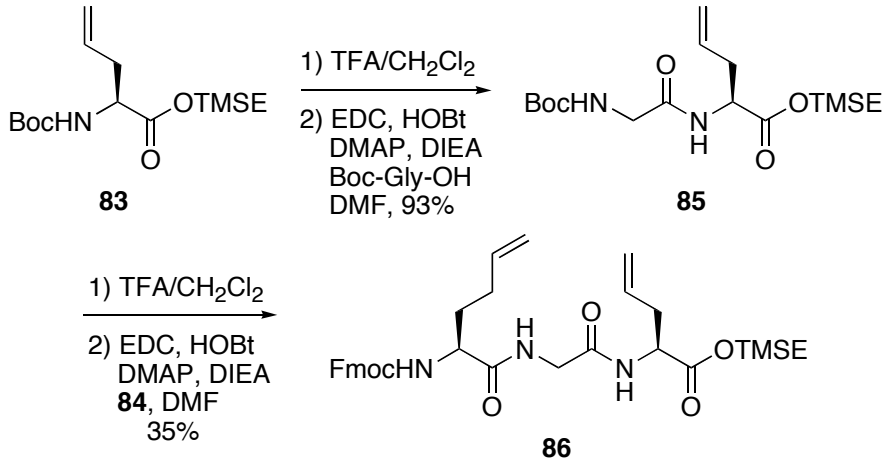
The trimethylsilylethyl (TMSE) group was used to protect the carboxylic end. Starting from the L-allylglycine **80**, Boc-allylglycine **82** was synthesized with Boc_2O in a saturated $NaHCO_3$ solution (Scheme 6.3). Then TMSE-OH was coupled to the carboxylic end, with EDC as the coupling reagent, to synthesize TMSE-protected allylglycine **83**. If glycine was mixed with allylglycine, the best time to separate the product species was at this point. Protected-glycine and allylglycine (or homoallylglycine) were significantly different in polarity, which made it easy to separate them with flash chromatography. Homoallylglycine **81** was protected with an Fmoc-group by stirring with Fmoc-OSu in a saturated $NaHCO_3$ solution, and Fmoc-homoallylglycine **84** was obtained as a light yellow solid in the same way.

Scheme 6.3 Synthesis of protected allylglycine **83** and homoallylglycine **84**.



The Boc-group was cleaved with 25 % TFA in CH_2Cl_2, and the resulting TMSE-protected allylglycine was coupled to Boc-Gly-OH with EDC-HOBt to give the dipeptide mimic **85** (Scheme 6.4). This coupling reaction was very clean, and no side products were observed. Then the Boc-protecting group of dipeptide **85** was removed by 25 % TFA in CH_2Cl_2. Fmoc-homoallylglycine **84** was coupled with EDC-HOBt, and tripeptide mimic **86** was obtained, but the yield for this step was only 35 %, which was much lower than expected. A model reaction with Fmoc-Phe-OTMSE showed that TFA could completely cleave the TMSE protecting group within 2 h.

Scheme 6.4 Synthesis of tripeptide **86**.

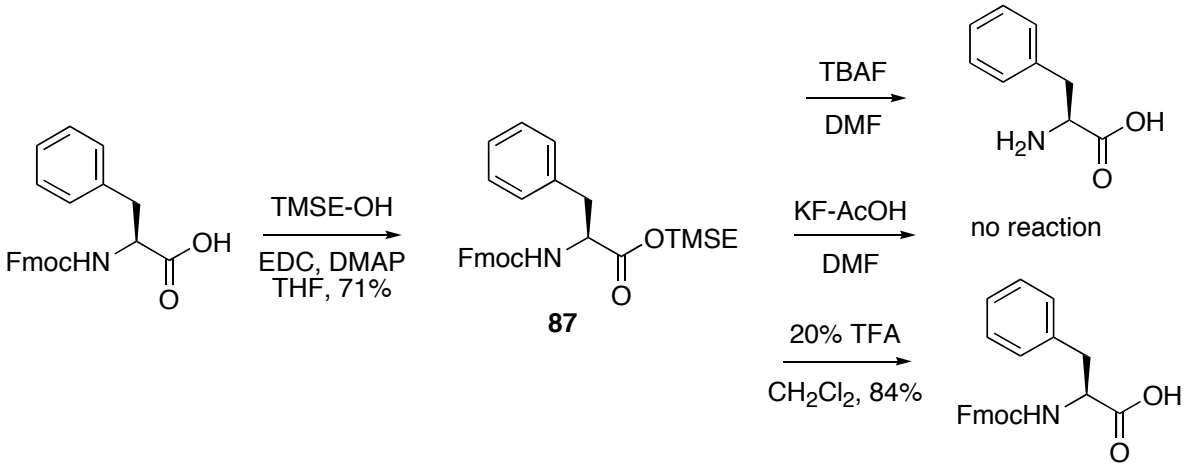


It was noticeable that the TMSE group was not removed when stirring Boc-allylglycine TMSE ester **83** in 25 % TFA in CH_2Cl_2. We hypothesized that the Boc-group and the TMSE-group were both cleaved by TFA, and the deprotection of the Boc protecting group had a faster reaction rate. When the reaction was performed over a long period of time, more of the TMSE protecting group was cleaved by TFA. Dipeptide **85** was synthesized again, and this time the TMSE-protected Boc-allylglycine **83** was stirred in 25 % TFA in CH_2Cl_2 for 40 min. The yield decreased to 76.4 %, but it was still much higher than for the

coupling reaction of tripeptide **86**. This result indicated that the cleavage of the TMSE-group was also substrate-dependent.

These results provided a new method to deprotect the TMSE protecting group. A model reaction with Fmoc-Phe-OH was designed, and three methods were used to deprotect the TMSE-group (Scheme 6.5). Experiments showed that TBAF in DMF[482] could cleave both the Fmoc and the TMSE-groups, and the Fmoc protecting group was completely removed within 15 min. KF with AcOH[483] did not cleave either protecting group, and the starting material Fmoc-Phe-OTMSE was recovered. TFA in CH_2Cl_2 (20 %) could cleave the TMSE-group in 2 h without affecting Fmoc-group in good yield (84 %).

Scheme 6.5 Deprotection of the TMSE-group.

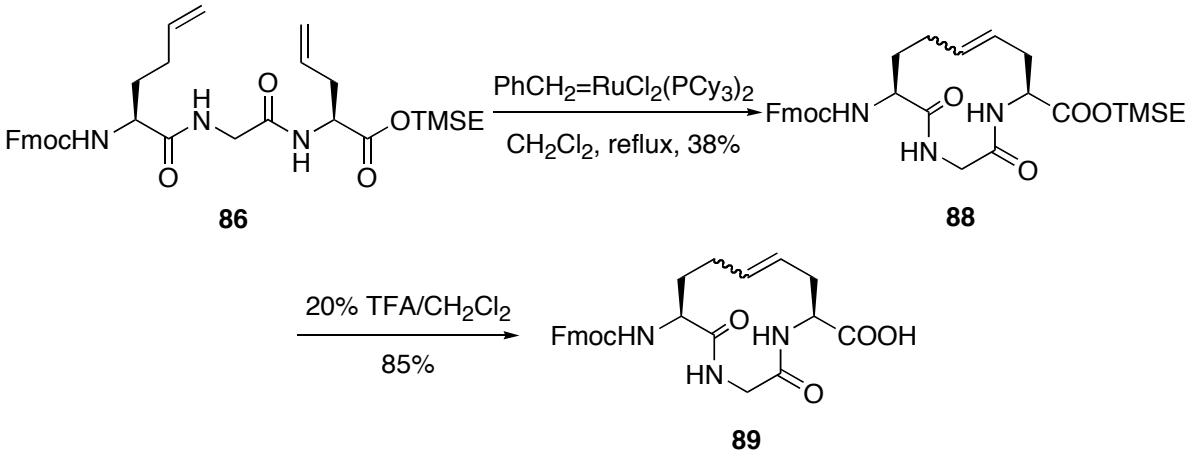


Starting from tripeptide **86**, the cyclic tripeptide mimic **88** was synthesized by ring closing metathesis (Scheme 6.6).[477,484] Grubbs catalyst generation I was used, and the reaction was heated at reflux in CH_2Cl_2 for 48 h, with fresh catalyst added every 12 h. Theoretically, both (*E*) and (*Z*)-alkenes should have been synthesized,[484] but no separation of the two isomers was observed by TLC and preparative TLC. Then, the TMSE protecting group was cleaved by 20 % TFA in CH_2Cl_2 within 3 h, and the HTH-turn mimic **89** was obtained as a yellow solid. The position of the alkene bond was one carbon different from the

designed HTH-turn mimic **75**. Either mimic could be used as the control for the HTH-turn mimic **74**.

The solubility of the HTH-turn mimic **89** was very poor. It did not dissolve in acetone, MeOH, EtOH, EtOAc, THF, CH_2Cl_2, H_2O or DMSO. It could be dissolved in $CHCl_3$ or DMSO in the presence of 2-5 % TFA or AcOH, but the solubility was still not good. Purification of the HTH-turn mimic **89** was a big problem due to its poor solubility.

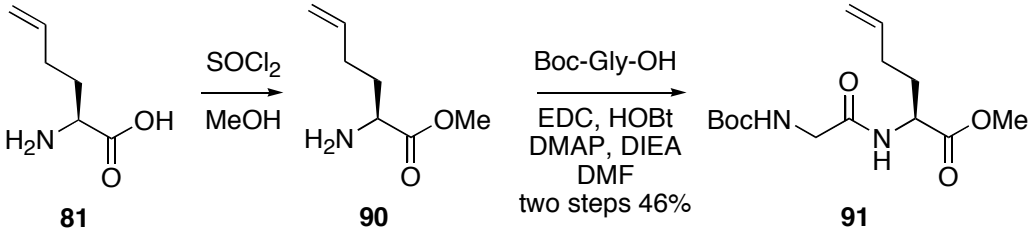
Scheme 6.6 Synthesis of the HTH-turn mimic **89**.



86 **88** **89**

Methyl protection

Scheme 6.7 Synthesis of dipeptide **91**.



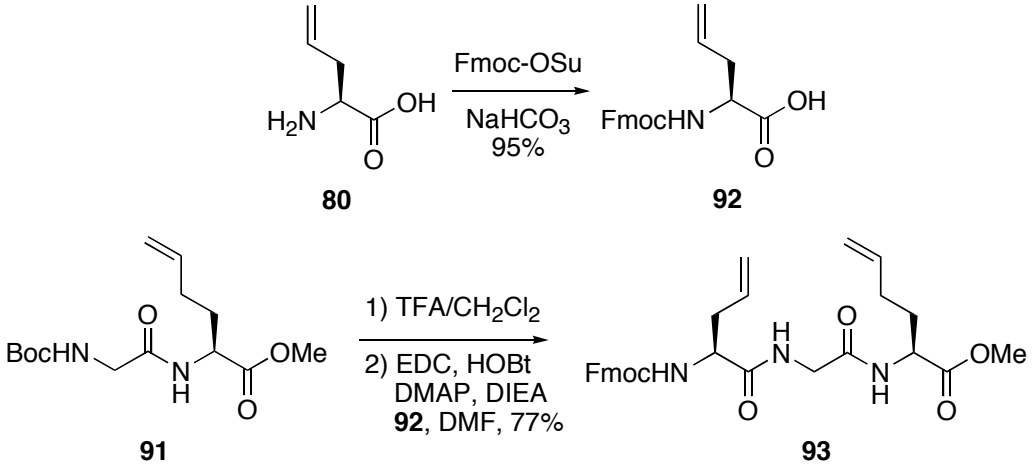
81 **90** **91**

A methyl group was also used to protect the carboxylic end. The homoallylglycine methyl ester **90** was synthesized from L-homoallylglycine **81** by Fisher esterification with $SOCl_2$ in MeOH (Scheme 6.7). The resulting yellow oil was carried to next step without

further purification, and coupled to Boc-Gly-OH with EDC-HOBt to synthesize dipeptide **91.** The yield of these two stes was low (46 %) due to the poor quality of $SOCl_2$.

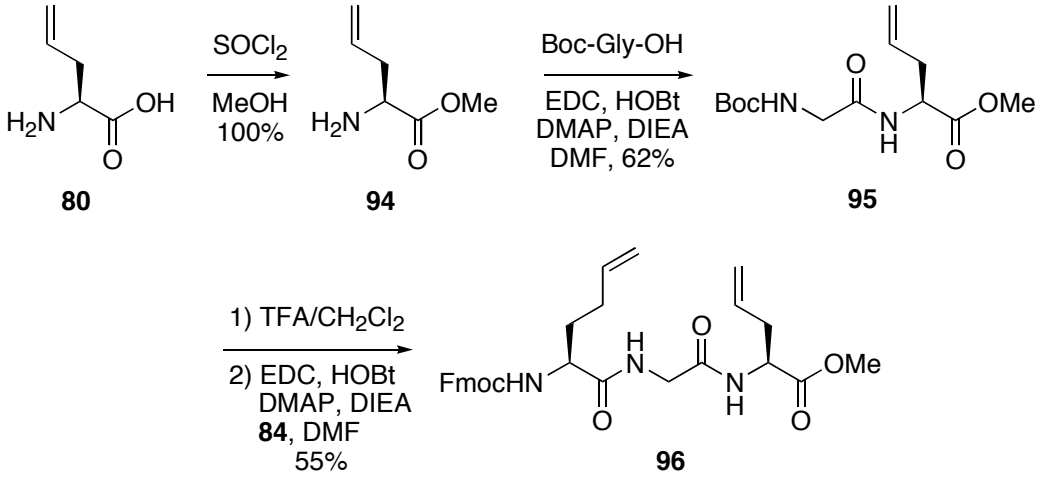
The Boc-protecting group of dipeptide **91** was cleaved with 25 % TFA in CH_2Cl_2 (Scheme 6.8), and the resulting amine was coupled to Fmoc-allylglycine **93**, which was synthesized from L-allylglycine **80** by stirring with Fmoc-OSu in a saturated $NaHCO_3$ solution. Tripeptide **93** was synthesized with EDC-HOBt as the coupling reagent. The purity of Fmoc-allylglycine **92** greatly affected the yield for this reaction, and slow precipitation of the Fmoc-allylglycine **92** in MeOH-Et_2O improved the yield significantly. If precipitation was inadequate to obtain pure product, flash chromatography with 5 % MeOH in chloroform was used to purify Fmoc-allylglycine **92**.

Scheme 6.8 Synthesis of tripeptide **93**.



Tripeptide mimic with the reverse sequence was also synthesized with the same method. Starting from L-allylglycine **80**, allylglycine methyl ester **94** was synthesized by Fisher esterification (Scheme 6.9). Allylglycine methyl ester **94** was coupled with Boc-Gly-OH, and dipeptide **95** was obtained as a colorless oil. After removing the Boc-protecting group with 25 % TFA, Fmoc-homoallylglycine **84** was coupled and tripeptide **96** was obtained as a colorless oil.

Scheme 6.9 Synthesis of tripeptide **96** with a reverse sequence.



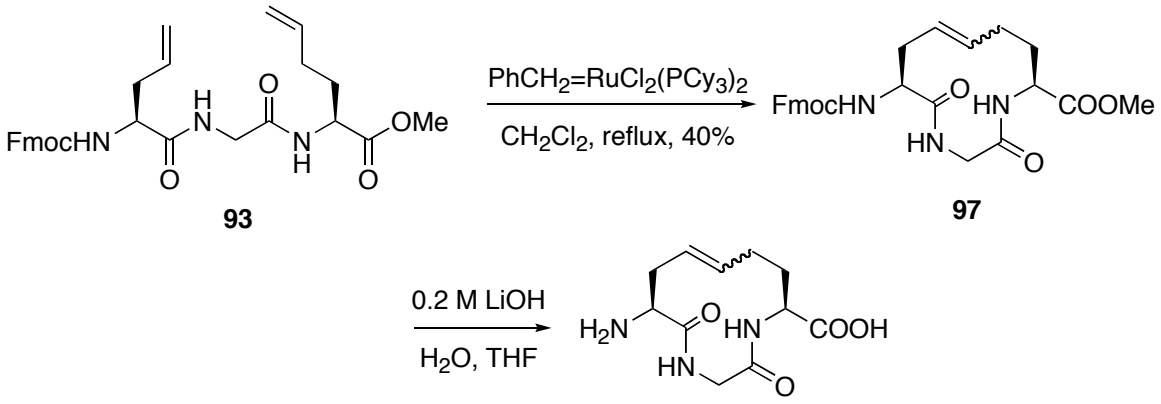
80 **94** **95** **96**

Compared with the syntheses of tripeptides **93** and **96**, there was no significant difference in the overall yield for these two routes. Since allylglycine **80** was easier to synthesize than homoallylglycine **81**, tripeptide **96** was a better target in terms of required effort.

The cyclic peptide mimic **97** was synthesized by heating tripeptide **93** at reflux in CH_2Cl_2 for 36-48 h in the presence of the Grubbs catalyst generation I (Scheme 6.10).[484,485] The yield was usually 30-40 %. Cyclic peptide **97** had very poor solubility. It did not dissolve in acetone, MeOH, EtOH, EtOAc and CH_2Cl_2, and had very low solubility in DMSO and THF. Preparative TLC was used to purify cyclic peptide **97** with 2:1 ratio of EtOAc-hexanes, although it had a very low solubility in this mixture. There was great product loss during the separation.

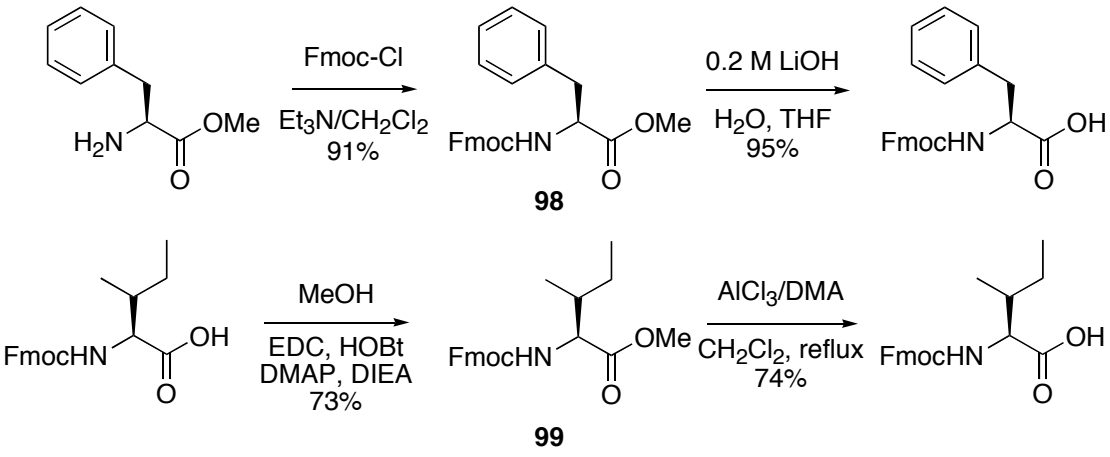
We planned to remove the methyl protecting group with 0.2 M LiOH solution in THF.[231,486] In a model reaction with Fmoc-Phe-OMe **98**, the methyl group was removed without affecting the Fmoc protecting group (Scheme 6.11). But for the cyclic peptide mimic **97**, the methyl protecting group and the Fmoc-group were cleaved simultaneously (Scheme 6.10).

Scheme 6.10 Ring-closing metathesis.



A new method to remove the methyl group by using the Lewis acid $AlCl_3$ and *N*, *N*-dimethyl aniline (DMA) was recently reported.[232] A model reaction with Fmoc-Ile-OMe **99** was performed (Scheme 6.11). Fmoc-Ile-OMe **99** was heated at reflux in CH_2Cl_2 in the presence of the $AlCl_3$/DMA complex, and the methyl group was successfully cleaved without affecting the Fmoc-protection.

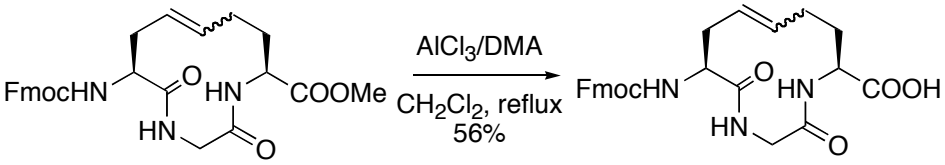
Scheme 6.11 Model reactions to deprotect the methyl protecting group.



The activity and concentration of the catalyst complex were essential for this methyl deprotection method. $AlCl_3$ was extremely moisture sensitive and white smoke was observed when it was exposed to air. It had to be stored, and weighed, in a dry-box. $AlCl_3$ and DMA had to be pre-mixed, or the reaction would not work. The $AlCl_3$/DMA complex was a clear

dark blue or dark green liquid, and did not contain precipitates. This method could remove the methyl protecting group without affecting the Fmoc protecting group. Even when the reaction did not work, all of the starting material was recovered.

Scheme 6.12 Synthesis of the cyclic tripeptide mimic **100**.



The methyl protecting group of cyclic tripeptide **97** was removed under the same conditions, and the desired cyclic tripeptide **100** was obtained as a yellow solid (Scheme 6.12). Like cyclic peptide **89**, cyclic tripeptide **100** had a very poor solubility. It did not dissolve in EtOAc, acetone, $CHCl_3$, DMSO, THF and MeOH, and had a limited solubility in $CHCl_3$ in the presence of 2 % AcOH or TFA.

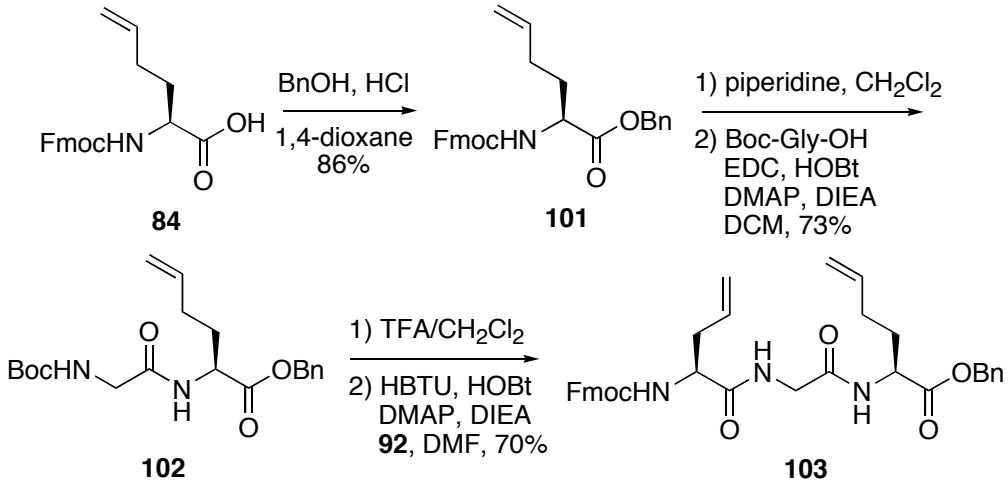
Benzyl protection

The tripeptide mimic with a benzyl protecting group was also synthesized by a similar method to the TMSE and methyl protected tripeptide mimics. Starting from Fmoc-homoallylglycine **84**, BnOH was coupled in the presence of HCl in anhydrous dioxane, or by EDC peptide coupling (Scheme 6.13). Then Fmoc-homoallylglycine benzyl ester **101** was treated with 20 % piperidine in CH_2Cl_2 to remove the Fmoc protecting group. Boc-Gly-OH was coupled to it with EDC, and dipeptide **102** was synthesized as a colorless oil. Then, the Boc protecting group was removed with 25 % TFA in CH_2Cl_2, and Fmoc-allylglycine **92** was coupled to **102** with EDC to synthesize tripeptide **103** in 70 % yield.

The benzyl protecting group could not be removed by hydrogenation in a H_2-Pd/C system due to the presence of the C=C double bond and Fmoc group, while it might be

removable with a Lewis acid such as BCl_3 or $AlCl_3$,[255,256] or under mild hydrogenation condition, such as using 1,4-cyclohexadiene as the hydrogen source.[262] The benzyl protecting group was more stable than the TMSE group, and was easier to remove than the methyl protecting group. It should be a good protecting group for this synthesis, but this step was not completed.

Scheme 6.13 Synthesis of tripeptide benzyl ester **103**.



6.3 Summary

A 12-membered ring, Fmoc-protected HTH-turn mimic was designed. Various protecting groups were tried, and both the methyl protection and TMSE protection gave the desired product. The benzyl protecting group was promising, but the synthesis was not finished yet.

The Fmoc-protected cyclic tripeptide was synthesized and was ready for solid phase peptide synthesis. The solubility of the cyclic peptide was very poor, and the purification of the final product was very difficult. The solubility problem might also affect solid phase peptide synthesis in the future. An alternative route, cyclizing peptides on resin will be explored in the future.

Experimental

General Information. Amino acid derivatives, resins, and reagents were purchased. Unless otherwise indicated, all reactions were carried out under N_2 sealed from moisture. Anhydrous THF was obtained by reflux from Na-benzophenone. Anhydrous CH_2Cl_2 was dried by passage through a dry alumina column. Other anhydrous solvents were used directly from sealed bottles, which were stored under Ar. Brine (NaCl), $NaHCO_3$, and NH_4Cl refer to saturated aqueous solutions unless otherwise noted. Flash chromatography was performed on 32-63 μm or 230-400 mesh silica gel with reagent grade solvents. NMR spectra were obtained at ambient temperature in $CDCl_3$ unless otherwise noted. ^1H and ^{13}C NMR spectra were obtained at 500 and 125 MHz, respectively, unless otherwise noted. Minor rotamer chemical shifts are shown in parenthesis.

H-Gly-OMe, 76. By the method of Myers, A. G. *et al.*[478] H-Gly-OMe•HCl (12.4 g, 99.2 mmol) was suspended in Et_2O (120 mL), and stirred at rt. NH_3 was bubbled in via a needle for 3 h. The resulting suspension was filtered, and the filtrate was concentrated in vacuo. A clear colorless oil (8.81 g, 99.8 %) was obtained. ^1H NMR: δ 3.68 (s, 3H), 3.38 (s, 2H), 1.52 (br s, 2H). ^{13}C NMR: δ 170.6, 44.7, 40.9.

(*R*,*R*)-(–)-Pseudoephedrine glycinamide, 77. (*1R*,*2R*)-(–)-Pseudoephedrine (11.0 g, 66.5 mmol) and LiCl (5.66 g, 133 mmol) was suspended in THF (50 mL), and stirred at 0 °C for 10 min. LiOMe (1.28 g, 33.6 mmol) was added and stirred for another 20 min. H-Gly-OMe **76** (6.54 g, 73.4 mmol) was added dropwise, and the solution was stirred at 0 ºC for 14 h. The reaction was quenched with H_2O (50 mL). The resulting solution was concentrated in vacuo, and extracted with CH_2Cl_2 (6 × 50 mL). The organic layers were combined, dried with K_2CO_3, and concentrated in vacuo. The resulting light yellow oil was dissolved in a

minimal amount of hot THF, and H_2O (1 mL) was added. The mixture was cooled to rt until the appearance of a white solid. The solution was cooled to –20 ºC for 24 h, and the white solid was collected by filtration. The resulting solid was suspended in CH_2Cl_2 (120 mL), and K_2CO_3 (11.7 g, 84.7 mmol) was added, and the mixture was stirred at rt for 60 min. The resulting solution was filtered through Celite™, and concentrated in vacuo. The resulting clear oil was dissolved in a minimal amount of hot toluene, cooled to rt until the appearance of a white plate-like crystal. The solution recrystallization was completed by cooling to –20 ºC for 24 h. The resulting crystal was washed with Et_2O (25 mL), and dried in vacuo. A white solid (12.1 g, 89.1 %) was obtained. mp 78-82 ºC, lit. mp 78-80 °C.[478] ^{1}H NMR: δ 7.35 (m, 5H), 4.60 (d, *J* = 8.3, 0.5H), 4.54 (d, *J* = 9.0, 0.5H), 3.87 (app. dq, *J* = 8.9, 7.1, 0.5H), 3.79 (d, *J* = 25.5, 0.5H), 3.53 (d, *J* = 8.4, 0.5H), 3.50 (d, *J* = 6.8, 0.5H), 3.41 (d, *J* = 16.7, 1H), 2.96 (s, 1H), 2.79 (s, 2H), 2.46 (br s, 2H), 1.06 (d, *J* = 6.4, 2H), 0.98 (d, *J* = 7.0, 1H). ^{13}C NMR: δ 174.4 (173.5), 142.0 (142.3), 128.6 (128.8), 128.0 (128.4), 126.7 (127.0), 76.2 (75.2), 57.7, 43.9 (43.6), 27.1 (30.6), 15.5 (15.4). Rotamers: 3:2.

Pseudoephedrine L-allylglycinamide, 78. *i*-Pr_2NH (9.75 mL, 69.0 mmol) was dissolved in THF (50 mL), and stirred at 0 °C for 20 min. n-BuLi (2.5 M, 26 mL, 65 mmol) was added dropwise and stirred at 0 °C for 30 min to generate LDA. Pseudoephedrine glycinamide **77** (7.29 g, 32.8 mmol) and LiCl (6.97 g, 164 mmol) were suspended in THF (150 mL) and stirred at 0 °C for 30 min, and the LDA solution was added via a cannula dropwise over 50 min. After stirring at 0 °C for another 45 min, allylbromide (3.1 mL, 36 mmol) was added dropwise. The solution was stirred at 0 °C for 3 h, and quenched with H_2O (50 mL). The solution was concentrated in vacuo, and diluted with a mixture of 3 M HCl (30 mL) and EtOAc (100 mL). The organic layer was extracted with 1 M HCl (2 × 50 mL). The aqueous

layers were combined, basified to pH 14 with 50 % NaOH at 0 °C, and extracted with CH_2Cl_2 (4 × 150 mL). The resulting organic layers were combined, concentrated in vacuo, dried over K_2CO_3, and purified by flash chromatography (3 % MeOH, 2 % Et_3N in CH_2Cl_2). A pale yellow oil (7.59 g, 88.2 %) was obtained. ^{1}H NMR: δ 7.33 (m, 5H), 5.70 (ddt, *J* = 17.2, 9.8, 7.2, 1H), 5.10 (m, 2H), 4.59 (m, 2H), 3.65 (dd, *J* = 7.6, 5.0, 1H), 2.86 (s, 3H), 2.23 (m, 1H), 2.11 (m, 1H), 1.06 (d, *J* = 6.9, 3H). ^{13}C NMR: δ 176.3 (175,4), 142.3 (141.7), 133.8 (134.9), 128.5 (128.9), 127.9, 126.7 (127.0), 118.5 (118.2), 75.9 (75.4), 57.9, 51.5, 40.1, 39.7, 14.6. HRMS calcd. for $C_{15}H_{23}N_2O_2$ (MH^+) *m/z* = 263.1760, found *m/z* = 263.1764. Rotamers: 3:1.

Pseudoephedrine L-homoallylglycinamide, 79. *i*-Pr_2NH (7.0 mL, 50 mmol) was dissolved in THF (20 mL), and stirred at 0 °C for 20 min. n-BuLi (2.5 M, 19 mL, 47 mmol) was added dropwise and stirred at 0 °C for 30 min to generate LDA. Pseudoephedrine glycinamide **77** (5.28 g, 23.8 mmol) and LiCl (5.33 g, 126 mmol) were suspended in THF (50 mL) and stirred at –78 °C for 30 min, and the LDA solution was added via a cannula dropwise over 50 min. After stirring at –78 °C for another 45 min, 4-bromo-1-butene (2.65 mL, 26.1 mmol) was added dropwise. The solution was stirred at –15 °C for 72 h, and quenched with H_2O (50 mL). The solution was concentrated in vacuo, and diluted with a mixture of 3 M HCl (30 mL) and EtOAc (100 mL). The organic layer was extracted with 1 M HCl (4 × 30 mL). The aqueous layers were combined, basified to pH 14 with 50 % NaOH at 0 °C, and extracted with CH_2Cl_2 (6 × 100 mL). The resulting organic layers were combined, concentrated in vacuo, dried over K_2CO_3, and purified by flash chromatography (3 % MeOH, 2 % Et_3N in CH_2Cl_2). A pale yellow oil (5.4 g, 82 %) was obtained. ^{1}H NMR: δ 7.34 (5H, m), 5.75 (ddt, *J* = 17.0, 10.3, 6.7, 1H), 4.99 (m, 2H), 4.63 (d, *J* = 7.8, 1H), 4.42 (m, 1H), 3.56

(dd, *J* = 8.2, 4.4, 1H), 2.83 (s, 3H), 2.11 (m, 2H), 1.50 (m, 1H), 1.42 (m, 1H), 1.12 (d, *J* = 7.2, 3H). ^{13}C NMR: δ 177.4, 142.4, 137.8 (138.3), 128.4 (128.7), 127.8 (128.2), 126.5 (127.1), 115.3 (115.5), 75.9, 57.8, 51.2, 34.3, 30.3, 30.0, 14.5. HRMS calcd. for $C_{16}H_{25}N_2O_2$ (MH^+) *m/z* = 277.1916, found *m/z* = 277.1916. Rotamers: 3:1.

H-L-allylGly-OH, 80. Pseudoephedrine L-allylglycinamide **78** (2.79 g, 10.6 mmol) was heated at reflux in H_2O (50 mL) for 14 h. The solution was cooled to rt, and NH_3•H_2O (5 mL) was added. The resulting colorless needle-like crystals were removed by filtration, and the filtrate was washed with EtOAc (3 × 20 mL). The organic layers were combined and washed with NH_3·H_2O solution (20 % NH_3·H_2O in H_2O, 25 mL). The aqueous layers were combined, and H_2O was removed with lyophilizer. The resulting yellow solid was washed with EtOH (3 × 5 mL), and dried in vacuo. A pale yellow solid (0.94 g, 77 %) was obtained. $[\alpha]_D^{20}$ = –35.6° (H_2O), lit. $[\alpha]_D^{23}$ = –37.2° (*c* 4, H_2O).[478] ^{1}H NMR (D_2O): δ 5.77 (ddt, *J* = 17.2, 10.0, 7.2, 1H), 5.28 (ddd, *J* = 17.2, 3.0, 1.5, 1H), 5.26 (ddd, *J* = 10.2, 2.6, 1.0, 1H), 3.80 (dd, *J* = 7.2, 4.8, 1H), 2.63 (m, 2H). ^{13}C NMR (D_2O): δ 174.3, 131.5, 120.7, 54.1, 35.0.

H-L-homoallylGly-OH, 81. Pseudoephedrine L-homoallylglycinamide **79** (3.35 g, 12.1 mmol) was heated at reflux in H_2O (50 mL) for 14 h. The solution was cooled to rt, and NH_3•H_2O (5 mL) was added. The resulting colorless needle-like crystals were removed by filtration, and the filtrate was washed with EtOAc (3 × 20 mL). The organic layers were combined and washed with NH_3·H_2O solution (20 % NH_3·H_2O in H_2O, 25 mL). The aqueous layers were combined, and H_2O was removed with lyophilizer. The resulting yellow solid was washed with EtOH (3 × 5 mL), and dried in vacuo. A pale yellow solid (1.32 g, 84 %) was obtained. ^{1}H NMR (D_2O): δ 5.89 (ddt, *J* = 17.1, 10.4, 6.6, 1H), 5.15 (ddd, *J* = 17.3, 3.4,

1.7, 1H), 5.08 (ddd, *J* = 10.3, 3.1, 1.2, 1H), 3.75 (dd, *J* = 7.2, 5.2, 1H), 2.18 (m, 2H), 1.97 (m, 2H). ^{13}C NMR (D_2O): δ 174.9, 137.2, 116.1, 54.5, 29.9, 28.8.

Boc-allylGly-OH, 82. H-allylGly-OH (53 mg, 0.46 mmol) and $NaHCO_3$ (78 mg, 0.93 mmol) were dissolved in H_2O (2.0 mL), and stirred at 0 °C. Boc_2O (113 mg, 0.516 mmol) in 1,4-dioxane (1.0 mL) was added dropwise, and the resulting solution was stirred at 0 °C for 3 h. The solvent was removed in vacuo, and the resulting white solid was dissolved in $NaHCO_3$ (25 mL). The resulting solution was washed with EtOAc (2 × 20 mL), and acidified to pH 1 with 1 M HCl and 1 M $NaHSO_4$, and extracted with EtOAc (3 × 50 mL) The organic layers were combined, dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (2 % MeOH, 0.5 % AcOH in $CHCl_3$). A pale light yellow oil (95.0 mg, 96.5 %) was obtained. ^{1}H NMR: δ 9.80 (br s, 1H), 6.23 (m, 1H), 5.67 (dt, *J* = 17.0, 7.0, 1H), 5.08 (m, 2H), 4.34 (q, *J* = 6.4, 1H), 2.53 (m, 1H), 2.46 (m, 1H), 1.38 (s, 9H). ^{13}C NMR: δ 176.7, 155.7, 132.3, 119.5, 80.4, 52.9, 36.5, 28.4.

Boc-allylGly-OTMSE, 83. Boc-allylGly-OH **82** (1.53 g, 7.12 mmol) was dissolved in THF (50 mL). EDC (1.65 g, 8.59 mmol) and DMAP (90 mg, 0.74 mmol) were added, and the solution was stirred at 0 °C for 5 min. TMSE-OH (7.0 mL, 0.049 mol) was added, and the mixture was stirred at 0 °C for 6 h. The solution was diluted with EtOAc (150 mL), washed with 1 M HCl (2 × 50 mL), H_2O (50 mL), $NaHCO_3$ (2 × 50 mL), and brine (50 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (10 % EtOAc in hexanes). A colorless oil (1.96 g, 87 %) was obtained. ^{1}H NMR: δ 5.64 (ddt, *J* = 16.8, 10.3, 7.2, 1H), 5.07 (m, 3H), 4.28 (dd, *J* = 13.6, 6.2, 1H), 4.16 (m, 2H), 2.49 (m, 1H), 2.41 (m, 1H), 1.38 (s, 9H), 0.95 (m, 2H), –0.01 (s, 9H). ^{13}C NMR: δ 172.2, 155.2, 132.5, 118.9, 79.7, 63.7, 53.0, 36.8, 28.3, 17.4, –1.5.

Fmoc-homoallylGly-OH, 84. H-homoallylGly-OH **81** (206 mg, 1.59 mmol) was dissolved in saturated $NaHCO_3$ solution (8 mL), and stirred at 0 °C for 10 min. Fmoc-OSu (634 mg, 1.88 mmol) in 1,4-dioxane (3 mL) was added. The mixture was stirred at 0 °C for 14 h, then acidified to pH 1 with 2 M HCl, and extracted with EtOAc (4 × 25 mL). The resulting organic layers were combined, washed with H_2O (2 × 25 mL) and brine (25 mL), dried over Na_2SO_4, and concentrated in vacuo. The resulting pale yellow oil was dissolved in a minimal amount of CH_2Cl_2, and petrolium ether was added. A white solid (515 mg, 92.2 %) was obtained by filtration. ^{1}H NMR: δ 10.22 (br s, 1H), 7.77 (d, *J* = 7.4, 2H), 7.61 (t, *J* = 6.1, 2H), 7.40 (t, *J* = 7.2, 2H), 7.32 (t, *J* = 7.3, 2H), 6.43 (d, *J* = 7.4, 1H), 5.80 (m, 1H), 5.38 (d, *J* = 8.4, 1H), 5.05 (m, 2H), 4.44 (d, *J* = 6.9, 2H), 4.23 (t, *J* = 6.9, 1H), 2.16 (dd, *J* = 14.2, 7.0, 2H), 2.02 (m, 1H), 1.82 (dt, *J* = 14.3, 7.4, 1H). ^{13}C NMR: δ 177.5, 156.2, 143.8 (143.9), 141.4, 136.8, 127.8, 127.2, 125.2, 120.1, 116.2, 67.2, 53.4, 47.3, 31.6, 29.5. MS calcd. for $C_{21}H_{22}NO_4$ (MH^+) *m/z* = 352, found *m/z* = 352.

Boc-Gly-allylGly-OTMSE, 85. Boc-allylGly-OTMSE **83** (28 mg, 0.090 mmol) was dissolved in CH_2Cl_2 (3 mL), and TFA (0.5 mL) was added. The solution was stirred at rt for 20 min, and concentrated in vacuo. The resulting yellow oil was dissolved in DMF (3 mL). EDC (31 mg, 0.16 mmol), HOBt (23 mg, 0.15 mmol), DMAP (2.1 mg, 0.017 mmol) and DIEA (50 μL, 0.29 mmol) were added, and the resulting mixture was stirred at 0 °C for 10 min. Boc-Gly-OH (27 mg, 0.16 mmol) was added, and the solution was stirred at 0 °C for 6 h. The solution was diluted with EtOAc (35 mL), washed with 1 M HCl (2 × 25 mL), H_2O (25 mL), $NaHCO_3$ (2 × 25 mL) and brine (25 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (25 % EtOAc in hexanes). A yellow oil was obtained (31.3 mg, 93 %). ^{1}H NMR: δ 6.70 (d, *J* =6.6, 1H), 5.66 (m, 1H), 5.24 (t, *J* = 5.7,

1H), 5.11 (m, 1H), 5.08 (m, 1H), 4.61 (m, 1H), 4.20 (ddd, *J* = 10.6, 5.8, 2.6, 2H), 3.79 (dt, *J* = 16.8, 5.6, 2H), 2.56 (m, 1H), 2.50 (m, 1H), 1.43 (s, 9H), 0.99 (m, 2H), 0.02 (s, 9H). ^{13}C NMR: δ 171.6, 169.2, 156.1, 132.1, 119.4, 80.3, 64.1, 51.7, 44.4, 36.5, 28.4, 17.5, –1.4. MS calcd. for $C_{17}H_{33}N_2O_5Si$ (MH^+) *m/z* = 373.2, found *m/z* = 373.4.

Fmoc-homoallylGly-Gly-allylGly-OTMSE, 86. Boc-Gly-allylGly-OTMSE **85** (138 mg, 0.372 mmol) was dissolved in CH_2Cl_2 (10 mL), and TFA (2 mL) was added. The solution was stirred at rt for 40 min, and concentrated in vacuo. The resulting dark yellow oil was dissolved in DMF (10 mL). HOBt (127 mg, 0.831 mmol), EDC (145 mg, 0.755 mmol), DMAP (9.2 mg, 0.075 mmol) and DIEA (0.20 mL, 1.1 mmol) were added. The solution was stirred at 0 °C for 20 min, and Fmoc-homoallylGly-OH (320 mg, 0.911 mmol) was added. The solution was stirred at rt for 8 h, diluted with EtOAc (100 mL), washed with 1 M HCl (2 × 50 mL), H_2O (50 mL), $NaHCO_3$ (50 mL) and brine (50 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (33 % EtOAc in hexanes). A white solid was obtained (78.6 mg, 35 %). ^{1}H NMR: δ 7.75 (d, *J* = 7.4, 2H), 7.57 (d, *J* = 7.5, 2H), 7.39 (t, *J* = 7.5, 2H), 7.29 (dt, *J* = 7.4, 1.0, 2H), 6.91 (m, 1H), 6.80 (m, 1H), 5.78 (m, 1H), 5.65 (m, 1H), 5.55 (d, *J* = 7.4, 1H), 5.09 (m, 2H), 5.01 (m, 2H), 4.61 (dt, *J* = 6.4, 6.4, 1H), 4.41 (m, 2H), 4.20 (m, 4H), 4.04 (m, 1H), 3.94 (m, 1H), 2.56 (m, 1H), 2.49 (m, 1H), 2.11 (m, 2H), 1.95 (m, 1H), 1.75 (m, 1H), 0.99 (m, 2H), 0.02 (s, 9H). ^{13}C NMR: δ 172.3, 171.6, 168.3, 156.4, 143.9, 141.4, 137.1, 132.2, 127.9, 127.2, 125.2, 120.1, 119.4, 116.1, 67.2, 64.2, 54.7, 52.0, 47.3, 43.1, 36.4, 31.7, 29.7, 17.5, –1.4.

Fmoc-Phe-OTMSE, 87. Fmoc-Phe-OH (513 mg, 1.32 mmol) was dissolved in THF (50 mL). EDC (501 mg, 2.61 mmol) and DMAP (16 mg, 0.13 mmol) were added. The solution was stirred at 0 °C for 30 min, and TMSE-OH (0.90 mL, 6.3 mmol) was added. The

resulting solution was stirred at rt for 6 h, diluted with EtOAc (50 mL), washed with 1 M HCl (2 × 25 mL), H_2O (25 mL), $NaHCO_3$ (2 × 25 mL) and brine (30 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (10 % EtOAc in hexanes). A colorless oil (453 mg, 71 %) was obtained. ^{1}H NMR: δ 7.76 (d, *J* = 7.6, 2H), 7.56 (t, *J* = 7.8, 2H), 7.40 (dt, *J* = 7.5, 0.5, 2H), 7.30 (m, 5H), 7.11 (d, *J* = 6.9, 2H), 5.27 (d, *J* = 8.3, 1H), 4.64 (dt, *J* = 8.2, 6.0, 1H), 4.43 (dd, *J* = 10.6, 7.2, 1H), 4.33 (dd, *J* = 10.8, 7.1, 1H), 4.20 (m, 3H), 3.15 (dd, *J* = 13.6, 5.8, 1H), 3.09 (dd, *J* = 13.6, 6.1, 1H), 0.97 (dd, *J* = 13.2, 6.6, 1H), 0.97 (t, *J* = 7.3, 1H), 0.04 (s, 9H). ^{13}C NMR: δ 171.7, 155.7, 143.9 (140.0), 141.5, 136.0, 129.5, 128.7, 127.8, 127.22, 127.18, 125.3 (125.2), 120.10 (120.12), 67.1, 64.1, 55.0, 47.3, 38.4, 17.5, –1.4.

Cyclic tripeptide TMSE ester, 88. Fmoc-homoallylGly-Gly-allylGly-OTMSE **86** (19 mg, 0.031 mmol) was dissolved in CH_2Cl_2 (60 mL). Grubbs Catalyst Generation I, $PhCH_2{=}RuCl_2(PCy_3)_2$, (4.0 mg, 0.0049 mmol) was added. The solution was heated at reflux for 36 h, and fresh catalyst was added every 12 h. The reaction was quenched with H_2O (1 mL), and the solvent was removed in vacuo. The resulting dark solid was purified by preparative TLC (50 % EtOAc in hexanes), and a white solid (6.2 mg, 38 %) was obtained. ^{1}H NMR: δ 7.79 (d, *J* = 7.5, 2H), 7.58 (d, *J* = 6.9, 2H), 7.43 (t, *J* = 7.8, 2H), 7.33 (m, 2H), 6.84 (m, 1H), 6.69 (m, 1H), 5.42 (m, 1H), 5.30 (m, 0.5H), 5.21 (m, 1.5H), 4.52 (m, 1H), 4.45 (m, 2H), 4.23 (m, 3H), 4.01 (m, 1H), 3.66 (m, 2H), 2.54 (m, 1H), 2.40 (m, 1H), 2.09 (m, 2H), 1.97 (m, 1H), 1.87 (m, 1H), 1.01 (m, 2H), 0.05 (s, 9H). HRMS calcd. for $C_{31}H_{40}N_3O_6Si$ (MH^+) *m/z* = 578.2686, found *m/z* = 578.2662.

Cyclic tripeptide mimic, 89. Cyclic tripeptide TMSE ester **88** (6.2 mg, 0.010 mmol) was dissolved in CH_2Cl_2 (5.0 mL), and TFA (1.0 mL) was added. The solution was stirred at rt

for 3 h, concentrated in vacuo, washed with $CHCl_3$, and dried in vacuo. The resulting brown oil was purified by preparative TLC (2 % MeOH, 0.5 % AcOH in $CHCl_3$), and a yellow solid (4.2 mg, 86 %) was obtained. ^{1}H NMR (DMSO-d_6, 400 MHz): δ 8.56 (m, 1H), 8.58 (t, *J* = 6.2, 0.5H), 7.89 (d, *J* = 7.4, 2H), 7.75 (t, *J* = 6.9, 2H), 7.65 (d, *J* =8.1, 1H), 7.42 (t, *J* = 7.4, 2H), 7.31 (t, *J* = 7.3, 2H), 7.06 (d, *J* = 7.1, 1H), 5.31 (m, 1H), 5.14 (m, 1H), 4.25 (m, 3H), 4.10 (m, 2H), 3.61 (m, 2H), 3.12 (dd, *J* = 15.0, 7.2, 1H), 2.32 (m, 2H), 1.98 (m, 2H), 1.83 (m, 2H). ^{13}C NMR (DMSO-d_6, 100 MHz): δ 172.0, 169.1, 168.9, 155.7, 143.9 (143.8), 140.7, 136.2, 127.7, 127.1, 125.4, 124.0, 120.1, 65.8, 53.7, 53.4, 46.7, 43.7, 34.3, 31.8, 29.0, 27.4. MS calcd. for $C_{26}H_{28}N_3O_6$ (MH^+) *m/z* = 478.5, found *m/z* = 478.6.

H-homoallylGly-OMe•HCl, 90. H-homoallylGly-OH **81** (101 mg, 0.800 mmol) was dissolved in MeOH (5 mL), and stirred at 0 °C for 10 min. $SOCl_2$ (65 μL, 0.89 mmol) was added dropwise. The solution was stirred at rt for 8 h, and concentrated in vacuo. A light yellow oil (112 mg, 98.0 %) was obtained. ^{1}H NMR (CD_3OD): δ 5.84 (m, 1H), 5.13 (d, *J* = 17.3, 1H), 5.07 (d, *J* = 10.2, 1H), 4.05 (t, *J* = 5.7, 1H), 3.84 (s, 3H), 2.23 (m, 2H), 2.00 (m, 2H). ^{13}C NMR (CD_3OD): δ 171.7 (170.8), 137.2 (137.4), 116.8 (116.9), 53.7 (53.5), 53.4 (53.3), 30.77 (30.83), 30.0 (20.9).

Boc-Gly-homoallylGly-OMe, 91. H-homoallylGly-OMe **90** (112 mg, 0.784 mmol) was dissolved in DMF (5 mL). EDC (312 mg, 1.63 mmol), HOBt (245 mg, 1.60 mmol), DMAP (17 mg, 0.14 mmol) and DIEA (1.00 mL, 5.72 mmol) were added, and the solution was stirred at 0 °C for 10 min. Boc-Gly-OH (283 mg, 1.62 mmol) was added. The resulting solution was stirred at 0 °C for 6 h, then diluted with EtOAc (30 mL), washed with 1 M HCl (2 × 20 mL), H_2O (20 mL), $NaHCO_3$ (20 mL), and brine (25 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (25 % EtOAc in hexanes). A

colorless oil was obtained (109 mg, 46 %). ^{1}H NMR: δ 6.89 (app. d, *J* = 5.7, 1H), 5.72 (ddt, *J* = 17.0, 10.3, 6.6, 1H), 5.41 (t, *J* = 5.6, 1H), 5.00 (dq, *J* = 17.3, 1.6, 1H), 4.96 (dd, *J* = 10.1, 1.3, 1H), 4.58 (dd, *J* = 13.0, 7.6, 1H), 3.79 (m, 2H), 3.70 (s, 3H), 2.04 (m, 2H), 1.92 (m, 1H), 1.75 (m, 1H), 1.41 (s, 9H). ^{13}C NMR: δ 172.7, 169.6, 156.2, 136.8, 115.9, 80.2, 52.4, 51.7, 44.3, 31.5, 29.4, 28.3.

Fmoc-allylGly-OH, 92. H-allylGly-OH **80** (296 mg, 2.57 mmol) was dissolved in $NaHCO_3$ (10 mL), and stirred at 0 °C for 10 min. Fmoc-OSu (1.08 g, 3.19 mmol) in 1,4-dioxane (8 mL) was added dropwise. The solution was stirred at rt for 10 h, then acidified to pH 1 with 3 M HCl. The aqueous solution was extracted with EtOAc (4 × 30 mL). The resulting organic layers were combined, washed with H_2O (2 × 20 mL) and brine (15 mL), dried over Na_2SO_4, and concentrated in vacuo. The resulting yellow oil was dissolved in a minimal amount of CH_2Cl_2, and petrolium ether was added. A white solid (843 mg, 97 %) was collected by filtration. ^{1}H NMR: δ 7.76 (d, *J* =, 7.6, 2H), 7.58 (m, 2H), 7.40 (t, *J* = 7.5, 2H), 7.31 (t, *J* = 7.4, 2H), 5.74 (m, 1H), 5.30 (d, *J* = 7.3, 1H), 5.19 (d, *J* = 13.2, 2H), 4.49 (app. dd, *J* = 11.7, 6.1, 1H), 4.41 (d, *J* = 6.9, 2H), 4.23 (t, *J* = 7.0, 1H), 2.65 (m, 1H), 2.57 (m, 1H). ^{13}C NMR: δ 168.7, 156.6, 143.9, 141.5 (141.4), 133.5, 127.8, 127.6, 127.2, 125.4, 120.1, 67.2, 47.2, 46.6, 25.6.

Fmoc-allylGly-Gly-homoallylGly-OMe, 93. Boc-Gly-homoallylGly-OMe **91** (392 mg, 1.30 mmol) was dissolved in CH_2Cl_2 (10 mL), and TFA (5 mL) was added. The solution was stirred at rt for 1 h, and the solvent and excess TFA was removed in vacuo. The resulting dark yellow oil was dissolved in DMF (15 mL). HOBt (305 mg, 1.99 mmol), EDC (383 mg, 2.00 mmol), DMAP (21 mg, 0.17 mmol) and DIEA (0.90 mL, 5.2 mmol) were added, and the solution was stirred at 0 °C for 20 min. Fmoc-allylGly-OH **92** (550 mg, 1.65 mmol) was

added. The resulting solution was stirred at rt for 8 h, then diluted with EtOAc (75 mL), washed with 1 M HCl (2 × 50 mL), H_2O (50 mL), $NaHCO_3$ (50 mL) and brine (50 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (50 % EtOAc in hexanes). A white solid was obtained (515 mg, 76 %). ^{1}H NMR: δ 7.76 (dt, *J* = 7.6, 0.9, 2H), 7.56 (d, *J* = 7.6, 2H), 7.40 (t, *J* = 7.5, 2H), 7.30 (dtt, *J* = 7.4, 2.2, 1.1, 2H), 6.87 (m, 2H), 5.73 (dtt, *J* = 17.0, 10.2, 6.6, 2H), 5.40 (m, 1H), 5.18 (m, 1H), 5.15(m, 1H), 5.00 (m, 2H), 4.58 (ddt, *J* = 8.0, 5.0, 3.0, 1H), 4.43 (m, 2H), 4.20 (dd, *J* = 11.8, 6.6, 2H), 4.07 (m, 1H), 3.90 (dd, *J* = 16.7, 5.5, 1H), 3.70 (s, 1H), 3.69 (s, 2H), 2.58 (m, 1H), 2.50 (m, 1H), 2.08 (m, 2H), 1.93 (m, 1H), 1.77 (m, 1H). ^{13}C NMR: δ 172.8, 172.05 (172.00), 168.9 (168.8), 156.5 (156.4), 143.7 (143.8), 141.3, 136.72 (136.71), 132.8, 127.7, 127.1, 125.1, 120.0, 119.23 (119.15), 115.90 (115.86), 67.2, 54.6 (54.5), 52.37 (52.39), 51.87 (51.88), 47.10 (47.08), 43.06 (43.11), 36.9 (36.8), 31.25 (31.21), 29.50 (29.48).

H-allylGly-OMe, 94. H-allylGly-OH **80** (206 mg, 1.79 mmol) was dissolved in MeOH (5 mL), and stirred at 0 °C for 10 min. $SOCl_2$ (0.20 mL, 2.7 mmol) was added dropwise. The solution was stirred at rt for 8 h, then concentrated and dried in vacuo. A pale yellow oil (281 mg, 94.9 %) was obtained. ^{1}H NMR: δ 8.66 (br s, 2H), 5.85 (dd, *J* = 14.5, 7.8, 1H), 5.32 (d, *J* = 16.6, 1H), 5.24 (d, *J* = 9.9, 1H), 4.26 (m, 1H), 3.80 (s, 3H), 2.84 (m, 2H). ^{13}C NMR: δ 169.5, 130.3, 121.6, 53.4, 53.1, 34.6.

Boc-Gly-allylGly-OMe, 95. H-allylGly-OMe **94** (280 mg, 1.69 mmol) was dissolved in DMF (10 mL), and EDC (980 mg, 5.11 mmol), HOBt (785 mg, 5.13 mmol), DMAP (26 mg, 0.21 mmol) and DIEA (3.00 mL, 17.2 mmol) were added. The solution was stirred at 0 °C for 10 min, and Boc-Gly-OH (893 mg, 5.10 mmol) was added. The resulting solution was stirred at rt for 6 h, then diluted with EtOAc (50 mL), washed with 1 M HCl (2 × 25 mL),

H_2O (25 mL), $NaHCO_3$ (25 mL) and brine (25 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (33 % EtOAc in hexanes). A clear oil was obtained (300 mg, 62.1 %). ^{1}H NMR: δ 6.77 (m, 1H), 5.64 (m, 1H), 5.32 (m, 1H), 5.09 (app. d, *J* = 13.2, 2H), 4.64 (dd, *J* = 12.5, 6.0, 1H), 3.79 (m, 2H), 3.71 (s, 3H), 2.55 (m, 1H), 2.48 (m, 1H), 1.42 (s, 9H). ^{13}C NMR: δ 172.0, 169.4, 156.1, 132.1, 119.4, 80.3, 52.5, 51.7, 44.3, 36.4, 28.4.

Fmoc-homoallylGly-Gly-allylGly-OMe, 96. Boc-Gly-allylGly-OMe **95** (300 mg, 1.05 mmol) was dissolved in CH_2Cl_2 (10 mL), and TFA (5 mL) was added. The solution was stirred at rt for 1 h. The solvent and excess TFA was removed in vacuo. The resulting dark yellow oil was dissolved in DMF (15 mL). HOBt (231 mg, 1.51 mmol), EDC (308 mg, 1.61 mmol), DMAP (24 mg, 0.19 mmol) and DIEA (0.75 mL, 4.3 mmol) were added. The solution was stirred at 0 °C for 20 min, and Fmoc-homoallylGly-OH **84** (480 mg, 1.36 mmol) was added. The resulting solution was stirred at rt for 8 h, then diluted with EtOAc (75 mL), washed with 1 M HCl (2 × 50 mL), H_2O (50 mL), $NaHCO_3$ (50 mL) and brine (50 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (50 % EtOAc in hexanes). A white solid was obtained (285 mg, 55 %). ^{1}H NMR: 7.74 (d, *J* = 7.6, 2H), 7.57 (dd, *J* = 11.8, 7.2, 2H), 7.38 (t, *J* = 7.4, 2H), 7.28 (dt, *J* = 7.5, 0.8, 2H), 7.12 (m, 1H), 7.03 (dd, *J* = 17.7, 7.8, 1H), 5.73 (m, 2H), 5.64 (m, 1H), 5.05 (m, 3H), 4.99 (d, *J* = 9.9, 1H), 4.64 (dd, *J* = 14.9, 6.8, 1H), 4.38 (m, 2H), 4.20 (m, 2H), 4.06 (dt, *J* = 17.0, 5.2, 1H), 3.95 (m, 1H), 3.68 (s, 2H), 3.67 (s, 1H), 2.55 (m, 1H), 2.46 (m, 1H), 2.10 (m, 2H), 1.93 (m, 1H), 1.74 (m, 1H). ^{13}C NMR: δ 172.5 (172.4), 172.02 (172.05), 168.6, 156.4 (156.5), 143.8 (143.9), 141.4, 137.07 (137.05), 132.1 (132.2), 127.9, 127.2, 125.1,

120.1, 119.3 (119.4), 116.01 (116.03), 67.2, 54.7 (54.8), 52.50 (52.49), 51.97 (52.03), 47.3, 43.2 (43.1), 36.36 (36.40), 31.7 (31.8), 29.7.

Cyclic tripeptide methyl ester, 97. Grubbs Catalyst Generation I, $PhCH_2=RuCl_2(PCy_3)_2$, (26 mg, 0.031 mmol) was dissolved in CH_2Cl_2 (250 mL). Fmoc-allylGly-Gly-homoallylGly-OMe **93** (103 mg, 0.198 mmol) was dissolved in CH_2Cl_2 (10 mL), and added dropwise within 40 min. The resulting solution was heated at reflux for 36 h, and fresh catalyst was added every 12 h. The reaction was quenched with H_2O (1 mL), and the solvent was removed in vacuo. The resulting dark brown solid was purified by preparative TLC (67 % EtOAc in hexanes), and a white solid (12 mg, 40 %) was obtained. ^{1}H NMR (DMSO-d_6): δ 8.61 (t, *J* = 5.8,1H) 7.89 (d, *J* = 7.5, 2H), 7.74 (m, 2.5H), 7.42 (t, *J* = 7.4, 2H), 7.34 (m, 2H), 7.18 (d, *J* = 6.8, 0.5H), 5.58 (m, 1H), 5.47 (m, 1H), 4.29 (m, 2H), 4.22 (m, 2H), 4.10 (m, 1H), 3.94 (m, 1H), 3.64 (s, 3H), 3.42 (m, 1H), 2.29 (m, 2H), 2.12 (m, 1H), 1.98 (m, 1H), 1.80 (m, 2H). ^{13}C NMR (DMSO-d_6): δ 171.8 (172.8, 172.3), 171.7 (171.92, 171.86), 169.9 (168.7, 168.2), 155.9 (155.6, 155.5), 143.75 (143.85, 143.88), 140.7 (143.82, 143.94), 136.8, 136.0, 127.7, 127.1, 125.3 (125.4, 125.5), 120.1, 65.72 (65.67), 54.3 (54.0), 53.4 (52.5), 52.0 (52.2, 51.9), 46.68 (46.64, 46.62), 42.8 (43.7, 42.2), 35.8 (36.1), 28.9 (28.5), 28.7 (28.0, 27.2). Mixture of three isomers: 11:5:4.

Fmoc-Phe-OMe, 98. H-Phe-OMe•HCl (140 mg, 0.648 mmol) was dissolved in CH_2Cl_2 (10 mL). Et_3N (1.50 mL, 1.08 mmol) and Fmoc-Cl (192 mg, 0.743 mmol) were added. The solution was stirred at rt for 4 h, then diluted with CH_2Cl_2 (75 mL), washed with H_2O (40 mL), 1 M HCl (2 × 40 mL) and brine (50 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (25 % EtOAc in hexanes). A light yellow oil (237 mg, 91 %) was obtained. ^{1}H NMR (400 MHz): δ 7.75 (d, *J* = 7.8, 2H), 7.55 (t, *J* =6.9, 2H), 7.38

(t, *J* = 7.4, 2H), 7.29 (m, 5H), 7.08 (d, *J* = 7.6, 2H), 5.32 (d, *J* = 8.2, 1H), 4.67 (dd, *J* = 13.9, 6.0, 1H), 4.43 (ddd, *J* = 10.8, 7.2, 1.4, 1H), 4.32 (m, 1H), 4.19 (t, *J* = 7.0, 1H), 3.70 (s, 3H), 3.14 (dd, *J* = 13.8, 5.8, 1H), 3.07 (dd, *J* = 13.9, 6.2, 1H). ^{13}C NMR (100 MHz): δ 172.0, 155.6, 143.8 (143.9), 141.4, 135.8, 129.4, 128.7, 127.8, 127.2, 127.1, 125.2 (125.1), 120.1, 67.0, 54.9, 52.4, 47.2, 38.3.

Fmoc-Ile-OMe, 99. Fmoc-Ile-OH (310 mg, 0.876 mmol) was dissolved in THF (15 mL). EDC (334 mg, 1.74 mmol) and DMAP (16 mg, 0.14 mmol) were added, and the solution was stirred at 0 °C for 30 min. Excess MeOH (6 mL) was added. The solution was stirred at rt for 6 h, then diluted with EtOAc (100 mL), washed with 1 M HCl (2 × 50 mL), H_2O (50 mL), $NaHCO_3$ (2 × 50 mL) and brine (50 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (10 % EtOAc in hexanes). A clear oil (236 mg, 73 %) was obtained. ^{1}H NMR: δ 7.75 (d, *J* = 7.3, 2H), 7.60 (dd, *J* = 7.2, 3.6, 2H), 7.39 (t, *J* =7.4, 2H), 7.31 (dt, *J* = 7.4, 1.2, 2H), 5.34 (d, *J* = 8.7, 1H), 4.33-4.40 (d, *J* = 7.1, 2H), 4.35 (dd, *J* = 9.2, 5.0, 1H), 4.22 (t, *J* = 7.2, 1H), 3.74 (s, 3H), 1.89 (m, 1H), 1.43 (m, 2H), 0.91 (m, 6H). ^{13}C NMR: δ 172.7, 156.2, 143.9 (144.1), 141.4, 127.8, 127.2, 125.2, 120.08 (120.10), 67.1, 58.4, 52.2, 47.3, 38.2, 25.2, 15.6, 11.7.

Cyclic tripeptide mimic, 100. Cyclic tripeptide methyl ester **97** (18 mg, 0.038 mmol) was dissolved in CH_2Cl_2 (6 mL), and pre-mixed $AlCl_3$/DMA solution (5:8, 70 μL) was added. The resulting blue solution was heated at reflux for 24 h, and the reaction was quenched with 1 M HCl (30 mL). The aqueous layer was extracted with CH_2Cl_2 (3 × 25 mL). The organic layers were combined, washed with brine (30 mL), dried over Na_2SO_4, concentrated in vacuo, and purified by flash chromatography (2 % MeOH, 0.5 % AcOH in $CHCl_3$). A pale yellow solid (10.1 mg, 57 %) was obtained. ^{1}H NMR (DMSO-d_6): δ 12.80 (br s, 1H), 8.30

(dd, *J* = 7.4, 5.0, 0.5H), 8.58 (t, *J* = 6.2, 0.5H), 7.89 (dd, *J* = 7.4, 3.8, 2H), 7.76 (m, 3H), 7.42 (dt, *J* = 7.5, 3.6, 2H), 7.33 (m, 2H), 7.23 (d, *J* = 8.5, 0.5H), 7.11 (d, *J* = 7.6, 0.5H), 5.58 (m, 0.5H), 5.52 (dd, *J* = 9.8, 4.0, 0.5H), 5.46 (m, 1H), 4.42 (app. dt, *J* = 5.0, 3.6, 1H), 4.23 (m, 3H), 4.09 (m, 1H), 3.98 (m, 0.5H), 3.84 (dd, *J* = 15.6, 6.8, 0.5H), 3.46 (dd, *J* = 15.6, 5.5, 0.5H), 3.36 (dd, *J* = 16.8, 5.0, 0.5H), 2.31 (m, 2H), 2.11 (m, 1H), 1.98 (m, 1H), 1.83 (m, 1.5H), 1.73 (m, 0.5H). ^{13}C NMR (DMSO-d_6): δ 172.8 (172.9), 171.9 (171.6), 167.9 (168.4), 155.5 (155.8), 143.8 (143.7), 140.7, 137.0, 136.3, 127.7, 127.1, 125.3 (125.4), 120.2 (120.0), 65.8 (65.7), 54.4, 53.2 (52.3), 46.6, 42.8, 36.2 (35.8), 30.8 (30.6), 28.5 (28.8). MS calcd. for $C_{26}H_{28}N_3O_6$ (MH^+) *m/z* = 478.5, found *m/z* = 478.7. Mixture of three isomers: 5:5:2.

Fmoc-homoallylGly-OBn, 101. Fmoc-homoallylGly-OH **84** (71 mg, 0.20 mmol) and BnOH (21 μL, 0.20 mmol) was dissolved in 1,4-dioxane (1.5 mL) and HCl in ether (2.0 M, 0.2 mL, 0.40 mmol) was added. The solution was heated at reflux for 14 h, then concentrated in vacuo, diluted with EtOAc (15 mL), washed with 1 M HCl (2 × 15 mL), $NaHCO_3$ (2×15 mL), and brine (15 mL), dried over Na_2SO_4, and purified by flash chromatography (20 % EtOAc in hexanes). A white solid (77 mg, 86 %) was obtained. ^{1}H NMR: δ 7.77 (d, *J* = 7.6, 2H), 7.60 (d, *J* = 7.5, 2H), 7.37 (m, 9H), 5.76 (ddt, *J* = 16.9, 10.3, 6.6, 1H), 5.33 (d, *J* = 8.3, 1H), 5.22 (d, *J* = 12.4, 1H), 5.17 (d, *J* = 12.3, 1H), 5.00 (m, 2H), 4.45 (m, 1H), 4.41 (d, *J* = 7.1, 2H), 4.23 (t, *J* = 7.0, 1H), 2.07 (m, 2H), 1.98 (m, 1H), 1.78 (m, 1H). ^{13}C NMR: δ 172.4, 156.0, 144.1 (143.9), 141.5, 137.0, 135.4, 133.2, 129.9, 128.8, 128.7, 128.5, 127.9, 127.2, 125.2, 120.1, 116.0, 67.4, 67.1, 53.7, 47.3, 32.0, 29.4.

Boc-Gly-homoallylGly-OBn, 102. Fmoc-homoallylGly-OBn **101** (71 mg, 0.20 mmol) was dissolved in THF (3 mL), and piperidine (0.5 mL) was added. The solution was stirred at rt for 30 min, then concentrated in vacuo. The resulting pale yellow solid was dissolved in

CH_2Cl_2 (3.0 mL) and stirred at 0 °C for 10 min. Boc-Gly-OH (142 mg, 0.810 mmol), HOBt (124 mg, 0.810 mmol), EDC (155 mg, 0.809 mmol) and DIEA (0.30 mL, 1.7 mmol) were added. The resulting solution was stirred at rt 8 h, then concentrated in vacuo, diluted with EtOAc (20 mL), washed with 1 M HCl (2 × 15 mL), $NaHCO_3$ (3 × 15 mL), H_2O (2 × 10 mL), and brine (15 mL), concentrated in vacuo, and purified by flash chromatography (25 % EtOAc in hexanes). A pale yellow oil (48 mg, 73 %) was obtained. ^{1}H NMR: δ 7.34 (m, 5H), 6.60 (d, *J* = 7.7, 1H), 5.72 (ddt, *J* = 17.0, 10.4, 6.5, 1H), 5.57 (m, 1H), 5.19 (d, *J* = 12.1, 1H), 5.14 (d, *J* = 12.0, 1H), 4.98 (dd, *J* = 17.1, 1.6, 1H), 4.96 (d, *J* = 10.1, 1H), 4.67 (dt, *J* = 7.6, 4.7, 1H), 3.84 (dd, *J* = 16.9, 5.3, 1H), 3.78 (dd, *J* = 16.6, 5.8, 1H), 2.01 (m, 3H), 1.77 (m, 1H), 1.45 (s, 9H). ^{13}C NMR: δ 172.1, 169.3, 156.1, 136.9, 135.4, 128.8, 128.7, 128.4, 116.0, 79.6, 67.3, 51.9, 45.5, 31.7, 28.4, 24.5. MS calcd. for $C_{20}H_{29}N_2O_5$ (MH^+) *m/z* = 377, found *m/z* = 377.

Fmoc-allylGly-Gly-homoallylGly-OBn, 103. Boc-Gly-homoallylGly-OBn **102** (110 mg, 0.249 mmol) was dissolved in CH_2Cl_2 (10 mL), and TFA (5 mL) was added. The solution was stirred at rt for 40 min. The solvent and excess was removed in vacuo. The resulting dark yellow oil was dissolved in DMF (8 mL), and stirred at 0 °C for 20 min. Fmoc-allylGly-OH **92** (120 mg, 0.355 mmol), HOBt (57 mg, 0.37 mmol), HBTU (145 mg, 0.382 mmol) and DIEA (0.25 mL, 1.4 mmol) were added. The solution was stirred at 0 °C for 2 h, diluted with EtOAc (30 mL), washed with 1 M HCl (2 × 20 mL), $NaHCO_3$ (3 × 15 mL), H_2O (2 × 20 mL) and brine (25 mL), concentrated in vacuo, and purified by flash chromatography (25 %, then 50 % EtOAc in hexanes). A white solid (70 mg, 44 %) was obtained. ^{1}H NMR: δ 7.75 (d, *J* = 7.9, 2H), 7.57 (m, 2H), 7.39 (t, *J* = 7.4, 2H), 7.31 (m, 7H), 7.20 (br s, 1H), 6.95 (app. t, *J* = 7.3, 1H), 5.70 (m, 2H), 5.52 (d, *J* = 7.5, 0.5H), 5.47 (dd, *J* = 16.9, 6.3, 0.5H), 5.14 (m, 2H),

5.09 (dd, *J* = 12.5, 4.4, 1H), 4.93 (m, 1H), 4.62 (m, 1H), 4.39 (m, 2H), 4.20 (dt, *J* = 13.8, 6.8, 2H), 4.07 (m, 2H), 3.89 (app. dt, *J* = 16.6, 5.3, 1H), 2.58 (m, 1H), 2.49 (m, 1H), 2.02 (m, 1H), 1.94 (m, 1H), 1.75 (m, 1H), 1.64 (m, 1H). ^{13}C NMR: δ 172.2, 171.8, 168.7, 156.4, 143.8, 141.4, 136.8, 132.8, 128.7, 128.6, 128.4, 127.9, 127.8, 127.2, 125.3 (125.2), 120.14, 120.09, 116.0, 68.1, 52.1, 47.2, 45.6, 43.2 (43.1), 41.3, 37.4, 31.4, 29.5.

Chapter 7 Conclusion and future work

7.1 Collagen-like polypeptides

7.1.1 Conclusion

The conformationally locked alkene isostere Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OH was synthesized. In this alkene isostere, one amide bond was locked in the *trans*-conformation with an (*E*)-alkene bond to prevent *cis*-*trans* isomerization, and the C=O of Pro and N-H of Gly were left intact to maintain the interchain hydrogen bonding. The synthesis of the Gly-*trans*-Pro alkene isostere had no stereochemical control, and the two diastereomers of the tripeptide isostere Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OBn were separated by normal phase HPLC.

The tripeptide alkene isostere Fmoc-Gly-Ψ[(*E*)CH=C]-Pro-Hyp(^{t}Bu)-OH was incorporated into a host-guest system: Ac-(Gly-Pro-Hyp)$_3$-Gly-Xaa-Yaa-(Gly-Pro-Hyp)$_4$-Gly-Gly-Tyr-NH_2. The Gly-*trans*-Pro alkene isostere locked one amide bond to the *trans*-conformation, and eliminated *cis*-*trans* isomerization, which could lower the entropic energy barrier during the collagen folding process. We expected this substitution to lead to an overall stabilization of the collagen triple helix. CD analysis showed that the peptide with Gly-Ψ[(*E*)CH=C]-L-Pro-Hyp as the guest triplet formed a stable triple helix, but the T_m value was 28.3 °C, which was 21.7 °C lower than its control peptide (T_m = 50.0 °C).[487]

To study to what degree *cis*-*trans* isomerization affects the stability of the collagen triple helix, another conformationally locked alkene isostere Fmoc-Pro-Ψ[(*E*)CH=C]-Pro-OH was synthesized. This Pro-*trans*-Pro alkene isostere locked one amide bond in the *trans*-

conformation, but also eliminated one interchain hydrogen bond by removing the C=O of Pro at the Xaa position. Although the stereoselectivity of the asymmetric reduction was not good for the Pro-*trans*-Pro alkene isostere, the resulting diastereomers could be separated by flash chromatography, and the absolute stereochemistry of the two diastereomers was determined by Mosher's method.

The Pro-*trans*-Pro alkene isostere was incorporated into another host-guest peptide H-(Pro-Pro-Gly)$_4$-Xaa-Yaa-Gly-(Pro-Pro-Gly)$_5$-OH. The T_m value of the Pro-Pro alkenyl peptide was determined to be –22.0 °C, a 54.6 °C decrease compared to its control peptide. By comparing this result with the Gly-*trans*-Pro and the published Pro-*trans*-Gly[223] alkene isosteres, we found that replacing any amide with an (*E*)-alkene bond caused a significant stability decrease in the resulting collagen triple helix. *Cis*-*trans* isomerization has a limited contribution to the triple helix stability. The backbone interchain hydrogen bond between the Gly N-H and the Xaa C=O the adjacent chain is one of the major forces in stabilizing the collagen triple helix. The intrinsic properties of the amide bond also have a significant influence on the stability of the collagen triple helix.

7.1.2 Future work

Previous studies showed that *cis*-*trans* isomerization was the rate-limiting step during the collagen propagation step; while our studies showed that *cis*-*trans* isomerization only had a limited contribution to the triple helix stability. We hypothesize that the amide bond, especially the carbonyl group of the amide bond has a significant contribution to the triple helix stability. New alkene isosteres will be designed to maintain the dipole moment and n-to-π* interactions of the amide bonds, which are two factors that might be responsible for the destabilizing effect when the amide bond was replaced with an (*E*)-alkene bond. A ketone

isostere will also be designed to study how much a carbonyl group can contribute to the stability of the collagen triple helix, although the geometry of this isostere is not expected to conform to the triple helix geometry.

When an alkenyl peptide with a similar (or higher) T_m value to the control peptide is obtained, a homogenous peptide with only alkene isosteres will be synthesized. This homogenous peptide can provide a better model to study the stability and interactions of alkene isosteres. More structural details about hydrogen bonding patterns, hydration networks, and other stabilizing effects of the homogenous alkenyl peptide can be obtained if the single crystal structure is obtained. This can greatly help us to understand the stabilizing factors of the collagen triple helix, and improve our ability to design more stable collagen-like polypeptides.

Our ultimate goal is to design and synthesize a synthetic collagen-like biomaterial, which has better stability than natural collagens, and does not have the drawbacks of natural collagens that are used currently in the clinic. To achieve this goal, a better polymerization method will be necessary. The polymerization conditions depend on the properties and stability of the tripeptide alkene isosteres, as well as the length and properties of the final collagen-like polymers. This requires a knowledge of polymer chemistry, material science and engineering, even biological and medicinal science.

7.2 Helix-turn-helix peptides and turn mimetics

7.2.1 Conclusion

Stepwise modifications of the Antennapedia HTH peptide (27-55) were performed to improve the helicity. The peptide with more side-chain ion-pairs was 4.28 times more helical

than the native Antp peptide, while the Ala-based peptide was over 9.27 times more helical than the native peptide.

The Fmoc-protected cyclic tripeptide HTH-turn mimic was synthesized in a form ready for solid phase peptide synthesis. The solubility of the cyclic tripeptide was very poor, and purification of the final product was very difficult. The solubility problem might also affect solid phase peptide synthesis in future.

7.2.2 Future work

DNA binding affinity tests will be performed for the Antp HTH peptides. CE or gel-electrophoresis will be used to obtain the equilibrium dissociation constant (K_d), which is used to evaluate the binding affinity of a peptide to DNA. If the Antp HTH peptides are still available in sufficient quantity, CD analysis will be performed again, so that the percentage helicity can be calculated to evaluate the helicity of these peptides quantitatively.

Due to the poor solubility of the cyclic tripeptide HTH-turn mimic, the synthesis of the HTH peptide with a cyclic turn mimic was not successful (Dr. Tao Liu, unpublished). On-resin click cyclization[488,489] and ring-closing metathesis[490-492] will be used in the synthesis of peptides with HTH-turn mimics. New tripeptide HTH-turn mimics will be designed. The new tripeptide mimics have linear structures, which will have better solubility in many common solvents, and make it much easier for solid phase peptide synthesis. The ring-closing metathesis or click chemistry will be performed on-resin after the whole sequence is synthesized.

The control peptides for HTH-turn mimics were already purified and characterized by CD. CD analysis for the peptides with cyclic HTH-turn isosteres will be performed. DNA-binding affinity tests will be performed for the peptides with cyclic HTH-turn isosteres and

their control peptides. We hope by covalently linking the side-chains of the residues in the HTH-turn, the resulting HTH peptides can be pre-organized to increase their binding affinities to DNA.

Reference

1. Ramachandran, G. N.; Reddi, A. H., *Biochemistry of Collagen*. Plenum Press: New York and London, 1976.
2. Piez, K. A., Collagen. In *Encyclopedia of polymer science and engineering*, Mark, H. F.; Kroschwitz, J. I., Eds. Wiley: New York, 1985; pp 699-727.
3. Birk, D. E.; Silver, F. H.; Trelstad, R. L., Matrix Assembly. In *Cell Biology of Extracellular Matrix*, 2nd ed.; Hay, E. D., Ed. Plenum Press: New York, 1991.
4. Koide, T., Designed Triple-Helical Peptides as Tools for Collagen Biochemistry and Matrix Engineering. *Phil. Trans. R. Soc. B* **2007,** *362*, (1484), 1281-1291.
5. Birk, D. E.; Bruckner, P., Collagen Suprastructures. *Top. Curr. Chem.* **2005,** *247*, (Collagen), 185-205.
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* **1994,** *266*, (5182), 75-81.
7. Ramachandran, G. N.; Bansal, M.; Bhatnagar, R. S., A Hypothesis on the Role of Hydroxyproline in Stabilizing Collagen Structure. *Biochim. Biophys. Acta* **1973,** *322*, (1), 166-171.
8. Bansal, M.; Ananthanarayanan, V. S., The Role of Hydroxyproline in Collagen Folding: Conformational Energy Calculations on Oligopeptides Containing Proline and Hydroxyproline. *Biopolymers* **1988,** *27*, (2), 299-312.
9. Kucharz, E. J., *The Collagens: Biochemistry and Pathophysiology*. Springer-Verlag: Berlin, 1992.
10. Kersteen, E. A.; Raines, R. T., Contribution of Tertiary Amides to the Conformational Stability of Collagen Triple Helices. *Biopolymers* **2001,** *59*, (1), 24-28.
11. Long, C. G.; Li, M. H.; Baum, J.; Brodsky, B., Nuclear Magnetic Resonance and Circular Dichroism Studies of a Triple-Helical Peptide with a Glycine Substitution. *J. Mol. Biol.* **1992,** *225*, (1), 1-4.
12. Kramer, R. Z.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M., Sequence Dependent Conformational Variations of Collagen Triple-Helical Structure. *Nat. Struct. Biol.* **1999,** *6*, (5), 454-457.
13. Adzhubei, A. A.; Sternberg, M. J. E., Left-Handed Polyproline II Helices Commonly Occur in Globular Proteins. *J. Mol. Biol.* **1993,** *229*, (2), 472-493.
14. Siligardi, G.; Drake, A. F., The Importance of Extended Conformations and, in Particular, the PII Conformation for the Molecular Recognition of Peptides. *Biopolymers* **1995,** *37*, (4), 281-292.
15. Shi, Z.; Woody, R. W.; Kallenbach, N. R., Is Polyproline II a Major Backbone Conformation in Unfolded Proteins? *Adv. Protein Chem.* **2002,** *62*, (Unfolded Proteins), 163-240.
16. Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A.-F.; Creamer, T. P., Host-Guest Study of Left-Handed Polyproline II Helix Formation. *Biochemistry* **2001,** *40*, (48), 14376-14383.

17. Rath, A.; Davidson, A. R.; Deber, C. M., The Structure of Unstructured Regions in Peptides and Proteins: Role of the Polyproline II Helix in Protein Folding and Recognition. *Biopolymers* **2005,** *80*, (2-3), 179-185.
18. Bochicchio, B.; Tamburro, A. M., Polyproline II Structure in Proteins: Identification by Chiroptical Spectroscopies, Stability, and Functions. *Chirality* **2002,** *14*, (10), 782-792.
19. Rucker, A. L.; Creamer, T. P., Polyproline II Helical Structure in Protein Unfolded States: Lysine Peptides Revisited. *Protein Sci.* **2002,** *11*, (4), 980-985.
20. Park, S. H.; Shalongo, W.; Stellwagen, E., The Role of PII Conformations in the Calculation of Peptide Fractional Helix Content. *Protein Sci.* **1997,** *6*, (8), 1694-1700.
21. Stapley, B. J.; Creamer, T. P., A Survey of Left-Handed Polyproline II Helices. *Protein Sci.* **1999,** *8*, (3), 587-595.
22. Hopfinger, A. J., *Conformational Properties of Macromolecules*. Academic Press: New York, 1973.
23. Kleywegt, G. J.; Jones, T. A., Phi/Psi-Chology: Ramachandran Revisited. *Structure* **1996,** *4*, (12), 1395-1400.
24. Cowan, P. M.; McGavin, S.; North, A. C. T., The Polypeptide Chain Configuration of Collagen. *Nature* **1955,** *176*, (4492), 1062-1064.
25. Cowan, P. M.; McGavin, S., Structure of Poly-L-Proline. *Nature* **1955,** *176*, (4480), 501-503.
26. Burge, R. F.; Harrison, P. M.; McGavin, S., The Structure of Poly-L-Proline II. *Acta Cryst.* **1962,** *15*, (9), 914-915.
27. Blanch, E. W.; Morozova-Roche, L. A.; Cochran, D. A. E.; Doig, A. J.; Hecht, L.; Barron, L. D., Is Polyproline II Helix the Killer Conformation? A Raman Optical Activity Study of the Amyloidogenic Prefibrillar Intermediate of Human Lysozyme. *J. Mol. Biol.* **2000,** *301*, (2), 553-563.
28. Gill, A. C.; Ritchie, M. A.; Hunt, L. G.; Steane, S. E.; Davies, K. G.; Bocking, S. P.; Rhie, A. G. O.; Bennett, A. D.; Hope, J., Post-Translational Hydroxylation at the *N*-Terminus of the Prion Protein Reveals Presence of PPII Structure *in vivo*. *EMBO J.* **2000,** *19*, (20), 5324-5331.
29. Blanch, E. W.; Robinson, D. J.; Hecht, L.; Barron, L. D., A Comparison of the Solution Structures of Tobacco Rattle and Tobacco Mosaic Viruses from Raman Optical Activity. *J. Gen. Virol.* **2001,** *82*, (6), 1499-1502.
30. Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R., Polyproline II Structure in a Sequence of Seven Alanine Residues. *Proc. Natl. Acad. Sci. USA* **2002,** *99*, (13), 9190-9195.
31. Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P., Host-Guest Scale of Left-Handed Polyproline II Helix Formation. *Proteins* **2003,** *53*, (1), 68-75.
32. Traub, W.; Shmueli, U., Structure of Poly-L-Proline I. *Nature* **1963,** *198*, (4886), 1165-1166.
33. Ramachandran, G. N.; Kartha, G., Structure of Collagen. *Nature* **1955,** *176*, (4482), 593-595.
34. Rich, A.; Crick, F. H. C., The Structure of Collagen. *Nature* **1955,** *176*, (4489), 915-916.
35. Rich, A.; Crick, F. H. C., The Molecular Structure of Collagen. *J. Mol. Biol.* **1961,** *3*, 483-506.

36. Ramachandran, G. N.; Kartha, G., Structure of Collagen. *Nature* **1954,** *174*, (4423), 269-270.
37. Traub, W.; Yonath, A.; Segal, D. M., On the Molecular Structure of Collagen. *Nature* **1969,** *221*, (5184), 914-917.
38. Daragan, V. A.; Ilyina, E.; Mayo, K. H., Effects of Molecular Association on Structure and Dynamics of a Collagenous Peptide. *Biopolymers* **1993,** *33*, (4), 521-533.
39. Hongo, C.; Nagarajan, V.; Noguchi, K.; Kamitori, S.; Okuyama, K.; Tanaka, Y.; Nishino, N., Average Crystal Structure of $(Pro\text{-}Pro\text{-}Gly)_9$ at 1.0 Å Resolution. *Polym. J. (Tokyo, Jpn.)* **2001,** *33*, (10), 812-818.
40. Engel, J.; Baechinger, H. P., Structure, Stability and Folding of the Collagen Triple Helix. *Top. Curr. Chem.* **2005,** *247*, (Collagen), 7-33.
41. Golbik, R.; Eble, J. A.; Ries, A.; Kühn, K., The Spatial Orientation of the Essential Amino Acid Residues Arginine and Aspartate within the $\alpha 1\beta 1$ Integrin Recognition Site of Collagen IV Has Been Resolved Using Fluorescence Resonance Energy Transfer. *J. Mol. Biol.* **2000,** *297*, (2), 501-509.
42. Tsai, M. I.-H.; Xu, Y.; Dannenberg, J. J., Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProDAla, and ProProDSer Triads. *J. Am. Chem. Soc.* **2005,** *127*, (41), 14130-14131.
43. Horng, J.-C.; Kotch, F. W.; Raines, R. T., Is Glycine a Surrogate for a D-Amino Acid in the Collagen Triple Helix? *Protein Sci.* **2007,** *16*, (2), 208-215.
44. Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B., Peptide Investigations of Pairwise Interactions in the Collagen Triple-Helix. *J. Mol. Biol.* **2002,** *316*, (2), 385-394.
45. Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B., Triple-Helix Propensity of Hydroxyproline and Fluoroproline: Comparison of Host-Guest and Repeating Tripeptide Collagen Models. *J. Am. Chem. Soc.* **2003,** *125*, (38), 11500-11501.
46. Piez, K. A.; Trus, B. L., A New Model for Packing of Type-I Collagen Molecules in the Native Fibril. *Biosci. Rep.* **1981,** *1*, (10), 801-810.
47. Bateman, J. F.; Lamande, S. R.; Ramshaw, J. A. M., Collagen Superfamily. In *Extracellular Matrix*, Comper, W. D., Ed. Harwood Academic: Amsterdam, 1996; Vol. 2, pp 22-67.
48. Ramachandran, G. N., Structure of Collagen at the Molecular Level. In *Treatise on Collagen*, Ramachandran, G. N., Ed. Academic Press: London, 1967; Vol. 1, pp 103-183.
49. Okuyama, K.; Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M., Crystal and Molecular Structure of a Collagen-like Polypeptide $(Pro\text{-}Pro\text{-}Gly)_{10}$. *J. Mol. Biol.* **1981,** *152*, (2), 427-443.
50. Fraser, R. D.; MacRae, T. P.; Suzuki, E., Chain Conformation in the Collagen Molecule. *J. Mol. Biol.* **1979,** *129*, (3), 463-481.
51. Cohen, C.; Bear, R. S., Helical Polypeptide Chain Configuration in Collagen. *J. Am. Chem. Soc.* **1953,** *75*, (11), 2783-2784.
52. Kramer, R. Z.; Bella, J.; Brodsky, B.; Berman, H. M., The Crystal and Molecular Structure of a Collagen-like Peptide with A Biologically Relevant Sequence. *J. Mol. Biol.* **2001,** *131*, (1), 131-147.
53. Kramer, R. Z.; Vitagliano, L.; Bella, J.; Berisio, R.; Mazzarella, L.; Brodsky, B.; Zagari, A.; Berman, H. M., X-Ray Crystallographic Determination of a Collagen-like Peptide with the Repeating Sequence (Pro-Pro-Gly). *J. Mol. Biol.* **1998,** *280*, (4), 623-638.

54. Nagarajan, V.; Kamitori, S.; Okuyama, K., Crystal Structure Analysis of Collagen Model Peptide (Pro-Pro-Gly)$_{10}$. *J. Biochem. (Tokyo)* **1998,** *124*, (6), 1117-1123.
55. Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A., Preferred Proline Puckerings in *cis* and *trans* Peptide Groups: Implications for Collagen Stability. *Protein Sci.* **2001,** *10*, (12), 2627-2632.
56. Vitagliano, L.; Berisio, R.; Mazzarella, L.; Zagari, A., Structural Bases of Collagen Stabilization Induced by Proline Hydroxylation. *Biopolymers* **2001,** *58*, (5), 459-464.
57. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A., Crystal Structure of the Collagen Triple Helix Model [(Pro-Pro-Gly)$_{10}$]$_3$. *Protein Sci.* **2002,** *11*, (2), 262-270.
58. Schumacher, M.; Mizuno, K.; Bachinger, H. P., The Crystal Structure of the Collagen-like Polypeptide (Glycyl-4(*R*)-Hydroxyprolyl-4(*R*)-Hydroxyprolyl)$_9$ at 1.55 Å Resolution Shows Up-puckering of the Proline Ring in the Xaa Position. *J. Biol. Chem.* **2005,** *280*, (21), 20397-20403.
59. Kadler, K., Extracellular Matrix 1: Fibril-Forming Collagens. *Protein Profile* **1995,** *2*, (5), 491-619.
60. Friess, W., Collagen - Biomaterial for Drug Delivery. *Eur. J. Pharm. Biopharm.* **1998,** *45*, (2), 113-136.
61. Miller, E. J., Collagen Types: Structure, Distribution and Functions. In *Collagen Vol I - Biochemsitry*, Nimni, M. E., Ed. CRC Press: Boca Raton, FL, 1988; pp 139-157.
62. Weiss, J. B.; Ayad, S., An Introduction to Collagen. In *Collagen in Health and Disease*, Weiss, J. B.; Jayson, M. I. V., Eds. Churchill Livingstone: New York, 1982; pp 1-17.
63. Lodish, H.; Berk, A.; Zipursky, L. S.; Matsudaira, P.; Baltimore, D.; Darnell, J., *Molecular Cell Biology*. Fourth Edition ed.; W. H. Freeman: New York, 2000.
64. Veit, G.; Kobbe, B.; Keene, D. R.; Paulsson, M.; Koch, M.; Wagener, R., Collagen XXVIII, A Novel von Willebrand Factor A Domain-Containing Protein with Many Imperfections in the Collagenous Domain. *J. Biol. Chem.* **2006,** *281*, (6), 3494-3504.
65. Koide, T.; Nagata, K., Collagen Biosynthesis. *Top. Curr. Chem.* **2005,** *247*, (Collagen), 85-114.
66. Timpl, R., Immunology of the Collagens. In *Extracellular Matrix Biochemistry*, Piez, K. A.; Reddi, A. H., Eds. Elsevier: New York, 1984; pp 159-190.
67. Viguet-Carrin, S.; Garnero, P.; Delmas, P. D., The Role of Collagen in Bone Strength. *Osteoporos. Int.* **2006,** *17*, (3), 319-336.
68. Niyibizi, C.; Eyre, D. R., Structural Characteristics of Cross-Linking Sites in type V Collagen of Bone. Chain Specificities and Heterotypic Links to Type I Collagen. *Eur. J. Biochem.* **1994,** *224*, (3), 943-950.
69. Bachra, B. N.; Fischer, H. R., Mineral Deposition in Collagen *in vitro*. *Calcif. Tissue Res.* **1968,** *2*, (4), 343-352.
70. Bachra, B. N.; Fischer, H. R., Recalcification of Decalcified Bone Collagen *in vitro* as a Model for Biologic Calcification. *Calcif. Tissue Res.* **1968**, Suppl:7.
71. Arsenault, A. L., Image Analysis of Mineralized and Non-Mineralized Type I Collagen Fibrils. *J. Electron Microsc. Tech.* **1991,** *18*, (3), 262-268.
72. Arsenault, A. L., A Comparative Electron Microscopic Study of Apatite Crystals in Collagen Fibrils of Rat Bone, Dentin and Calcified Turkey Leg Tendons. *Bone Miner.* **1989,** *6*, (2), 165-177.

73. Ricard-Blum, S.; Ruggiero, F.; Rest, M. v. d., The Collagen Superfamily. *Top. Curr. Chem.* **2005,** *247*, (Collagen), 35-84.
74. Brinckmann, J., Collagens at a Glance. *Top. Curr. Chem.* **2005,** *247*, (Collagen), 1-6.
75. Wu, J. J.; Woods, P. E.; Eyre, D. R., Identification of Cross-Linking Sites in Bovine Cartilage Type IX Collagen Reveals an Antiparallel Type II-Type IX Molecular Relationship and Type IX to Type IX Bonding. *J. Biol. Chem.* **1992,** *267*, (32), 23007-23014.
76. Eyre, D. R.; Apon, S.; Wu, J. J.; Ericsson, L. H.; Walsh, K. A., Collagen Type IX: Evidence for Covalent Linkages to Type II Collagen in Cartilage. *FEBS Lett.* **1987,** *220*, (2), 337-341.
77. Trentham, D. E.; Dynesius, R. A.; Rocklin, R. E.; David, J. R., Cellular Sensitivity to Collagen in Rheumatoid Arthritis. *Trans. Assoc. Am. Physicians* **1978,** *91*, 118-128.
78. Trentham, D. E.; Dynesius, R. A.; Rocklin, R. E.; David, J. R., Cellular Sensitivity to Collagen in Rheumatoid Arthritis. *N. Engl. J. Med.* **1978,** *299*, (7), 327-332.
79. Menashi, S.; Finch, A.; Gardner, P. J.; Ledward, D. A., Enthalphy Changes Associated with the Denaturation of Collagens of Different Imino Acid Content. *Biochim. Biophys. Acta* **1976,** *444*, (2), 623-625.
80. Mumby, S. M.; Raugi, G. J.; Bornstein, P., Interactions of Thrombospondin with Extracellular Matrix Proteins: Selective Binding to Type V Collagen. *J. Cell Biol.* **1984,** *98*, (2), 646-652.
81. Galvin, N. J.; Vance, P. M.; Dixit, V. M.; Fink, B.; Frazier, W. A., Interaction of Human Thrombospondin with Types I-V Collagen: Direct Binding and Electron Microscopy. *J. Cell Biol.* **1987,** *104*, (5), 1413-1422.
82. Kadler, K. E.; Holmes, D. F.; Trotter, J. A.; Chapman, J. A., Collagen Fibril Formation. *Biochem. J.* **1996,** *316 (Pt 1)*, 1-11.
83. Broek, D. L.; Madri, J.; Eikenberry, E. F.; Brodsky, B., Characterization of the Tissue Form of Type V Collagen from Chick Bone. *J. Biol. Chem.* **1985,** *260*, (1), 555-562.
84. Bonaventure, J.; Zylberberg, L.; Cohen-Solal, L.; Allain, J. C.; Lasselin, C.; Maroteaux, P., A New Lethal Brittle Bone Syndrome with Increased Amount of Type V Collagen in a Patient. *Am. J. Med. Genet.* **1989,** *33*, (3), 299-310.
85. Michalickova, K.; Susic, M.; Willing, M. C.; Wenstrup, R. J.; Cole, W. G., Mutations of the α2(V) Chain of Type V Collagen Impair Matrix Assembly and Produce Ehlers-Danlos Syndrome Type I. *Hum. Mol. Genet.* **1998,** *7*, (2), 249-255.
86. Nicholls, A. C.; Oliver, J. E.; McCarron, S.; Harrison, J. B.; Greenspan, D. S.; Pope, F. M., An Exon Skipping Mutation of a Type V Collagen Gene (*COL5A1*) in Ehlers-Danlos Syndrome. *J. Med. Genet.* **1996,** *33*, (11), 940-946.
87. De Paepe, A.; Nuytinck, L.; Hausser, I.; Anton-Lamprecht, I.; Naeyaert, J. M., Mutations in the *COL5A1* Gene Are Causal in the Ehlers-Danlos Syndromes I and II. *Am. J. Hum. Genet.* **1997,** *60*, (3), 547-554.
88. Toriello, H. V.; Glover, T. W.; Takahara, K.; Byers, P. H.; Miller, D. E.; Higgins, J. V.; Greenspan, D. S., A Translocation Interrupts the *COL5A1* Gene in a Patient with Ehlers-Danlos Syndrome and Hypomelanosis of Ito. *Nat. Genet.* **1996,** *13*, (3), 361-365.
89. Glanville, R. W., Type IV Collagen. In *Structure and Function of Collagen Types*, Mayne, R.; Burgeson, R. E., Eds. Academic Press: London, 1987; pp 43-79.
90. Hudson, B. G.; Reeders, S. T.; Tryggvason, K., Type IV Collagen: Structure, Gene Organization, and Role in Human Diseases. Molecular Basis of Goodpasture and Alport Syndromes and Diffuse Leiomyomatosis. *J. Biol. Chem.* **1993,** *268*, (35), 26033-26036.

91. Haralson, M. A.; Federspiel, S. J.; Martinez-Hernandez, A.; Rhodes, R. K.; Miller, E. J., Synthesis of [Pro α1(IV)]$_3$ Collagen Molecules by Cultured Embryo-Derived Parietal Yolk Sac Cells. *Biochemistry* **1985,** *24*, (21), 5792-5797.
92. Saus, J.; Wieslander, J.; Langeveld, J. P.; Quinones, S.; Hudson, B. G., Identification of the Goodpasture Antigen as the α3(IV) Chain of Collagen IV. *J. Biol. Chem.* **1988,** *263*, (26), 13374-13380.
93. Langeveld, J. P.; Wieslander, J.; Timoneda, J.; McKinney, P.; Butkowski, R. J.; Wisdom, B. J., Jr.; Hudson, B. G., Structural Heterogeneity of the Noncollagenous Domain of Basement Membrane Collagen. *J. Biol. Chem.* **1988,** *263*, (21), 10481-10488.
94. Gunwar, S.; Noelken, M. E.; Hudson, B. G., Properties of the Collagenous Domain of the α3(IV) Chain, the Goodpasture Antigen, of Lens Basement Membrane Collagen. Selective Cleavage of α(IV) Chains with Retention of Their Triple Helical Structure and Noncollagenous Domain. *J. Biol. Chem.* **1991,** *266*, (21), 14088-14094.
95. Barker, D. F.; Hostikka, S. L.; Zhou, J.; Chow, L. T.; Oliphant, A. R.; Gerken, S. C.; Gregory, M. C.; Skolnick, M. H.; Atkin, C. L.; Tryggvason, K., Identification of Mutations in the *COL4A5* Collagen Gene in Alport Syndrome. *Science* **1990,** *248*, (4960), 1224-1227.
96. Zhou, J.; Barker, D. F.; Hostikka, S. L.; Gregory, M. C.; Atkin, C. L.; Tryggvason, K., Single Base Mutation in α5(IV) Collagen Chain Gene Converting a Conserved Cysteine to Serine in Alport Syndrome. *Genomics* **1991,** *9*, (1), 10-18.
97. Tryggvason, K.; Zhou, J.; Hostikka, S. L.; Shows, T. B., Molecular Genetics of Alport Syndrome. *Kidney Int.* **1993,** *43*, (1), 38-44.
98. Timpl, R., Structure and Biological Activity of Basement Membrane Proteins. *Eur. J. Biochem.* **1989,** *180*, (3), 487-502.
99. Blumberg, B.; MacKrell, A. J.; Fessler, J. H., Drosophila Basement Membrane Procollagen α1(IV). II. Complete cDNA Sequence, Genomic Structure, and General Implications for Supramolecular Assemblies. *J. Biol. Chem.* **1988,** *263*, (34), 18328-18337.
100. Brodsky, B.; Ramshaw, J. A., The Collagen Triple-Helix Structure. *Matrix Biol.* **1997,** *15*, (8-9), 545-554.
101. Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J., Structure of the Integrin α2β1-Binding Collagen Peptide. *J. Mol. Biol.* **2004,** *335*, (4), 1019-1028.
102. Kramer, R. Z.; Venugopal, M. G.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M., Staggered Molecular Packing in Crystals of a Collagen-like Peptide with a Single Charged Pair. *J. Mol. Biol.* **2000,** *301*, (5), 1191-1205.
103. Emsley, J.; Knight, C.; Farndale, R.; Barnes, M.; Liddington, R., Structural Basis of Collagen Recognition by Integrin α2β1. *Cell* **2000,** *101*, (1), 47-56.
104. Brodsky, B.; Persikov, A. V., Molecular Structure of the Collagen Triple Helix. *Adv. Protein Chem.* **2005,** *70*, 301-339.
105. Bella, J.; Brodsky, B.; Berman, H. M., Hydration Structure of a Collagen Peptide. *Structure* **1995,** *3*, (9), 893-906.
106. Suzuki, E.; Fraser, R. D. B.; MacRae, T. P., Role of Hydroxyproline in the Stabilization of the Collagen Molecule via Water Molecules. *Int. J. Biol. Macromol.* **1980,** *2*, 54-56.
107. Vitagliano, L.; Nemethy, G.; Zagari, A.; Scheraga, H. A., Stabilization of the Triple-Helical Structure of Natural Collagen by Side-Chain Interactions. *Biochemistry* **1993,** *32*, (29), 7354-7359.

108. Mechling, D. E.; Bächinger, H. P., The Collagen-like Peptide $(GER)_{15}GPCCG$ Forms pH-dependent Covalently Linked Triple Helical Trimers. *J. Biol. Chem.* **2000,** *275*, (19), 14532-14536.
109. Jiravanichanun, N.; Mizuno, K.; Bachinger, H. P.; Okuyama, K., Threonine in Collagen Triple-Helical Structure. *Polym. J. (Tokyo, Jpn.)* **2006,** *38*, (4), 400-403.
110. Bann, J. G.; Bachinger, H. P., Glycosylation/Hydroxylation-Induced Stabilization of the Collagen Triple Helix. 4-*trans*-Hydroxyproline in the Xaa Position Can Stabilize the Triple Helix. *J. Biol. Chem.* **2000,** *275*, (32), 24466-9.
111. Doi, M.; Nishi, Y.; Uchiyama, S.; Nishiuchi, Y.; Nishio, H.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y., Collagen-like Triple Helix Formation of Synthetic $(Pro\text{-}Pro\text{-}Gly)_{10}$ Analogues: $(4(S)\text{-}Hydroxyprolyl\text{-}4(R)\text{-}Hydroxyprolyl\text{-}Gly)_{10}$, $(4(R)\text{-}Hydroxyprolyl\text{-}4(R)\text{-}Hydroxyprolyl\text{-}Gly)_{10}$ and $(4(S)\text{-}Fluoroprolyl\text{-}4(R)\text{-}Fluoroprolyl\text{-}Gly)_{10}$. *J. Pept. Sci.* **2005,** *11*, (10), 609-616.
112. Krimm, S.; Kuroiwa, K., Low Temperature Infrared Spectra of Polyglycines and C-H···O=C Hydrogen Bonding in Polyglycine II. *Biopolymers* **1968,** *6*, (3), 401-407.
113. Bella, J.; Berman, H. M., Crystallographic Evidence for Cα–H···O=C Hydrogen Bonds in a Collagen Triple Helix. *J. Mol. Biol.* **1996,** *264*, (4), 734-742.
114. Kobayashi, Y.; Sakai, R.; Kakiuchi, K.; Isemura, T., Physicochemical Analysis of $(Pro\text{-}Pro\text{-}Gly)_n$ with Defined Molecular Weight-Temperature Dependence of Molecular Weight in Aqueous Solution. *Biopolymers* **1970,** *9*, (4), 415-425.
115. Shaw, B. R.; Schurr, J. M., The Association Reaction of Collagen Model Polypeptides $(Pro\text{-}Pro\text{-}Gly)_n$. *Biopolymers* **1975,** *14*, (9), 1951-1985.
116. Sakakibara, S.; Inouye, K.; Shudo, K.; Kishida, Y.; Kobayashi, Y.; Prockop, D. J., Synthesis of $(Pro\text{-}Hyp\text{-}Gly)_n$ of Defined Molecular Weights. Evidence for the Stabilization of Collagen Triple Helix by Hydroxypyroline. *Biochim. Biophys. Acta* **1973,** *303*, (1), 198-202.
117. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T., Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *J. Am. Chem. Soc.* **2001,** *123*, (4), 777-778.
118. Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T., A Hyperstable Collagen Mimic. *Chem. Biol.* **1999,** *6*, (2), 63-70.
119. Feng, Y.; Melacini, G.; Taulane, J. P.; Goodman, M., Acetyl-Terminated and Template-Assembled Collagen-Based Polypeptides Composed of Gly-Pro-Hyp Sequences. 2. Synthesis and Conformational Analysis by Circular Dichroism, Ultraviolet Absorbance, and Optical Rotation. *J. Am. Chem. Soc.* **1996,** *118*, (43), 10351-10358.
120. Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B., Amino Acid Propensities for the Collagen Triple-Helix. *Biochemistry* **2000,** *39*, (48), 14960-14967.
121. Berg, R. A.; Prockop, D. J., The Thermal Transition of a Non-Hydroxylated Form of Collagen. Evidence for a Role for Hydroxyproline in Stabilizing the Triple-Helix of Collagen. *Biochem. Biophys. Res. Commun.* **1973,** *52*, (1), 115-120.
122. Engel, J.; Chen, H.-T.; Prockop, D. J.; Klump, H., The Triple Helix Coil Conversion of Collagen-like Polytripeptides in Aqueous and Nonaqueous Solvents. Comparison of the Thermodynamic Parameters and the Binding of Water to $(L\text{-}Pro\text{-}L\text{-}Pro\text{-}Gly)_n$ and $(L\text{-}Pro\text{-}L\text{-}Hyp\text{-}Gly)_n$. *Biopolymers* **1977,** *16*, (3), 601-622.
123. O'Neil, K. T.; DeGrado, W. F., A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science* **1990,** *250*, (4981), 646-651.

124. Smith, C. K.; Withka, J. M.; Regan, L., A Thermodynamic Scale for the β-Sheet Forming Tendencies of the Amino Acid. *Biochemistry* **1994,** *33*, (18), 5510-5517.
125. Shah, N. K.; Ramshaw, J. A.; Kirkpatrick, A.; Shah, C.; Brodsky, B., A Host-Guest Set of Triple-Helical Peptides: Stability of Gly-X-Y Triplets Containing Common Nonpolar Residues. *Biochemistry* **1996,** *35*, (32), 10262-10268.
126. Chan, V. C.; Ramshaw, J. A.; Kirkpatrick, A.; Beck, K.; Brodsky, B., Positional Preferences of Ionizable Residues in Gly-X-Y Triplets of the Collagen Triple-Helix. *J. Biol. Chem.* **1997,** *272*, (50), 31441-31446.
127. Shah, N. K.; Sharma, M.; Kirkpatrick, A.; Ramshaw, J. A.; Brodsky, B., Gly-Gly-Containing Triplets of Low Stability Adjacent to a Type III Collagen Epitope. *Biochemistry* **1997,** *36*, (19), 5878-5883.
128. Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A.; Brodsky, B., Destabilization of Osteogenesis Imperfecta Collagen-like Model Peptides Correlates with the Identity of the Residue Replacing Glycine. *Proc. Natl. Acad. Sci. USA* **2000,** *97*, (8), 4273-4278.
129. Persikov, A. V.; Ramshaw, J. A.; Brodsky, B., Collagen Model Peptides: Sequence Dependence of Triple-Helix Stability. *Biopolymers* **2000,** *55*, (6), 436-450.
130. Jenkins, C. L.; Bretscher, L. E.; Guzei, I. A.; Raines, R. T., Effect of 3-Hydroxyproline Residues on Collagen Stability. *J. Am. Chem. Soc.* **2003,** *125*, (21), 6422-6427.
131. Jones, E. Y.; Miller, A., Analysis of Structural Design Features in Collagen. *J. Mol. Biol.* **1991,** *218*, (1), 209-219.
132. Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J., Synthesis and Physical Properties of (Hydroxyproline-Proline-Glycine)$_{10}$: Hydroxyproline in the X-position Decreases the Melting Temperature of the Collagen Triple Helix. *Arch. Biochem. Biophys.* **1982,** *219*, (1), 198-203.
133. Inouye, K.; Sakakibara, S.; Prockop, D. J., Effects of the Stereo-Configuration of the Hydroxyl Group in 4-Hydroxyproline on the Triple-Helical Structures Formed by Homogenous Peptides Resembling Collagen. *Biochim. Biophys. Acta* **1976,** *420*, (1), 133-141.
134. Privalov, P. L., Stability of Proteins. Proteins Which Do Not Present a Single Cooperative System. *Adv. Protein Chem.* **1982,** *35*, 1-104.
135. Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T., Code for Collagen's Stability Deciphered. *Nature* **1998,** *392*, (6677), 666-667.
136. Dunitz, J. D., Organic Fluorine: Odd Man Out. *Chembiochem* **2004,** *5*, (5), 614-621.
137. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A., Recent Progress on Collagen Triple Helix Structure, Stability and Assembly. *Protein Pept. Lett.* **2002,** *9*, (2), 107-116.
138. Eberhardt, E. S.; Panisik, N., Jr.; Raines, R. T., Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. *J. Am. Chem. Soc.* **1996,** *118*, (49), 12261-12266.
139. Panasik, N., Jr.; Eberhardt, E. S.; Edison, A. S.; Powell, D. R.; Raines, R. T., Inductive Effects on the Structure of Proline Residues. *Int. J. Pept. Protein Res.* **1994,** *44*, (3), 262-269.
140. Milner-White, E. J.; Bell, L. H.; Maccallum, P. H., Pyrrolidine Ring Puckering in *cis* and *trans*-Proline Residues in Proteins and Polypeptides. Different Puckers are Favoured in Certain Situations. *J. Mol. Biol.* **1992,** *228*, (3), 725-734.

141. Nishi, Y.; Uchiyama, S.; Doi, M.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y., Different Effects of 4-Hydroxyproline and 4-Fluoroproline on the Stability of Collagen Triple Helix. *Biochemistry* **2005,** *44*, (16), 6034-6042.
142. Shoulders, M. D.; Hodges, J. A.; Raines, R. T., Reciprocity of Steric and Stereoelectronic Effects in the Collagen Triple Helix. *J. Am. Chem. Soc.* **2006,** *128*, (25), 8112-8113.
143. Hodges, J. A.; Raines, R. T., Stereoelectronic Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers. *J. Am. Chem. Soc.* **2003,** *125*, (31), 9262-9263.
144. Shoulders, M. D.; Guzei, I. A.; Raines, R. T., 4-Chloroprolines: Synthesis, Conformational Analysis, and Effect on the Collagen Triple Helix. *Biopolymers* **2008,** *89*, (5), 443-454.
145. Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L., Fluoroprolines as Tools for Protein Design and Engineering. *Angew. Chem. Int. Ed.* **2001,** *40*, (5), 923-925.
146. Barth, D.; Milbradt, A. G.; Renner, C.; Moroder, L., A (4*R*)- or a (4*S*)-Fluoroproline Residue in Position Xaa of the (Xaa-Yaa-Gly) Collagen Repeat Severely Affects Triple-Helix Formation. *Chembiochem* **2004,** *5*, (1), 79-86.
147. Mizuno, K.; Hayashi, T.; Peyton, D. H.; Bächinger, H. P., Hydroxylation-Induced Stabilization of the Collagen Triple Helix. Acetyl-(Glycyl-4(*R*)-Hydroxyprolyl-4(*R*)-Hydroxyprolyl)$_{10}$-NH_2 Forms a Highly Stable Triple Helix. *J. Biol. Chem.* **2004,** *279*, (36), 38072-38078.
148. Mizuno, K.; Hayashi, T.; Bächinger, H. P., Hydroxylation-Induced Stabilization of the Collagen Triple Helix. Further Characterization of Peptides with 4(*R*)-Hydroxyproline in the Xaa Position. *J. Biol. Chem.* **2003,** *278*, (34), 32373-32379.
149. Kawahara, K.; Nishi, Y.; Nakamura, S.; Uchiyama, S.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y., Effect of Hydration on the Stability of the Collagen-like Triple-Helical Structure of [4(*R*)-Hydroxyprolyl-4(*R*)-Hydroxyprolylglycine]$_{10}$. *Biochemistry* **2005,** *44*, (48), 15812-15822.
150. Nishi, Y.; Doi, M.; Uchiyama, S.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y., Stabilization Mechanism of Triple Helical Structure of Collagen Molecules. *Int. J. Pept. Res. Therapeut.* **2003,** *10*, (5-6), 533-537.
151. Mizuno, K.; Hayashi, T.; Peyton, D. H.; Bächinger, H. P., The Peptides Acetyl-(Gly-3(*S*)Hyp-4(*R*)Hyp)$_{10}$-NH_2 and Acetyl-(Gly-Pro-3(*S*)Hyp)$_{10}$-NH_2 Do Not Form a Collagen Triple Helix. *J. Biol. Chem.* **2004,** *279*, (1), 282-287.
152. Hodges, J. A.; Raines, R. T., Stereoelectronic and Steric Effects in the Collagen Triple Helix: Toward a Code for Strand Association. *J. Am. Chem. Soc.* **2005,** *127*, (45), 15923-15932.
153. Schumacher, M. A.; Mizuno, K.; Bachinger, H. P., The Crystal Structure of a Collagen-like Polypeptide with 3(*S*)-Hydroxyproline Residues in the Xaa Position Forms a Standard 7/2 Collagen Triple Helix. *J. Biol. Chem.* **2006,** *281*, (37), 27566-27574.
154. Pauling, L.; Corey, R. B.; Branson, H. R., The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. USA* **1951,** *37*, (4), 205-211.
155. Radzicka, A.; Pedersen, L.; Wolfenden, R., Influences of Solvent Water on Protein Folding: Free Energies of Solvation of *cis* and *trans* Peptides Are Nearly Identical. *Biochemistry* **1988,** *27*, (12), 4538-4541.

156. Li, P.; Chen, X. G.; Shulin, E.; Asher, S. A., UV Resonance Raman Ground and Excited State Studies of Amide and Peptide Isomerization Dynamics. *J. Am. Chem. Soc.* **1997,** *119*, (5), 1116-1120.
157. Jorgensen, W. L.; Gao, J., *Cis-trans* Energy Difference for the Peptide Bond in the Gas Phase and in Aqueous Solution. *J. Am. Chem. Soc.* **1988,** *110*, (13), 4212-4216.
158. Yu, H. A.; Karplus, M.; Pettitt, B. M., Aqueous Solvation of *N*-Methylacetamide Conformers: Comparison of Simulations and Integral Equation Theories. *J. Am. Chem. Soc.* **1991,** *113*, (7), 2425-2434.
159. Scherer, G.; Kramer, M. L.; Schutkowski, M.; Reimer, U.; Fischer, G., Barriers to Rotation of Secondary Amide Peptide Bonds. *J. Am. Chem. Soc.* **1998,** *120*, (22), 5568-5574.
160. Gratwohl, C.; Wuthrich, K., The X-Pro Peptide Bond as an NMR Probe for Conformational Studies of Flexible Linear Peptides. *Biopolymers* **1976,** *15*, (10), 2025-2041.
161. Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G., Side-Chain Effects on Peptidyl-Prolyl *cis/trans* Isomerisation. *J. Mol. Biol.* **1998,** *279*, (2), 449-460.
162. Schoetz, G.; Trapp, O.; Schurig, V., Determination of the *cis-trans* Isomerization Barrier of Several L-Peptidyl-L-Proline Dipeptides by Dynamic Capillary Electrophoresis and Computer Simulation. *Electrophoresis* **2001,** *22*, (12), 2409-2415.
163. Sarkar, S. K.; Young, P. E.; Sullivan, C. E.; Torchia, D. A., Detection of *cis* and *trans* X-Pro Peptide Bonds in Proteins by ^{13}C NMR: Application to Collagen. *Proc. Natl. Acad. Sci. USA* **1984,** *81*, (15), 4800-4803.
164. Mann, K.; Mechling, D. E.; Bächinger, H. P.; Eckerskorn, C.; Gaill, F.; Timpl, R., Glycosylated Threonine but Not 4-Hydroxyproline Dominates the Triple Helix Stabilizing Positions in the Sequence of a Hydrothermal Vent Worm Cuticle Collagen. *J. Mol. Biol.* **1996,** *261*, (2), 255-266.
165. Gaill, F.; Mann, K.; Wiedemann, H.; Engel, J.; Timpl, R., Structural Comparison of Cuticle and Interstitial Collagens from Annelids Living in Shallow Sea-Water and at Deep-Sea Hydrothermal Vents. *J. Mol. Biol.* **1995,** *246*, (2), 284-294.
166. Bann, J. G.; Peyton, D. H.; Bächinger, H. P., Sweet is Stable: Glycosylation Stabilizes Collagen. *FEBS Lett.* **2000,** *473*, (2), 237-240.
167. Bann, J. G.; Bachinger, H. P.; Peyton, D. H., Role of Carbohydrate in Stabilizing the Triple-Helix in a Model for a Deep-Sea Hydrothermal Vent Worm Collagen. *Biochemistry* **2003,** *42*, (14), 4042-8.
168. Bachinger, H. P.; Engel, J., The Thermodynamics and Kinetics of Collagen Folding. In *Protein Folding Handbook*, Buchner, J.; Kiefhaber, T., Eds. Wiley-VCH: Weinheim, 2005; Vol. 2, pp 1059-1110.
169. Brandts, J. F.; Halvorson, H. R.; Brennan, M., Consideration of the Possibility that the Slow Step in Protein Denaturation Reactions is due to *cis-trans* Isomerism of Proline Residues. *Biochemistry* **1975,** *14*, (22), 4953-4963.
170. Baldwin, R. L., Finding Intermediates in Protein Folding. *Bioessays* **1994,** *16*, (3), 207-210.
171. Oka, M.; Montelione, G. T.; Scheraga, H. A., Chain-Folding Initiation Structures in Ribonuclease A: Conformational Analysis of *trans*-Ac-Asn-Pro-Tyr-NHMe and *trans*-Ac-Tyr-Pro-Asn-NHMe in Water and in the Solid State. *J. Am. Chem. Soc.* **1984,** *106*, (25), 7959-7969.

172. Houry, W. A.; Scheraga, H. A., Nature of the Unfolded State of Ribonuclease A: Effect of *cis-trans* X-Pro Peptide Bond Isomerization. *Biochemistry* **1996,** *35*, (36), 11719-11733.
173. Fischer, G.; Schmid, F. X., The Mechanism of Protein Folding. Implications of *in vitro* Refolding Models for *De Novo* Protein Folding and Translocation in the Cell. *Biochemistry* **1990,** *29*, (9), 2205-2212.
174. Bachinger, H. P.; Bruckner, P.; Timpl, R.; Engel, J., The Role of *cis-trans* Isomerization of Peptide Bonds in the Coil Triple Helix Conversion of Collagen. *Eur. J. Biochem.* **1978,** *90*, (3), 605-613.
175. Bachinger, H. P.; Bruckner, P.; Timpl, R.; Prockop, D. J.; Engel, J., Folding Mechanism of the Triple Helix in Type-III Collagen and Type-III pN-Collagen. Role of Disulfide Bridges and Peptide Bond Isomerization. *Eur. J. Biochem.* **1980,** *106*, (2), 619-632.
176. Bruckner, P.; Eikenberry, E. F.; Prockop, D. J., Formation of the Triple Helix of Type I Procollagen in Cellulo. A Kinetic Model Based on *cis-trans* Isomerization of Peptide Bonds. *Eur. J. Biochem.* **1981,** *118*, (3), 607-613.
177. Bruckner, P.; Eikenberry, E. F., Formation of the Triple Helix of Type I Procollagen in Cellulo. Temperature-Dependent Kinetics Support a Model Based on *cis-trans* Isomerization of Peptide Bonds. *Eur. J. Biochem.* **1984,** *140*, (2), 391-395.
178. Traub, W.; Piez, K. A., The Chemistry and Structure of Collagen. *Adv. Protein Chem.* **1971,** *25*, 243-352.
179. Temussi, P. A.; Tancredi, T.; Quadrifoglio, F., Conformational Rigidity of the Amide Bond. Variable-Temperature Nuclear Magnetic Resonance Study of the System Ag^{+}-*N,N*-Dimethylacetamide. *J. Phys. Chem.* **1969,** *73*, (12), 4227-4232.
180. Steinberg, I. Z.; Harrington, W. F.; Berger, A.; Sela, M.; Katchalski, E., The Configurational Changes of Poly-L-Proline in Solution. *J. Am. Chem. Soc.* **1960,** *82*, (20), 5263-5279.
181. Cheng, H. N.; Bovey, F. A., *Cis-trans* Equilibrium and Kinetic Studies of Acetyl-L-Proline and Glycyl-L-Proline. *Biopolymers* **1977,** *16*, (7), 1465-1472.
182. McLaughlin, S. H.; Bulleid, N. J., Molecular Recognition in Procollagen Chain Assembly. *Matrix Biol.* **1998,** *16*, (7), 369-377.
183. Baum, J.; Brodsky, B., Folding of Peptide Models of Collagen and Misfolding in Disease. *Curr. Opin. Struct. Biol.* **1999,** *9*, (1), 122-128.
184. Engel, J.; Prockop, D. J., The Zipper-Like Folding of Collagen Triple Helices and the Effects of Mutations that Disrupt the Zipper. *Annu. Rev. Biophys. Biophys. Chem.* **1991,** *20*, 137-152.
185. Beck, K.; Boswell, B. A.; Ridgway, C. C.; Bachinger, H. P., Triple Helix Formation of Procollagen Type I Can Occur at the Rough Endoplasmic Reticulum Membrane. *J. Biol. Chem.* **1996,** *271*, (35), 21566-21573.
186. Kielty, C. M.; Grant, M. E., The Collagen Family: Structure, Assembly, and Organization in the Extracellular Matrix. In *Connective Tissue and Its Heritable Disorders (Second Edition)*, Royce, P. M.; Steinmann, B., Eds. John Wiley & Sons: New York, 2003; pp 159-221.
187. Fields, G. B.; Prockop, D. J., Perspectives on the Synthesis and Application of Triple-Helical, Collagen-Model Peptides. *Biopolymers* **1996,** *40*, (4), 345-357.

188. Li, M. H.; Fan, P.; Brodsky, B.; Baum, J., Two-Dimensional NMR Assignments and Conformation of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and a Designed Collagen Triple-Helical Peptide. *Biochemistry* **1993,** *32*, (29), 7377-7387.
189. Ottl, J.; Battistutaa, R.; Pieper, M.; Tschesche, H.; Bode, W.; Kühn, K.; Moroder, L., Design and Synthesis of Heterotrimeric Collagen Peptides with a Built-in Cystine-Knot - Models for Collagen Catabolism by Matrix-Metalloproteases. *FEBS Lett.* **1996,** *398*, (1), 31-36.
190. Buevich, A.; Baum, J., Nuclear Magnetic Resonance Characterization of Peptide Models of Collagen-Folding Diseases. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2001,** *356*, (1406), 159-168.
191. Bulleid, N. J.; Dalley, J. A.; Lees, J. F., The *C*-Propeptide Domain of Procollagen Can Be Replaced with a Transmembrane Domain without Affecting Trimer Formation or Collagen Triple Helix Folding During Biosynthesis. *EMBO J.* **1997,** *16*, (22), 6694–6701.
192. Areida, S. K.; Reinhardt, D. P.; Muller, P. K.; Fietzek, P. P.; Kowitz, J.; Marinkovich, M. P.; Notbohm, H., Properties of the Collagen Type XVII Ectodomain: Evidence for *N*- to *C*-Terminal Triple Helix Folding. *J. Biol. Chem.* **2001,** *276*, (2), 1594-1601.
193. Frank, S.; Boudko, S.; Mizuno, K.; Schulthess, T.; Engel, J.; Bachinger Hans, P., Collagen Triple Helix Formation Can Be Nucleated at Either End. *J. Biol. Chem.* **2003,** *278*, (10), 7747-7750.
194. Davis, J. M.; Boswell, B. A.; Bachinger, H. P., Thermal Stability and Folding of Type IV Procollagen and Effect of Peptidyl-Prolyl *cis-trans* Isomerase on the Folding of the Triple Helix. *J. Biol. Chem.* **1989,** *264*, (15), 8956-8962.
195. Bachinger, H. P., The Influence of Peptidyl-Prolyl *cis-trans* Isomerase on the *in vitro* Folding of Type III Collagen. *J. Biol. Chem.* **1987,** *262*, (35), 17144-17148.
196. Bachmann, A.; Kiefhaber, T.; Boudko, S.; Engel, J.; Bächinger, H. P., Collagen Triple-Helix Formation in All-*trans* Chains Proceeds by a Nucleation/Growth Mechanism with a Purely Entropic Barrier. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, (39), 13897-13902.
197. Leikina, E.; Mertts, M. V.; Kuznetsova, N.; Leikin, S., Type I Collagen Is Thermally Unstable at Body Temperature. *Proc. Natl. Acad. Sci. USA* **2002,** *99*, (3), 1314-1318.
198. Bächinger, H. P.; Engel, J., Thermodynamic vs. Kinetic Stability of Collagen Triple Helices. *Matrix Biol.* **2001,** *20*, (4), 267-269.
199. Persikov, A. V.; Xu, Y.; Brodsky, B., Equilibrium Thermal Transitions of Collagen Model Peptides. *Protein Sci.* **2004,** *13*, (4), 893-902.
200. Davis, J. M.; Bachinger, H. P., Hysteresis in the Triple Helix-Coil Transition of Type III Collagen. *J. Biol. Chem.* **1993,** *268*, (34), 25965-25972.
201. Morris, N. P.; Watt, S. L.; Davis, J. M.; Bachinger, H. P., Unfolding Intermediates in the Triple Helix to Coil Transition of Bovine Type XI Collagen and Human Type V Collagens $\alpha 12\alpha 2$ and $\alpha 1\alpha 2\alpha 3$. *J. Biol. Chem.* **1990,** *265*, (17), 10081-10087.
202. Piez, K. A.; Sherman, M. R., Equilibrium and Kinetic Studies of the Helix-Coil Transition in 1-CB2, a Small Peptide from collagen. *Biochemistry* **1970,** *9*, (21), 4134-4140.
203. Suto, K.; Noda, H., Conformational Change of the Triple-Helical Structure. II. Conformation of $(Pro\text{-}Pro\text{-}Gly)_n$ and $(Pro\text{-}Pro\text{-}Gly)_n(Ala\text{-}Pro\text{-}Gly)_m(Pro\text{-}Pro\text{-}Pro\text{-}Gly)_n$ in an Aqueous Solution. *Biopolymers* **1974,** *13*, (11), 2391-2404.
204. Boudko, S.; Frank, S.; Kammerer, R. A.; Stetefeld, J.; Schulthess, T.; Landwehr, R.; Lustig, A.; Bachinger, H. P.; Engel, J., Nucleation and Propagation of the Collagen Triple

Helix in Single-Chain and Trimerized Peptides: Transition from Third to First Order Kinetics. *J. Mol. Biol.* **2002,** *317*, (3), 459-470.
205. Ackerman, M. S.; Bhate, M.; Shenoy, N.; Beck, K.; Ramshaw, J. A.; Brodsky, B., Sequence Dependence of the Folding of Collagen-like Peptides. Single Amino Acids Affect the Rate of Triple-Helix Nucleation. *J. Biol. Chem.* **1999,** *274*, (12), 7668-7673.
206. Xu, Y.; Bhate, M.; Brodsky, B., Characterization of the Nucleation Step and Folding of a Collagen Triple-Helix Peptide. *Biochemistry* **2002,** *41*, (25), 8143-8151.
207. Stultz, C. M., The Folding Mechanism of Collagen-like Model Peptides Explored through Detailed Molecular Simulations. *Protein Sci.* **2006,** *15*, (9), 2166-2177.
208. Gorham, S. D., Collagen. In *Biomaterials: Novel Materials from Biological Sources*, Byrom, D., Ed. Stockton Press: New York, 1991; pp 55-122.
209. Chvapil, M.; Kronenthal, R. L.; Walton van Winkle, J., Medical and Surgical Applications of Collagen. In *International review of connective tissue research*, Hall, D. A.; Jackson, D. S., Eds. Academic Press: New York, 1973; Vol. 6, pp 1-61.
210. Lu, M.-Y. F.; Thies, C., Collagen-Based Drug Delivery Devices. In *Polymers for Controlled Drug Delivery*, Tarcha, P. J., Ed. CRC Press: Boca Raton, FL, 1991; pp 149-161.
211. Sakaguchi, M.; Hori, H.; Hattori, S.; Irie, S.; Imai, A.; Yanagida, M.; Miyazawa, H.; Toda, M.; Inouye, S., IgE Reactivity to α1 and α2 Chains of Bovine Type 1 Collagen in Children with Bovine Gelatin Allergy. *J. Allergy Clin. Immunol.* **1999,** *104*, (3), 695-699.
212. Lynn, A. K.; Yannas, I. V.; Bonfield, W., Antigenicity and Immunogenicity of Collagen. *J. Biomed. Mater. Res. B* **2004,** *71*, (2), 343-354.
213. Cooperman, L.; Michaeli, D., The Immunogenicity of Injectable Collagen. I. A 1-Year Prospective Study. *J. Am. Acad. Dermatol.* **1984,** *10*, (4), 638-646.
214. Koide, T., Triple Helical Collagen-Like Peptides: Engineering and Applications in Matrix Biology. *Connect. Tissue Res.* **2005,** *46*, (3), 131-141.
215. 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.* **1968,** *41*, (5), 1273.
216. Wang, A. Y.; Mo, X.; Chen, C. S.; Yu, S. M., Facile Modification of Collagen Directed by Collagen Mimetic Peptides. *J. Am. Chem. Soc.* **2005,** *127*, (12), 4130-4131.
217. Fields, G. B.; Lauer, J. L.; Dori, Y.; Forns, P.; Yu, Y.-C.; Tirrell, M., Proteinlike Molecular Architecture: Biomaterial Applications for Inducing Cellular Receptor Binding and Signal Transduction. *Biopolymers* **1998,** *47*, (2), 143-151.
218. Qian, J. J.; Bhatnagar, R. S., Enhanced Cell Attachment to Anorganic Bone Mineral in the Presence of a Synthetic Peptide Related to Collagen. *J. Biomed. Mater. Res. A* **1996,** *31*, (4), 545-554.
219. Johnson, G.; Jenkins, M.; McLean, K. M.; Griesser, H. J.; Kwak, J.; Goodman, M.; Steele, J. G., Peptoid-Containing Collagen Mimetics with Cell Binding Activity. *J. Biomed. Mater. Res. A* **2000,** *51*, (4), 612-624.
220. Kotch, F. W.; Raines, R. T., Self-Assembly of Synthetic Collagen Triple Helices. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, (9), 3028-3033.
221. Koide, T.; Homma, D. L.; Asada, S.; Kitagawa, K., Self-Complementary Peptides for the Formation of Collagen-like Triple Helical Supramolecules. *Bioorg. Med. Chem. Lett.* **2005,** *15*, (23), 5230-5233.

222. Buevich, A. V.; Dai, Q.-H.; Liu, X.; Brodsky, B.; Baum, J., Site-Specific NMR Monitoring of *cis-trans* Isomerization in the Folding of the Proline-Rich Collagen Triple Helix. *Biochemistry* **2000,** *39*, (15), 4299-4308.
223. 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.* **2005,** *7*, (13), 2619-2622.
224. Wang, X. J.; Xu, B.; Mullins, A. B.; Neiler, F. K.; Etzkorn, F. A., Conformationally Locked Isostere of PhosphoSer-*cis*-Pro Inhibits Pin1 23-Fold Better than PhosphoSer-*trans*-Pro Isostere. *J. Am. Chem. Soc.* **2004,** *126*, (47), 15533-15542.
225. Chang, C. D.; Meienhofer, J., Solid-Phase Peptide Synthesis Using Mild Base Cleavage of *N* α-Fluorenylmethyloxycarbonylamino Acids, Exemplified by a Synthesis of Dihydrosomatostatin. *Int. J. Pept. Protein Res.* **1978,** *11*, (3), 246-249.
226. Barany, G.; Kneib-Cordonier, N.; Mullen, D. G., Solid-Phase Peptide Synthesis: A Silver Anniversary Report. *Int. J. Pept. Protein Res.* **1987,** *30*, (6), 705-739.
227. Merrifield, R. B., Solid Phase Peptide Synthesis. II. The Synthesis of Bradykinin. *J. Am. Chem. Soc.* **1964,** *86*, (2), 304-305.
228. Merrifield, R. B., Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963,** *85*, (14), 2149-2154.
229. Enders, D.; Kirchhoff, J. H.; Kobberling, J.; Peiffer, T. H., Asymmetric Synthesis of α-Branched Primary Amines on Solid Support via Novel Hydrazine Resins. *Org. Lett.* **2001,** *3*, (8), 1241-1244.
230. Rump, E. T.; Rijkers, D. T. S.; Hilbers, H. W.; de Groot, P. G.; Liskamp, R. M. J., Cyclotriveratrylene (CTV) as a New Chiral Triacid Scaffold Capable of Inducing Triple Helix Formation of Collagen Peptides Containing Either a Native Sequence or Pro-Hyp-Gly Repeats. *Chem. Eur. J.* **2002,** *8*, (20), 4613-4621.
231. West, C. W.; Estiarte, M. A.; Rich, D. H., New Methods for Side-Chain Protection of Cysteine. *Org. Lett.* **2001,** *3*, (8), 1205-1208.
232. Gioia, M. L. D.; Leggio, A.; Pera, A. L.; Siciliano, C.; Liguori, A.; Sindona, G., α-Amino Acid and Lipophilic *N*-Fmoc-Dipeptide Methyl Esters with Aluminium Trichloride and *N,N*-Dimethylaniline. *J. Pept. Res.* **2004,** *63*, (4), 383-387.
233. Holden, K. G.; Mattson, M. N.; Cha, K. H.; Rapoport, H., Synthesis of Chiral Pilocarpine Analogues via a C-8 Ketone Intermediate. *J. Org. Chem.* **2002,** *67*, (17), 5913-5918.
234. Wang, X. J.; Hart, S. A.; Xu, B.; Mason, M. D.; Goodell, J. R.; Etzkorn, F. A., Serine-*cis*-Proline and Serine-*trans*-Proline Isosteres: Stereoselective Synthesis of (*Z*)- and (*E*)-Alkene Mimics by Still-Wittig and Ireland-Claisen Rearrangements. *J. Org. Chem.* **2003,** *68*, (6), 2343-2349.
235. Bischofberger, N.; Waldmann, H.; Saito, T.; Simon, E. S.; Lees, W.; Bednarski, M. D.; Whitesides, G. M., Synthesis of Analogs of 1,3-Dihydroxyacetone Phosphate and Glyceraldehyde 3-Phosphate for Use in Studies of Fructose-1,6-Diphosphate Aldolase. *J. Org. Chem.* **1988,** *53*, (15), 3457-3465.
236. Ireland, R. E.; Mueller, R. H., Claisen Rearrangement of Allyl Esters. *J. Am. Chem. Soc.* **1972,** *94*, (16), 5897-5898.
237. Ireland, R. E.; Mueller, R. H.; Willard, A. K., The Ester Enolate Claisen Rearrangement. Stereochemical Control Through Stereoselective Enolate Formation. *J. Am. Chem. Soc.* **1976,** *98*, (10), 2868-2877.

238. Jones, J., *Amino Acid and Peptide Synthesis*. 2 ed.; Oxford University Press: New York, 2002.
239. Ottl, J.; Musiol, H. J.; Moroder, L., Heterotrimeric Collagen Peptides Containing Functional Epitopes. Synthesis of Single-Stranded Collagen Type I Peptides Related to the Collagenase Cleavage Site. *J. Pept. Sci.* **1999,** *5*, (2), 103-110.
240. Vigneaud, V. d.; Meyer, C. E., The Temporary Formation of the Azlactone Ring in the Racemization of Acyl Derivatives of Amino Acids with Acetic Anhydride. *J. Biol. Chem.* **1932,** *99*, (1), 143-151.
241. Sheradsky, T.; Silcoff, E. R., Synthesis of (*2R,5R*)-2-Amino-5-Hydroxyhexanoic Acid by Intramolecular Cycloaddition. *Molecules* **1998,** *3*, (3), 80-87.
242. Rapaka, R. S.; Bhatnagar, R. S.; Nitecki, D. E., Racemization in the Synthesis of Sequential Polypeptides Using *N*-Hydroxysuccinimide Esters. *Biopolymers* **1976,** *15*, (8), 1585-1590.
243. Benedetti, F.; Berti, F.; Norbedo, S., Epoxyalcohol Route to Hydroxyethylene Dipeptide Isosteres. Stereodivergent Synthesis of the Diamino Alcohol Core of Ritonavir and Its C-2 Epimer. *J. Org. Chem.* **2002,** *67*, (24), 8635-8643.
244. Vabeno, J.; Brisander, M.; Lejon, T.; Luthman, K., Diastereoselective Reduction of a Chiral *N*-Boc-Protected δ-Amino-α,β-unsaturated γ-Keto Ester Phe-Gly Dipeptidomimetic. *J. Org. Chem.* **2002,** *67*, (26), 9186-9191.
245. Corey, E. J.; Bakshi, R. K.; Shibata, S., Highly Enantioselective Borane Reduction of Ketones Catalyzed by Chiral Oxazaborolidines. Mechanism and Synthetic Implications. *J. Am. Chem. Soc.* **1987,** *109*, (18), 5551-5553.
246. Corey, E. J.; Shibata, S.; Bakshi, R. K., An Efficient and Catalytically Enantioselective Route to (*S*)-(−)-Phenyloxirane. *J. Org. Chem.* **1988,** *53*, (12), 2861-2863.
247. Noyori, R.; Tomino, I.; Tanimoto, Y.; Nishizawa, M., Asymmetric Synthesis via Axially Dissymmetric Molecules. 6. Rational Designing of Efficient Chiral Reducing Agents. Highly Enantioselective Reduction of Aromatic Ketones by Binaphthol-Modified Lithium Aluminum Hydride Reagents. *J. Am. Chem. Soc.* **1984,** *106*, (22), 6709-6716.
248. Noyori, R.; Tomino, I.; Yamada, M.; Nishizawa, M., Asymmetric Synthesis via Axially Dissymmetric Molecules. 7. Synthetic Applications of the Enantioselective Reduction by Binaphthol-Modified Lithium Aluminum Hydride Reagents. *J. Am. Chem. Soc.* **1984,** *106*, (22), 6717-6725.
249. Wang, X. J. Design, Synthesis and Bioactivities of Conformationally Locked Pin1 Ground State Inhibitors. Virginia Polytechnic Institute and State University, Blacksburg, VA, 2005.
250. El Fangour, S.; Guy, A.; Despres, V.; Vidal, J.-P.; Rossi, J.-C.; Durand, T., Total Synthesis of the Eight Diastereomers of the *syn-anti-syn* Phytoprostanes F1 Types I and II. *J. Org. Chem.* **2004,** *69*, (7), 2498-2503.
251. Reetz, M. T.; Li, X., An Efficient Catalyst System for the Asymmetric Transfer Hydrogenation of Ketones: Remarkably Broad Substrate Scope. *J. Am. Chem. Soc.* **2006,** *128*, (4), 1044-1045.
252. Goertz, W.; Kamer, P. C. J.; Leeuwen, P. W. N. M. v.; Vogt, D., Asymmetric Nickel-Catalyzed Hydrocyanation of Vinylarenes by Applying Homochiral Xantphos Ligands. *Chem. Eur. J.* **2001,** *7*, (8), 1614-1618.

253. van der Vlugt, J. I.; Paulusse, J. M. J.; Zijp, E. J.; Tijmensen, J. A.; Mills, A. M.; Spek, A. L.; Claver, C.; Vogt, D., Coordination Chemistry and Asymmetric Catalysis with a Chiral Diphosphonite. *Eur. J. Inorg. Chem.* **2004,** *2004*, (21), 4193-4201.
254. Mughal, R. K.; Davey, R. J.; Blagden, N., Application of Crystallization Inhibitors to Chiral Separations. 1. Design of Additives to Discriminate between the Racemic Compound and the Pure Enantiomer of Mandelic Acid. *Cryst. Growth Des.* **2007,** *7*, (2), 218-224.
255. Tsuji, T.; Kataoka, T.; Yoshioka, M.; Sendo, Y.; Nishitani, Y.; Hirai, S.; Maeda, T.; Nagata, W., Synthetic Studies on β-Lactam Antibiotics. VII. Mild Removal of the Benzyl Ester Protecting Group with Aluminum Trichloride. *Tetrahedron Lett.* **1979,** *20*, (30), 2793-2796.
256. Schmidt, U.; Kroner, M.; Griesser, H., Total Synthesis of the Didemnins; IV.1 Synthesis of the Peptolide Ring and Construction of the Side Chain. *Synthesis* **1991,** *1991*, (4), 294-300.
257. Huffman, W. F.; Hall, R. F.; Grant, J. A.; Holden, K. G., Nuclear Analogs of β-Lactam Antibiotics. 4. Total Synthesis of Bisnorisopenicillins from Antibacterially Active Monocyclic β-Lactam Precursor. *J. Med. Chem.* **1978,** *21*, (5), 413-415.
258. Paryzek, Z.; Koenig, H.; Tabaczka, B., Ammonium Formate/Palladium on Carbon: A Versatile System for Catalytic Hydrogen Transfer Reductions of Carbon - Carbon Double Bonds. *Synthesis* **2003,** *2003*, (13), 2023-2026.
259. Johnstone, R. A. W.; Wilby, A. H.; Entwistle, I. D., Heterogeneous Catalytic Transfer Hydrogenation and its Relation to Other Methods for Reduction of Organic Compounds. *Chem. Rev.* **1985,** *85*, (2), 129-170.
260. Ahmed, N.; van Lier, J. E., Pd-C/Ammonium Formate: A Selective Catalyst for the Hydrogenation of Chalcones to Dihydrochalcones. *J. Chem. Res.* **2006,** *2006*, (9), 584-585.
261. Etzkorn, F. A.; Guo, T.; Lipton, M. A.; Goldberg, S. D.; Bartlett, P. A., Cyclic Hexapeptides and Chimeric Peptides as Mimics of Tendamistat. *J. Am. Chem. Soc.* **1994,** *116*, (23), 10412-10425.
262. Felix, A. M.; Heimer, E. P.; Lambros, T. J.; Tzougraki, C.; Meienhofer, J., Rapid Removal of Protecting Groups from Peptides by Catalytic Transfer Hydrogenation with 1,4-Cyclohexadiene. *J. Org. Chem.* **1978,** *43*, (21), 4194-4196.
263. Koh, J. T.; Cornish, V. W.; Schultz, P. G., An Experimental Approach to Evaluating the Role of Backbone Interactions in Proteins Using Unnatural Amino Acid Mutagenesis. *Biochemistry* **1997,** *36*, (38), 11314-11322.
264. Bandur, N. G.; Harms, K.; Koert, U., First Stereoselective Synthesis of a Pro-Pro *E*-Alkene Dipeptide Isostere. *Synlett* **2005,** *2005*, (5), 773-776.
265. Williams, T. M.; Crumbie, R.; Mosher, H. S., dl-Isoserine and Related Compounds. *J. Org. Chem.* **1985,** *50*, (1), 91-97.
266. Hart, S. A.; Sabat, M.; Etzkorn, F. A., Enantio- and Regio-selective Synthesis of a (*Z*)-Alkene *cis*-Proline Mimic. *J. Org. Chem.* **1998,** *63*, (22), 7580-7581.
267. Chérest, M.; Felkin, H., Torsional Strain Involving Partial Bonds. The Steric Course of the Reaction between Allyl Magnesium Bromide and 4-*t*-Butyl-Cyclohexanone. *Tetrahedron Lett.* **1968,** *9*, (18), 2205-2208.
268. Chérest, M.; Felkin, H.; Prudent, N., Torsional Strain Involving Partial Bonds. The Stereochemistry of the Lithium Aluminium Hydride Reduction of Some Simple Open-Chain Ketones. *Tetrahedron Lett.* **1968,** *9*, (18), 2199-2204.

269. Anh, N. T.; Eisenstein, O., Theoretical Interpretation of 1,2-Asymmetric Induction. The Importance of Antiperiplanarity. *Nouv. J. Chim.* **1977,** *1*, (1), 61-70.
270. Luche, J. L., Lanthanides in Organic Chemistry. 1. Selective 1,2-Reductions of Conjugated Ketones. *J. Am. Chem. Soc.* **1978,** *100*, (7), 2226-2227.
271. Bächinger, H. P.; Davis, J. M., Sequence Specific Thermal Stability of the Collagen Triple Helix. *Int. J. Biol. Macromol.* **1991,** *13*, (3), 152-156.
272. Venugopal, M. G.; Ramshaw, J. A.; Braswell, E.; Zhu, D.; Brodsky, B., Electrostatic Interactions in Collagen-like Triple-Helical Peptides. *Biochemistry* **1994,** *33*, (25), 7948-7956.
273. Shah, N. K.; Brodsky, B.; Kirkpatrick, A.; Ramshaw, J. A., Structural Consequences of D-Amino Acids in Collagen Triple-Helical Peptides. *Biopolymers* **1999,** *49*, (4), 297-302.
274. Brandts, J. F.; Kaplan, L. J., Derivative Spectroscopy Applied to Tyrosyl Chromophores. Studies on Ribonuclease, Lima Bean Inhibitors, Insulin, and Pancreatic Trypsin Inhibitor. *Biochemistry* **1973,** *12*, (10), 2011-2024.
275. Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L., Aromatic Side-Chain Contribution to Far-Ultraviolet Circular Dichroism of Helical Peptides and Its Effect on Measurement of Helix Propensities. *Biochemistry* **1993,** *32*, (21), 5560-5565.
276. Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I., Color Test for Detection of Free Terminal Amino Groups in the Solid-Phase Synthesis of Peptides. *Anal. Biochem.* **1970,** *34*, (2), 595-598.
277. Kelly, S. M.; Jess, T. J.; Price, N. C., How to Study Proteins by Circular Dichroism. *Biochim. Biophys. Acta* **2005,** *1751*, (2), 119-139.
278. Kelly, S. M.; Price, N. C., The Use of Circular Dichroism in the Investigation of Protein Structure and Function. *Curr. Protein Pept. Sci.* **2000,** *1*, (4), 349-384.
279. Kelly, S. M.; Price, N. C., The Application of Circular Dichroism to Studies of Protein Folding and Unfolding. *Biochim. Biophys. Acta* **1997,** *1338*, (2), 161-185.
280. Greenfield, N. J., Methods to Estimate the Conformation of Proteins and Polypeptides from Circular Dichroism Data. *Anal. Biochem.* **1996,** *235*, (1), 1-10.
281. Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T., How to Measure and Predict the Molar Absorption Coefficient of a Protein. *Protein Sci.* **1995,** *4*, (11), 2411-2423.
282. Gill, S. C.; von Hippel, P. H., Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem.* **1989,** *182*, (2), 319-326.
283. Scott, R. A.; Scheraga, H. A., Structural Studies of Ribonuclease. XI. Kinetics of Denaturation. *J. Am. Chem. Soc.* **1963,** *85*, (23), 3866-3873.
284. Anfinsen, C. B., Principles That Govern the Folding of Protein Chains. *Science* **1973,** *181*, (4096), 223-230.
285. Long, C. G.; Braswell, E.; Zhu, D.; Apigo, J.; Baum, J.; Brodsky, B., Characterization of Collagen-like Peptides Containing Interruptions in the Repeating Gly-X-Y Sequence. *Biochemistry* **1993,** *32*, (43), 11688-11695.
286. Bruckner, P.; Bachinger, H. P.; Timpl, R.; Engel, J., Three Conformationally Distinct Domains in the Amino-Terminal Segment of Type III Procollagen and Its Rapid Triple Helix Leads to and Comes from Coil Transition. *Eur. J. Biochem.* **1978,** *90*, (3), 595-603.
287. Makarov, A. A.; Adzhubei, I. A.; Protasevich, I. I.; Lobachov, V. M.; Fasman, G. D., Melting of the Left-Handed Helical Conformation of Charged Poly-L-Lysine. *Biopolymers* **1994,** *34*, (8), 1123-1124.

288. Makarov, A. A.; Adzhubei, I. A.; Protasevich, I. I.; Lobachov, V. M.; Esipova, N. G., Scanning Microcalorimetry and Circular Dichroism Study of Melting of the Natural Polypeptides in the Left-Handed Helical Conformation. *J. Protein Chem.* **1993,** *12*, (1), 85-91.
289. Baskakov, I.; Bolen, D. W., Forcing Thermodynamically Unfolded Proteins to Fold. *J. Biol. Chem.* **1998,** *273*, (9), 4831-4834.
290. Liu, Y.; Bolen, D. W., The Peptide Backbone Plays a Dominant Role in Protein Stabilization by Naturally Occurring Osmolytes. *Biochemistry* **1995,** *34*, (39), 12884-12891.
291. Wang, A.; Bolen, D. W., A Naturally Occurring Protective Aystem in Urea-Rich Cells: Mechanism of Osmolyte Protection of Proteins against Urea Denaturation. *Biochemistry* **1997,** *36*, (30), 9101-9108.
292. Bolen, D. W.; Baskakov, I. V., The Osmophobic Effect: Natural Selection of a Thermodynamic Force in Protein Folding. *J. Mol. Biol.* **2001,** *310*, (5), 955-963.
293. Jiravanichanun, N.; Hongo, C.; Wu, G.; Noguchi, K.; Okuyama, K.; Nishino, N.; Silva, T., Unexpected Puckering of Hydroxyproline in the Guest Triplets, Hyp-Pro-Gly and Pro-alloHyp-Gly Aandwiched between Pro-Pro-Gly Sequence. *Chembiochem* **2005,** *6*, (7), 1184-1187.
294. McClellan, A. L., *Tables of Experimental Dipole Moments*. W. H. Freeman: San Francisco, 1963.
295. Nicholson, H.; Becktel, W. J.; Matthews, B. W., Enhanced Protein Thermostability from Designed Mutations that Interact with α-Helix Dipoles. *Nature* **1988,** *336*, (6600), 651-656.
296. Serrano, L.; Fersht, A. R., Capping and α-Helix Stability. *Nature* **1989,** *342*, (6247), 296-299.
297. Ramachandran, G. N.; Doyle, B. B.; Bloot, E. R., Single-Chain Triple Helical Structure. *Biopolymers* **1968,** *6*, (12), 1771-1775.
298. Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A., Crystal Structure of a Collagen-like Polypeptide with Repeating Sequence Pro-Hyp-Gly at 1.4 Å Resolution: Implications for Collagen Hydration. *Biopolymers* **2001,** *56*, (1), 8-13.
299. Desiraju, G. R.; Steiner, T., *The Weak Hydrogen Bond: In Structural Chemistry and Biology*. Oxford University Press: Oxford, 1999.
300. Hinderaker, M. P.; Raines, R. T., An Electronic Effect on Protein Structure. *Protein Sci.* **2003,** *12*, (6), 1188-1194.
301. Okuyama, K.; Hongo, C.; Fukushima, R.; Wu, G.; Narita, H.; Noguchi, K.; Tanaka, Y.; Nishino, N., Crystal Structures of Collagen Model Peptides with Pro-Hyp-Gly Repeating Sequence at 1.26 Å Resolution: Implications for Proline Ring Puckering. *Biopolymers* **2004,** *76*, (5), 367-377.
302. Paramonov, S. E.; Gauba, V.; Hartgerink, J. D., Synthesis of Collagen-like Peptide Polymers by Native Chemical Ligation. *Macromolecules* **2005,** *38*, (18), 7555-7561.
303. Kishimoto, T.; Morihara, Y.; Osanai, M.; Ogata, S.-i.; Kamitakahara, M.; Ohtsuki, C.; Tanihara, M., Synthesis of Poly(Pro-Hyp-Gly)$_n$ by Direct Polycondensation of (Pro-Hyp-Gly)$_n$, Where n = 1, 5, and 10, and Stability of the Triple-Helical Structure. *Biopolymers* **2005,** *79*, (3), 163-172.
304. Barton, M. A.; Singh, B.; Fraga, E., Synthetic Polypeptide Antigens of Defined Geometry. *J. Am. Chem. Soc.* **1977,** *99*, (26), 8491-8498.

305. DeTar, D. F.; Honsberg, W.; Honsberg, U.; Wieland, A.; Gouge, M.; Bach, H.; Tahara, A.; Brinigar, W. S.; F. F. Rogers, Jr., Synthesis of Peptide Polymers with Repeating Sequences. *J. Am. Chem. Soc.* **1963,** *85*, (18), 2873-2874.
306. DeTar, D. F.; Vajda, T., Sequence Peptide Polymers. II. Poly Glu(OH)-Gly, Poly Glu(OH)-Ser(H)-Gly, and Their Benzyl Esters. *J. Am. Chem. Soc.* **1967,** *89*, (4), 998-1004.
307. Bloom, S. M.; Dasgupta, S. K.; Patel, R. P.; Blout, E. R., The Synthesis of Glycyl-L-prolylglycyl and Glycyl-L-prolyl-L-alanyl Oligopeptides and Sequential Polypeptides. *J. Am. Chem. Soc.* **1966,** *88*, (9), 2035-2041.
308. Kovacs, J.; Giannotti, R.; Kapoor, A., Polypeptides with Known Repeating Sequence of Amino Acids. Synthesis of Poly-L-Glutamyl-L-Alanyl-L-Glutamic Acid and Polyglycylglycyl-L-Phenylalanine through Pentachlorophenyl Active Ester. *J. Am. Chem. Soc.* **1966,** *88*, (10), 2282-2292.
309. Becktel, W. J.; Schellman, J. A., Protein Stability Curves. *Biopolymers* **1987,** *26*, (11), 1859-1877.
310. Horng, J. C.; Hawk, A. J.; Zhao, Q.; Benedict, E. S.; Burke, S. D.; Raines, R. T., Macrocyclic Scaffold for the Collagen Triple Helix. *Org. Lett.* **2006,** *8*, (21), 4735-4738.
311. Barlow, D. J.; Thornton, J. M., Helix Geometry in Proteins. *J. Mol. Biol.* **1988,** *201*, (3), 601-619.
312. Parthasarathy, R.; Chaturvedi, S.; Go, K., Design of α-Helical Peptides: Their Role in Protein Folding and Molecular Biology. *Prog. Biophys. Molec. Biol.* **1995,** *64*, (1), 1-54.
313. Kim, P. S.; Baldwin, R. L., A Helix Stop Signal in the Isolated S-Peptide of Ribonuclease A. *Nature* **1984,** *307*, (5949), 329-334.
314. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V., Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963,** *7*, 95-99.
315. Blanc, J. P.; Kaiser, E. T., Biological and Physical Properties of a β-Endorphin Analog Containing Only D-Amino Acids in the Amphiphilic Helical Segment 13-31. *J. Biol. Chem.* **1984,** *259*, (15), 9549-9556.
316. Branden, C.; Tooze, J., *Introduction to Protein Structure*. Garland Publishing Inc.: New York and London, 1991.
317. Errington, N.; Iqbalsyah, T. M.; Doig, A. J., Structure and Stability of the α-Helix, Lessons for Design. In *Protein Design: Methods and Applications*, Guerois, R.; de la Paz, M. L., Eds. Humana Press: Totowa, NJ, 2006.
318. Petsko, G. A.; Ringe, D., *Protein Structure and Function*. New Science Press: Waltham, MA, 2004.
319. Glover, I.; Haneef, I.; Pitts, J.; Wood, S.; Moss, D.; Tickle, I.; Blundell, T., Conformational Flexibility in a Small Globular Hormone: X-Ray Analysis of Avian Pancreatic Polypeptide at 0.98 Å Resolution. *Biopolymers* **1983,** *22*, (1), 293-304.
320. Doig, A. J.; Errington, N.; Iqbalsyah, T. M., Stability and Design of α-Helices. In *Protein Folding Handbook*, Buchner, J.; Kiefhaber, T., Eds. Wiley-VCH: Weinheim, 2005; Vol. 1, pp 247-313.
321. Guzzo, A. V., The Influence of Amino Acid Sequence on Protein Structure. *Biophys. J.* **1965,** *5*, (6), 809-822.
322. Davies, D. R., A Correlation Between Amino Acid Composition and Protein Structure. *J. Mol. Biol.* **1964,** *9*, 605-609.
323. Levitt, M., Conformational Preferences of Amino Acids in Globular Proteins. *Biochemistry* **1978,** *17*, (20), 4277-4285.

324. Chou, P. Y.; Fasman, G. D., Conformational Parameters for Amino Acids in Helical, Sheet, and Random Coil Regions Calculated from Proteins. *Biochemistry* **1974,** *13*, (2), 211-222.
325. Chou, P. Y.; Fasman, G. D., Prediction of Protein Conformation. *Biochemistry* **1974,** *13*, (2), 222-245.
326. Park, S. H.; Shalongo, W.; Stellwagen, E., Residue Helix Parameters Obtained from Dichroic Analysis of Peptides of Defined Sequence. *Biochemistry* **1993,** *32*, (27), 7048-7053.
327. Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R., Side Chain Contributions to the Stability of α-Helical Structure in Peptides. *Science* **1990,** *250*, (4981), 669-673.
328. Lyu, P. C.; Wang, P. C.; Liff, M. I.; Kallenbach, N. R., Local Effect of Glycine Substitution in a Model Helical Peptide. *J. Am. Chem. Soc.* **1991,** *113*, (9), 3568-3572.
329. Myers, J. K.; Pace, C. N.; Scholtz, J. M., Helix Propensities Are Identical in Proteins and Peptides. *Biochemistry* **1997,** *36*, (36), 10923 -10929.
330. Myers, J. K.; Pace, C. N.; Scholtz, J. M., A Direct Comparison of Helix Propensity in Proteins and Peptides. *Proc. Natl. Acad. Sci. USA* **1997,** *94*, (7), 2833-2837.
331. Pace, C. N.; Scholtz, J. M., A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophy. J.* **1998,** *75*, (1), 422-427.
332. Lifson, S.; Roig, A., Theory of Helix-Coil Transition in Polypeptides. *J. Chem. Phys.* **1961,** *34*, (6), 1963-1974.
333. Chakrabartty, A.; Doig, A. J.; Baldwin, R. L., Helix Capping Propensities in Peptides Parallel Those in Proteins. *Proc. Natl. Acad. Sci. USA* **1993,** *90*, (23), 11332-11336.
334. Doig, A. J.; Chakrabartty, A.; Klingler, T. M.; Baldwin, R. L., Determination of Free Energies of *N*-Capping in α-Helices by Modification of the Lifson-Roig Helix-Coil Theory To Include *N*- and *C*-Capping. *Biochemistry* **1994,** *33*, (11), 3396-3403.
335. Stapley, B. J.; Rohl, C. A.; Doig, A. J., Addition of Side Chain Interactions to Modified Lifson-Roig Helix-Coil Theory: Application to Energetics of Phenylalanine-Methionine Interactions. *Protein Sci.* **1995,** *4*, (11), 2383-2391.
336. Zimm, B. H.; Bragg, J. K., Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. *J. Chem. Phys.* **1959,** *31*, (2), 526-535.
337. Schellman, J. A., The Stability of Hydrogen-Bonded Peptide Structures in Aqueous Solution. *C. R. Trav. Lab. Carlsberg [Chim]* **1955,** *29*, (14-15), 230-259.
338. Baldwin, R. L.; Rose, G. D., Is Protein Folding Hierarchic? II. Folding Intermediates and Transition States. *Trends Biochem. Sci.* **1999,** *24*, (2), 77-83.
339. Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W., Calorimetric Determination of the Enthalpy Change for the α-Helix to Coil Transition of an Alanine Peptide in Water. *Proc. Natl. Acad. Sci. USA* **1991,** *88*, (7), 2854-2858.
340. Fernández, A.; Colubri, A.; Stephen Berry, R., Three-Body Correlations in Protein Folding: The Origin of Cooperativity. *Physica A* **2002,** *307*, (1-2), 235-259.
341. Richardson, J. M.; Lopez, M. M.; Makhatadze, G. I., Enthalpy of Helix–Coil Transition: Missing Link in Rationalizing the Thermodynamics of Helix-Forming Propensities of the Amino Acid Residues. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, (5), 1413-1418.
342. Myers, J. K.; Pace, C. N., Hydrogen Bonding Stabilizes Globular Proteins. *Biophy. J.* **1996,** *71*, (4), 2033-2039.

343. Fleming, P. J.; Rose, G. D., Do All Backbone Polar Groups in Proteins Form Hydrogen Bonds? *Protein Sci.* **2005,** *14*, (7), 1911-1917.
344. Kresheck, G. C.; Klotz, I. M., Thermodynamics of Transfer of Amides from an Apolar to an Aqueous Solution. *Biochemistry* **1969,** *8*, (1), 8-12.
345. Ben-Tal, N.; Sitkoff, D.; Topol, I. A.; Yang, A.-S.; Burt, S. K.; Honig, B., Free Energy of Amide Hydrogen Bond Formation in Vacuum, in Water, and in Liquid Alkane Solution. *J. Phys. Chem. B* **1997,** *101*, (3), 450-457.
346. Mitchell, J. B. O.; Price, S. L., On the Relative Strengths of Amide···Amide and Amide···Water Hydrogen-Bonds. *Chem. Phys. Lett.* **1991,** *180*, (6), 517-523.
347. Mitchell, J. B. O.; Price, S. L., The Nature of the N-H···O=C Hydrogen Bond: An Intermolecular Perturbation Theory Study of the Formamide / Formaldehyde Complex. *J. Comp. Chem.* **1990,** *11*, (11), 1217-1233.
348. Makhatadze, G. I.; Privalov, P. L., Contribution of Hydration to Protein Folding Thermodynamics. I. The Enthalpy of Hydration. *J. Mol. Biol.* **1993,** *232*, (2), 639-659.
349. Hol, W. G. J., The Role of the α-Helix Dipole in Protein Function and Structure. *Prog. Biophys. Molec. Biol.* **1985,** *45*, (3), 149-195.
350. Aurora, R.; Rose, G. D., Helix Capping. *Protein Sci.* **1998,** *7*, (1), 21-38.
351. Richardson, J. S.; Richardson, D. C., Amino Acid Preferences for Specific Locations at the Ends of α-Helices. *Science* **1988,** *240*, (4859), 1648-1652.
352. Presta, L. G.; Rose, G. D., Helix Signals in Proteins. *Science* **1988,** *240*, (4859), 1632-1641.
353. McGregor, M. J.; Islam, S. A.; Sternberg, M. J. E., Analysis of the Relationship between Side-Chain Conformation and Secondary Structure in Globular Proteins. *J. Mol. Biol.* **1987,** *198*, (2), 295-310.
354. Dunbrack, R. L.; Karplus, M., Conformational Analysis of the Backbone-Dependent Rotamer Preferences of Protein Sidechains. *Nat. Struct. Biol.* **1994,** *1*, (5), 334-340.
355. Swindells, M. B.; MacArthur, M. W.; Thornton, J. M., Intrinsic ϕ, ψ Propensities of Amino Acids, Derived from the Coil Regions of Known Structures. *Nat. Struct. Biol.* **1995,** *2*, (7), 596-603.
356. Doig, A. J.; MacArthur, M. W.; Stapley, B. J.; Thornton, J. M., Structures of *N*-Termini of Helices in Proteins. *Protein Sci.* **1997,** *6*, (1), 147-155.
357. Kumar, S.; Nussinov, R., Close-Range Electrostatic Interactions in Proteins. *Chembiochem* **2002,** *3*, (7), 604-617.
358. Ermolenko, D. N.; Thomas, S. T.; Aurora, R.; Gronenborn, A. M.; Makhatadze, G. I., Hydrophobic Interactions at the Ccap Position of the *C*-Capping Motif of α-Helices. *J. Mol. Biol.* **2002,** *322*, (1), 123-135.
359. Aurora, R.; Srinivasan, R.; Rose, G. D., Rules for α-Helix Termination by Glycine. *Science* **1994,** *264*, (5162), 1126-1130.
360. Berezovsky, I. N.; Kilosanidze, G. T.; Tumanyan, V. G.; Kisselev, L. L., Amino Acid Composition of Protein Termini are Biased in Different Manners. *Protein Eng.* **1999,** *12*, (1), 23-30.
361. Bhattacharyya, R.; Pal, D.; Chakrabarti, P., Secondary Structures at Polypeptide-Chain Termini and Their Features. *Acta Crystallogr. D Biol. Crystallogr.* **2002,** *D58*, (2), 1793-1802.
362. Doig, A. J.; Baldwin, R. L., *N*- and *C*-Capping Preferences for All 20 Amino Acids in α-Helical Peptides. *Protein Sci.* **1995,** *4*, (7), 1325-1335.

363. Forood, B.; Feliciano, E. J.; Nambiar, K. P., Stabilization of α-Helical Structures in Short Peptides via End Capping. *Proc. Natl. Acad. Sci. USA* **1993,** *90*, (3), 838-842.
364. Marqusee, S.; Baldwin, R. L., Helix Stabilization by $Glu^{-}\cdots Lys^{+}$ Salt Bridges in Short Peptides of *De Novo* Design. *Proc. Natl. Acad. Sci. USA* **1987,** *84*, (24), 8898-8902.
365. Butterfield, S. M.; Patel, P. R.; Waters, M. L., Contribution of Aromatic Interactions to α-Helix Stability. *J. Am. Chem. Soc.* **2002,** *124*, (33), 9751-9755.
366. Shi, Z.; Olson, C. A.; Jr., A. J. B.; Kallenbach, N. R., Stabilization of α-Helix Structure by Polar Side-Chain Interactions: Complex Salt Bridges, Cation-π Interactions, and C-H···O H-Bonds. *Biopolymers* **2001,** *60*, (5), 366-380.
367. Andrew, C. D.; Bhattacharjee, S.; Kokkoni, N.; Hirst, J. D.; Jones, G. R.; Doig, A. J., Stabilizing Interactions between Aromatic and Basic Side Chains in α-Helical Peptides and Proteins. Tyrosine Effects on Helix Circular Dichroism. *J. Am. Chem. Soc.* **2002,** *124*, (43), 12706-12714.
368. Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W., pH-Induced Denaturation of Proteins: A Single Salt Bridge Contributes 3-5 kcal/mol to the Free Energy of Folding of T4 Lysozyme. *Biochemistry* **1990,** *29*, (9), 2403-2408.
369. Lyu, P. C.; Marky, L. A.; Kallenbach, N. R., The Role of Ion Pairs in α-Helix Stability: Two New Designed Helical Peptides. *J. Am. Chem. Soc.* **1989,** *111*, (7), 2733-2734.
370. Merutka, G.; Stellwagen, E., Effect of Amino Acid Ion Pairs on Peptide Helicity. *Biochemistry* **1991,** *30*, (6), 1591-1594.
371. Huyghues-Despointes, B. M. P.; Klingler, T. M.; Baldwin, R. L., Measuring the Strength of Side-Chain Hydrogen Bonds in Peptide Helixes: The Gln···Asp (*i*, *i* + 4) Interaction. *Biochemistry* **1995,** *34*, (41), 13267-13271.
372. Smith, J. S.; Scholtz, J. M., Energetics of Polar Side-Chain Interactions in Helical Peptides: Salt Effects on Ion Pairs and Hydrogen Bonds. *Biochemistry* **1998,** *37*, (1), 33-40.
373. Lyu, P. C.; Gans, P. J.; Kallenbach, N. R., Energetic Contribution of Solvent-Exposed Ion Pairs to α-Helix Structures. *J. Mol. Biol.* **1992,** *223*, (1), 343-350.
374. Fairman, R.; Shoemaker, K. R.; York, E. J.; Stewart, J. M.; Baldwin, R. L., The Glu $2^{-}\cdots$Arg 10^{+} Side-Chain Interaction in the *C*-Peptide Helix of Ribonuclease A. *Biophys. Chem.* **1990,** *37*, (1-3), 107-119.
375. Kumar, S.; Nussinov, R., Relationship between Ion Pair Geometries and Electrostatic Strengths in Proteins. *Biophys. J.* **2002,** *83*, (3), 1595-1612.
376. Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. K., Strength of Solvent-Exposed Salt-Bridges. *J. Phys. Chem. B* **1999,** *103*, (4), 727-736.
377. Luo, P.; Baldwin, R. L., Interaction Between Water and Polar Groups of the Helix Backbone: An Important Determinant of Helix Propensities. *Proc. Natl. Acad. Sci. USA* **1999,** *96*, (9), 4942-4946.
378. Huyghues-Despointes, B. M. P.; Baldwin, R. L., Ion-Pair and Charged Hydrogen-Bond Interactions Between Histidine and Aspartate in a Peptide Helix. *Biochemistry* **1997,** *36*, (8), 1965-1970.
379. Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L., The Energetic of Ion-Pair hand Hydrogen-Bonding Interactions in a Helical Peptide. *Biochemistry* **1993,** *32*, (37), 9668-9676.
380. Marqusee, S.; Sauer, R. T., Contributions of a Hydrogen Bond/Salt Bridge Network to the Stability of Secondary and Tertiary Structure in λ-Repressor. *Protein Sci.* **1994,** *3*, (12), 2217-2225.

381. Errington, N.; Doig, A. J., A Phosphoserine-Lysine Salt Bridge within an α-Helical Peptide, the Strongest α-Helix Side-Chain Interaction Measured to Date. *Biochemistry* **2005,** *44*, (20), 7553-7558.
382. Liehr, S.; Chenault, H. K., A Comparison of the α-Helix Forming Propensities and Hydrogen Bonding Properties of Serine Phosphate and α-Amino-γ-Phosphonobutyric Acid. *Bioorg. Med. Chem. Lett.* **1999,** *9*, (18), 2759-2762.
383. Huyghues-Despointes, B. M. P.; Scholtz, J. M.; Baldwin, R. L., Helical Peptides with Three Pairs of Asp-Arg and Glu-Arg Residues in Different Orientations and Spacings. *Protein Sci.* **1993,** *2*, (1), 80-85.
384. Horovitz, A.; Serrano, L.; Avron, B.; Bycroft, M.; Fersht, A. R., Strength and Co-operativity of Contributions of Surface Salt Bridges to Protein Stability. *J. Mol. Biol.* **1990,** *216*, (4), 1031-1044.
385. Spek, E. J.; Bui, A. H.; Lu, M.; Kallenbach, N. R., Surface Salt Bridges Stabilize the GCN4 Leucine Zipper. *Protein Sci.* **1998,** *7*, (11), 2431-2437.
386. Iqbalsyah, T. M.; Doig, A. J., Anticooperativity in a Glu-Lys-Glu Salt Bridge Triplet in an Isolated α-Helical Peptide. *Biochemistry* **2005,** *44*, (31), 10449-10456.
387. Armstrong, K. M.; Baldwin, R. L., Affects α-Helix Stability at All Positions in the Helix by Interacting with the Backbone Charges. *Proc. Natl. Acad. Sci. USA* **1993,** *90*, (23), 11337-11340.
388. Avbelj, F.; Luo, P.; Baldwin, R. L., Energetics of the Interaction between Water and the Helical Peptide Group and Its Role in Determing Helix Propensities. *Proc. Natl. Acad. Sci. USA* **2000,** *97*, (20), 10786-10791.
389. Stapley, B. J.; Doig, A. J., Hydrogen Bonding Interactions between Glutamine and Asparagine in α-Helical Peptides. *J. Mol. Biol.* **1997,** *272*, (3), 465-473.
390. Wu, X.; Wang, S., Helix Folding of an Alanine-Based Peptide in Explicit Water. *J. Phys. Chem. B* **2001,** *105*, (11), 2227-2235.
391. Dill, K. A., Dominant Forces in Protein Folding. *Biochemistry* **1990,** *29*, (31), 7133-7155.
392. Padmanabhan, S.; Baldwin, R. L., Helix-Stabilizing Interaction Between Tyrosine and Leucine or Valine when the Spacing is *i*, *i*+4. *J. Mol. Biol.* **1994,** *241*, (5), 706-713.
393. ten Wolde, P. R., Hydrophobic Interactions: An Overview. *J. Phys. Condens. Matter* **2002,** *14*, (40), 9445-9460.
394. Fernández, A.; Scheraga, H. A., Insufficiently Dehydrated Hydrogen Bonds as Determinants of Protein Interactions. *Proc. Natl. Acad. Sci. USA* **2003,** *100*, (1), 113-118.
395. Creamer, T. P.; Rose, G. D., Interactions between Hydrophobic Side Chains within α-Helices. *Protein Sci.* **1995,** *4*, (7), 1305-1314.
396. Padmanabhan, S.; Baldwin, R. L., Tests for Helix-Stabilizing Interactions Between Various Nonpolar Side Chains in Alanine-Based Peptides. *Protein Sci.* **1994,** *3*, (11), 1992-1997.
397. Padmanabhan, S.; Jiménez, M. A.; Laurents, D. V.; Rico, M., Helix-Stabilizing Nonpolar Interactions between Tyrosine and Leucine in Aqueous and TFE Solutions: 2D-^{1}H NMR and CD Studies in Alanine-Lysine Peptides. *Biochemistry* **1998,** *37*, (49), 17318-17330.
398. Luo, P.; Baldwin, R. L., Origin of the Different Strengths of the (*i*, *i*+4) and (*i*, *i*+3) Leucine Pair Interactions in Helices. *Biophys. Chem.* **2002,** *96*, (2-3), 103-108.

399. Shalongo, W.; Stellwagen, E., Incorporation of Pairwise Interactions into the Lifson-Roig Model for Helix Prediction. *Protein Sci.* **1995,** *4*, (6), 1161-1166.
400. Armstrong, K. M.; Fairman, R.; Baldwin, R. L., The (*i*, *i*+4) Phe-His Interaction Studied in an Alanine-Based α-Helix. *J. Mol. Biol.* **1993,** *230*, (1), 284-291.
401. Gallivan, J. P.; Dougherty, D. A., Cation-π Interactions in Structural Biology. *Proc. Natl. Acad. Sci. USA* **1999,** *96*, (17), 9459-9464.
402. Stenkamp, R. E.; Sieker, L. C.; Jensen, L. H., Crystallographic Studies of Azide, Thiocyanate and Perchlorate Complexes of Methemerythrin. *J. Mol. Biol.* **1978,** *126*, (3), 457-466.
403. Deakyne, C. A.; Meot-Ner, M., Unconventional Ionic Hydrogen Bonds. 2. $NH^{+}\cdots\pi$ Complexes of Onium Ions with Olefins and Benzene Dericatives. *J. Am. Chem. Soc.* **1985,** *107*, (2), 474-479.
404. Meot-Ner, M.; Deakyne, C. A., Unconventional Ionic Hydrogen Bonds. 1. $CH\delta^{+}\cdots X$ Complexes of Quaternary Ions with n- and π-Donors. *J. Am. Chem. Soc.* **1985,** *107*, (2), 469-474.
405. Perutz, M. F.; Fermi, G.; Abraham, D. J.; Poyart, C.; Bursaux, E., Hemoglobin as a Receptor of Drugs and Peptides: X-Ray Studies of the Stereochemistry of Binding. *J. Am. Chem. Soc.* **1986,** *108*, (5), 1064-1078.
406. Burley, S. K.; Petsko, G. A., Amino-Aromatic Interactions in Proteins. *FEBS Lett.* **1986,** *203*, (2), 139-143.
407. Okada, A.; Miura, T.; Takeuchi, H., Protonation of Histidine and Histidine-Tryptophan Interaction in the Activation of the M2 Ion Channel from Influenza A Virus. *Biochemistry* **2001,** *40*, (20), 6053-6060.
408. Gallivan, J. P.; Dougherty, D. A., A Computational Study of Cation-π Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering. *J. Am. Chem. Soc.* **2000,** *122*, (5), 870-874.
409. Tsou, L. K.; Tatko, C. D.; Waters, M. L., Simple Cation-π Interaction between a Phenyl Ring and a Protonated Amine Stabilizes an α-Helix in Water. *J. Am. Chem. Soc.* **2002,** *124*, (50), 14917-14921.
410. Dougherty, D. A., Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr and Trp. *Science* **1996,** *271*, (5246), 163-168.
411. Ma, J. C.; Dougherty, D. A., The Cation-π Interaction. *Chem. Rev.* **1997,** *97*, (5), 1303-1324.
412. Shepodd, T. J.; Petti, M. A.; Dougherty, D. A., Molecular Recognition in Aqueous Media: Donor-Acceptor and Ion-Dipole Interactions Produce Tight Binding for Highly Soluble Guests. *J. Am. Chem. Soc.* **1988,** *110*, (6), 1983-1985.
413. Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R., The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and its Complexes with Inhibitors. *Acta Crystallogr., Sect. B* **1983,** *39*, (4), 480-490.
414. Gorina, S.; Pavletich, N. P., Structure of the P53 Tumor Suppressor Bound to the Ankyrin and SH3 Domains of 53BP2. *Science* **1996,** *274*, (5289), 1001-1005.
415. Hubbard, S. R.; Wei, L.; Ellis, L.; Hendrickson, W. A., Crystal Structure of the Tyrosine Kinase Domain of the Human Insulin Receptor. *Nature* **1994,** *372*, (6508), 746-754.
416. Brocchieri, L.; Karlin, S., Geometry of Interplanar Residue Contacts in Protein Structures. *Proc. Natl. Acad. Sci. USA* **1994,** *91*, (20), 9297-9301.

417. Flocco, M. M.; Mowbray, S. L., Planar Stacking Interactions of Arginine and Aromatic Side-Chains in Proteins. *J. Mol. Biol.* **1994,** *235*, (2), 709-717.
418. Shi, Z.; Olson, C. A.; Kallenbach, N. R., Cation-π Interaction in Model α-Helical Peptides. *J. Am. Chem. Soc.* **2002,** *124*, (13), 3284-3291.
419. Olympia, P. L., Some Theoretical Aspects of C-H···O Bonding. *Chem. Phys. Lett.* **1970,** *5*, (9), 593-596.
420. Taylor, R.; Kennard, O., Crystallographic Evidence for the Existence of C-H···O, C-H···N and C-H···Cl Hydrogen Bonds. *J. Am. Chem. Soc.* **1982,** *104*, (19), 5063-5070.
421. Derewenda, Z. S.; Lee, L.; Derewenda, U., The Occurence of C-H···O Hydrogen Bonds in Proteins. *J. Mol. Biol.* **1995,** *252*, (2), 248-262.
422. Weiss, M. S.; Brandl, M.; Sühnel, J.; Pal, D.; Hilgenfeld, R., More Hydrogen Bonds for the (Structural) Biologist. *Trends Biochem. Sci.* **2001,** *26*, (9), 521-523.
423. Thomas, K. A.; Smith, G. M.; Thomas, T. B.; Feldmann, R. J., Electronic Distributions within Protein Phenylalanine Aromatic Rings Are Reflected by the Three-Dimensional Oxygen Atom Environments. *Proc. Natl. Acad. Sci. USA* **1982,** *79*, (16), 4843-4847.
424. Steiner, T.; Koellner, G., Hydrogen Bonds with π-Acceptors in Proteins: Frequencies and Role in Stabilizing 3D Structures. *J. Mol. Biol.* **2001,** *305*, (3), 535-557.
425. Hough, E.; Hansen, L. K.; Birknes, B.; Jynge, K.; Hansen, S.; Hordvik, A.; Little, C.; Dodson, E.; Derewenda, Z., High-Resolution (1.5 Å) Crystal Structure of Phospholipase C from Bacillus Cereus. *Nature* **1989,** *338*, (6213), 357-360.
426. Chakrabarti, P.; Chakrabarti, S., C-H···O Hydrogen Bonds Involving Proline Residues in α-Helices. *J. Mol. Biol.* **1998,** *284*, (4), 867-873.
427. Shi, Z.; Olson, C. A.; Bell, A. J.; Kallenbach, N. R., Non-Classical Helix-Stabilizing Interactions: C-H···O H-Bonding between Phe and Glu Side Chains in α-Helical Peptides. *Biophys. Chem.* **2002,** *101-102*, 267-279.
428. Scheiner, S.; Kar, T.; Gu, Y., Strength of the Cα-H···O Hydrogen Bond of Amino Acid Residues. *J. Biol. Chem.* **2001,** *276*, (13), 9832-9837.
429. Forristal, I.; Lowman, J.; Afarinkia, K.; Steed, J. W., Helical Structure Held Together by Phosphonate C-H···O Hydrogen Bonds. *CrystEngComm* **2001,** *3*, 53-56.
430. Hill, R. B.; Raleigh, D. P.; Lombardi, A.; DeGrado, W. F., *De Novo* Design of Helical Bundles as Models for Understanding Protein Folding and Function. *Acc. Chem. Res.* **2000,** *33*, (11), 745-754.
431. Scholtz, J. M.; Barrick, D.; York, E. L.; Stewart, J. M.; Baldwin, R. L., Urea Unfolding of Peptide Helices as a Model for Interpreting Protein Unfolding. *Proc. Natl. Acad. Sci. USA* **1995,** *92*, (1), 185-189.
432. Kohtani, M.; Jarrold, M. F., The Initial Steps in the Hydration of Unsolvated Peptides: Water Molecule Adsorption on Alanine-Based Helices and Globules. *J. Am. Chem. Soc.* **2002,** *124*, (37), 11148-11158.
433. Spek, E. J.; Olson, C. A.; Shi, Z.; Kallenbach, N. R., Alanine is an Intrinsic α-Helix Stabilizing Amino Acid. *J. Am. Chem. Soc.* **1999,** *121*, (23), 5571-5572.
434. Marqusee, S.; Robbins, V. H.; Baldwin, R. L., Usually Stable Helix Formation in Short Alanine-Based Peptides. *Proc. Natl. Acad. Sci. USA* **1989,** *86*, (14), 5286-5290.
435. Munoz, V.; Serrano, L., Helix Design, Prediction and Stability. *Curr. Opin. Biotechnol.* **1995,** *6*, (4), 382-386.

436. Thompson, P. A.; Eaton, W. A.; Hofrichter, J., Laser Temperature Jump Study of the Helix-Coil Kinetics of an Alanine Peptide Interpreted with a 'Kinetic Zipper' Model. *Biochemistry* **1997,** *36*, (30), 9200-9210.
437. Matthews, B. W.; Ohlendorf, D. H.; Anderson, W. F.; Takeda, Y., Structure of the DNA-Binding Region of lac Repressor Inferred from Its Homology with Cro Repressor. *Proc. Natl. Acad. Sci. USA* **1982,** *79*, (5), 1428-1432.
438. Steitz, T. A.; Ohlendorf, D. H.; McKay, D. B.; Anderson, W. F.; Matthews, B. W., Structural Similarity in the DNA-Binding Domains of Catabolite Gene Activator and Cro Repressor Proteins. *Proc. Natl. Acad. Sci. USA* **1982,** *79*, (10), 3097-3100.
439. Ohlendorf, D. H.; Anderson, W. F.; Fisher, R. G.; Takeda, Y.; Matthews, B. W., The Molecular Basis of DNA−Protein Recognition Inferred from the Structure of Cro Repressor. *Nature* **1982,** *298*, (5876), 718-723.
440. Sauer, R. T.; Yocum, R. R.; Doolittle, R. F.; Lewis, M.; Pabo, C. O., Homology among DNA-Binding Proteins Suggests Use of a Conserved Super-Secondary Structure. *Nature* **1982,** *298*, (5873), 447-451.
441. Ohlendorf, D. H.; Anderson, W. F.; Matthews, B. W., Many Gene-Regulatory Proteins Appear to Have a Similar α-Helical Fold that Binds DNA and Evolved from a Common Precursor. *J. Mol. Evol.* **1983,** *19*, (2), 109-114.
442. Aravind, L.; Anantharaman, V.; Balaji, S.; Babu, M. M.; Iyer, L. M., The Many Faces of the Helix-Turn-Helix Domain: Transcription Regulation and Beyond. *FEMS Microbiol. Rev.* **2005,** *29*, (2), 231-262.
443. Brennan, R. G.; Matthews, B. W., The Helix-Turn-Helix DNA Binding Motif. *J. Biol. Chem.* **1989,** *264*, (4), 1903-1906.
444. Chou, P. Y.; Fasman, G. D., β-Turns in Proteins. *J. Mol. Biol.* **1977,** *115*, (2), 135-175.
445. Etzkorn, F. A.; Hart, S. A.; Travins, J. M., Rare Protein Turns: Helix-Turn-Helix, γ-Turn and *cis*-Proline Mimics. In *Advances Amino Acid Mimetics and Peptidomimetics*, Abell, A., Ed. JAI Press Inc.: Greenwich, CT, 1999; Vol. 2, pp 125-163.
446. Milner-White, E. J.; Ross, B. M.; Ismail, R.; Belhadj-Mostefa, K.; Poet, R., One Type of γ-Turn, Rather Than the Other Gives Rise to Chain-Reversal in Proteins. *J. Mol. Biol.* **1988,** *204*, (3), 777-782.
447. Milner-White, E. J., Situations of γ-Turns in proteins: Their Relation to α-Helices, β-Sheets and Ligand Binding Sites. *J. Mol. Biol.* **1990,** *216*, (2), 385-397.
448. Travins, J. M.; Etzkorn, F. A., Design and Enantioselective Synthesis of a Peptidomimetic of the Helix-Turn-Helix DNA-Binding Protein Motif. *J. Org. Chem.* **1997,** *62*, (24), 8387-8393.
449. Richardson, J. S., The Anatomy and Taxonomy of Protein Structure. *Adv. Protein Chem.* **1981,** *34*, 167-339.
450. Venkatachalam, C. M., Stereochemical Criteria for Polypeptides and Proteins. V. Conformation of a System of Three Linked Peptide Units. *Biopolymers* **1968,** *6*, (10), 1425-1436.
451. Smith, J. A.; Pease, L. G., Reverse Turns in Peptides and Proteins. *CRC Crit. Rev. Biochem.* **1980,** *8*, (4), 315-399.
452. Vass, E.; Hollosi, M.; Besson, F.; Buchet, R., Vibrational Spectroscopic Detection of β- and γ-Turns in Synthetic and Natural Peptides and Proteins. *Chem. Rev.* **2003,** *103*, (5), 1917-1954.

453. Gierasch, L. M.; Deber, C. M.; Madison, V.; Niu, C.-H.; Blout, E. R., Conformations of (X-L-Pro-Y)$_2$ Cyclic Hexapeptides. Preferred β-turn Conformers and Implications for β-turns in Proteins. *Biochemistry* **1981,** *20*, (16), 4730-4738.
454. Mondragón, A.; Harrison, S. C., The Phage 434 Cro/OR1 Complex at 2.5 Å Resolution. *J. Mol. Biol.* **1991,** *219*, (2), 321-334.
455. Klemm, J. D.; Rould, M. A.; Aurora, R.; Herr, W.; Pabo, C. O., Crystal Structure of the Oct-1 POU Domain Bound to an Octamer Site: DNA Recognition with Tethered DNA-Binding Modules. *Cell* **1994,** *77*, (1), 21-32.
456. Harrison, S. C.; Aggarwal, A. K., DNA Recognition by Proteins with the Helix-Turn-Helix Motif. *Annu. Rev. Biochem.* **1990,** *59*, 933-969.
457. Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O., Crystal Structure of an Engrailed Homeodomain-DNA Complex at 2.8 Å Resolution: A Framework for Understanding Homeodomain-DNA Interactions. *Cell* **1990,** *63*, (3), 579-590.
458. Aggarwal, A. K.; Rodgers, D. W.; Drottar, M.; Ptashne, M.; Harrison, S. C., Recognition of a DNA Operator by the Repressor of Phage 434: A View at High Resolution. *Science* **1988,** *242*, (4880), 899-907.
459. Jordan, S. R.; Pabo, C. O., Structure of the Lambda Complex at 2.5 Å Resolution: Details of the Repressor-Operator Interactions. *Science* **1988,** *242*, (4880), 893-899.
460. Wolberger, C.; Dong, Y.; Ptashne, M.; Harrison, S. C., Structure of a Phage 434 Cro/DNA Complex. *Nature* **1988,** *335*, (6193), 789-795.
461. Trelsman, J.; Gönczy, P.; Vashishtha, M.; Harris, E.; Desplan, C., A Single Amino Acid Can Determine the DNA Binding Specificity of Homeodomain Proteins. *Cell* **1989,** *59*, (3), 553-562.
462. Hanes, S. D.; Brent, R., DNA Specificity of the Bicoid Activator Protein is Determined by Homeodomain Recognition Helix Residue 9. *Cell* **1989,** *57*, (7), 1275-1283.
463. Fraenkel, E.; Pabo, C. O., Comparison of X-Ray and NMR Structures for the Antennapedia Homeodomain−DNA Complex. *Nat. Struct. Biol.* **1998,** *5*, (8), 692-697.
464. Kaufman, T. C.; Lewis, R.; Wakimoto, B., Cytogenetic Analysis of Chromosome 3 in *Drosophila Melanogaster*: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. *Genetics* **1980,** *94*, (1), 115-133.
465. Scott, M. P.; Weiner, A. J., Structural Relationships among Genes That Control Development: Sequence Homology between the Antennapedia, Ultrabithorax, and Fushi Tarazu Loci of Drosophila. *Proc. Natl. Acad. Sci. USA* **1984,** *81*, (13), 4115-4119.
466. Mihara, H.; Kaiser, E. T., A Chemically Synthesized Antennapedia Homeo Domain Binds to a Specific DNA Sequence. *Science* **1988,** *242*, (4880), 925-927.
467. Qian, Y. Q.; Otting, G.; Billeter, M.; Muller, M.; Gehring, W.; Wuthrich, K., Nuclear Magnetic Resonance Spectroscopy of a DNA Complex with the Uniformly ^{13}C-Labeled Antennapedia Homeodomain and Structure Determination of the DNA-Bound Homeodomain. *J. Mol. Biol.* **1993,** *234*, (4), 1070-1083.
468. Qian, Y. Q.; Billeter, M.; Otting, G.; Müller, M.; Gehring, W. J.; Wüthrich, K., The Structure of the Antennapedia Homeodomain Determined by NMR Spectroscopy in Solution: Comparison with Prokaryotic Repressors. *Cell* **1989,** *59*, (3), 573-580.
469. Saviano, M.; Isernia, C.; Bassarello, C.; P. Di Lello; Galdiero, S.; Mierke, D. F.; Benedetti, E.; Pedone, C., Conformational Analysis by NMR and Distance Geometry Techniques of a Peptide Mimetic of the Third Helix of the Antennapedia Homeodomain. *J. Pept. Res.* **2005,** *65*, (2), 200-208.

470. Müller, M.; Affolter, M.; Leupin, W.; Otting, G.; Wüthrich, K.; Gehring, W. J., Isolation and Sequence-Specific DNA Binding of the Antennapedia Homeodomain. *EMBO J.* **1988,** *7*, (13), 4299-4304.
471. Holzwarth, G.; Doty, P., The Ultraviolet Circular Dichroism of Polypeptides. *J. Am. Chem. Soc.* **1965,** *87*, (2), 218-228.
472. Venyaminov, S. Y.; Yang, J. T., Determination of Protein Secondary Structure. In *Circular Dichroism and the Conformational Analysis of Biomolecules*, Fasman, G. D., Ed. Plenum Press: New York, 1996; pp 69-108.
473. Luo, P.; Baldwin, R. L., Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/Water Mixtures. *Biochemistry* **1997,** *36*, (27), 8413-8421.
474. Edelhoch, H., Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry* **1967,** *6*, (7), 1948-1954.
475. Kohno, T.; Kusunoki, H.; Sato, K.; Wakamatsu, K., A New General Method for the Biosynthesis of Stable Isotope-Enriched Peptides Using a Decahistidine-Tagged Ubiquitin Fusion System: An Application to the Production of Mastoparan-X Uniformly Enriched with ^{15}N and $^{15}N/^{13}C$. *J. Biomol. NMR* **1998,** *12*, (1), 109-121.
476. Liu, T. Synthesis of a Potent Histone Deacetylase Inhibitor; Studies towards a Stabilized Helix-Turn-Helix Peptide. Virginia Polytechnic Institute and State University, Blacksburg, VA, 2007.
477. Travins, J. M. Turn Mimics for the Helix-Turn-Helix Protein Motif. University of Virginia, Charlottesville, VA, 1999.
478. Myers, A. G.; Gleason, J. L., Asymmetric Synthesis of α-Amino Acids by the Alkylation of Pseudoephedrine Glycinamide: L-Allylglycine and *N*-Boc-L-Allylglycine. *Org. Synth.* **1999,** *76*, 57-76.
479. Swali, V.; Matteucci, M.; Elliot, R.; Bradley, M., The Stereospecific Synthesis of 'Orthogonally' Protected Lanthionines. *Tetrahedron* **2002,** *58*, (44), 9101-9109.
480. Krasnov, V. P.; Levit, G. L.; Bukrina, I. M.; Demin, A. M.; Chupakhin, O. N.; Yoo, J. U., Efficient Large (Ca. 40 g) Laboratory Scale Preparation of (*S*)- and (*R*)-Valine *tert*-Butyl Esters. *Tetrahedron: Asymmetry* **2002,** *13*, (17), 1911-1914.
481. Graffner-Nordberg, M.; Fyfe, M.; Brattsand, R.; Mellgård, B.; Hallberg, A., Design and Synthesis of Dihydrofolate Reductase Inhibitors Encompassing a Bridging Ester Group. Evaluation in a Mouse Colitis Model. *J. Med. Chem.* **2003,** *46*, (16), 3455-3462.
482. Sieber, P., Der 2-Trimethylsilyläthyl-Rest als Selektiv Abspaltbare Carboxy-Schutzgruppe. *Helv. Chim. Acta* **1977,** *60*, (8), 2711-2716.
483. Roush, W. R.; Madar, D. J.; Coffey, D. S., Synthesis of Highly Functionalized Naphthoate Precursors to Damavaricin D - Observation of Kinetically Stable Benzocyclohexadienones in the Bromination Reactions of Highly Functionalized β-Naphthol Derivatives. *Can. J. Chem.* **2001,** *79*, (11), 1711-1726.
484. Boruah, A.; Rao, I. N.; Nandy, J. P.; Kumar, S. K.; Kunwar, A. C.; Iqbal, J., Synthesis of a Novel *cis*-Proline-Derived Cyclic Type VI β-Turn Mimic via Ring-Closing Metathesis. *J. Org. Chem.* **2003,** *68*, (12), 5006-5008.
485. Schwab, P.; Grubbs, R. H.; Ziller, J. W., Synthesis and Applications of $RuCl_2(=CHR')(PR_3)_2$: The Influence of the Alkylidene Moiety on Metathesis Activity. *J. Am. Chem. Soc.* **1996,** *118*, (1), 100-110.

486. Young, T.; Kiessling, L. L., A Strategy for the Synthesis of Sulfated Peptides. *Angew. Chem. Int. Ed.* **2002,** *114*, (18), 3449-3451.
487. Dai, N.; Wang, X. J.; Etzkorn, F. A., The Effect of a *trans*-Locked Gly-Pro Alkene Isostere on Collagen Triple Helix Stability. *J. Am. Chem. Soc.* **2008,** *130*, (16), 5396-5397.
488. Lober, S.; Rodriguez-Loaiza, P.; Gmeiner, P., Click Linker: Efficient and High-Yielding Synthesis of a New Family of SPOS Resins by 1,3-Dipolar Cycloaddition. *Org. Lett.* **2003,** *5*, (10), 1753-1755.
489. Turner, R. A.; Oliver, A. G.; Lokey, R. S., Click Chemistry as a Macrocyclization Tool in the Solid-Phase Synthesis of Small Cyclic Peptides. *Org. Lett.* **2007,** *9*, (24), 5011-5014.
490. Oishi, S.; Shi, Z.-D.; Worthy, K. M.; Bindu, L. K.; Fisher, R. J.; Terrence R. Burke, Jr., Ring-Closing Metathesis of *C*-Terminal Allylglycine Residues with an *N*-Terminal β-Vinyl-Substituted Phosphotyrosyl Mimetic as an Approach to Novel Grb2 SH2 Domain-Binding Macrocycles. *Chembiochem* **2005,** *6*, (4), 668-674.
491. Tang, Q.; Wareing, J. R., Ring-Closing Metathesis versus Cross Metathesis of Resin-Bound Olefins. *Tetrahedron Lett.* **2001,** *42*, (8), 1399-1401.
492. Pattabiraman, V. R.; Stymiest, J. L.; Derksen, D. J.; Martin, N. I.; Vederas, J. C., Multiple On-Resin Olefin Metathesis to Form Ring-Expanded Analogues of the Lantibiotic Peptide, Lacticin 3147 A2. *Org. Lett.* **2007,** *9*, (4), 699-702.