# The fully-extended conformation in peptides and proteins

**Marco Crisma[1], Fernando Formaggio[1,2], Carlos Alemán[3,4], Joan Torras[3,4], Chandrasekharan Ramakrishnan[5], Neha Kalmankar[6], Padmanabhan Balaram[5], Claudio Toniolo[1,2]**


[1]Institute of Biomolecular Chemistry, Padova Unit, CNR, 35131 Padova, Italy

[2]Department of Chemistry, University of Padova, 35131 Padova, Italy

[3]Departament d'Enginyeria Química, EEBE, Universitat Politècnica de Catalunya, 08019 Barcelona, Spain

[4]Barcelona Research Center in Multiscale Science and Engineering, Universitat Politècnica de Catalunya, 08019 Barcelona, Spain

[5]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560065, India

[6]National Centre for Biological Sciences (TIFR), GKVK Campus, Bangalore 560065, India


**Correspondence:**

Dr. Marco Crisma

Institute of Biomolecular Chemistry, Padova Unit, CNR, via Marzolo 1, 35131 Padova, Italy.

E-mail: marco.crisma@unipd.it

*This Review Article has been submitted for the Special Issue " Emerging Peptide Science from Italy"*

**Abstract**

The intramolecularly H-bonded, fully-extended conformation ($C_5$) of an $\alpha$-amino acid residue (and the resulting $2.0_5$-helix obtained *via* its propagation) is one of the least extensively investigated types of peptide and protein backbone secondary structure. This situation does still currently occur despite its unique ability to enjoy by far the largest separation per residue among peptide conformations. In this article, we offer a detailed update of our present knowledge on this intriguing 3D-structure of peptides in the crystal state as obtained from recently published investigations, complemented by a statistical analysis for its occurrence in the crystal structures of $\alpha$-amino acid derivatives and peptides available in the Cambridge Structural Database. We have expanded this useful information to the results of a bioinformatics analysis performed on this (so far largely unappreciated) conformation authenticated in all proteins solved by X-ray diffraction to a resolution of $\leq 1.5$ Å. In the last section, we describe the results of our DFT calculations on the conformational preferences of a set of homo-peptides (from monomer to octamer) based on as many as six protein and two non-coded, carefully selected, $\alpha$-amino acids. From this literature survey integrated by new energy calculations we have definitely provided strong support to the thesis that this polypeptide 3D-structure does indeed exist, it should be not neglected in future studies by structural biochemists, and it represents a very attractive, novel backbone for applications for organic, medicinal, and biomaterials chemists.

# 1. INTRODUCTION

To further expand the arsenal of molecular platforms offered to colleagues working in synthesis, spectroscopic analysis, and host-guest interaction studies, among structural biochemists there is a currently growing interest to the rarely observed (backbone) N-H…O=C (backbone) intramolecularly H-bonded peptide conformations where the direction of the H-bonds goes "the other way around" with respect to the classical, overwhelmingly observed, direction, namely from an upstream N-H group donor to a downstream C=O acceptor.[1]

The three possible conformations of the former type in a system of four covalently linked peptide units are presented in Figure 1A. According to the two "numerical" nomenclatures most frequently employed, starting from the smallest *pseudocycle* (a cyclic structure where one covalent bond is replaced by an intramolecular H-bond) they are termed: (i) $C_5$, $C_8$, and $C_{11}$, where C stands for cycle and the subscript number corresponds to that of the atoms involved in the ring system, and (ii) $2 \rightarrow 2$ (or the equivalent $3 \rightarrow 3$ or $4 \rightarrow 4$), $2 \rightarrow 3$ (or $3 \rightarrow 4$) and $2 \rightarrow 4$, where in each case the number *before* the arrow indicates the position in the sequence of the upstream N-H H-bonding donor while the number *after* the arrow indicates that of the downstream C=O acceptor.

At the level of the two largest conformations ($C_8$, and $C_{11}$) a reversal of the peptide main chain direction does take place, in contrast to the $C_5$ conformation (Figure 1B) where the main chain is fully extended. The $C_8$ and $C_{11}$ turn conformations are characterized by one and two peptide -CO-NH- groups, respectively, internal to the annular structure (in the $C_8$ conformation the peptide group is forced by the small size of the cycle to be *cis*, $\omega$ torsion angle $\cong 0°$, whereas in the $C_{11}$ conformation each peptide group can accommodate either in the *cis* or in the *trans*, $\omega$ torsion angle $\cong 180°$, conformation). Very recently, our research group published detailed review articles on each of these previously unappreciated conformations where the additional, "non-numerical" nomenclature $\delta$-turn[2] and $\varepsilon$-turn,[3] respectively, was proposed.

# 2. BRIEF SUMMARY ON PEPTIDES BEFORE 2012

Double or multiple, consecutive, *intra*residue H-bonded $C_5$ conformations generate a flat helical structure, termed $2.0_5$-helix,[4] where 2.0 indicates its exact twofold symmetry. This peptide conformation is positioned in the major *n* =2 (two amino acid residues per basic helical unit)

diagonal of the universally known Ramachandran map,[5,6] more specifically in its upper-left corner ($\phi,\psi$ backbone torsion angles $\cong 180°,180°$). It is spatially close to the second most frequently reported polypeptide conformation in peptides and proteins (pleated $\beta$-sheet, with either parallel- or antiparallel-strand orientation) where the corresponding ranges of $\phi,\psi$ angles are $-120° \div -140°$, $115° \div 135°$, respectively.

In a previous article,[4] partially reviewing the available literature on the $2.0_5$-helix to 2011 (included), the Padova group described:

(i) A brief historical retrospective of this planar sheet structure, beginning with the original, modelling-based proposal from Pauling and Corey,[7] where these authors predicted a reasonable stability only for homo-peptides from Gly among all protein amino acids, because this residue lacks any side chain $R$ group which could interfere unfavorably through *inter*molecular steric interactions involving $C^\beta$-atoms of adjacent and in register main chains.

(ii) Results from numerous conformational energy computations on very short model compounds, usually $\alpha$-amino acid derivatives, which *inter alia* highlighted the potentially negative role on the stability of the $C_5$ conformation played by the bulky lateral chain(s) $R$ and $R'$ generating unfavorable *intra*molecular non-bonded interactions, in particular with the C=O oxygen atom of the preceding residue (Figure 1B). From this Figure, it is also evident that the $C_5$ conformation does not contain any internal peptide bond. Here, it is worth noting that more recently Newberry and Raines[8] reported the results of their quantum-mechanical analysis on the $C_5$ conformation of the minimal peptide backbone model Ac-Gly-NHMe (Ac, acetyl; NHMe, methylamino), demonstrating that the operative intra-residue interaction, characterized by overlap of the *p*-type C=O lone pair and the N-H $\sigma^*$ orbital, is endowed with the properties of a canonical H-bond.

(iii) Few, selected examples of X-ray diffraction structures of peptides based on coded amino acids (Gly, Ala, and Phe), and $C^\alpha$-tetrasubstituted (Figure 2) and $\Delta$Ala ($C^{\alpha,\beta}$-didehydroalanine) residues. Now, we wish to complete this part by briefly presenting and discussing two additional, relevant examples related to the $2.0_5$-helix formed exclusively by $C^\alpha$-tetrasubstituted residues and not treated in ref. [4]. An X-ray diffraction analysis revealed that in the crystalline state a terminally-blocked, homochiral, homo-tripeptide based on L-($\alpha$Me)Leu ($C^\alpha$-methyl leucine) adopts a multiple fully-extended conformation (Figure 3).[9] This represents the first unambiguous finding of such a 3D-structure in a peptide sequence formed exclusively by $C^\alpha$-*methylated* $\alpha$-amino acids, otherwise typically folded in the $3_{10}$-helical conformation.[10-14] Moreover, an explanation for the published crystallographic results on the terminally-protected homo-pentapeptide sequence from Deg ($C^{\alpha,\alpha}$-diethyl glycine) (Figure 2), a $C^\alpha$-tetrasubstituted residue generally considered the simplest one of

the sub-class largely favoring the fully-extended conformation,[4,15,16] is not clear-cut (Figure 4). Indeed, the originally reported 3D-structure of the Tfa (trifluoroacetyl) / OtBu (tert-butyloxy) terminally protected -(Deg)$_5$- was found to adopt the fully-extended conformation,[15] whereas that of Tfa-(Deg)$_5$-OEt (OEt, ethoxy), published years later, is 3$_{10}$-helical.[16,17] Interestingly, this latter observation was recently corroborated by the results of two additional X-ray diffraction analyses, specifically on Ac-(Deg)$_5$-OtBu tetrahydrofuran solvate[18] and PyrAc-(Deg)$_5$-O($p$NO$_2$)Bzl,[19] where PyrAc is 1-pyrenylacetyl and O($p$NO$_2$)Bzl is *para*-nitrobenzyloxy. It remains to be seen whether crystal packing forces and/or the different nature of the terminal (blocking / protecting) groups would have played any role in biasing these peptide backbone conformations. In any case, we are inclined to suggest that the 2.0$_5$-helix peptide conformation would not be a particularly robust peptide 3D-structure and its transformation into the 3$_{10}$-helix could take place relatively easily. Furthermore, an interestingly potential application of the 2.0$_5$-helical structure (simply termed by the authors "β-sheet" without any specification), only formed by C$^\alpha$-tetrasubstituted α-amino acids was proposed more than 20 years ago[20] (Figure 5). This supramolecular system represents an autocatalytic replicator of peptide bond formation with a self-complementary structure. The three residues of this class involved in this construct are Deg and two hybrid α-amino acids each bearing two identical nucleo-alkyl side chains. The two original nucleobases were adenine (A) and uridine (U), the latter subsequently replaced by thymine (T). One of the foreseen advantages arising from this highly conformationally constrained peptide was the possibility to control the 3D-structure of the system. The base-pairing information (molecular recognition) along the sequence could facilitate formation of the central peptide bond promoted by the partly amino acid / partly nucleic acid molecules. In this sense, this appealing system might also have a prebiotic significance. Unfortunately, although these achiral nucleo-amino acids with double side chains would indeed be prepared by the authors,[21] synthesis and utilization of the corresponding peptides were never reported.

(iv) A comparison between the characteristically diverging IR absorption and $^1$H NMR signatures in solvents of low polarity of the peptide 2.0$_5$-helix and the (sometimes) competing 3$_{10}$-helix reported primarily by the Padova group.[4,16,19,22-27] In the last few years, these investigations were further expanded and extended to fluorescence spectroscopy using 2.0$_5$-helix peptide stretches as spacers between two N- and C-terminal photoprobes.[19,24,28]

In the next three sections, we will focus on the results of:

• A detailed literature survey on peptide C$_5$ conformations and 2.0$_5$-helices published in the last 5-6 years (and therefore not included in the previous review article[4]) as extracted exclusively from X-

5

ray diffraction analyses, complemented by a statistical analysis for its occurrence in the crystal structures of α-amino acid derivatives and peptides available in the Cambridge Structural Database.

- An exhaustive bioinformatics investigation on the aforementioned 3D-structures found in high-resolution X-ray diffraction analyses of globular proteins.

- An updated conformational energy calculation study on model homo-peptides to the octamer level based on the aliphatic Gly, Ala, Aib, and Deg, the β-hydroxylated Ser, and the β-aromatic Phe residues in absence external forces (i.e. in vacuum), aimed at evaluating the intrinsic tendency of such residues to stabilize the extended conformation and the relative stabilities of the fully-extended and helical conformations.


## 3. DETAILED LITERATURE SURVEY ON PEPTIDES SINCE 2012

In Tables 1-5[29-70] we listed all published α-amino acid derivatives and (either linear or, more rarely, cyclic) peptides to the decamer level where single or consecutive fully-extended ($C_5$) structures were identified by X-ray diffraction in the last five/six years. A few, still unpublished 3D-structures, characterized by the $C_5$ form and solved in the Padova laboratory, are presented in Table 6. [18]

The α-amino acids involved range from the by far most commonly found $C^\alpha$-*di*substituted Gly (Table 1) to a very limited number of $C^\alpha$-*tri*substituted residues, where one example each is provided by the small-sized representatives Cys, Ala, and Dap (Table 2). Moreover, we surprisingly noted a large variety of $C^\alpha$-*tetra*substituted α-amino acids (Tables 3 and 4), most of them being well known for their extremely effective helicogenic properties.[71-73] Examples of two derivatives and a very short peptide with an α,β-didehydro-α-amino acid are also given in Table 5. Finally, still unpublished $C_5$-containing peptides (Table 6) are based on each of the classes of α-amino acids mentioned above.

Compared to the overwhelmingly "isolated" $C_5$ conformations, only in five peptides back-to-back conformations of this type ($2.0_5$-helices) were uncovered in the course of our survey on recently solved X-ray diffraction structures. They are: 1) $H_2^+$-(Gly)$_4$-O$^-$, 2) $H_2^+$-(Gly)$_5$-O$^-$, 3) Z-Gly-ψ[CSNH]-Gly-ψ[CSNH]-Gly-OMe, 4) Tfa-(*S*)-Epg-Deg-Dp$_n$g-Deg-O*t*Bu, and 5) *p*CNBz-(ΔAla)$_3$-OMe (abbreviations: Z, benzyloxycarbonyl; ψ[CSNH], thionamide; Epg, $C^\alpha$-ethyl, $C^\alpha$-*n*-pentyl glycine; Dp$_n$g, $C^{\alpha,\alpha}$-di-*n*-propyl glycine; *p*CNBz, *para*-cyanobenzoyl).

Three detailed crystal structures of Gly homo-peptides were recently reported.[30,40] This emerging interest is not surprising in view of the special biological role of some of these compounds. The most relevant is certainly that played by the -(Gly)$_5$- bridges cross-linking pentaglycan layers in *Staphylococcus aureus* cell walls. Interestingly, in these Gram-positive bacterial cells this pentapeptide system is largely folded in a compact conformation, which however tends to convert into a fully-extended 3D-structure in response to a temperature variation. The ultimate result is a swelling of the cell walls.[74] In the crystal structure of the R-(Gly)$_2$-OMe (OMe, methoxy) homo-dipeptide, only the C-terminal residue is fully extended (Table 1).[30] In both crystal structures of the terminally zwitterionic $H_2^+$-(Gly)$_4$-O$^-$ and $H_2^+$-(Gly)$_5$-O$^-$ (for the latter peptide, see Figure 6) two independent molecules (**A** and **B**) in the asymmetric unit were observed.[40] In each case, all Gly residues except the N-terminal ones are part of ideal, flat 2.0$_5$-helices (Table 1). Each of the N-H and C=O groups involved in the C$_5$ intramolecularly H-bonded forms is also involved in intermolecular H-bonds with a flanking strand. It should be noted that the proposed conformation of poly(Gly)$_n$ form I does not correspond to this secondary structure, but rather to the classical β-sheet. This structural difference has been tentatively reconciled by invoking the much higher concentration of terminal charges of opposite sign occurring in the short homo-oligomers as compared to the polymer. Also, the X-ray diffraction structure of the $H_2^+$-(Gly)$_3$-O$^-$ β-polymorph has been long ago reported to be very similar to those of its two higher homologs.[75] Unfortunately, no higher Gly homo-oligomers have been investigated for their crystal structure.

In the crystal structure of the terminally protected *bis*-endothioxo Gly homo-tripeptide (Table 1 and Figure 7) an intriguing type of 2.0$_5$-helix is observed.[41] The C-terminal C$_5$ conformation involving Gly$^3$ is slightly distorted from the φ,ψ 180°,180° values. Much more unusual is the observation that its preceding -Gly$^2$-ψ[CSNH]- residue is also fully extended but the H-bond acceptor is a *thionamide* sulfur atom. This is the first case where an intramolecular N-H…S=C H-bond is seen in a peptidomimetic C$_5$ structure.

Among the C$^\alpha$-tetrasubstituted α-amino acids the simplest and most common Aib was identified in the C$_5$ conformation only very rarely before 2012,[76-82] despite the huge number of its derivative and peptide X-ray diffraction structures solved. It is worth recalling that this secondary structure was found sterically and energetically allowed for Aib, although much less favored than the helical conformations (for a specific review article on this topic, see ref. [76]). In this connection, the crystal structure of the "N$_3$"-(Aib)$_3$-O$t$Bu (where "N$_3$" stands for the azido precursor of the α-amino group) was recently solved.[45] Only the central (Aib$^2$) residue was found in the regular fully-extended form (Table 3). Very interestingly, the first example in the literature of an -

Aib-Aib- $2.0_5$-helix is almost a specific property of this compound in that the value of the $\psi$ torsion angle of Aib$^3$ (-147.8°) is close to, but not within, the acceptable range for this conformation.

Short peptides characterized by the immediately higher Aib homologs Deg and Dp$_n$g, and the chiral Epg residue were also recently analyzed by X-ray diffraction (Table 4). In the dipeptide Tfa-Dp$_n$g-Deg-O$t$Bu only the N-terminal Dp$_n$g was found to be in the $C_5$ form, the C-terminal Deg being helical.[25] Conversely, in the tetrapeptide Tfa-(S)-Epg-Deg-Dp$_n$g-Deg-O$t$Bu as many as three residues (at positions 1, 2, and 4) are fully extended (Figure 8).[25] However, the helical Dp$_n$g$^3$ residue breaks down the potentially full stretch of $2.0_5$-helix, leaving only the N-terminal dipeptide sequence in this disposition. Finally, two Deg- and Gly-containing dipeptides were recently analyzed in the crystal state by X-ray diffraction.[18] In both of them the Deg residue is fully extended (Table 6) while Gly is not (Figure 9 **A**-**C**). Specifically, in both crystallographically independent molecules of Z-Deg-Gly-O$t$Bu (Figure 9 **A** and **B**) the Gly residues adopt a *semi-extended* conformation characterized by $\phi,\psi$ = -77.9°,165.0° in molecule **A** and -91.0°,166.8° in molecule **B**. In the structure of Ac-Gly-Deg-O$t$Bu (Figure 9 **C**) the set of $\phi,\psi$ backbone torsion angles of the N-terminal Gly residue (88.8°,17.0°) fall in the "bridge" region of the conformational map.[83] In Table 6 and Figure 9, panels **D** and **E**, we also show that in both of the two crystallographically independent molecules of the endothioxopeptide Boc-Gly-$\psi$[CS-NH]-L-Leu-OMe the Leu residue is involved in an intramolecularly H-bonded $C_5$ conformation, slightly distorted at the level of the $\phi$ backbone torsion angle.[18] Conversely, the thionated Gly residues, adopting the "bridge" conformation, lack the intramolecular H-bond with the C=S group potential acceptor.

A few examples of $C_5$ structures involving a $C^{\alpha,\beta}$-didehydro-$\alpha$-amino acid were reported since 2012 (Table 5). One of them accommodates a $\Delta$Pyr ($C^{\alpha,\beta}$-didehydro-pyridin-2-yl-alanyl) residue,[69] whereas all of the remaining $C_5$ structures are formed by $\Delta$Ala.[66,67,70] Interestingly, in the structure of the macrocyclic antibiotic thiostrepton, two out of the three $\Delta$Ala residues occurring in the molecule are fully extended, one of them within the macrocyclic part and the other C-terminal to the "tail".[70]

In the last crystal structure listed in Table 6, an additional example of a terminally blocked $\Delta$Ala homo-tripeptide is reported,[18] following the exciting, original one published many years ago (1999).[84] All three $\Delta$Ala residues adopt sets of $\phi,\psi$ torsion angles very close to the ideal values (180°,180°) for a $2.0_5$-helix (Figure 10). The backbone of this molecule is fully extended and essentially flat, and the amino acid *side chains* are *coplanar* to it. The packing mode of the molecule generates layers without any contribution from intermolecular N-H…O=C H-bonds. This

structure is stabilized by two types of intramolecular H-bonds, namely the three intraresidue N-H…O=C H-bonds, typical of the $C_5$ conformation, and the three unusual $C^{\beta}_{i+1}$-H…$O_i$=$C_i$, found only in ΔAla peptides and giving rise to "$C_6$-ring" systems. We are still confident that this unique peptide disposition will soon catalyze applications in biochemistry and materials science.

We exploited the opportunity offered by this review article to perform a search for the occurrence of the $C_5$ conformation in the X-ray diffraction structures of *all* $N^{\alpha}$-acylated α-amino acid derivatives and peptides available in the Cambridge Crystallographic Database[85] (CSD version 5.38 of November 2016, updated to May 2017, organics only) by using the values of the backbone torsion angles $\phi,\psi = 180°\pm20°,180°\pm20°$ as the search criterion. The search was carried out separately for Gly, $C^{\alpha}$-trisubstituted α-amino acids, $C^{\alpha}$-tetrasubstituted α-amino acids, and ΔAla. Among the four subsets returned, the most populated is that of $C^{\alpha}$-tetrasubstituted α-amino acids, with 188 residues in 139 structures. The Gly subset, containing 123 residues in 94 entries, exceeds in size that formed by all (coded and noncoded) $C^{\alpha}$-trisubstituted α-amino acids (99 residues in 93 entries). Since the CSD search carried out for ΔAla returned only 18 residues in 13 structures, we included in the ΔAla subset the data of the three residues of our unpublished structure of *p*CNBz-(ΔAla)$_3$-OMe[18] (Table 6), thus increasing the statistical sample to 21 residues. The average geometrical parameters characterizing the $C_5$ conformation were derived for each subset. The results are summarized in Table 7.

In general, the values of bond distances are far from unexpected. In particular, the N-$C^{\alpha}$ and $C^{\alpha}$-C bond distances for ΔAla are shorter than the standard values typical for coded amino acids,[86,87] as a result of the $sp^2$ hybridization of the α-carbon and the likely electron delocalization involving the $C^{\alpha}$=$C^{\beta}$ double bond and the flanking amide units. Interestingly, the value of the N-$C^{\alpha}$-C ($\tau$) bond angle averaged for ΔAla, 110.1°, is remarkably narrowed compared to that typical for an $sp^2$-hybridized carbon atom (120°). Similarly, the $\tau$ bond angle averaged for the fully-extended $C^{\alpha}$-tetrasubstituted α-amino acid residues (104.5°) is about 6° smaller than the nearly tetrahedral value normally adopted by these residues when helical (*e.g.*, 110.5° for Aib). Overall, the present statistical results fully corroborate early observations on Deg and ΔAla homo-peptides about the narrowing of the $\tau$ bond angle which accompanies the onset of the $C_5$ conformation.[4,84] From our analysis, a similar trend emerges also for Gly and the $C^{\alpha}$-trisubstituted residues. Specifically: (i) the value of $\tau$ averaged from the 123 fully-extended Gly residues is 110.2°, whereas that averaged from 774 Gly residues in conformations other than fully-extended (selected through the search criterion $\phi,\psi = 0°\pm160°,0°\pm160°$) found in the CSD is 114.0°; (ii) the value of $\tau$ averaged from the fully-extended $C^{\alpha}$-trisubstituted residues is 108.1°, as compared to the value of 112.6° averaged from 792

helical Ala residues ($\phi,\psi$ = -60°±20°,-30°±20°) found in the CSD. The sensitivity of the valence geometry of amino acid residues, and in particular of the $\tau$ bond angle, to the peptide/protein backbone conformation is a well-established phenomenon.[88-93]

As for the H-bond geometry, the N…O distance is the result of the interplay among the values of N-C$^{\alpha}$, C$^{\alpha}$-C, and C=O bond lengths, N-C$^{\alpha}$-C and C$^{\alpha}$-C=O bond angles, and the $\psi$ torsion angle. In the four subsets of fully-extended residues reported in Table 7, the average intraresidue N…O distance increases slightly on going from the C$^{\alpha}$-tetrasubstituted $\alpha$-amino acids to $\Delta$Ala and from $\Delta$Ala to the C$^{\alpha}$-trisubstituted residues, more markedly between the C$^{\alpha}$-trisubstituted residues and Gly.

The H…O distance depends, in addition to the parameters governing the N…O distance listed above, on the values of the $\phi$ torsion angle, the N-C$^{\alpha}$ and N-H bond lengths, and the C$^{\alpha}$-N-H bond angle. It is worth recalling that X-ray diffraction experiments do not provide routinely accurate and precise positions for H-atoms which, in our statistical dataset, might be affected by the way they have been located/calculated and subsequently handled in the course of the crystallographic refinements. In particular, in recent years H-atoms are commonly allowed to ride on the carrying atom by imposing a value for the N-H bond distance which depends on the temperature. Although the true inter-nuclear N-H bond distances (as determined by neutron diffraction) do not vary with temperature and are longer than those employed for this riding mode of H-atom refinement, such a practice takes into account libration effects and allows for a better fit to the X-ray diffraction data. Despite these limitations about the positioning of the H atom, our statistical analysis returns for the four subsets of fully-extended residues average values of the N-H bond distances that are identical within 0.02 Å, and values of the C$^{\alpha}$-N-H bond angles that do not differ by more than 3.2°. On this basis, it is worth comparing the average H…O distances characterizing the intramolecularly H-bonded C$_5$ form among the four subsets. The H…O separations follow the same rank order of the N…O distances discussed above, in that the smallest average value is observed for the C$^{\alpha}$-tetrasubstituted $\alpha$-amino acids (2.15 Å), followed by $\Delta$Ala (2.20 Å), the C$^{\alpha}$-trisubstituted residues (2.28 Å), and Gly (2.34 Å). In any case, all of these values are well within the commonly accepted limit of 2.50 Å (which is less than the sum of the van der Waals radii of H and O, 2.72 Å) for the occurrence of an N-H…O=C H-bond.[94-97]

As for the N-H…O angle, the values range from 103.3° for Gly to 110.7° for the C$^{\alpha}$-tetrasubstituted residues. Interestingly, an inverse correlation is observed between the average value of the N-H…O angle and that of the $\tau$ bond angle for three out of the four subsets investigated, the exception being the C$^{\alpha}$,C$^{\beta}$ unsaturated $\Delta$Ala. The general view for the occurrence of an N-H…O=C

H-bond is that two conditions should be concomitantly fulfilled: (i) H…O distance $\leq 2.50$ Å, and (ii) N-H…O angle $\geq 120°$. The N-H…O angles returned from our analysis are significantly narrower than $120°$. However, there is ample evidence from IR absorption and NMR spectroscopies,[8,26,98-103] corroborated in the last ten years by computational analyses,[8,104,105] that the intraresidue N-H…O interaction characterizing the $C_5$ conformation is a true H-bond, thus supporting the view that in this particular case, in which the N-H and C=O bonds are nearly parallel to each other and equiverse, the angular criterion reported above should be relaxed. Not surprisingly, for the large majority of the X-ray diffraction structures covered by our survey, the occurrence of intramolecular H-bonds of the $C_5$ type was not pinpointed by the various authors themselves. In this connection, it is worth mentioning that for the structures belonging to the subsets comprising Gly and the $C^\alpha$-trisubstituted residues, the N-H and C=O groups engaged in a $C_5$ form are usually involved in *inter*molecular H-bonds as well. Conversely, in the case of $C^\alpha$-tetrasubstituted residues and $\Delta$Ala, the participation of the N-H and C=O groups forming a $C_5$ structure in *inter*molecular H-bonds is a rare event.

## 4. ANALYSIS ON PROTEINS

A non-homologous dataset of 2154 independent protein chains (sequence identity cut-off of $\leq 30\%$ and resolution cut-off of $\leq 1.5$ Å) was used for the analysis. Individual amino acid residues with backbone torsion angles of $\varphi = 180\pm20°$ and $\psi = 180\pm20°$ were extracted from the dataset. Figure 11 shows the distribution of the propensity values for the individual amino acids to occur in the fully-extended region of the Ramachandran map. Propensities were calculated using the following formula:

$$Propensity = \frac{\left(\dfrac{No.\ of\ fully-extended\ individual\ amino\ acid\ (X)}{Total\ no.\ of\ fully-extended\ amino\ acid}\right)}{\left(\dfrac{No.\ of\ individual\ amino\ acid\ (X)\ in\ dataset}{Total\ no.\ of\ amino\ acid\ in\ dataset}\right)}$$

Values of approximately one are anticipated for residues which have no specific preference for this region. Values significantly greater than one indicates a preference for the fully-extended

conformation, while values ≪ 1 suggest exclusion. Table 8 provides the relevant statistics for the individual amino acid residues. Clearly, Gly has by far the highest propensity for the fully-extended conformation. Ser also shows a significant preference, closely followed by its sulfur mimics Cys. The aromatic amino acids Phe, His, Trp, and Tyr appear to indicate some preference (particularly weak for the last residue). In their recent paper on the intraresidue H-bonded $C_5$ form, Newberry and Raines[8] performed a similar search on a different dataset of nonredundant protein structures (resolution < 1.6 Å) by using a less stringent criterion for the identification of fully-extended residues, $\varphi = 180\pm40°$ and $\psi = 180\pm40°$. Not surprisingly, their propensity values show a much smoother distribution. However, the amino acids possessing a propensity value > 1 are the same returned from our analysis, apart from the inclusion of Ala. Moreover, both analyses basically share the same rank order of propensity, *i.e.*, Gly > Ser > Cys > aromatic residues.

Protein segments in which three or more sequential $\varphi,\psi$ values fall within the fully-extended region ($\varphi = 180\pm20°$ and $\psi = 180\pm20°$) were extracted from an expanded protein dataset which utilized a resolution cut-off of $\leq 2.0$Å and sequence identity cut-off of $\leq 30\%$. A total of 36 three-residue segments and 3 four-residue segments were obtained. Remarkably attractive are the findings of 3-residue stretches formed by consecutive Gly or Ser residues, displayed in Figures 12 and 13, respectively.[106,107] Segments which flank the fully-extended strands are also shown. It is evident that the fully-extended, flattened strand is almost invariably flanked by conventional β-strand segments providing opportunities for interstrand H-bonding. The flattened fully-extended strand appears to be stabilized in many examples [including the four-residue segments -Gly[198]-Tyr[199]-Phe[200]-His[201]- in the aurora-A kinase domain (PDB code: 3O50)[108] and -Tyr[306]-Phe[307]-Gly[308]-Ala[309]- in the extracellular domain of the murine protein RANKL (PDB code: 1S55)[109]] by aromatic side chains which may contribute both intrastrand and interstrand aromatic⋯amide interactions. However, such interactions do not represent a necessary condition for the onset of long fully-extended strands, as exemplified by the four-residue segment -Gly[64]-Ser[65]-Gly[66]-Ala[67]- in the enzyme catalase from *Micrococcus lysodeikticus* (PDB code: 1HBZ)[110]. Prior to the present analysis, the only fully-extended stretch encompassing four residues identified in a protein crystal structure was the -(Gly)$_4$- segment (residues 283-286) of the enzyme His-*t*RNA-synthetase in complex with histidine at 2.7 Å resolution (PDB code: 1ADJ),[111] shown in Figure 14.

A final point to note is that the extent of strand flattening may also be estimated from the virtual bond angles $C^\alpha_i$-$C^\alpha_{i+1}$-$C^\alpha_{i+2}$ ($C^\alpha_{i+1}$-$C^\alpha_{i+2}$-$C^\alpha_{i+3}$) determined over the strand segments. Virtual bond angles for the flanking strands fall in the range $101° \div 140°$, while those for the fully-extended strands are found in the range $137° \div 153°$.

# 5. CONFORMATIONAL ENERGY CALCULATIONS ON MODEL PEPTIDES

Density functional theory (DFT) calculations were performed with the Gaussian 09[112] computer package using the M06L[113,114] functional combined with the 6-31+G(d,p) basis set. The ability of the M06L functional to account for van der Waals and $\pi$-$\pi$ stacking interactions have been largely by comparing the geometries and interaction energies obtained for dispersion-dominated model dimers with those resulting from sophisticated ab initio Coupled-Cluster calculations.[113-116] The homo-peptides studied in this work, which are denoted -(Xxx)$_n$-, correspond to Ac-(Xxx)$_n$-NHMe, where Ac = acetyl, NHMe = methylamino, Xxx = Gly, Ala, Ser, Phe, Aib or Deg, and *n* is the number of residues, which ranges from 1 to 8. Homo-peptides were built in fully-extended and $3_{10}$-helical conformations, the backbone dihedral angles before geometry optimization being $\varphi,\psi$ = 180º,180º and -60º,-35º, respectively. All these structures were minimized at the M06L/6-31+G(d,p) level without any restriction. To ascertain the intrinsic conformational preferences of Xxx, all geometry optimizations were performed *in vacuo*.

Cooperative energy effects were examined by calculating the energy increment (EI) associated to the addition of a residue Xxx in a given conformation as the difference between the energies of -(Xxx)$_2$- and -(Xxx)$_1$-. Therefore, for a given homo-peptide, -(Xxx)$_n$-, the theoretically predicted energy (E$^{theor}$) can be estimated as:

$$E^{theor}(Xxx)_n = E(Xxx)_1 + (n-1) \cdot EI \qquad (1)$$

where E(Xxx)$_1$ is the energy of -(Xxx)$_1$-. The difference between the theoretically predicted energy and the quantum mechanical energy gives the cooperative energy effect:

$$E^{coop}(Xxx)_n = E(Xxx)_n - E^{theor}(Xxx)_n \qquad (2)$$

Considering that the $2.0_5$-helix is based on the propagation of the fully-extended conformation that involves a five-membered intramolecularly H-bonded ring (C$_5$), different homo-peptides were constructed to ascertain if the stability of this secondary structure is due to the existence of favorable cooperative energy effects intrinsically associated with the C$_5$ backbone···backbone (bb-bb) interaction or to the complementary assistance of side-chain···side-chain (sc-sc) and backbone···side-chain···backbone (bb-sc) interactions. According to their intramolecular bb-bb, sc-sc, and/or bb-sc interactions, the -(Xxx)$_n$- homo-peptides built up considering a fully-extended conformation were categorized as follows:

(1)     -(Gly)$_n$- homo-peptides, exclusively stabilized by C$_5$ interactions.

(2)     -(Ala)$_n$-, -(Aib)$_n$-, and -(Deg)$_n$- homo-peptides that, in addition to the C$_5$ H-bonds, exhibit van der Waals sc-sc interactions [*i.e.* methyl···methyl for -(Ala)$_n$- and -(Aib)$_n$-, or ethyl···ethyl for -(Deg)$_n$-].

(3)     -(Ser)$_n$- homo-peptides in which the backbone peptide bonds are not only involved in C$_5$ interactions but also form bb-sc H-bonds with the β-hydroxyl side groups: C$_{i-1}$=O$_{i-1}$···H$_i$–O$_i$ in the conformation hereafter termed Ser$_n$-a, and H$_i$–O$_i$·· H$_{i+1}$– N$_{i+1}$ in the conformation hereafter termed Ser$_n$-b.

(4)     -(Phe)$_n$- homo-peptides (hereafter termed Phe$_n$-a) in which the C$_5$ interactions coexist with the N–H···π interactions (bb-sc) involving the peptide bond of residue *i* and the β-phenyl side group of residue *i+1*.

(5)     -(Phe)$_n$- homo-peptides (hereafter termed Phe$_n$-b) in which the β-phenyl side groups of residues *i* and *i+1* are faced to form sc-sc interactions (π···π stacking).

Figure 15 displays the variation of E$^{coop}$(Xxx)$_n$ (Eq 2) against *n* for the fully-extended conformation of all the calculated homo-peptides, while Figure 16 shows the axial and equatorial projections of -(Xxx)$_8$- after geometry optimization, details of the corresponding bb-bb, sc-bb, and/or sc-sc intramolecular interactions, and the average values of the φ,ψ backbone torsion angles. It is worth noting that E$^{coop}$(Xxx)$_n$ values provide information about how non-additive interactions, such as polarization and/or charge transfer contributions, may enhance C$_5$, van der Waals, bb-sc hydrogen bonding and π···π stacking attractive forces, as a function of the peptide composition and length, when the backbone adopts a fully-extended conformation. However, the energy gain associated to such non-additive effects cannot be related with the stability of the fully-extended conformation with respect to other secondary motifs.

The results obtained for the -(Gly)$_n$- homo-peptides, which perfectly preserve the fully-extended conformation, indicate that cooperative energy effects are small when the only stabilizing interaction is that of the C$_5$ form. Thus, E$^{coop}$(Gly)$_n$ decreases from -0.2 to -1.7 kcal/mol when *n* increases from 3 to 8. We attribute this difference, representing a small gain of energy (approximately -0.3 kcal/mol per Gly residue incorporated into the peptide chain), to the specific geometry of the C$_5$ interaction, which is severely restricted by the molecular architecture. These constraints induce an elongation of the H-bonding distance H···O (*i.e.*, 2.14 Å) with respect to the ideal value of 1.90 Å and a remarkable reduction of the ∠N–H···O angle (*i.e.,* 108.1º) with respect to the ideal value of 180º. Consequently, the peculiar H-bonding geometry of the C$_5$ interaction apparently limits the occurrence of large cooperative effects in the fully-extended conformation. It

is worth noting that cooperative effects were found to be very pronounced in other peptide secondary structures with intra- or intermolecular H-bonding geometries close to the ideal ones, as for example in the $\alpha$- and $3_{10}$-helices[117-119] and in the $\beta$-sheets.[120]

The two fully-extended conformations considered for -(Ser)$_n$- differ in the type of bb-sc H-bond. Our results show that the highest energy difference between the dispositions with the $C_{i-1}=O_{i-1}\cdots H_i-O_i$ (Ser$_n$-a) and $H_i-O_i\cdots H_{i+1}-N_{i+1}$ (Ser$_n$-b) H-bonds is 1.0 kcal/mol ($n = 8$). In spite of this small energy difference, the cooperative behaviors of Ser$_n$-a and Ser$_n$-b are drastically different. Specifically, Ser$_n$-a exhibits an anticooperative behavior with $E^{coop}$(Ser$_n$-a) increasing from +0.2 ($n = 3$) to +1.7 kcal/mol ($n = 8$), whereas Ser$_n$-b displays stabilizing cooperative effects with $E^{coop}$(Ser$_n$-b) decreasing from -0.1 ($n= 3$) to -0.9 kcal/mol ($n= 8$). These opposite behaviors can be understood by examining the intramolecular interactions shown in Figure 16. Thus, the $O_i-H_i\cdots O_{i-1}=C_{i-1}$ interactions in Ser$_n$-a occur when the distance between the oxygen atoms of the hydroxyl and carbonyl groups (sc and bb moieties, respectively) is only 2.8 Å. This situation generates repulsive $O\cdots O$ interactions, thus explaining the anticooperative effects found for Ser$_n$-a. For Ser$_n$-b, the distance between the closest side-chain and backbone oxygen atoms increases to 3.5 Å and, most importantly, the H-atom of the NH group is between them. This stabilizing atomic disposition is responsible for the small cooperative energy effects found for the latter peptide. On the other hand, it is worth noting that the optimized geometries are slightly distorted with respect to the initial fully-extended conformation, exhibiting a partial evolution towards the pleated-sheet conformation. Thus, the optimized conformations for both Ser$_n$-a and Ser$_n$-b should be considered intermediate between such two arrangements.

Conversely, the fully-extended conformation is completely lost for the -(Phe)$_n$- homo-peptides. Indeed, optimized geometries of Phe$_n$-a and Phe$_n$-b correspond to the canonical pleated-sheet conformations. The $N-H\cdots\pi$ interactions found for the former geometry (Phe$_n$-a) provide a linear energy gain, $E^{coop}$(Phe$_n$-a) increasing from -0.6 kcal/mol ($n= 3$) to -6.1 ($n= 8$). Similarly, the $\pi\cdots\pi$ stacking interactions exhibited by Phe$_n$-b result in favorable cooperative energy effects. However, in this latter case the energy gain is not linear for $n \geq 6$ since the amount of interactions depends not only on the number of residues but also on the organization in dimers or trimers of the interacting rings. For example, the insets in Figure 15a show that the six phenyl rings of Phe$_6$-b organize in two interacting groups with three rings per group, while the sc-sc interactions in Phe$_7$-b are distributed in one group with three rings and two groups with two rings each. This situation produces an increment of +0.2 kcal/mol in $E^{coop}$(Phe$_n$-b) when $n$ increases from 6 to 7. A similar case is observed by comparing Phe$_7$-b and Phe$_8$-b, the latter displaying four pairs of interacting rings (Figure 16). In this case, in spite of the increase of $n$ and, therefore, of the incorporation of a

new interacting ring, $E^{coop}(Phe_n$-b) decreases by only -0.1 kcal/mol. In general, the results displayed in Figure 15a for $Phe_n$-b indicate that elimination of trimers of interacting rings has a negative effect on the cooperative phenomena.

Calculations on -$(Aib)_n$- and -$(Deg)_n$- reflect the stability of the $2.0_5$-helix. Thus, in the absence of specific sc-bb or sc-sc interactions, like those of $(Phe)_n$-, the fully-extended conformation does not evolve towards the pleated-sheet structure. Comparison of the $E^{coop}$ values calculated for -$(Aib)_n$- and -$(Deg)_n$- with those obtained for -$(Ala)_n$- and -$(Gly)_n$- suggests that the attractive cooperative effects increase with both the $C^{\alpha,\alpha}$-dialkylation and increase of the size of the alkyl groups. These effects should be attributed to the enhancing strength of the van der Waals interactions between residue $i$ and both residues $i$-2 and $i$+2. In spite of this phenomenon, the $E^{coop}$ values are significantly lower than those reported for the pleated-sheet and helical structures. For example, $E^{coop}(Xxx_n)$= -1.7, -1.8, -2.3 and -2.8 kcal/mol for -$(Xxx)_n$- = -$(Gly)_8$-, -$(Ala)_8$-, -$(Aib)_8$- and -$(Deg)_8$-, respectively. Taken together, these calculations corroborate the conclusion that cooperative energy effects do not play any important role in the stability of the $2.0_5$-helix. Probably, other features are responsible for the experimental observation of this specific structural motif.

A detailed inspection of the results shown in Figure 15a suggest that the experimental observation of the $2.0_5$-helix in -$(Deg)_n$- homo-peptides is not due to the existence of favorable cooperative energy effects but rather to the lack of repulsive interactions. In order to corroborate this hypothesis, additional calculations were performed on the -$(Xxx)_n$- homo-peptides with $n = 4$, 6, and 8. The starting structure in all these calculations was the $3_{10}$-helical conformation with $\varphi,\psi$ = -60°,-35°. It should be emphasized that, in a large number of theoretical studies, the $\alpha$- and $3_{10}$-helices have been found to be very stable 3D-structures due to the formation of intramolecular H-bonds.

Figure 15b represents the relative energy of the optimized helical structure with respect to that of the optimized fully-extended conformation for all the considered homo-peptides ($\Delta E_{fe/h}$). Averaged torsion angles, which are included in Figure 15b for the homo-peptides with $n$= 8, indicate that the helical conformation is perfectly maintained for -$(Gly)_n$-, -$(Ala)_n$-, -$(Ser)_n$-, and -$(Aib)_n$-. Furthermore, the helix is considerably favored with respect to the fully-extended conformation, since the energy gap between the two optimized 3D-structures increases rapidly with $n$. Obviously, this finding should be ascribed to the intramolecular H-bonds occurring in the helical conformation, which exhibit a much more favorable geometry than those of the $C_5$ interactions of the fully-extended structure. This feature is fully consistent with $E^{coop}(Xxx)_n$ values reported for the helical conformation of different homo-peptides, which were found to significantly more favorable than those obtained in this work for the fully-extended conformation.[117-119] Thus, attractive non-

16

additive energy contributions increased with the strength of the H-bond, which in turn was related with its geometry. In contrast, non-additive contributions were slightly lower for secondary structures stabilized by intermolecular H-bonds, as for example the β-sheet.[120].

On the other hand, comparison of Gly, Ala, and Aib homo-peptides indicates that, for a given $n$, the absolute $\Delta E_{fe/h}$ values increase as follows: -(Gly)$_n$- < -(Ala)$_n$- < -(Aib)$_n$-. This relative energy order proves that van der Waals interactions induced by the methyl side groups are slightly more stabilizing in the helical conformation than in the fully-extended conformation. This trend changes drastically when the comparison includes the Deg homo-peptides. Although -(Deg)$_n$-, independently of $n$, maintain the helical conformation upon geometry optimization starting from $\varphi,\psi$ = -60º,-35º, the relative stability of the helix with respect to the fully-extended arrangement is significantly smaller, and even not always of the same sign, than that obtained for -(Gly)$_n$-, -(Ala)$_n$-, -(Aib)$_n$-. More specifically, the $\Delta E_{fe/h}$ values are +1.1, -1.2 and -14.1 kcal/mol for $n$= 4, 6 and 8, respectively, while those obtained for -(Aib)$_n$- are -8.4, -17.6 and -27.8 kcal/mol. Therefore, for Deg homo-peptides *in vacuo*, the fully-extended conformation is slightly favored over the $3_{10}$-helix at the tetramer level, slightly disfavored at the hexamer, while it is significantly disfavored only at the octamer level. For the Deg homo-octaoligomer, however, the energy difference between fully-extended and helical conformations is about one half of that calculated for -(Aib)$_8$-. These results support the view that the attractive van der Waals interactions associated with the two Aib C$^\alpha$-methyl substituents transform into repulsive when these substituents are replaced by ethyl groups (as in Deg). Thus, the size of the two C$^\alpha$-ethyl substituents in Deg is not appropriate for the geometry of the helical conformation, which becomes exclusively stabilized by the intramolecular H-bonds (*i.e.*, sc-sc and bb-sc interactions are unfavorable). In contrast, the fully-extended conformation exhibits bb-bb C$_5$ and sc-sc van der Waals interactions which are both attractive.

A completely different situation was obtained for -(Phe)$_n$- since the helical conformation is disfavored with respect to the above discussed pleated-sheet structure (Phe$_n$-a) for each value of $n$. This result should be attributed to the fact that the sc-sc and bb-sc interactions involving the phenyl side groups are less favorable in the helix than in the pleated-sheet conformation.

In summary, our results demonstrate that the preferences of -(Deg)$_n$- towards the fully-extended conformation do not arise from the existence of very stabilizing cooperative effects in this specific structure. Thus, the van der Waals interactions between the C$^\alpha$-ethyl groups of residues $i$ and $i$+2 give rise to cooperative phenomena that are significantly smaller than those reported for other peptide secondary structures. In contrast, the steric conflicts produced by the two C$^\alpha$-ethyl substituents severely affect the conformational preferences of -(Deg)$_n$- homo-peptides, which exhibit a drastic destabilization of the classical conformations (as for example the helix) typically

favored for sequences constructed using coded α-amino acids. This effect is expected to be more pronounced in the condensed phases, in which the cooperative effects associated to the helical structures are much less significant than *in vacuo*.[117-119] Accordingly, in the condensed phases the preference of -(Deg)$_n$- for the fully-extended conformation is expected to be greater than *in vacuo* since the stability of the other structures decreases.

On the basis of DFT calculations on models constituted by Gly-homopeptides, Dannenberg and co-workers[121] proved that H-bond cooperativity plays an important role in the energetic of the association of β-strands to form β-sheets. However, such inter-strand cooperativity was found to be highly influenced by the strength of the H-bonded $C_5$ conformation, which was also identified as a cooperative interaction. Thus, the inherent cooperativity of the β-sheets was eliminated by weakening the $C_5$ H-bond. These results are fully consistent with those displayed in Figures 15 and 16. Moreover, our results on homopeptides made of residues different from Gly prove that the chemical nature of the side group also plays an essential role in the cooperativity of the $C_5$ H-bond because of the formation of sc-sc and bb-sc interactions.

## 6. SUMMARY AND OUTLOOK

In the present contribution, we initially reviewed all recently published literature data on the planar, fully-extended peptide $2.0_5$-helical structure and its $C_5$-conformation basic unit, unambiguously obtained by X-ray diffraction analyses. If all publications before 2012 surveyed in our previous review[4] are also taken into consideration, it turns out clearly that the overwhelming (but certainly not exclusive) preference should be assigned to Gly among coded amino acids and to the $C^\alpha$-tetrasubstituted α-amino acids with both side chains longer than methyl (*e.g.*, Deg). Remarkably, crystal structures of $2.0_5$-helices from both Gly and Deg homo-oligopeptides to the pentamer level were reported when carrying well selected terminal groups and under appropriate crystallization conditions. In any case, the utilization of achiral residues is particularly useful, but it is not a strict prerequisite, as originally believed. Moreover, a statistical survey on the data obtained from globular proteins to high resolution confirms Gly as the residue with the significantly highest propensity to give rise to this flat backbone structure. In this specific rank order, it is followed by amino acids with small (*e.g.*, Ser) or aromatic (*e.g.*, Phe) side chains. Interestingly, $2.0_5$-helices as long as tri- or tetrapeptide sequences were identified. Moreover, our DFT computational study on a number of homo-oligopeptides provided valuable information on the relative stabilities of the $2.0_5$-helices and the role played by cooperative energy effects. Most importantly, attractive effects are enhanced by both $C^{\alpha,\alpha}$-dialkylation and increase of the size of the alkyl groups. In general,

however, this phenomenon is remarkably less significant for the $2.0_5$-helix than those reported for the α- and $3_{10}$-helices, and the pleated β-sheet conformation as well. In the case of the preferential formation of the -(Deg)$_n$- $2.0_5$-helices, it appears that the unfavorable steric interactions induced by the two $C^\alpha$-ethyl substituents of each residue in the most common polypeptide conformations are the major factors responsible for this experimental observation.

At the end of this article, it is appropriate to mention that a novel peptide conformation, termed "$C_5$i" (where "i" stands for *imidate*), was recently reported in a series of papers by Prabhakaran and colleagues.[122-124] Typically, it does occur at the C-terminus of a peptide *amide* backbone, where the last -NH-CHR-CO-NHR amide group is artificially modified to a cyclic imidate (oxazine). This chemical conversion was specifically incorporated to remove the amide proton and concomitantly to introduce an H-bonding acceptor (the $sp^2$ N atom of the imidate group). The net result is a new type of $i \rightarrow i$ ($C_5$) intramolecular H-bond, from the N-H group of the C-terminal amino acid (donor) to the imidate nitrogen atom (acceptor). This stabilization, in turn, requires a rotation of the "ψ" torsion angle of the C-terminal residue of the peptidomimetic to near 0°, a value in a disallowed region of the Ramachandran map for protein residues.

After many years of an almost exclusive focus on the polypeptide backbone *spaceland* (long-range peptide 3D-structures: α-helix, pleated β-sheet conformation, $3_{10}$-helix), it is evident that structural biochemists and organic chemists are becoming increasingly interested in the lower dimensional patterning (polypeptide *flatland*: 2D-structures) with the aim at expanding their overall general knowledge, in particular at synthesizing suitable α-amino acid building blocks (and the resulting planar peptide platform thereof) which could ultimately enable the construction of novel chemical and biochemical compounds and materials.

# REFERENCES

[1]     C. Toniolo, *CRC Crit. Rev. Biochem.* **1980**, *9*, 1.

[2]     C. Toniolo, M. Crisma, A. Moretto, C. Peggion, F. Formaggio, C. Alemán, C. Cativiela, C. Ramakrishnan, P. Balaram, *Chem. Eur. J.* **2015**, *21*, 13866.

[3]     C. Toniolo, M. Crisma, F. Formaggio, C. Alemán, C. Ramakrishnan, N. Kalmankar, P. Balaram, *Biopolymers (Pept. Sci.)* **2017**, *108*, e22911.

[4]     C. Peggion, A. Moretto, F. Formaggio, M. Crisma, C. Toniolo, *Biopolymers (Pept. Sci.)* **2013**, *100*, 621.

[5]     G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* **1963**, *7*, 95.

[6]     C. M. Venkatachalam, G. N. Ramachandran, *Annu. Rev. Biochem.* **1969**, *38*, 687.

[7]     L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 729.

[8]     R. W. Newberry, R. T. Raines, *Nat. Chem. Biol.* **2016**, *12*, 1084.

[9]     A. Aubry, V. Del Duca, M. Pantano, F. Formaggio, M. Crisma, C. Toniolo, J. Kamphuis, *Lett. Pept. Sci.* **1994**, *1*, 157.

[10]    N. Shamala, R. Nagaraj, P. Balaram, *J. Chem. Soc., Chem. Commun.* **1978**, 996.

[11]    C. Toniolo, E. Benedetti, *Trends Biochem. Sci.* **1991**, *16*, 350.

[12]    I. L. Karle, P. Balaram, *Biochemistry* **1990**, *29*, 6747.

[13]    G. R. Marshall, E. E. Hodgkin, D. A. Langs, G. D. Smith, J. Zabrocki, M. T. Leplawy, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 487.

[14]    K. A. Bolin, G. L. Millhauser, *Acc. Chem. Res.* **1999**, *32*, 1027.

[15]    E. Benedetti, V. Barone, A. Bavoso, B. Di Blasio, F. Lelj, V. Pavone, C. Pedone, G. M. Bonora, C. Toniolo, M. T. Leplawy, K. Kaczmarek, A. Redlinsky, *Biopolymers* **1988**, *27*, 357.

[16]    M. Tanaka, *Chem. Pharm. Bull. (Tokyo)* **2007**, *55*, 349.

[17]    M. Tanaka, N. Imawaka, M. Kurihara, H. Suemune, *Helv. Chim. Acta* **1999**, *82*, 494.

[18]    The X-ray diffraction structures of Ac-Gly-Deg-O*t*Bu, Z-Deg-Gly-O*t*Bu, Boc-Gly-ψ[CS-NH]-L-Leu-OMe, *p*CNBz-(ΔAla)$_3$-OMe, and Ac-(Deg)$_5$-O*t*Bu tetrahydrofuran solvate have been deposited with the Cambridge Crystallographic Data Centre with deposition numbers CCDC 1560166-1560170, respectively. The data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/structures.

[19]    F. Formaggio, M. Crisma, G. Ballano, C. Peggion, M. Venanzi, C. Toniolo, *Org. Biomol. Chem.* **2012**, *10*, 2413.

[20]  C. Andreu, R. Beerli, N. Branda, M. Conn, J. de Mendoza, A. Galan, I. Huc, Y. Kato, M. Tymoschenko, C. Valdez, E. Wintner, R. Wyler, J. Rebek, Jr., *Pure Appl. Chem.* **1993**, *65*, 2313.

[21]  S. Kubik, R. S. Meissner, J. Rebek, Jr., *Tetrahedron Lett.* **1994**, *35*, 6635.

[22]  C. Toniolo, G. M. Bonora, A. Bavoso, E. Benedetti, B. Di Blasio, V. Pavone, C. Pedone, V. Barone, F. Lelj, M. T. Leplawy, K. Kaczmarek, A. Redlinski, *Biopolymers* **1988**, *27*, 373.

[23]  C. Peggion, M. Crisma, C. Toniolo, F. Formaggio, *Tetrahedron* **2012**, *68*, 4429.

[24]  F. Formaggio, M. Crisma, C. Peggion, A. Moretto, M. Venanzi, C. Toniolo, *Eur. J. Org. Chem.* **2012**, 167.

[25]  M. Crisma, C. Peggion, A. Moretto, R. Banerjee, S. Supakar, F. Formaggio, C. Toniolo, *Biopolymers (Pept. Sci.)* **2014**, *102*, 145.

[26]  H. Maekawa, G. Ballano, C. Toniolo, N.-H. Ge, *J. Phys. Chem. (B)* **2011**, *115*, 5168.

[27]  M. Crisma, A. Moretto, C. Peggion, L. Panella, B. Kaptein, Q.B. Broxterman, C. Toniolo, *Amino Acids* **2011**, *41*, 629.

[28]  R. Lettieri, M. Bischetti, E. Gatto, A. Palleschi, E. Ricci, F. Formaggio, M. Crisma, C. Toniolo, M. Venanzi, *Biopolymers (Pept. Sci.)* **2013**, *100*, 51.

[29]  B. K. Saronjini, B. Narayana, H. S. Yathirajan, T. Gerber, E. Hosten, R. Betz, *Z. Kristallogr. NCS* **2013**, *228*, 245.

[30]  H. Zhu, J. Shi, Z. Huang, L. Lv, J. Duan, *J. Mol. Struct.* **2015**, *1089*, 170.

[31]  T. Kiyotani, Y. Sugawara, *Acta Crystallogr. C* **2012**, *68*, o498.

[32]  R. Francis, P. Raghavaiah, K. Pius, *Acta Crystallogr. B* **2014**, *70*, 942.

[33]  C.-F. Ng, H.-F. Chow, *Chem. Commun.* **2015**, *51*, 2349.

[34]  Y.-C. Zhang, X.-Q. Zhang, K. Wang, Q. Chen, *Acta Crystallogr. E* **2013**, *69*, o 1712.

[35]  Y.-X. Yang, S. W. Ng, *Acta Crystallogr. E* **2012**, *68*, o2170.

[36]  C. Huo, H. Xie, F. Chen, J. Tang, Y. Wang, *Adv. Synth. Catal.* **2016**, *358*, 724.

[37]  K. Katayama, T. Okamura, T. Sunadome, K. Nakagawa, H. Takeda, M. Shiro, A. Matsuda, S. Ichikawa, *J. Org. Chem.* **2014**,*79*, 2580.

[38]  J. Voskuhl, M. Waller, S. Bandaru, B. A. Tkachenko, C. Fregonese, B. Wibbeling, P. R. Schreiner, B. J. Ravoo, *Org. Biomol. Chem.* **2012**, *10*, 4524.

[39]  I. S. Bushmarinov, A. O. Dmitrienko, A. A. Korlyukov, M. Yu. Antipin, *J. Appl. Crystallogr.* **2012**, *45*, 1187.

[40]  A. J. Smith, F. I. Ali, D. V. Soldatov, *CrystEngComm* **2014**, *16*, 7196.

[41]  F. Formaggio, M. Crisma, C. Toniolo, C. Peggion, *Eur. J. Org. Chem.* **2013**, 3455.

[42]  S. S. Kumar, A. Nangia, *CrystEngComm* **2013**, *15*, 6498.

[43]   Y. Hitotsuyanagi, M. Odagiri, S. Kato, J.-I. Kusano, T. Hasuda, H. Fukaya, K. Takeya, *Chem. Eur. J.* **2012**, *18*, 2839.

[44]   P. Paoli, P. Rossi, L. Chelazzi, M. Altamura, V. Fedi, D. Giannotti, *Cryst. Growth Des.* **2016**, *16*, 5294.

[45]   S. J. Pike, J. Raftery, S. J. Webb, J. Clayden, *Org. Biomol. Chem.* **2014**, *12*, 4124.

[46]   M. J. Leonard, A. R. Lingham, J. O. Niere, N. R. C. Jackson, P. G. Mckay, H. M. Hugel, *RSC Adv.* **2014**, *4*, 14143.

[47]   M. A. Al-Omar, A.-G. E. Amr, H. A. Ghabbour, C. K. Quah, H.-K. Fun, *Acta Crystallogr. E* **2012**, *68*, o1377.

[48]   S. S. Kale, S. M. Kunjir, R. I. Gawade, V. G. Puranik, P. R. Rajamohanan, G. J. Sanjayan, *Chem. Commun.* **2014**, *50*, 2886.

[49]   S. H. Gellman, I. A. Guzei, *CSD Communication*, **2016**, DOI: 10.5517/ccdc.csd.ccrsd74.

[50]   J. Wang, R. P. Hammer, F. R. Fronczek, *CSD Communication*, **2015**, DOI: 10.5517/cc1jfqk7.

[51]   B. Ghebremariam, M. L. McLaughlin, F.R.Fronczek, *CSD Communication*, **2015**, DOI: 10.5517/cc1jc8rx.

[52]   Y.-H. Liu, L. Shui, M. L. McLaughlin, F. R. Fronczek, *CSD Communication*, **2015**, DOI: 10.5517/cc1jvjgc.

[53]   J. R. Karr, F. R. Fronczek, *CSD Communication*, **2014**, DOI: 10.5517/cc13sd5g.

[54]   A. Scroggs, R. P. Hammer, F. R. Fronczek , *CSD Communication*, **2015**, DOI: 10.5517/cc1438wd.

[55]   J. Giraldes, M. L. McLaughlin, F. R. Fronczek, *CSD Communication*, **2015**, DOI: 10.5517/cc14nv7w.

[56]   S. Dörrich, S. Fallgner, S. Schweeberg, C. Burschka, P. Brodin, B. M. Wissing, B. Basta, P. Schell, U. Bauer, R. Tacke, *Organometallics* **2012**, *31*, 5903.

[57]   J. Xu, C. Mon, T. Zhu, B.-A. Song, Y. R. Chi, *Org. Lett.* **2014**, *16*, 3272.

[58]   L. Wei, X. Yang, Z.-Q. Gao, D.-Q. Wang, J.-Q. Xue, W.-J. Yang, J.-J. Wang, Y.-H. Zhang, G.-L. Fang, Y. Liu, *Synth. React. Inorg., Met.-Org., Nano-Met.Chem.* **2015**, *45*, 455.

[59]   L. Hroch, M. Hruskova, J. Schmitz, G. Schnakenburg, M. Gutschow, *Synthesis* **2012**, *44*, 1907.

[60]   K. Li, G. Tan, J. Huang, F. Song, J. You, *Angew. Chem. Int. Ed.* **2013**, *52*, 12942.

[61]   D. Uraguchi, Y. Ueki, A. Sugiyama, T. Ooi, *Chem. Sci.* **2013**, *4*, 1308.

[62]   T. Buyck, Q. Wang, J. Zhu, *Angew. Chem. Int. Ed.* **2013**, *52*, 12714.

[63] M. W. Ha, M. Lee, S. Choi, S. Kim, S. Hong, Y. Park, M.-H. Kim, T.-S. Kim, J. Lee, J. K. Lee, H.-G. Park, *J.Org.Chem.* **2015**, *80*, 3270.

[64] M. A. Marsini, J. T. Reeves, J.-N. Desrosiers, M. A. Herbage, J. Savoie, Z. Li, K. R. Fandrick, C. A. Sader, B. McKibben, D. A. Gao, J. Cui, N. C. Gonnella, H. Lee, X. Wei, F. Roschangar, B. Z. Lu, C. H. Senanayake, *Org. Lett.* **2015**, *17*, 5614.

[65] D. Uraguchi, K. Yamada, T. Ooi, *Angew. Chem. Int. Ed.* **2015**, *54*, 9954.

[66] C. S. Sit, A. C. Ruzzini, E. B. Van Arnam, T. R. Ramadhar, C. R. Currie, J. Clardy, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 13150.

[67] D. Siodlak, M. Bujak, M. Stas, *J. Mol. Struct.* **2013**, *1047*, 229.

[68] P. Lenartowicz, M. Makowski, B. Zarychta, K. Ejsmont, *Acta Crystallogr. E* **2014**, *70*, 599.

[69] Z. Q. Guo, W.-Q. Chen, X.-M. Duan, *Dyes Pigm.* **2011**, *92*, 619.

[70] M. Doi, *CSD Communication*, **2016**, DOI: 10.5517/ccdc.csd.cc1lnxtz.

[71] C. Toniolo, M. Crisma, F. Formaggio, C. Peggion, *Biopolymers (Pept. Sci.)* **2001**, *60*, 396.

[72] M. Crisma, M. De Zotti, F. Formaggio, C. Peggion, A. Moretto, C. Toniolo, *J. Pept. Sci.* **2015**, *21*, 148.

[73] M. Crisma, C. Toniolo, *Biopolymers (Pept. Sci.)* **2015**, *104*, 46.

[74] A. Lapidot, C. S. Irving, *Biochemistry* **1979**, *18*, 1788.

[75] T. Srikrishnan, N. Winiewicz, R. Parthasarathy, *Int. J. Pept. Protein Res.* **1982**, *19*, 103.

[76] S. Aravinda, N. Shamala, P. Balaram, *Chem. Biodivers.* **2008**, *5*, 1238.

[77] A. Bosch, K.-P. Voges, G. Jung, W. Winter, *Acta Crystallogr. C* **1983**, *39*, 481.

[78] A. N. Jayashree, C. G. Suresh, M. Vijayan, C. Toniolo, G. M. Bonora, *Acta Crystallogr. C* **1987**, *43*, 1618.

[79] J. T. Jeremic, A. Linden, K. Moehle, H. Heimgartner, *Tetrahedron* **2005**, *61*, 1871.

[80] J. L. Flippen-Anderson, C. George, J. R. Deschamps, *Acta Crystallogr. C* **1994**, *50*, 1095.

[81] I. Dannecker-Dörig, A. Linden, H. Heimgartner, *Coll. Czech. Chem. Commun.* **2009**, *74*, 901.

[82] D. Ranganathan, S. Kurur, A. C. Kunwar, A. V. S. Sarma, M. Vairamani, I. L. Karle, *J. Pept. Res.* **2000**, *56*, 416.

[83] S. S. Zimmerman, M. S. Pottle, G. Némethy, H. A. Scheraga, *Macromolecules* **1977**, *10*, 1.

[84] M. Crisma, F. Formaggio, C. Toniolo, T. Yoshikawa, T. Wakamiya, *J. Am. Chem. Soc.* **1999**, *121*, 3272.

[85] C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, *Acta Crystallogr. B* **2016**, *72*, 171.

[86] R. A. Engh, R. Huber, *Acta Crystallogr. A* **1991**, *47*, 392

[87] D. E. Tronrud, P.A. Karplus, *Acta Crystallogr. D* **2011**, *67*, 699.

[88] X. Jiang, M. Cao, B. Teppen, S.Q. Newton, L. Schaefer, *J. Phys. Chem.* **1995**, *99*, 10521.

[89] P.A. Karplus, *Protein Sci.* **1996**, *5*, 1406.

[90] S. M. Malathy Sony, S. Saraboji, N. Sukumar, M. N. Ponnuswamy, *Biophys. Chem.* **2006**, *120*, 24.

[91] D.S. Berkholz, M.V. Shapovalov, R.L. Dunbrack, Jr., P.A. Karplus, *Structure* **2009**, *17*, 1316.

[92] A. Q. Zhou, C. S. O'Hern, L. Regan, *Protein Sci.* **2011**, *20*, 1186.

[93] R. Improta, L. Vitagliano, L. Esposito, *Proteins: Struct., Funct., Bioinform.* **2015**, *83*, 1973.

[94] C. Ramakrishnan, N. Prasad, *Int. J. Pept. Protein Res.* **1971**, *3*, 209.

[95] R. Taylor, O. Kennard, W. Versichel, *Acta Crystallogr. B* **1984**, *40*, 280.

[96] C. H. Görbitz, *Acta Crystallogr. B* **1989**, *45*, 390.

[97] I. Y. Torshin, I. T. Weber, R. W. Harrosin, *Protein Eng.* **2002**, *15*, 359.

[98] A. I. Gonzalez Florez, E. Mucha, D.-S. Ahn, S. Gewinner, W. Schöllkopf, K. Pagel, G. von Helden, *Angew. Chem. Int. Ed.* **2016**, *55*, 3295.

[99] C. M. Leavitt, A. F. DeBlase, C. J. Johnson, M. van Stipdonk, A. B. McCoy, M. A. Johnson, *J. Phys. Chem. Lett.* **2013**, *4*, 3450.

[100] H. Torii, M. Kawanaka, *J. Phys. Chem. (B)* **2016**, *120*, 1624.

[101] H. Torii, *J. Phys. Chem. Lett.* **2012**, *3*, 112.

[102] P. S. Walsh, J. C. Dean, C. McBurney, H. Kang, S. H. Gellman, T. S. Zwier, *Phys. Chem. Chem. Phys.* **2016**,*18*, 11306.

[103] M. H. Abraham, R. J. Abraham, *New J. Chem.* **2017**, *41*, 6064.

[104] J. Casanovas, G. Revilla-Lopez, M. Crisma, C. Toniolo, C. Alemán, *J. Phys. Chem. (B)* **2012**, *116*, 13297.

[105] J. Torras, D. Zanuy, M. Crisma, C. Toniolo, O. Betran, C. Alemán, *Biopolymers (Pept. Sci.)* **2008**, *90*, 695.

[106] B. Bae, S. Ohene-Adjei, S. Kocherginskaya, R. I. Mackie, M. A. Spies, I. K. Cann, S. K. Nair, *J. Biol. Chem.* **2008**, *283*, 12415.

[107] L. Coates, P. T. Erskine, M. P. Crump, S. P. Wood, J. B. Cooper, *J. Mol. Biol.* **2002**, *318*, 1405.

[108] V. J. Cee, L. B. Schenkel, B. L. Hodous, H. L. Deak, H. N. Nguyen, P. R. Olivieri, K. Romero, A. Bak, X. Be, S. Bellon, T. L. Bush, A. C. Cheng, G. Chung, S. Coats, P. M. Eden, K. Hanestad, P. L. Gallant, Y. Gu, X. Huang, R. L. Kendall, M.-H. J. Lin, M. J. Morrison, V. F. Patel, R. Radinsky, P. E. Rose, S. Ross, J.-R. Sun, J. Tang, H. Zha, M. Payton, S. D. Geuns-Meyer, *J. Med. Chem.* **2010**, *53*, 6368.

[109] M.J. Teale, X. Feug, L. Chen, T. Bice, E. J. Meehan, Mouse RANKL Structure at 1.9 Å Resolution (**2004**). PDB ID: 1S55. DOI: 10.2210/pdb1s55/pdb.

[110] G. N. Murshudov, W. R. Melik-Adamyan, A. I. Grebenko, V. V. Barynin, A. A. Vagin, B. K. Vainshtein, Z. Dauter, K.S. Wilson, *FEBS Lett*. **1992**, *312*, 127.

[111] A. Åberg , A. Yaremchuk , M. Tukalo , B. Rasmussen , S. Cusack, *Biochemistry* **1997**, *36*, 3084.

[112] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford, CT, **2009**.

[113] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.

[114] Y. Zhao, D. G. Truhlar, *J. Chem. Phys.* **2006**, *125*, 194101.

[115] K. Remya, Ch. H. Suresh, *J. Comput. Chem.* **2013**, *34*, 1341.

[116] Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157

[117] C. Alemán, R. Roca, F. J. Luque, M. Orozco, *Proteins: Struct. Funct. Bioinform.* **1997**, *28*, 83.

[118] C. Alemán, *Proteins: Struct. Funct. Bioinform.* **1997**, *29*, 575.

[119] C. Alemán, J. J. Navas, S. Muñoz-Guerra, *Biopolymers* **2017**, *41*, 721.

[120] E. Mayans, G. Ballano, J. Casanovas, L. J. del Valle, M. M. Pérez-Madrigal, F. Estrany, A. I. Jiménez, J. Puiggalí, C. Cativiela, C. Alemán, *Soft Matter* **2016**, *12*, 5475.

[121] R. Viswanathan, A Asensio, J. J. Dannenberg, *J. Phys. Chem. A* **2004**, *108*, 9205.

[122] D. N. Reddy, R. Thirupathy, E. N. Prabhakaran, *Chem. Commun.* **2011**, *47*, 9417.

[123] R. Thirupathy, E. N. Prabhakaran, *Tetrahedron Lett.* **2014**, *55*, 3418.

[124] S. Tummiakatti, D. N. Reddy, E. N. Prabhakaran, *Biopolymers (Pept. Sci.)* **2015**, *104*, 21.

**TABLE 1** Fully-extended (C$_5$) structures found in Gly derivatives and peptides published since 2012

| Entry | Gly derivative / peptide | $\phi$ (°) | $\psi$ (°) | H-Bond geometry | | | Ref. |
|---|---|---|---|---|---|---|---|
| | | | | D…A (Å) | H…A (Å) | D-H…A (°) | |
| 1 | derivative | -178.2 | -177.2 | 2.705 | 2.38 | 102.6 | 29 |
| 2 | dipeptide [a] | 173.6 | -171.1 | 2.753 | 2.43 | 102.1 | 30 |
| 3 | dipeptide [b] | 173.6 | -177.5 | 2.624 | 2.25 | 105.9 | 31 |
| 4 | derivative [c] | 177.0 | -176.6 | 2.715 | 2.36 | 107.3 | 32 |
| | derivative [d] | 177.2 | -177.5 | 2.727 | 2.38 | 106.0 | |
| 5 | derivative | -176.4 | 177.5 | 2.651 | 2.27 | 107.0 | 33 |
| 6 | derivative [e] | -171.7 | 174.9 | 2.618 | 2.24 | 106.3 | 34 |
| | derivative [f] | -174.4 | 176.7 | 2.619 | 2.23 | 107.9 | |
| 7 | derivative | 179.2 | 172.2 | 2.676 | 2.31 | 105.7 | 35 |
| 8 | derivative | -171.8 | -179.7 | 2.679 | 2.33 | 104.6 | 36 |
| 9 | cyclic decapeptide [g] | -178.9 | -178.1 | 2.629 | 2.23 | 107.0 | 37 |
| 10 | derivative | 167.5 | -169.3 | 2.638 | 2.20 | 112.7 | 38 |
| 11 | dipeptide [h] | -176.3 | -173.9 | 2.604 | 2.26 | 103.8 | 39 |
| 12 | tetrapeptide [i] | 178.1 | -171.2 | 2.681 | 2.33 | 104.6 | 40 |
| | tetrapeptide [j] | 173.7 | -177.5 | 2.676 | 2.32 | 105.1 | |
| | tetrapeptide [k] | -172.3 | 169.1 | 2.660 | 2.29 | 105.7 | |
| | tetrapeptide [l] | -165.2 | 172.0 | 2.640 | 2.27 | 105.9 | |
| | tetrapeptide [m] | -175.6 | 172.8 | 2.729 | 2.40 | 103.4 | |
| | tetrapeptide [n] | -174.6 | 174.5 | 2.682 | 2.31 | 106.2 | |
| 13 | pentapeptide [o] | 163.9 | 171.5 | 2.664 | 2.30 | 105.6 | 40 |
| | pentapeptide [p] | -175.0 | 173.0 | 2.683 | 2.33 | 105.2 | |
| | pentapeptide [q] | 176.0 | 177.3 | 2.693 | 2.34 | 105.2 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | pentapeptide [r] | 174.2 | -170.0 | 2.671 | 2.30 | 105.9 |
| | pentapeptide [s] | -178.7 | 172.2 | 2.686 | 2.33 | 105.2 |
| | pentapeptide [t] | 172.6 | 174.6 | 2.673 | 2.32 | 105.2 |
| | pentapeptide [u] | -175.1 | 172.1 | 2.714 | 2.37 | 104.4 |
| | pentapeptide [v] | 173.7 | 174.3 | 2.687 | 2.31 | 106.5 |
| **14** | *bis*-endothioxo tripeptide [x] | 169.0 | -159.9 | 2.935 | 2.44 | 117.5 | 41 |
| | *bis*-endothioxo tripeptide [y] | -159.4 | 167.6 | 2.715 | 2.39 | 103.2 |

[a]-Gly-Gly*-. [b]-Leu-Gly- (molecule B). [c]Molecule A. [d]Molecule B. [e]Molecule A. [f]Molecule B. [g]-Gly-Sar-. [h]-Gly-Trp-. [i]Gly[2] in -(Gly)$_4$-, molecule A. [j]Gly[3] in -(Gly)$_4$-, molecule A. [k]Gly[4] in -(Gly)$_4$-, molecule A. [l]Gly[2] in -(Gly)$_4$-, molecule B. [m]Gly[3] in -(Gly)$_4$-, molecule B. [n]Gly[4] in -(Gly)$_4$-, molecule B. [o]Gly[2] in -(Gly)$_5$-, molecule A. [p]Gly[3] in -(Gly)$_5$-, molecule A. [q]Gly[4] in -(Gly)$_5$-, molecule A. [r]Gly[5] in -(Gly)$_5$-, molecule A. [s]Gly[2] in -(Gly)$_5$-, molecule B. [t]Gly[3] in -(Gly)$_5$-, molecule B. [u]Gly[4] in -(Gly)$_5$-, molecule B. [v]Gly[5] in -(Gly)$_5$-, molecule B. [x]Gly[2] in Z-Gly-ψ[CSNH]-Gly-ψ[CSNH]-Gly-OMe. [y]Gly[3] in Z-Gly-ψ[CSNH]-Gly-ψ[CSNH]-Gly-OMe.

**TABLE 2** Fully-extended ($C_5$) structures found in $C^\alpha$-trisubstituted $\alpha$-amino acid derivatives and peptides published since 2012

| Entry | Derivative / peptide | Amino acid | $\phi$ (°) | $\psi$ (°) | H-Bond geometry | | | Ref. |
|---|---|---|---|---|---|---|---|---|
| | | | | | D…A (Å) | H…A (Å) | D-H…A (°) | |
| 1 | derivative | Cys | -167.0 | -174.8 | 2.614 | 2.24 | 106.3 | 42 |
| 2 | cyclic hexapeptide [a] | Ala | -168.8 | 169.8 | 2.582 | 2.18 | 107.6 | 43 |
| 3 | bicyclic hexapeptide [b] | Dap | -165.2 | 173.9 | 2.655 | 2.32 | 102.6 | 44 |

[a] Ala[4]. [b] Dap (2,3-diaminopropionic acid)[5].

**TABLE 3** Fully-extended ($C_5$) structures found in Aib derivatives and peptides published since 2012

| Entry | Derivative / peptide | $\phi$ (°) | $\psi$ (°) | H-Bond geometry | | | Ref. |
|---|---|---|---|---|---|---|---|
| | | | | D…A (Å) | H…A (Å) | D-H…A (°) | |
| 1 | tripeptide [a] | 179.6 | -175.2 | 2.573 | 2.10 | 114.5 | 45 |
| 2 | derivative | 176.5 | -169.7 | 2.587 | 2.15 | 110.4 | 46 |
| 3 | derivative | 176.8 | 176.8 | 2.614 | 2.08 | 120.1 | 47 |
| 4 | dipeptide [b] | 176.5 | -177.9 | 2.555 | 2.10 | 112.7 | 48 |
| 5 | $\alpha,\beta$-tetrapeptide [c] | -179.7 | 177.3 | 2.637 | 2.22 | 108.4 | 49 |

[a] Aib[2] in -(Aib)$_3$-. [b] -Pro-Aib-. [c] Aib[4] in one of the two independent molecules of Boc-Phe-(ACPC)$_2$-Aib-O(pBr)Bzl, where ACPC is *trans*-2-aminocyclopentanecarboxylic acid and O(pBr)Bzl is *para*-bromobenzyloxy.

**TABLE 4**  Fully-extended ($C_5$) structures found in $C^\alpha$-tetrasubstituted $\alpha$-amino acid derivatives and peptides published since 2012

| Entry | Derivative / peptide | Amino acid | $\phi$ (°) | $\psi$ (°) | H-Bond geometry D...A (Å) | H...A (Å) | D-H...A (°) | Ref. |
|-------|---------------------|------------|-----------|-----------|----------|----------|-----------|------|
| **1** | derivative [a] | $(\alpha Bu^i)^b Phe$ | -174.7 | -175.9 | 2.616 | 2.19 | 109.2 | 50 |
| | derivative [c] | $(\alpha Bu^i)^b Phe$ | 176.7 | -178.3 | 2.671 | 2.22 | 111.7 | |
| **2** | derivative | $(\alpha,\beta\text{-diBzl})Asp(OMe)$ [d] | -177.7 | -179.8 | 2.591 | 2.14 | 113.5 | 51 |
| **3** | derivative | $(\gamma,\delta\text{-dehydro})\text{-}(\gamma,\delta\text{-}$ disubstituted)$Ac_5c$ [e] | 179.7 | 177.2 | 2.633 | 2.26 | 108.6 | 52 |
| **4** | derivative [f] | $(\alpha Pr^i)^g Ser$ (R) | -168.6 | 174.7 | 2.578 | 2.09 | 115.8 | 53 |
| | derivative [h] | $(\alpha Pr^i)^g Ser$ (S) | 171.0 | -170.0 | 2.562 | 2.07 | 116.6 | |
| **5** | derivative | $(\alpha Bzl)^i Phe$ | 179.7 | -178.6 | 2.624 | 2.17 | 114.8 | 54 |
| **6** | derivative [j] | $(\alpha Bzl)Asp$ | -175.0 | -170.4 | 2.649 | 2.24 | 110.4 | 55 |
| | derivative [k] | $(\alpha Bzl)Asp$ | -175.9 | -170.3 | 2.611 | 2.15 | 112.8 | |
| **7** | dipeptide [l] | $(\alpha TMSM)^m Ala$ | 173.6 | 167.8 | 2.640 | 2.24 | 109.2 | 56 |
| | dipeptide [l] | $(\alpha TMSM)^m Ala$ | 177.0 | 178.6 | 2.640 | 2.23 | 108.7 | |
| **8** | derivative [n] | $(\alpha,\beta,\gamma,\delta\text{-tetrasubstituted})Ac_5c$ | 167.3 | 176.9 | 2.587 | 2.19 | 107.3 | 57 |
| **9** | dipeptide [o] | $Deg$ [p] | 171.3 | 169.5 | 2.567 | 2.13 | 110.7 | 19 |
| **10** | derivative | $(\alpha CH_2\text{-NH-Boc})^q Phe$ | 177.4 | 177.6 | 2.634 | 2.21 | 110.4 | 58 |
| **11** | derivative [r] | $(\alpha CO\text{-NH-Ph})^s Apr$ | 179.5 | -178.5 | 2.619 | 2.22 | 109.2 | 59 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | derivative [t] | $(\alpha CO-NH-Ph)^s$Apr | 169.4 | -175.8 | 2.630 | 2.26 | 105.1 | |
| **12** | derivative | $(\alpha Bzl)$Trp | 178.0 | 175.3 | 2.632 | 2.20 | 110.5 | 60 |
| **13** | derivative [u] | $(\alpha CH=CH-COOCH_3)$Phe | -175.5 | -169.2 | 2.553 | 2.08 | 115.7 | 61 |
| | derivative [v] | $(\alpha CH=CH-COOCH_3)$Phe | -175.1 | 179.5 | 2.552 | 2.05 | 119.8 | |
| **14** | derivative | $(\alpha Ph\text{-}pBr)$Azb [w] | -175.6 | 169.8 | 2.662 | 2.24 | 108.0 | 62 |
| **15** | derivative | $(\alpha CO-NH$ alkyl)Phe | -173.9 | 178.7 | 2.596 | 2.07 | 119.1 | 63 |
| **16** | tripeptide [x] | $(\alpha Ph\text{-}pCl)$Cys(Ph-$p$Br) | 179.6 | -173.7 | 2.597 | 2.16 | 110.0 | 64 |
| | tripeptide [y] | $(\alpha Ph\text{-}pCl)$Cys(Ph-$p$Br) | -179.2 | -176.1 | 2.595 | 2.16 | 110.1 | |
| **17** | derivative | $(\alpha$-epoxyketone)Phe | -179.8 | -177.2 | 2.616 | 2.21 | 115.6 | 65 |
| **18** | cyclic hexadepsipeptide | $(\alpha Me)^{z'}$Ser | 175.3 | -169.9 | 2.623 | 2.18 | 110.6 | 66 |
| **19** | dipeptide [z''] | Dp$_n$g | -174.6 | 178.1 | 2.546 | 2.09 | 112.8 | 25 |
| **20** | derivative | Dp$_n$g | 177.9 | 177.5 | 2.575 | 2.13 | 111.9 | 25 |
| **21** | tetrapeptide | Epg[1] [z'''] | 178.9 | -173.5 | 2.553 | 2.09 | 112.9 | 25 |
| | tetrapeptide | Deg[2] | -179.0 | 179.1 | 2.552 | 2.09 | 112.7 | |
| | tetrapeptide | Deg[4] | 178.0 | 179.0 | 2.593 | 2.15 | 111.9 | |

[a] Molecule A. [b] $C^\alpha$-Isobutyl. [c] Molecule B. [d] $C^{\alpha,\beta}$-Dibenzyl-$C^\beta$-methoxy. [e] Ac$_5$c is 1-aminocyclopentane-1-carboxylic acid. [f] Molecule A. [g] $C^\alpha$-Isopropyl. [h] Molecule B. [i] $C^\alpha$-Benzyl. [j] Molecule A. [k] Molecule B. [l] Molecule A. [m] $C^\alpha$-(Trimethylsilyl)methyl. [n] Molecule B. [o] Deg[1] in -(Deg)$_2$-. [p] Deg is $C^{\alpha,\alpha}$-diethylglycine. [q] Boc is *tert*-butyloxycarbonyl. [r] Molecule A. [s] Ph is phenyl; Apr is $\alpha$-aminopropionic acid. [t] Molecule B. [u] Molecule A. [v] Molecule B. [w] ($C^\alpha$-$p$-Bromophenyl)-$C^\gamma$-azido-butyric acid. [x] Molecule A. [y] Molecule B. [z'] $C^\alpha$-Methyl. [z''] -Dp$_n$g-Deg- (Dp$_n$g is $C^{\alpha,\alpha}$-di-$n$-propylglycine). [z'''] Epg is $C^\alpha$-ethyl, $C^\alpha$-$n$-pentylglycine.

**TABLE 5** Fully-extended ($C_5$) structures found in $\alpha,\beta$-didehydro-$\alpha$-amino acid derivatives and peptides published since 2012

| Entry | Derivative / peptide | Amino acid | $\phi$ (°) | $\psi$ (°) | H-Bond geometry D…A (Å) | H…A (Å) | D-H…A (°) | Ref. |
|---|---|---|---|---|---|---|---|---|
| **1** | derivative | $\Delta$Ala | -175.5 | -177.8 | 2.684 | 2.26 | 110.2 | 67 |
| **2** | dipeptide [a] | $\Delta$Ala | 175.1 | 178.0 | 2.636 | 2.22 | 108.2 | 68 |
| **3** | derivative [b] | $\Delta$Pyr | -172.1 | 178.4 | 2.522 | 1.96 | 114.5 | 69 |
| **4** | thiostrepton | $\Delta$Ala [c] | 163.5 | -170.7 | 2.570 | 2.16 | 109.1 | 70 |
| | | $\Delta$Ala [d] | -179.0 | -169.5 | 2.534 | 2.11 | 110.0 | |

[a] -Phe-$\Delta$Ala-. [b] $C^{\alpha,\beta}$-Didehydro-pyridin-2-yl-alanyl. [c] -Ala-$\Delta$Ala-Ala-. [d] -$\Delta$Ala-$NH_2$.

**TABLE 6**  Fully-extended ($C_5$) structures found in peptides still unpublished by the Padova group[18]

| Entry | Peptide | Amino acid | $\phi$ (°) | $\psi$ (°) | H-Bond geometry D…A (Å) | H…A (Å) | D-H…A (°) |
|---|---|---|---|---|---|---|---|
| 1 | Ac-Gly-Deg-O*t*Bu | Deg | -179.1 | -179.8 | 2.612 | 2.18 | 110.6 |
| 2 | Z-Deg-Gly-O*t*Bu [a] | Deg | -168.8 | 179.0 | 2.576 | 2.14 | 110.8 |
|   | Z-Deg-Gly-O*t*Bu [b] | Deg | 178.8 | 177.5 | 2.572 | 2.14 | 111.0 |
| 3 | Boc-Gly-ψ[CS-NH]-L-Leu-OMe [c] | Leu | -150.5 | 172.6 | 2.659 | 2.33 | 102.9 |
|   | Boc-Gly-ψ[CS-NH]-L-Leu-OMe [d] | Leu | -151.4 | 170.9 | 2.661 | 2.34 | 102.5 |
| 4 | *p*CNBz-(ΔAla)$_3$-OMe | ΔAla$^1$ | 180.0 | 176.5 | 2.585 | 2.15 | 111.2 |
|   |   | ΔAla$^2$ | -178.0 | -176.8 | 2.579 | 2.14 | 111.5 |
|   |   | ΔAla$^3$ | -177.5 | 176.5 | 2.629 | 2.21 | 109.9 |

[a] Molecule A.  [b] Molecule B.  [c] Molecule A.  [d] Molecule B.  [e] *p*CNBz, *para*-cyanobenzoyl.

**Table 7** Average geometrical parameters characterizing the $C_5$ intramolecularly H-bonded conformation from a statistical analysis of all X-ray diffraction structures of amino acid derivatives and peptides known to date for Gly, $C^\alpha$-trisubstituted $\alpha$-amino acids, $C^\alpha$-tetrasubstituted $\alpha$-amino acids, and $\Delta$Ala

| | Gly [a] | $C^\alpha$-trisubstituted $\alpha$-amino acids [b] | $C^\alpha$-tetrasubstituted $\alpha$-amino acids [c] | $\Delta$Ala [d] |
|---|---|---|---|---|
| Bond distances (Å) | | | | |
| N-$C^\alpha$ | 1.446 | 1.451 | 1.463 | 1.401 |
| $C^\alpha$-C | 1.510 | 1.525 | 1.539 | 1.494 |
| C=O | 1.221 | 1.219 | 1.211 | 1.222 |
| N-H | 0.89 | 0.88 | 0.89 | 0.87 |
| Bond angles (°) | | | | |
| N-$C^\alpha$-C | 110.2 | 108.1 | 104.5 | 110.1 |
| $C^\alpha$-C=O | 122.1 | 121.5 | 122.8 | 120.8 |
| $C^\alpha$-N-H | 119.3 | 117.7 | 116.6 | 116.1 |
| H-bond parameters (Å, °) | | | | |
| Distance N…O | 2.684 | 2.645 | 2.597 | 2.619 |
| Distance H…O | 2.34 | 2.28 | 2.15 | 2.20 |
| Angle N-H…O | 103.3 | 105.2 | 110.7 | 109.2 |

[a] Values averaged over 123 residues. [b] Values averaged over 99 residues. [c] Values averaged over 188 residues. [d] Values averaged over 21 residues.

**Table 8**   Propensity distribution of the coded amino acids for the fully-extended $C_5$ conformation from our statistical analysis of protein X-ray diffraction structures (resolution cut-off: 1.5 Å; homology cut-off: 30%)

| Residue | No. in data set | No. in the fully-extended region [a] | Propensity |
|---------|-----------------|--------------------------------------|------------|
| Ala | 39941 | 354 | 0.936 |
| Leu | 41791 | 50 | 0.126 |
| Ile | 26065 | 7 | 0.028 |
| Val | 33283 | 6 | 0.019 |
| Asp | 28195 | 246 | 0.922 |
| Asn | 20450 | 209 | 1.080 |
| Glu | 29774 | 93 | 0.330 |
| Gln | 17446 | 81 | 0.490 |
| Cys | 5784 | 106 | 1.936 |
| Met | 7702 | 45 | 0.617 |
| Ser | 27376 | 539 | 2.080 |
| Thr | 26356 | 93 | 0.373 |
| Arg | 23025 | 147 | 0.674 |
| Lys | 25643 | 95 | 0.391 |
| His | 10937 | 127 | 1.227 |
| Phe | 19006 | 235 | 1.306 |
| Tyr | 16921 | 176 | 1.099 |
| Trp | 7145 | 82 | 1.212 |
| Pro | 20925 | 0 | 0.000 |
| Gly | 36294 | 1702 | 4.954 |
| **Total** | 464059 | 4393 | |

[a] $\phi = 180\pm20°$, $\psi = 180\pm20°$.

# Figure Legends

**FIGURE 1**    (**A**) Possible conformations in a system of four covalently linked peptide units where each intramolecular H-bond occurs between an "upstream" N-H hydrogen and a "downstream" carbonyl oxygen. (**B**) The basic unit ($C_5$ conformation) of the $2.0_5$-helix formed by Gly (R = R′ = H), and by $C^\alpha$-tri- (R or R′ = H) and $C^\alpha$-tetra- (both R and R′ ≠ H) substituted $\alpha$-amino acid residues with the $C^\alpha$ $sp^3$-configuration.

**FIGURE 2**    The most representative achiral and chiral $C^\alpha$-tetrasubstituted $\alpha$-amino acids known to adopt the $C_5$ conformation in the crystal state.

**FIGURE 3**    X-Ray diffraction structure of $p$BrBz-[D-($\alpha$Me)Leu]$_3$-O$t$Bu (adapted from ref. [9]). The $C_5$ intramolecular H-bonds are marked.

**FIGURE 4**    Comparison among the X-ray diffraction structures of Tfa-(Deg)$_5$-O$t$Bu (**A**) (adapted from ref. [15]),    Tfa-(Deg)$_5$-OEt (**B**) (adapted from ref. [16]), Ac-(Deg)$_5$-O$t$Bu (**C**),[18] and PyrAc-(Deg)$_5$-O($p$NO$_2$)Bzl (**D**) (adapted from ref. [19]). The $C_5$  and $3_{10}$-helical intramolecular H-bonds are marked.

**FIGURE 5**    Proposed informational 12-mer peptide replicator based on six alternating Deg residues (adapted from ref. [20]).

**FIGURE 6**    The two independent molecules (**A** and **B**) in the asymmetric unit of $H_2^+$-(Gly)$_5$-O$^-$ from the X-ray diffraction analysis  (adapted from ref. [40]). The $C_5$ intramolecular H-bonds are marked.

**FIGURE 7**    X-Ray diffraction structure of Z-Gly-$\psi$[CSNH]-Gly-$\psi$[CSNH]-Gly-OMe (adapted from ref. [41]). The $C_5$ intramolecular H-bonds are marked.

**FIGURE 8**    X-Ray diffraction structure of Tfa-($S$)-Epg-Deg-Dp$_n$g-Deg-O$t$Bu (adapted from ref. [25]). The $C_5$ intramolecular H-bonds are marked.

**FIGURE 9**    Unpublished X-ray diffraction structures of: Z-Deg-Gly-O$t$Bu (**A** and **B**, two independent molecules), Ac-Gly-Deg-O$t$Bu (**C**), and Boc-Gly-$\psi$[CS-NH]-L-Leu-

OMe (**D** and **E**, two independent molecules). The $C_5$ intramolecular H-bonds are marked. For a reference, see [18].

**FIGURE 10**    X-Ray diffraction structure of $p$CNBz-($\Delta$Ala)$_3$-OMe. The ($C_5$) N-H…O=C and the accompanying $C^{\beta}_{i+1}$-H…$O_i$=$C_i$ intramolecular H-bonds are marked. For a reference, see [18].

**FIGURE 11**    Propensity distribution of the coded amino acids for the fully-extended $C_5$ conformation from our statistical analysis of protein X-ray diffraction structures (resolution cut-off: 1.5 Å; homology cut-off: 30%).

**FIGURE 12**    The fully-extended 2.0$_5$-helix formed by the -Gly$^{49}$-Gly$^{50}$-Gly$^{51}$- sequence from the X-ray diffraction structure of the protein S-layer associated multidomain endoglucanase (PDB code: 2ZEX; resolution: 1.20 Å) (adapted from ref. [106]). Only the *interstrand* H-bonds are marked.

**FIGURE 13**    The fully-extended 2.0$_5$-helix formed by the -Ser$^{84}$-Ser$^{85}$-Ser$^{86}$- sequence from the X-ray diffraction structure of the protein endothiapepsin (PDB code: 1GVU; resolution: 0.94 Å) (adapted from ref. [107]). Only the *interstrand* H-bonds are marked.

**FIGURE 14**    The fully-extended 2.0$_5$-helix formed by the -Gly$^{283}$-Gly$^{284}$-Gly$^{285}$-Gly$^{286}$- sequence from the X-ray diffraction structure of the protein histidyl-tRNA synthetase (from *Thermus thermophilus*) (PDB code: 1ADJ; resolution: 2.70 Å) (adapted from ref. [111]).

**FIGURE 15**    (**a**) Cooperative effects for the fully-extended conformation of the -(Xxx)$_n$- homo-peptides studied in this work: E$^{coop}$ (Eq 2) vs. *n*. Insets: Images displaying the optimized conformations of -(Phe)$_6$- and -(Phe)$_7$-. The disposition of the phenyl rings explains the change of the E$^{coop}$ tendency observed for -(Phe)$_n$-. (**b**) Relative energy of the optimized helical structure with respect to the optimized fully-extended conformation ($\Delta$E$_{fe/h}$) for the -(Xxx)$_n$- homo-peptides with *n*= 4, 6, and 8.

**FIGURE 16**    For -(Xxx)$_n$- homo-peptides: (left) axial projection of the optimized conformation and averaged backbone dihedral angles with their respective standard deviations;

(center) equatorial projection of the optimized conformation; and (right) details of the $C_5$, sc-bb, and/or sc-sc interactions with averaged geometric parameters. The H-bonds, and N–H$\cdots\pi$ and $\pi$-$\pi$ stacking interactions are represented by dashed red lines, green arrows and green double-arrows, respectively (distances in Å and angles in °).
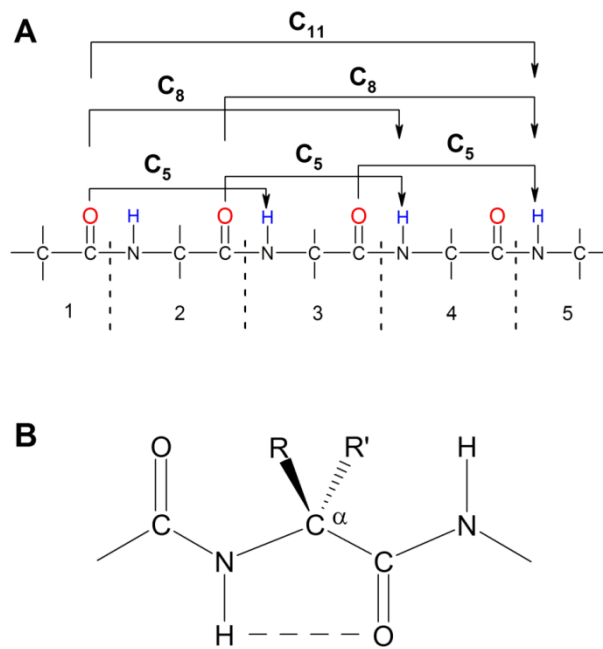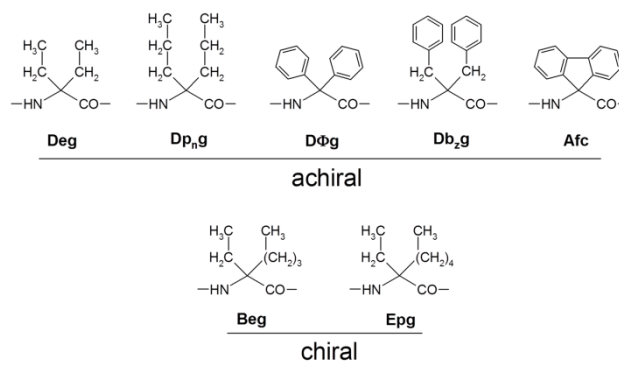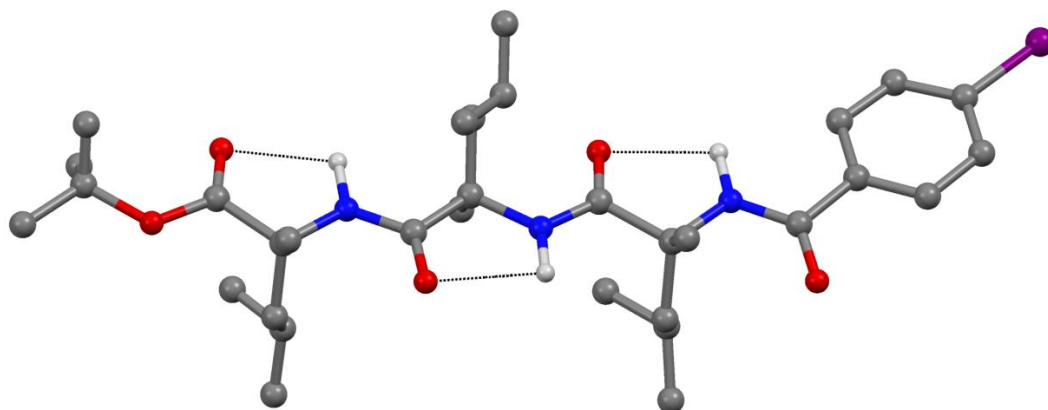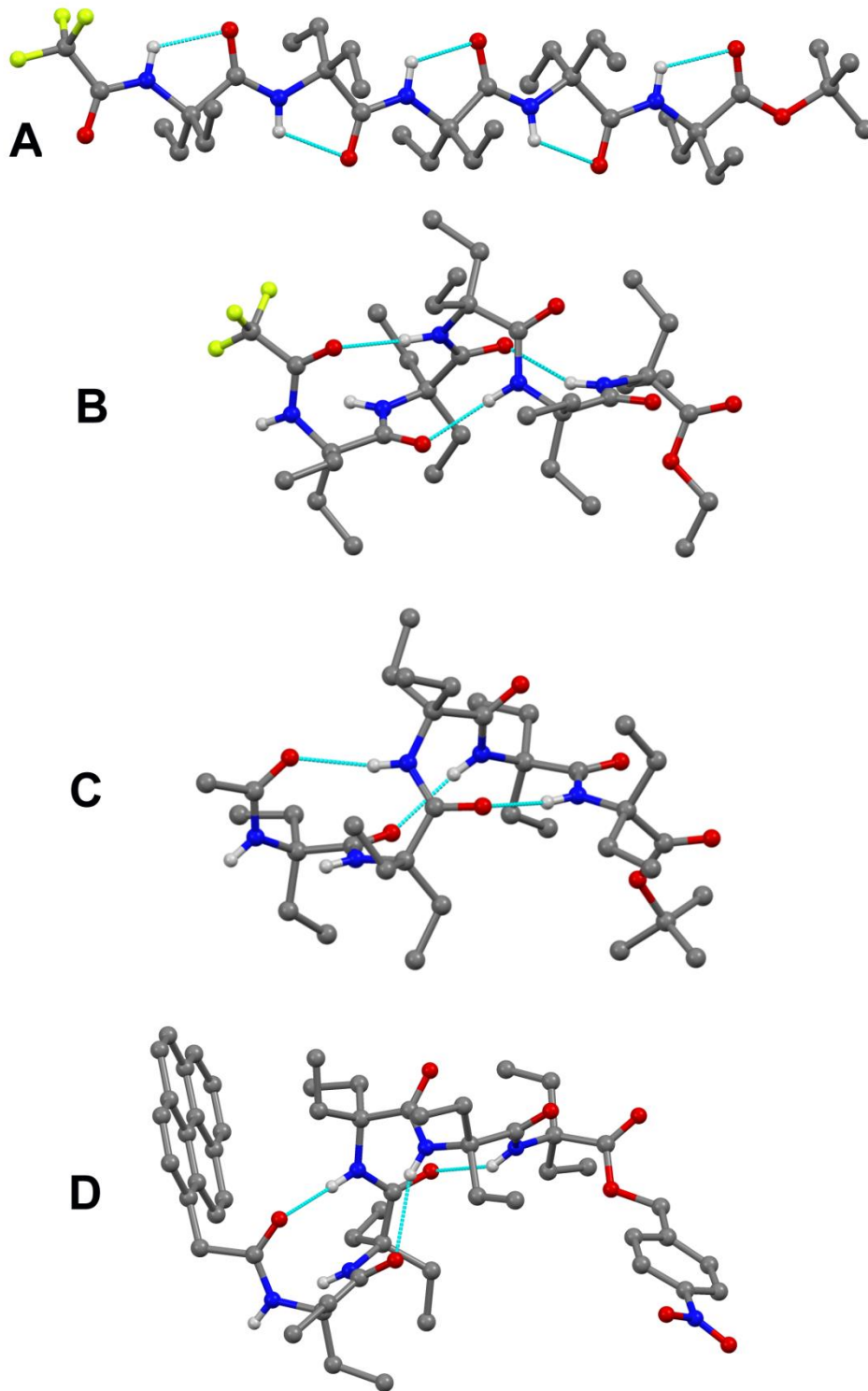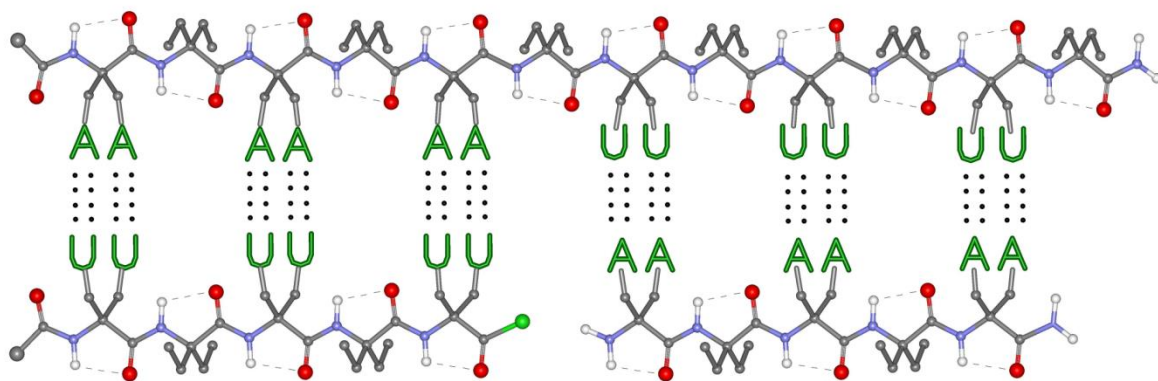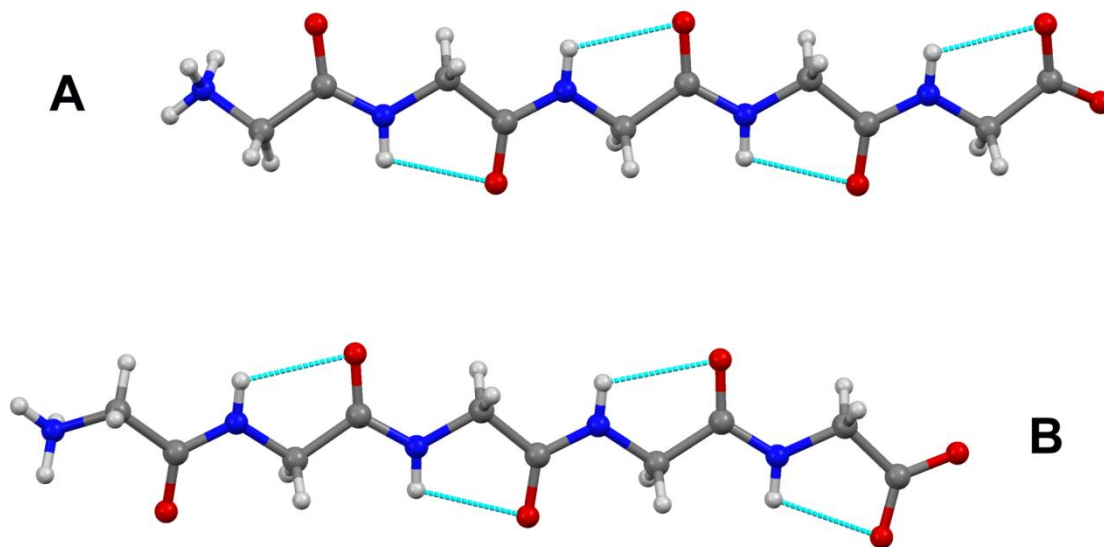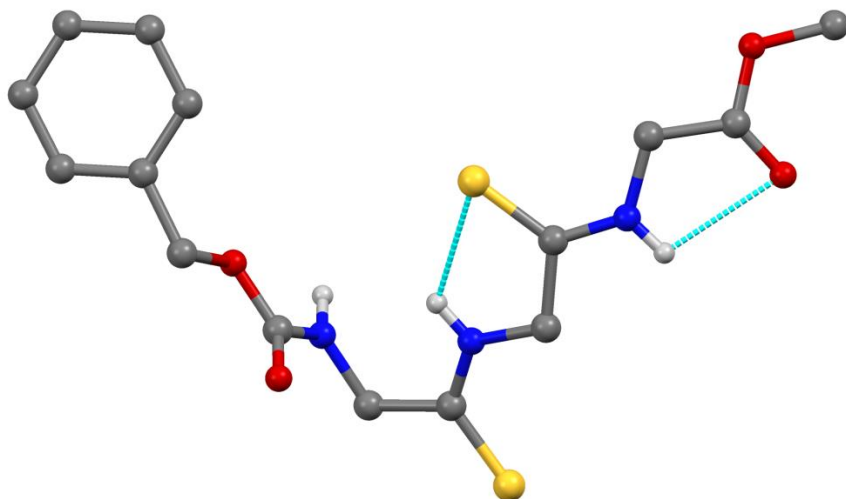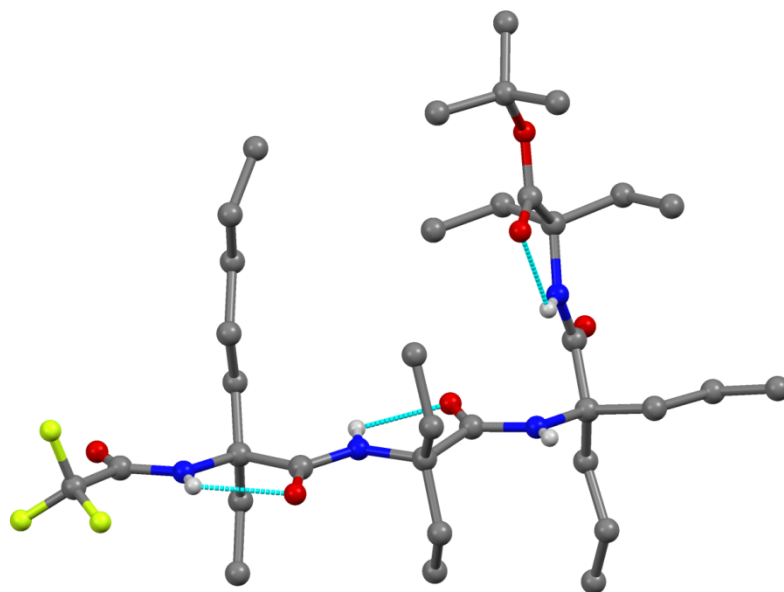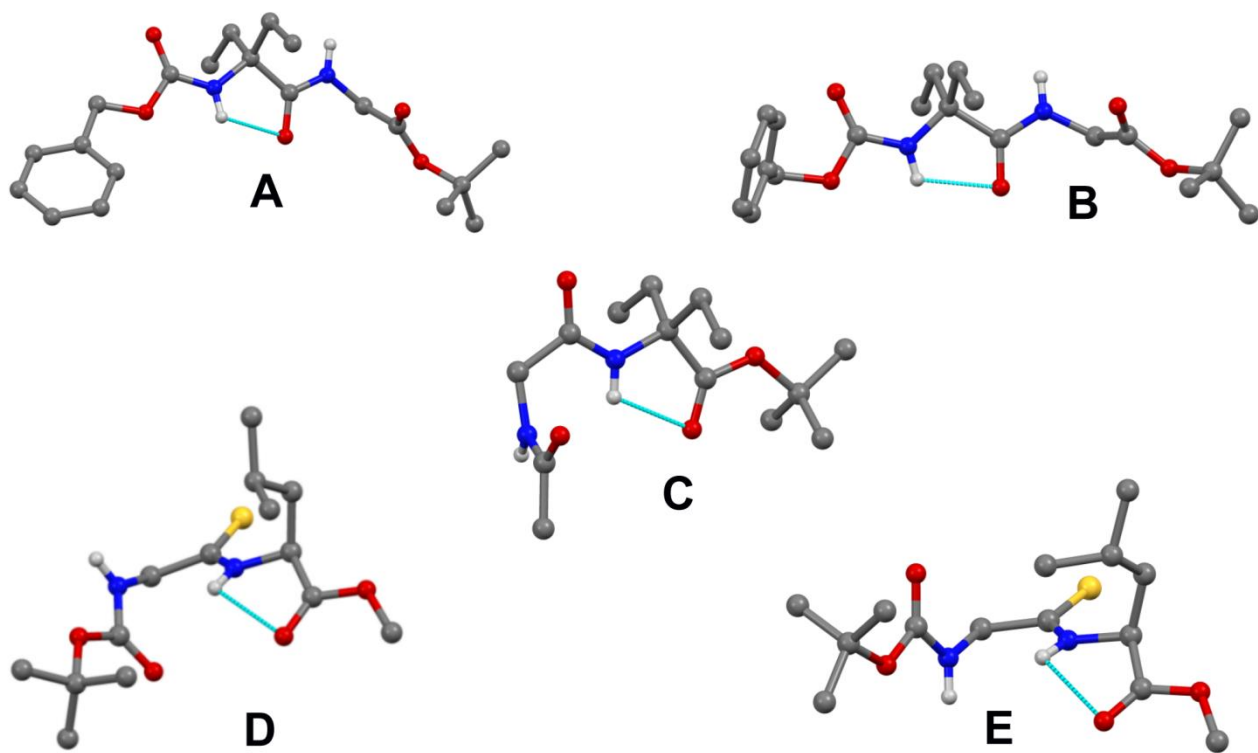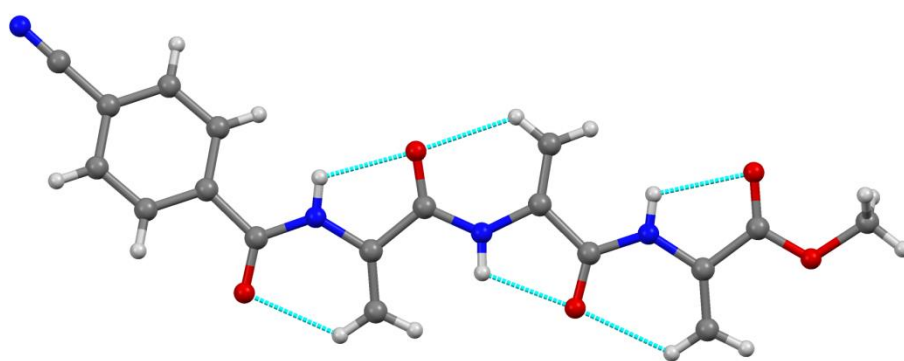
Figure 1



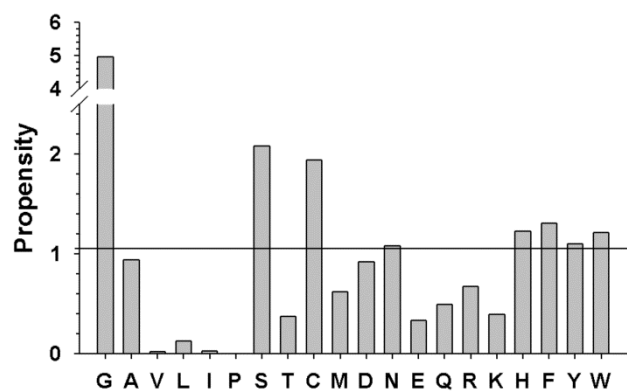Figure 2

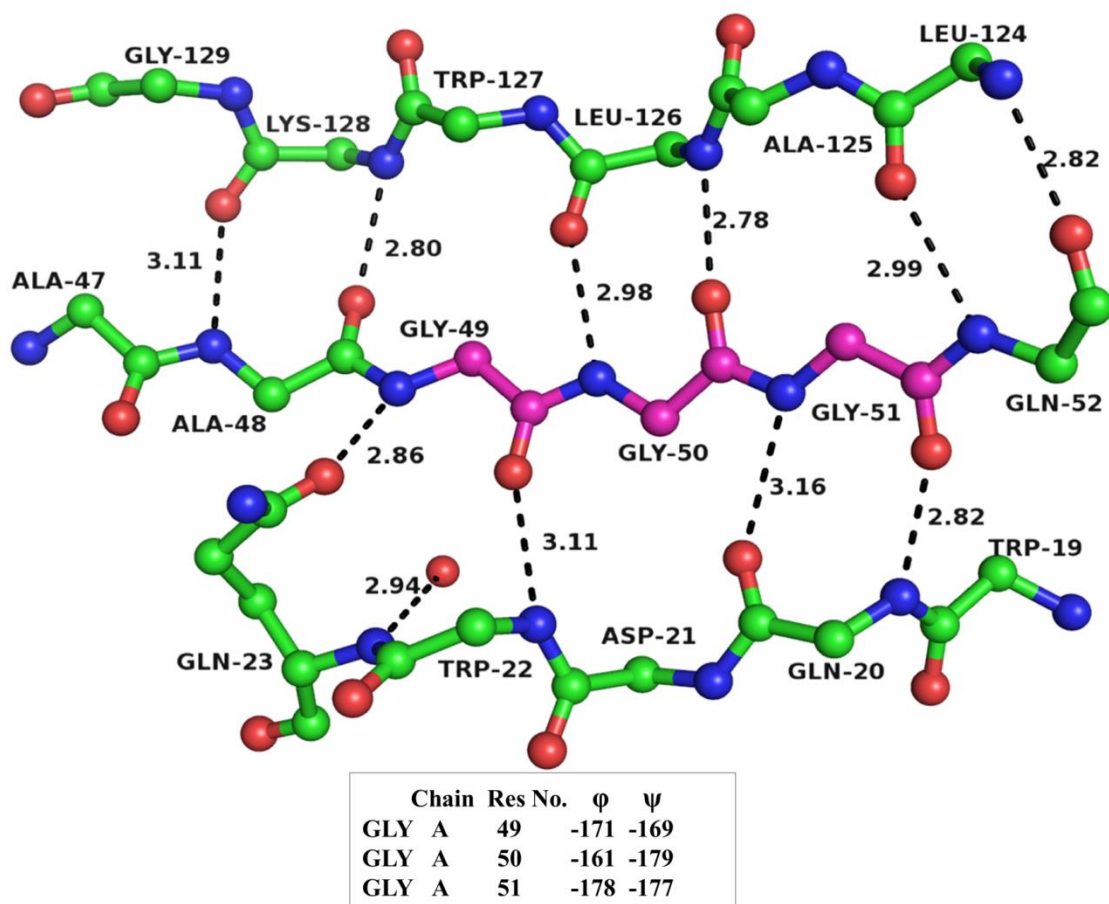Figure 3

Figure 4

Figure 5



Figure 6

Figure 7



Figure 8

Figure 9



Figure 10

Figure 11



| Chain | Res No. | φ | ψ |
|-------|---------|------|------|
| GLY A | 49 | -171 | -169 |
| GLY A | 50 | -161 | -179 |
| GLY A | 51 | -178 | -177 |

Figure 12

| | Chain | Res. No. | φ | ψ |
|---|---|---|---|---|
| SER | A | 84 | -164 | 166 |
| SER | A | 85 | -166 | 167 |
| SER | A | 86 | -163 | 166 |
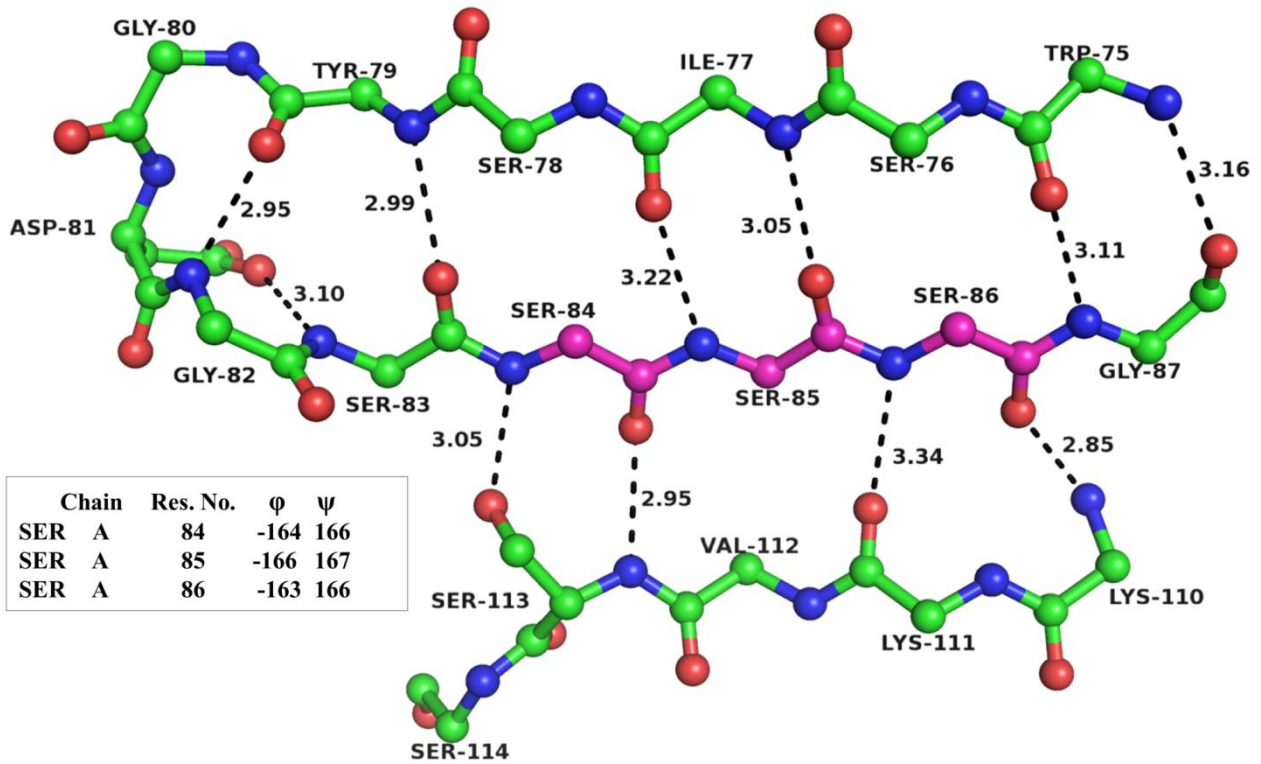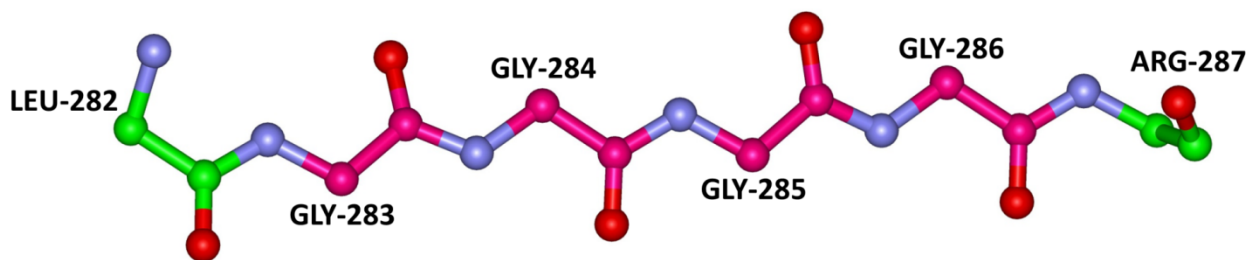
Figure 13



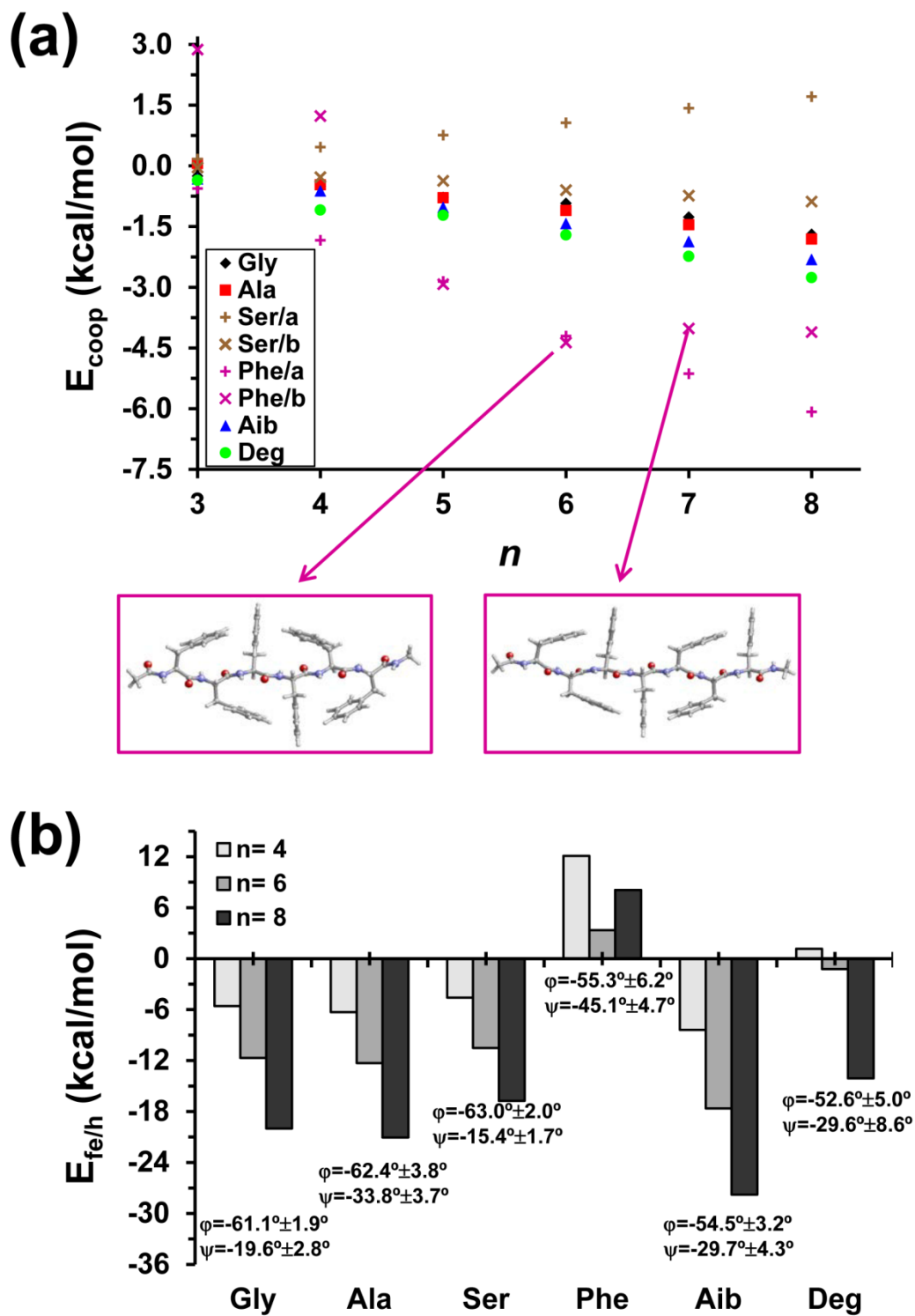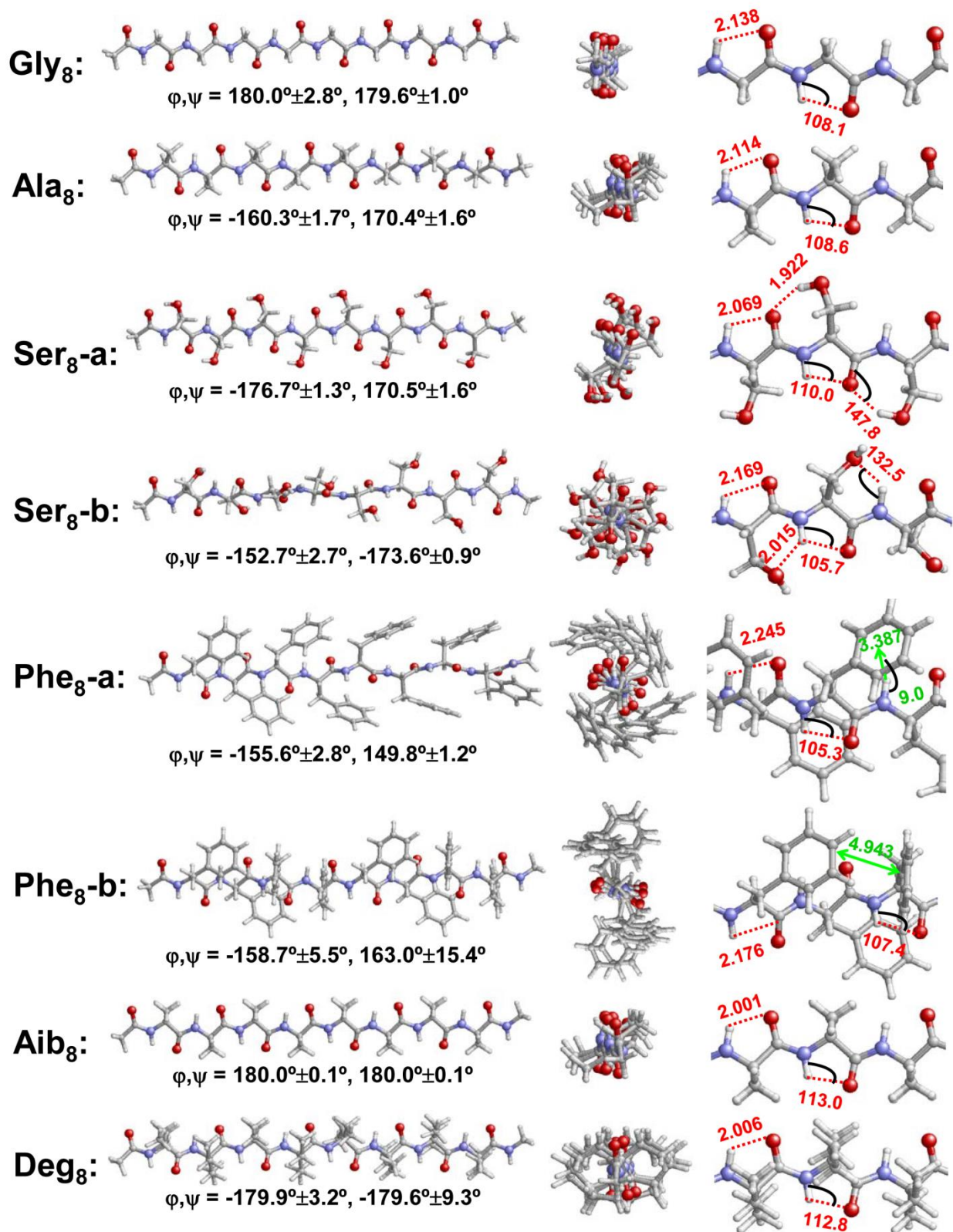| | Chain | Res. No. | φ | ψ |
|---|---|---|---|---|
| GLY | A | 283 | -177 | 175 |
| GLY | A | 284 | -170 | -173 |
| GLY | A | 285 | 173 | -166 |
| GLY | A | 286 | 168 | -175 |

Figure 14

Figure 15

**Gly$_8$:**

$\varphi, \psi = 180.0°\pm2.8°, 179.6°\pm1.0°$

**Ala$_8$:**

$\varphi, \psi = -160.3°\pm1.7°, 170.4°\pm1.6°$

**Ser$_8$-a:**

$\varphi, \psi = -176.7°\pm1.3°, 170.5°\pm1.6°$

**Ser$_8$-b:**

$\varphi, \psi = -152.7°\pm2.7°, -173.6°\pm0.9°$

**Phe$_8$-a:**

$\varphi, \psi = -155.6°\pm2.8°, 149.8°\pm1.2°$

**Phe$_8$-b:**

$\varphi, \psi = -158.7°\pm5.5°, 163.0°\pm15.4°$

**Aib$_8$:**

$\varphi, \psi = 180.0°\pm0.1°, 180.0°\pm0.1°$

**Deg$_8$:**

$\varphi, \psi = -179.9°\pm3.2°, -179.6°\pm9.3°$

Figure 16

**Graphical Abstract Image**