

Predictable Conformational Diversity in Foldamers of Sugar Amino Acids

Dóra K.Menyhárd^{1,2}, Ilona Hudaky¹, Imre Jákl^{1,2}, György Juhász³ and András Perczel^{1,2,*}

¹ *Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány P. stny. 1/A, 1117 Budapest, Hungary*

² *MTA-ELTE Protein Modeling Research Group, Pázmány P. sétány. 1/A, 1117 Budapest, Hungary*

³ *Department Of Chemistry, Faculty Of Education, J. Selye University, Bratislavská 3322, Komárno, Slovakia*

* *corresponding author: Perczel@chem.elte.hu*

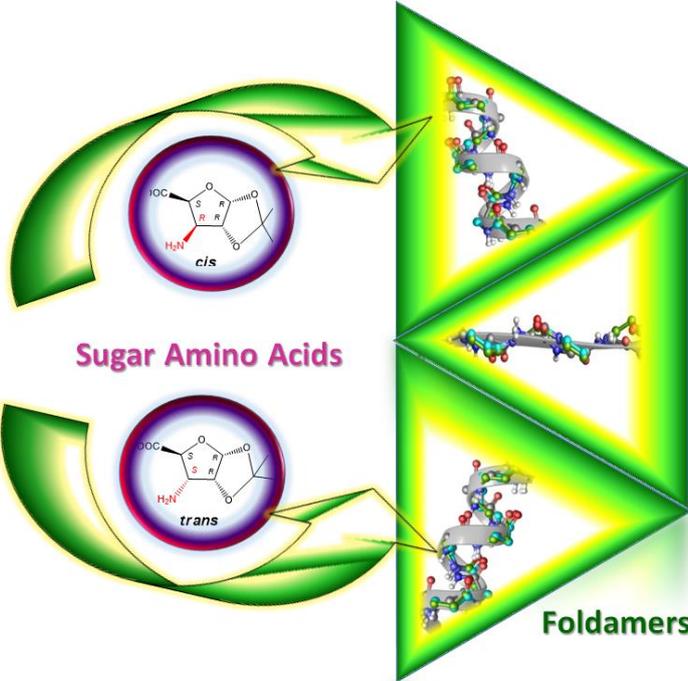
Abstract

Systematic conformational search was carried out for monomers and homo-hexamers of furanoid β -amino acids: *cis*-(*S,R*) and *trans*-(*S,S*) stereoisomers of aminocyclopentane carboxylic acid (ACPC), two different aminofuranuronic-acids (AFU ^{α} and AFU ^{β}), their isopropylidene derivatives (AFU(ip)) as well as the key intermediate β -aminotetrahydrofuran-carboxylic acid (ATFC). Stereochemistry of the building blocks was chosen to match with that of natural sugar amino acid (xylose and ribose) precursors (XylAFU and RibAFU). Results show that hexamers of *cis* furanoid β -amino acids show great variability: while hydrophobic cyclopentane (*cis*(ACPC)₆), and hydrophilic (XylAFU ^{α/β})₆ foldamers favor two different zigzagged conformation as hexamers, the backbone fold turns into a helix in case of (*cis*ATFC)₆ (10-helix) and (XylAFU(ip))₆ (14-helix). *Trans* stereochemistry resulted in hexamers exclusively of right-handed helix conformation, (H₁₂^P)₆, regardless of their polarity. We found that the preferred oligomeric structure of XylAFU ^{α/β} is conformationally compatible with β -pleated sheets, while that of the *trans*-(*S,S*) units match with α -helices of α -proteins.

Graphical abstract

formázott: Betűtípus: Dólt

formázott: Betűtípus: Dólt



Foldamers

Introduction

Backbone modified β -amino acids are prime candidates to increase the half-life of peptide drugs/hormones under physiological conditions. Their incorporation into peptides and proteins is possible without disrupting secondary structural element of proteins.¹⁻³ These α/β -chimera resist against proteolysis quite more effectively compared to the parent macromolecule.⁴ However, introducing additional main chain atom(s) can add to the complexity of the conformational landscape and to flexibility of these molecules, a *Janus*-faced characteristic, not necessarily beneficial. It has recently been shown that replacing α -amino acids with conformationally constrained cyclic β -residues instead of their open-chain variants is, in some cases at least, preferred for sustaining the bioactive conformation and the *in vivo* activity of the parent compound.⁵⁻⁸ ACPC (β -amino-cyclopentanecarboxylic acid) and ACHC (β -amino-cyclohexanecarboxylic acid) are benchmarks of cyclic β -amino acids, the monomeric building blocks of foldamers whose stereochemistry defines the 3D-fold of homo-oligomers built from them. Homochiral *cis*-ACPC (*c*ACPC) monomers form oligomers of extended/elongated backbone structure, while oligomers of *trans*-ACPC (*t*ACPC) adopt helical conformation. Using these monomers, it is possible to build a wide range of structural motifs.⁹⁻¹²

However, both monomers and homo-oligomers of ACPC are only soluble in organic solvents and not in water - a serious drawback of their potential physiological application. This can be amended by using their more polar derivatives, such as sugar amino acids, more specifically furanoid β -sugar amino acids or 3-amino-3-deoxy-furanuronic acids (AFU),¹³⁻¹⁴ of which both D-xylo and D-ribo moieties are especially interesting due to their broad bio-availability and bio-compatibility. This however prompts questions, namely, *i*) whether these monomers, with the appearance of the ring oxygen and the hydroxyl groups, would sustain the beneficial foldamer nature of ACPC, *ii*) would monomeric structure still determine those of the oligomers derived from it, and *iii*) whether these building blocks – by themselves or in form of longer stretches – would comply with the secondary structural elements of α -peptides and proteins.

The basic building block of the series (see **Figure 1**), ACPC, thus contains a completely apolar cyclopentane ring as a sidechain. Systems with either (*R,R*) or (*R,S*) configuration, has been studied thoroughly both by experimental and theoretical methods.^{1-3, 15-18} ACPCs of (*S,R*) and (*S,S*) configuration (corresponding to D-xylose and D-ribose) considered here are new building blocks, which however are expected to behave similarly to their mirror image variants, and form mirror image structures both in their monomeric and oligomeric forms. Substituting the $-C^{\epsilon}H_2-$ segment of the five-membered ring by an O-atom affords amino-tetrahydrofuran carboxylic acids (ATFC), a system which has been synthesized and characterized in the (*R,S*) configuration only.¹⁹ Attaching two hydroxyl groups to the remaining unsubstituted carbons of the cyclopentane ring leads to 3-amino-3-deoxy-D-xylofuranuronic acid (XylAFU) in the α/β anomeric forms (depending on the configuration of C1) and to 3-amino-3-deoxy- α/β -D-ribofuranuronic acid (RibAFU). We have recently published multi-gram synthetic routes to obtain these AFUs starting from D-glucose.²⁰ (The

1,2-isopropylidene derivatives of the AFUs (XylAFU(ip) and RibAFU(ip)) are also water soluble and key intermediates of the synthesis and have already been compiled into a –Gly-Gly-AFU-Gly-Gly- α/β -chimera peptide,²¹ studied in homo-oligomeric form by NMR spectroscopy¹⁴ and molecular dynamics simulation¹⁷ techniques.

In this paper we undertake the determination of secondary structural element forming propensities of five cyclic β -amino acids, each in two different configurations, thus in total a subset composed of 10 building blocks (**Figure 1**). By determining and comparing the 3D-structures of these homo-oligomeric foldamers with respect to their monomeric units, the extent of fold-conservation or conversion is to be established.

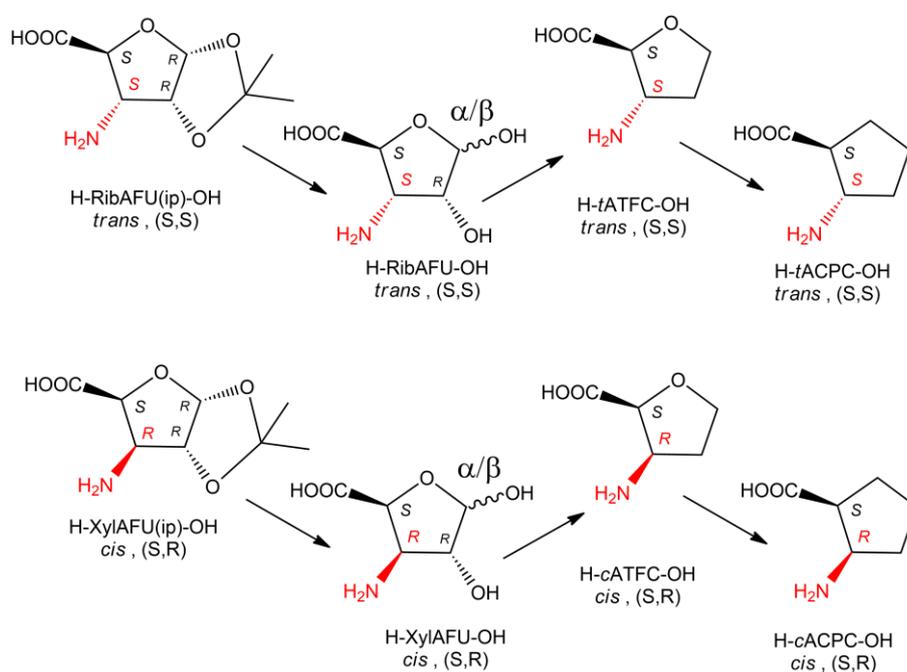


Figure 1 Cis and trans stereoisomers of β -amino acids. Both systematic names and abbreviations depend on the complexity of the molecular system. In this study both α and β anomers of RibAFU and XylAFU are considered, in total 10 model systems are studied.

Methods and Nomenclature

All backbone torsion angles defining the conformation of β -amino acids (**Figure 2**) can be classified as either *gauche+* ($0^\circ \leq g \leq +120^\circ$), *gauche-* ($-120^\circ \leq g \leq 0^\circ$) or *anti(a)*, $+120^\circ \leq a \leq -120^\circ$). Not including isomers of the amide planes, there are still 3 backbone torsions (ϕ , θ , ψ ; **Figure 2**) per β -amino acid residue and thus, in total 27 ($=3^3$) different conformer classes

were identified as described previously (see **Figure 3**).²² Among these backbone prototypes, four intrinsically different secondary structural elements were distinguished, namely zigzag (Z), helical (H), spiral (S) and elongated (E) forms. While E- and S-types have none, Z-types incorporate intra-residual, while H-types form inter-residual H-bonds in their oligomeric states. These structure codes (Z/H/S/E) were used to name each conformational state, along with subscript describing the resultant H-bond motif, superscripts indicating “enantiomeric” state (mirror imagery in the torsional space, for example: (g-, g+, g-) vs. (g+, g-, g+) conformers)

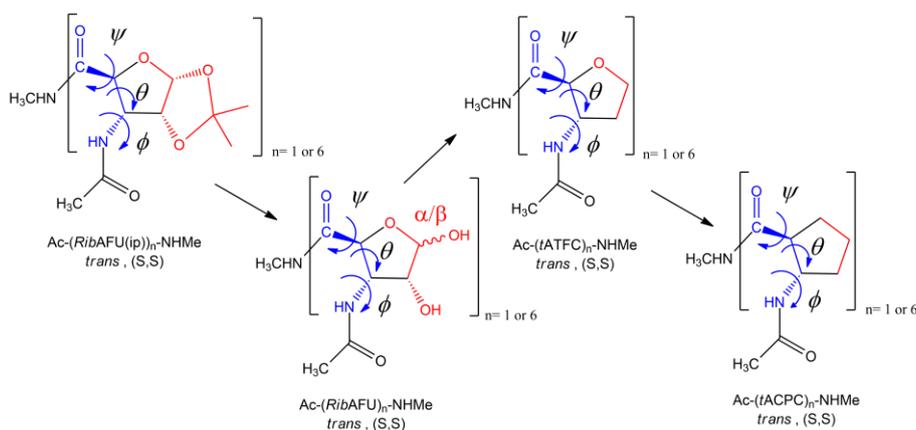


Figure 2 Mono- ($n=1$) and hexameric ($n=6$) polyamide models formed from the parent *trans* stereoisomeric β -amino acids. Both the backbone structure (blue) and related stereochemistry of all models are identical, differences (shown in red) however influence both backbone structural preferences as well as foldamer solubility.

and a star differentiating (in case of Z and S conformers) two topologically similar arrangements. Not counting the fully extended or (a,a,a) form, the remaining 26 conformers form mirror image pairs, distinguished by M and P superscripts (P: positive or clockwise and M: minus or anti-clockwise). Thus, H_{10}^M – a right-left-handed helical conformer has 10-membered H-bonds stabilizing its backbone (10-helix), which fold is the conformational mirror image of H_{10}^P , the left-right-handed version of the same foldamer. This 10 helical conformation was first described for oxetane β -amino acids hexamers.³² Similarly, Z_6^{*M} is the enantiomeric pair of Z_6^{*P} , both having a “zigzagged” backbone conformation with 6-membered intra-residual H-bonds. In naming these foldamers, the $+120^\circ$ and -120° boundaries between backbone prototypes were used more as guidelines, augmented by the apparent H-bond motifs and the actual structure of the resultant oligomer. In total 12 helical (H), 8 zigzag (Z), 4 spiral (S) and 2 extended (E) forms could be identified. As for any cyclic β -amino acids, favoring $\theta \sim 0 \pm 30^\circ$, the middle plain of the dihedral space - or 3D-Ramachandran cube – corresponding to $\theta \sim 180^\circ$, should be sparsely populated (see **Figure 3**).

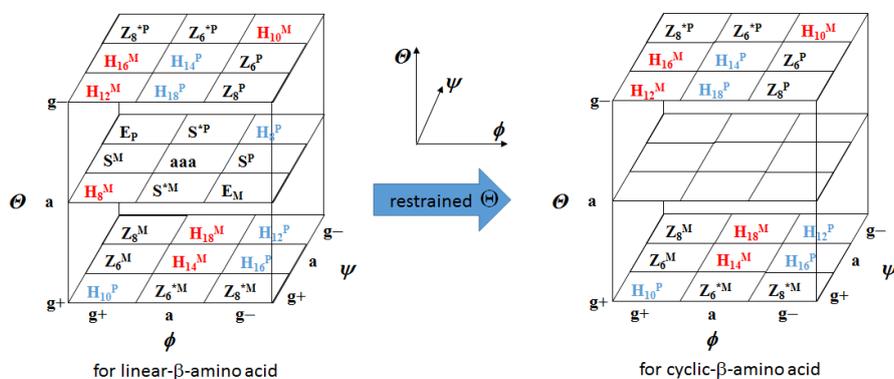


Figure 3 Location of backbone prototypes of β -peptides in the dihedral space defined by the ϕ , θ , ψ backbone angles (see Figure 2 for description). Enantiomeric conformers have the same code-name with different superscript: P (positive or clockwise) or M (minus or anti-clockwise). Helical foldamer building units printed in color: P-type blue and M-type red.

For each of the 10 monomeric units, furanoid β -amino acid diamides, all the 27 ideally possible backbone prototypes (including consideration of the puckering angles of the saturated five-membered rings) were geometry optimized by density functional method in aqueous medium. Calculations were carried out using the Gaussian 09 program package;²³ the PCM reaction field method for water at the B3LYP/6-31G(d) level of theory was applied. Computational results agreeing with the experimentally well described *cis* and *trans* ACPC systems confirmed the choice of the above level of theory. However, for the sake of reinforcement, selected calculations were conducted using a triple-zeta basis set (6-311++G(d,p)) too and for two special cases the M06-2X functional was also tested. Any conformer type established was considered as unique (or different from others) if at least one of the 3 backbone torsional angles (ϕ , θ or ψ) was different by more than 10° from those of another and thus retained and tabulated. These stable and unique monomers were used to build homo-hexamers which were studied at the same level of theory. *N*- and *C*- capped hexamers were only accepted for further consideration if, they remained conformationally homogeneous after full geometry optimization, meaning that the ϕ , θ , ψ backbone torsion angles along the main chain did not show an rms deviation of more than 10° from their respective average values.

All calculations were carried out in water solvated state, which might seem peculiar in those cases where the studied foldamer is insoluble in water (e.g. ACPC). However, in this work we set out to aid the design of incorporating these foldamer units into naturally occurring, or physiologically/medically relevant proteins and peptides which will result in the transfer of these foldamers to the solvated phase, whence their successful compliance with the surrounding α -protein matrix will depend on their willingness to participate in the given secondary structure - in water.

To describe the conformational similarity of the derived β -peptides to α -peptidic secondary structures, pseudo-Ramachandran backbone torsional angles were introduced: referred to hereafter as ϕ_d and ψ_d (where d stands for dummy). The dummy atom (C_{ad}) was introduced halfway between the C^α and C^β , which allows ϕ_d to be measured as the torsion defined by ($C_{i-1} - N_i - C_{ad} - C_i$) atoms and ψ_d along ($N_i - C_{ad} - C_i - N_{i+1}$) (**Figure 4**). The pseudo-Ramachandran dihedral angles thus determined were compared to the 2D-Ramachandran surface created by plotting the backbone dihedrals of high resolution structures of α -proteins and peptides and thus classified based on falling into the preferred (α -helix, β -stand, left-handed helix, *etc.*) and rarely sampled (disallowed) regions of the potential surface²⁴ as well as to the theoretically derived conformer classes of α -proteins.²⁵

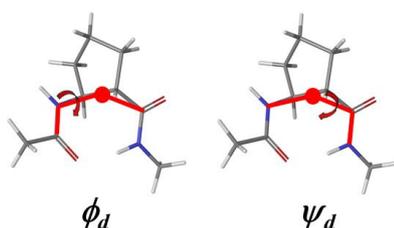


Figure 4 Definition of ϕ_d and ψ_d dihedrals, with the dummy atom introduced halfway between C_α and C_β shown as a red dot.

Results

Systematic search followed by geometry optimization of the initial conformers lead to 9-13 unique low energy minima in case of monomers of furanoid β -amino-acids, affording 5-12 different homohexamers for each studied system. First the simplest and experimentally well described ACPC elements were investigated, then the conformer selection and preservation of *cis/trans*ATFC, continued by the appropriate RibAFU and XylAFU derivatives. Detailed results of conformer preferences of the different β -sugar amino acids both as monomers or hexamers can be found in the Tables of the Supporting Information, while a compact overview is reported in **Table 1**, where letter size corresponds to the relative weight of a backbone conformer prototype.

Table 1. Conformational preferences of the different β -sugar amino acids as a monomer or hexamer (letter size corresponds to the relative weight of the conformers, see **Table S1-27** for the data). The results shown were obtained using the 6-31G(d) basis set. For the *trans*AFU $^\alpha$ hexamer, due to the small energy gap between the low-lying states, calculations were repeated using the larger basis set 6-311++G(d,p) and those results are indicated.

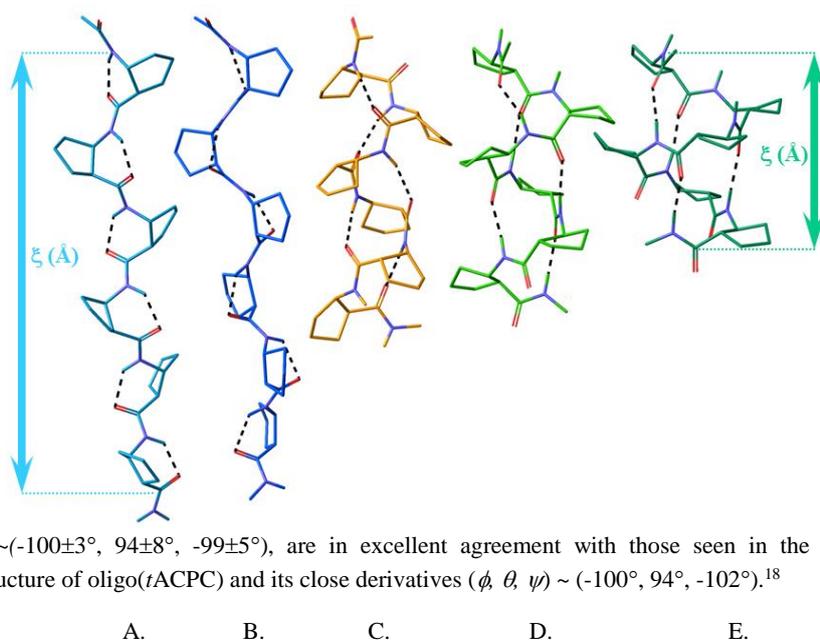
configuration/ constitution		ACPC	ATFC	AFU $^\alpha$	AFU $^\beta$	AFU $^\alpha$ (ip)
<i>cis</i> , (S,R), D-Xylo	monomer	Z ₆ ^{*P} Z ₆ ^M Z ₆ ^M	Z ₆ ^{*M}	Z ₆ ^{*M}	Z ₆ ^{*M}	H ₁₄ ^P Z ₆ ^{*P}
	hexamer ^b	Z ₆ ^{*P}	H ₁₀ ^P	Z ₆ ^{*M}	Z ₆ ^{*M}	H ₁₄ ^P
<i>trans</i> , (S,S), D-Ribo	monomer	H ₈ ^P Z ₈ ^M	H ₈ ^P Z ₈ ^M Z ₆ ^M	Z ₆ ^{*M} H ₁₀ ^P	H ₈ ^P Z ₆ ^M	Z ₈ ^{*M} H ₈ ^P
	hexamer ^a	H ₁₂ ^P	H ₁₂ ^P	H ₁₂ ^P Z ₆ ^M	H ₁₂ ^P	H ₁₂ ^P

Configuration induced conformers of (*c/t*ACPC)_n: the bimodal nature of the homochiral states

Seven different conformers populate the conformational landscape of monomeric Ac-*c*ACPC-NHMe (see **Table S1** of the Supporting Information and **Figure 5**). The Z₆^{*P} backbone conformer ($\phi, \theta, \psi \sim (171^\circ, -44^\circ, -116^\circ)$) is the most preferred one (37.7% of the equilibrium population), with additional extended forms of Z₆^M, Z₈^M, and Z₈^{*P} of notable weight. ~~Stability order of the different backbone conformers correlates with the~~ The number of intra-residual H-bonds ~~conformer has a significant effect on the energetic ordering of the monomers: not surprisingly, the conformers of the low energy regions tend to have the most number of H-bonds within the ensemble (see Tables of the Supporting Information).~~ The conformational heterogeneity, once forming a homohexamer, Ac-(*c*ACPC)₆-NHMe (in short: (*c*ACPC)₆), is significantly reduced: it adopts almost exclusively (91.3% of the equilibrium population at room temperature) the (Z₆^{*P})₆ conformer type (**Table S3** and **Figure 5**). This elongated backbone structure with average backbone angles (ϕ, θ, ψ) of (172°, -44°, -117°) and of about 23Å length (measured between the *N*-terminal amide N-atom and the *C*-terminal carbonyl C) is stabilized by six, 6-membered intra-residual backbone/backbone H-bonds – in perfect agreement with its solution phase structure determined by NMR.³ Thus, the same type of extended conformer and H-bond motif determines both the monomeric and the hexameric fold of *c*ACPC. For the hexamer there are two additional conformers of notable weight: amounting to 6.9 and 1.8% of the equilibrium population. The former one is also an extended conformation stabilized by similar six-membered H-bond loops as seen in the major

conformer, while the latter is a helical conformation (H_{10}^P), featuring 10-membered, $NH_i \rightarrow C=O_{i+1}$ type backbone H-bonds. Unlike for the *t*ACPC, the formation of H12 helix in case of the *cis* stereoisomer is quite disfavored both in the monomeric and hexameric state (at an excess of 23.1 kJ/mol and 99.2 kJ/mol as compared to the appropriate global minimum). Such a backbone arrangement places the carbonyl group in the unfavorable, nearly eclipsed, position with respect to the five-membered ring: in the monomer ($O-C-C^\alpha-C^\epsilon$) $\sim -20.5^\circ$, while in the hexamer even closer to zero: $-10.5 \pm 5.7^\circ$, which results in an $O-C^\epsilon$ distance is $\sim 2.8 \text{ \AA}$ in both forms – thus in strained geometries.

The *trans* or (*S,S*)-stereoisomer, on the other hand, favors negative values of ϕ torsional angles ($\sim -75^\circ$), with H_8^P and Z_8^{*M} conformers sampled the most by the monomer, with the H_{12}^P element also represented (**Table S6 and Figure 5**). For the hexamer, $Ac-(tACPC)_6-NHMe$ ($(tACPC)_6$, from here on), the $(H_{12}^P)_6$ helical fold becomes overwhelmingly dominant, stabilized by five, 12-membered, $C=O_i \leftarrow NH_{i+3}$ type backbone H-bonds (**Table S8 and Figure 5**). This H-bond motif and the calculated average backbone dihedrals of (ϕ , θ ,



ψ) $\sim (-100 \pm 3^\circ, 94 \pm 8^\circ, -99 \pm 5^\circ)$, are in excellent agreement with those seen in the crystal structure of oligo(*t*ACPC) and its close derivatives (ϕ , θ , ψ) $\sim (-100^\circ, 94^\circ, -102^\circ)$.¹⁸

Figure 5 The calculated structure of Z_6^{*P} (A.), Z_6^M (B.) and H_{10}^P (C.) conformations of $(cACPC)_6$ and the H_{12}^P (D.) and H_{16}^P conformers (E.) of $(tACPC)_6$ homo-hexamers. Color scheme is the same as that used in the Tables of the Supporting Information. The arrows on the side show how chain-length was measured between the N-terminal amide N atom and the C-terminal carbonyl C (referred to as ξ in the text).

Interestingly, the second best conformer, populated ~1%, ($\Delta E=17.6$ kJ/mol), is also a helical foldamer, $(H_{16}^P)_6$, but significantly wider and shorter than the $(H_{12}^P)_6$ variant (with chain-end separations (ζ) of $\zeta(H_{12}^P)_6$ is ~13.2 Å and $\zeta(H_{16}^P)_6$ is ~10.2 Å) (see **Figure 5**). Here the H-bond motif is shifted by a residue, resulting in four $C=O_i \leftarrow NH_{i+4}$ type H-bonds, linked by 16-membered loops – so, formally this makes a 16-helix. Rewinding of the helix ($(H_{12}^P)_n \rightarrow (H_{16}^P)_n$) requests the re-pairing of the H-bonds but it does not require a large backbone conformational shift (~20° shift in terms of torsional angles). Accordingly, the average pseudo-Ramachandran dihedrals of these two conformers fall in the same region of the Ramachandran surface (see **Figure 6**), corresponding to the α -helical structures of α -proteins. Interestingly enough, based on NMR NOE-data, the turn up of a 16-helix conformer in solutions of substituted *t*ACPC oligomers was already postulated.¹⁸ In other words, the herein calculated helical variant, the $(H_{16}^P)_n$ foldamer, can indeed be present in solution if stabilized in a suitable manner.

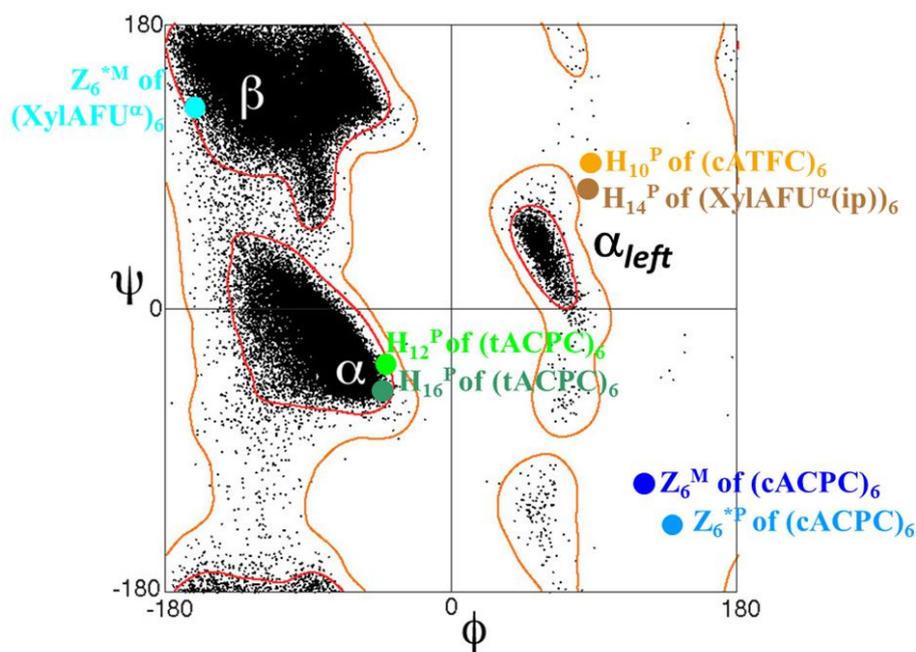


Figure 6 Illustrating location of calculated pseudo-Ramachandran dihedrals of selected β -foldamer conformers on the Ramachandran plot based on high-resolution structures of α -peptides and proteins, preferred regions are contoured in red.²⁶ Color scheme of the major foldamers are in line with that used in the Tables of the Supporting Information and the Figures below.

The 3rd and 4th most stable conformers, helical H_8^P and the zigzagged Z_8^{*M} , both have intra-residual 8-membered H-bonds (**Table S8**). The $(H_8^P)_6$ helix, built from the most stable

monomer, might be derived from the $(H_{12}^P)_6$, by shifting the hydrogen bond motif in the “opposite” direction, resulting in a tighter and longer helix with $C=O_i \leftarrow NH_{i+2}$ type H-bonds. Though a $(H_8^P)_6$ helix contains six of such backbone H-bonds (one extra over the global minimum energy structure), it nevertheless is 33.7 kJ/mol less stable than $(H_{12}^P)_6$. The lower stability stems from the fact, that the $C=O_i \leftarrow NH_{i+2}$ type H-bonds are more strained than $C=O_i \leftarrow NH_{i+3}$ type H-bonds present in $(H_{12}^P)_6$. Although in both helices the average NH--O distances and N—H--O angles are quite similar ($\sim 2.9\text{\AA}$ and $171.5\pm 1^\circ$), the acceptor angles (C=O--H) are different and far from the optimum value in case of $(H_8^P)_6$ (123.1° in $(H_8^P)_6$ and 148.5° in $(H_{12}^P)_6$, respectively).

We thus found that configuration of ACPC strongly determines the overall conformational properties of both its monomeric and homo-oligomeric foldamer, e.g. $(ACPC)_6$. The *cis* stereoisomer exclusively adopts an **elongated** or Z_6 -type **foldamer**, disfavoring any helical structure. On the contrary, the *tACPC* building block is intrinsically helical even in its monomeric state – a nice example of the configuration induced conformer selection – which property is inherited in its homo-oligomeric state: $(tACPC)_6$ forms a stable and robust $(H_{12}^P)_6$ helix. These findings are in agreement with those calculated by others at different levels of theory and determined in solution by various spectroscopic methods.¹⁻³

Inspecting pseudo-Ramachandran angles (ϕ_i , ψ_i) of the $(cACPC)_6$ and $(tACPC)_6$ conformers, (see **Tables S3, S8** and **Figure 6**), it is obvious that while both *cis* and *trans* stereoisomers form well-defined backbone conformers, only that of $(H_{12}^P)_6$ helix is compatible with its counterpart in proteins (an α -helix formed by α -L-amino acid residues) (see **Figure 6**). Moreover, since the directionality of the two helices is the same {right handed helices with $C=O_i \leftarrow NH_{i+3}$ and $C=O_i \leftarrow NH_{i+4}$ type H-bonds in $(H_{12}^P)_6$ and $(\alpha_L)_n$, respectively} the (S,S) -foldameric unit $(H_{12}^P)_n$ homo-oligomers might be inserted into an α -helix of a protein, without seriously distorting its backbone structure. On the other hand, both Z_6^{*P} and Z_6^M extended backbone conformers of $(cACPC)_6$ have characteristically different backbones from those of β -pleated sheets of proteins. In fact, the ϕ_i , ψ_i dihedral values of both $Z_6^{*P} \sim (+135^\circ, -134^\circ)$ and $Z_6^M \sim (+126^\circ, -110^\circ)$ fall on a disallowed region of the Ramachandran map (indicating that these arrangements are not sampled significantly by proteins, **Figure 6**). This means that the insertion of $(cACPC)_6$ into a β -strand within a protein would be difficult, without seriously distorting it, and disrupting the “original” inter-strand H-bond pattern. Nevertheless, as both $(Z_6^{*P})_6$ and $(Z_6^M)_6$ are autonomous and self-determining secondary structural elements, they could be applied as stand-alone new structural motifs for foldamer and protein design. Interestingly, while *cACPC* was indeed found to be rather a β -strand breaker, monomers of *cis*-ACHC ((1*R*,2*S*)-2-aminocyclohexanecarboxylic acid) were successfully incorporated into betabellin-14 β -sheet core structure.²⁸

Results pertaining to ACPC were also used to find the optimum QM level of theory applicable for similar systems, since the conformational preferences of *cis* and *trans* ACPC monomers and oligomers have already been unequivocally established by spectroscopic and X-ray crystallographic methods. As mentioned above, B3LYP/6-31G(d) derived conformers are in good agreement with the experimental data. Accuracy of this DFT/dzp approach was

challenged either by using a larger basis set (6-31G(d) → 6-311++G(d,p)) or a different density function (e.g. M06-2X) (compare **Tables S1 - S5**). In case of the *c*ACPC monomer, the basic stability trend determined using the 6-31G(d) basis set, that of $Z_6 > Z_8 > H$, was reproduced using 6-311++G(d,p) basis set too. Z_6^{*P} , Z_6^M , Z_8^M and Z_8^{*P} conformers appear in the low-energy region using either approach (also in agreement with our previous results,²² however, while Z_6^{*P} and Z_6^M remain the two major conformers present, their energetic ordering is reversed. Z_6^{*P} is the global minimum of the monomer state when calculated using the 6-31G(d) basis set, followed by Z_6^M at +0.7 kJ/mol, while using 6-311++G(d,p) affords Z_6^M as the lowest energy conformation, more favorable by 1.1 kJ/mol than Z_6^{*P} . This difference propagates to the hexamer state in case of both, resulting in a ~ 6 kJ/mol energy gap between the first two states using either basis set, but in reversed order. However, *i*) the energy difference falls in the range of the expected accuracy and *ii*) we have to note, that $(Z_6^{*P})_6$ and $(Z_6^M)_6$ are quite similar conformers as they are close “neighbors” on the Ramachandran surface (**Figure 6**), as discussed above. In other words, although alteration of the basis set and level of theory alters gently the overall hypersurface, it results in rather similar backbone folds. The M06-2x functional with the larger basis set (6-311++G(d,p)) was also applied for $(Z_6^{*P})_6$ and $(Z_6^M)_6$ hexamers, which lead to results more in agreement with those of B3LYP/6-31G(d), showing a 13.0 kJ/mol energy difference in favor of the Z_6^{*P} state (see **Table S5**). H_8^P , Z_8^{*M} and H_{12}^P are the preferred conformers of the monomeric *t*ACPC, an order established and not reshuffled by using either of the two basis sets. For $(tACPC)_6$ exclusively the $(H_{12}^P)_6$ conformer is determined as the global minimum by all methods tested here. (See **Tables S6-S9**)

These calculations showed that using either the smaller or the large basis set, basic conclusions hold: *i*) a Z_6 -type hexameric structure is selected for the *cis* configuration, *ii*) while a H_{12} -helix for the *trans* one. In line with these data, in this paper we will compare structures and relative stabilities established at B3LYP/6-31G(d) level of theory (keeping in mind their semi-quantitative nature), while in cases where the energy separation of key states is low, calculations will be repeated at a higher level of theory too.

The exclusive helices of $(ATFC)_n$: left-handed-like for the *cis*, while right-handed for the *trans* stereoisomers

To understand the origin and inheritance of folding properties of amino furanuronic-acids (AFU^α and AFU^β) first those of tetrahydrofuran- β -amino-acids, ATFC, were considered (**Figure 1 and 2**). All $-(c/tATFC)_n-$ type elements differ from those of $-(c/tACPC)_n-$ by the oxygen atom of the tetrahydrofuran ring. To elaborate the influence of the ring O-atom and the $-CONH-$ group on each other, both the *cis* and *trans* stereoisomers as well as their hexamers, $Ac-(c/tACPC)_6-NHMe$, were studied.

In the case of the *cis* monomeric form, $Ac-(cATFC)-NHMe$, the Z_6^{*M} backbone conformer is by far (93% of the equilibrium population) the most stable structure (**Table S10**). This elongated zigzag form of *c*ATFC is the mirror image structure of Z_6^{*P} (**Figure 3**), found significant for *c*ACPC: this difference in behavior, however, is easily explained by the

presence of the furanoid ring oxygen. In the Z_6^{*M} conformer of *c*ACPC the C^ϵ carbon atom would be too close (2.5 Å) to the N-atom of the C-terminal amide group, while in Z_6^{*P} the same distance is longer (~3.6 Å), making this molecular arrangement preferred. However, in *c*ATFC where an O-atom stands for the $-C^\epsilon H_2-$ group of *c*ACPC, the Z_6^{*M} fold becomes favorable as the spatial vicinity of the NH and the ring O-atom results in a stable inter-residual H-bond. (The N-atom in question corresponds to the backbone amide N-atom of residue (*i*+1) present in any oligo- and polypeptides composed of these β -amino acids.)

In the hexameric state, Ac-(*c*ATFC)₆-NHMe (in short: (*c*ATFC)₆) adopts exclusively (100%) the (H_{10}^P)₆ conformer (see **Table S11 and Figure 9**), which fold type is indeed poorly sampled (2.7%) by the monomer. This helical backbone fold of the hexamer contains 5 main-chain H-bonds (forming 10-membered pseudo-rings), plus 6 additional intra-residue H-bonds ($NH_i \rightarrow (O_{\text{furanoid}})_i$). It is interesting to note, that in the (Z_6^{*M})₆ fold even more, 6 backbone H-bonds and 6 $NH_i \rightarrow (O_{\text{furanoid}})_i$ H-bonds are present. What makes the former, helical fold still more stable in spite of this, is that the H-bonds of the helix are better oriented than those of (Z_6^{*M})₆. In fact, donor-acceptor H-bond angles in (H_{10}^P)₆ are quite a bit more favorable ($157.2 \pm 6.6^\circ$ and $147.6 \pm 5.6^\circ$) than those of (Z_6^{*M})₆ ($132.8 \pm 1.3^\circ$ and $98.3 \pm 0.5^\circ$).

Note, that in $(H_{10}^P)_6$ the directionality of the H-bonds, $NH_i \rightarrow C=O_{i+1}$, is the opposite of that seen in α -helices of proteins. This “reversal” is also depicted by the ϕ_i, ψ_i dihedral angles which place this conformer near the left-handed helix region, α_D , of the Ramachandran-map (see [Figure 6](#)), in spite of the fact that its wind is actually right-handed, just as those of a regular α -helix or the H_{12}^P conformer of *t*ACPC oligomers (see [Figure 87](#)).

formázott: Betűtípus: Félkövér

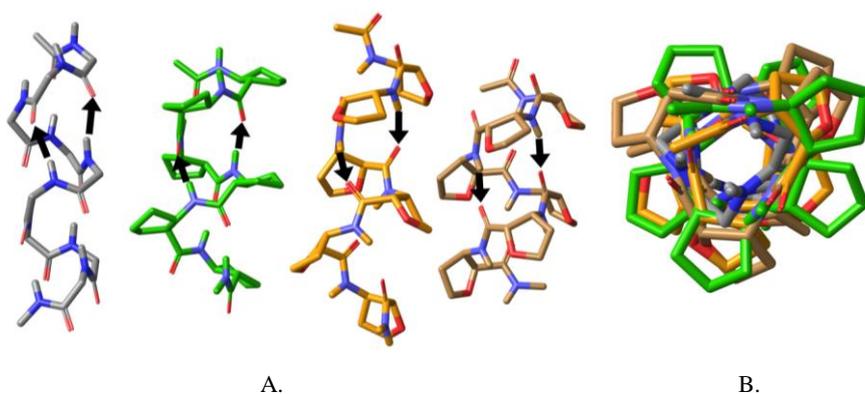


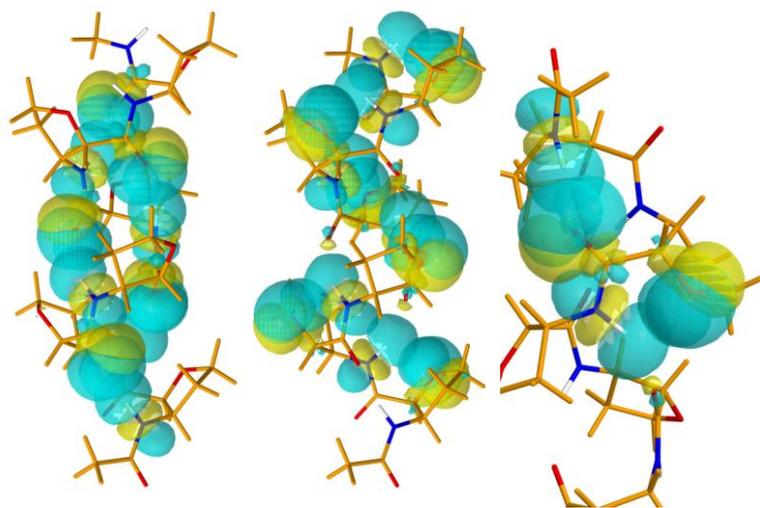
Figure 8-7 A) H-bond directionality of an α -helix backbone (only the backbone shown, grey), H_{12}^P helix of $(tACPC)_6$ (green), the H_{10}^P helix of $(cATFC)_6$ (yellow) and the H_{14}^P helix conformer of the same (brown). B) Overlaid structures viewed from the top, showing the “internal” holes and similar wind of the four helices.

Beside the 10-helix of $(H_{10}^P)_6$, another helical fold, that of $(H_{14}^P)_6$ was also found to be a relatively low energy conformation of $(cATFC)_6$ ($\Delta E=36.6$ kJ/mol), in fact the second in the stability order. This wider and shorter helix is quite similar to $(H_{10}^P)_6$, with its H-bond motif simply shifted by a residue (4 atoms), now encompassing $NH_i \rightarrow C=O_{i+2}$ H-bonds. Directionality and thus, the macrodipole of both helices are similar, as well as their ϕ_i, ψ_i angles ($\sim 96^\circ, 100^\circ$ and $88^\circ, 86^\circ$ for $(H_{10}^P)_6$ and $(H_{14}^P)_6$, respectively). The H_{12}^P fold, which mimics the α -helical conformation, is quite disfavored in case of *c*ATFC (by 190 kJ/mol as compared to $(H_{10}^P)_6$) since it places the furanoid ring oxygens in the vicinity of the carbonyl oxygen (at 2.7Å) creating a strongly destabilizing “repulsion”.

In summary, in the monomeric state of *c*ATFC, the zigzagged conformer Z_6^{*M} (stabilized by **two** H-bonds) is selected over all the others comprising a **single** H-bond only. In the hexameric form of *c*ATFC however, unlike for *c*ACPC, there is the possibility of forming better oriented inter-residue H-bonds in helices than in any extended or zigzag foldamer type. The slender $(H_{10}^P)_6$ helix dominates the pool ([Table S11](#)) but the second best

structure the $(H_{14}^P)_6$ is also a helix of similar widening. Experimental results on the tetra- and octamers of (R,S) -ATFC – the mirror image structure of the presently studied (S,R) -ATFC – shown that they adopt a H_{14} helical conformation.¹⁹ However the $(NH)_i \rightarrow (C_\beta H)_{i+2}$ type NOEs used to characterize the H_{14} -helix are fully consistent with the narrower $(H_{10}^P)_n$ molecular packing also. (Note that the average distances of these two H atoms in the $(H_{10}^P)_6$ and $(H_{14}^P)_6$ folds are too similar in magnitude ($3.9 \pm 0.1 \text{ \AA}$ and $3.7 \pm 0.1 \text{ \AA}$, respectively) to be distinguished by 1H - 1H NOEs. To further support our preference of H_{10} - over H_{14} -helix additional calculations were carried out. Minima of hexamers were re-optimized on the larger basis set 6-311++G(d,p) confirming $(H_{10}^P)_6$ as the favored state (see **Table S12**).

The intramolecular hydrogen bond network was studied on conformers of Ac -($cATFC$)₆-NHMe with Natural Bond Orbital (NBO) analysis method.²⁷ The selected NBOs on the skeleton of $(H_{10}^P)_6$ and $(Z_6^{*M})_6$ clearly show (see **Figure 78**) two type of hydrogen bond network between the amide-amide and the amide-THF-ring. The amide-amide (left columns), the amide-“THF-ring” H-bonds (middle columns), and those combined per residue (right columns) are depicted. H-bond energies as estimated by NBO are as follows: amide-amide/ H_{10}^P : 8.4 ± 2.6 kcal/mol, amide-THF-ring/ H_{10}^P : 2.8 ± 0.6 kcal/mol as well as amide-amide/ Z_6^{*M} : 7.4 ± 0.2 kcal/mol, amide-THF-ring/ Z_6^{*M} : 3.5 ± 0.2 kcal/mol, respectively. Note that although the former one, more stable has fewer H-bonds compared to $(Z_6^{*M})_6$ structure, in the latter one the donor-acceptor H-bond angles are less favorable.



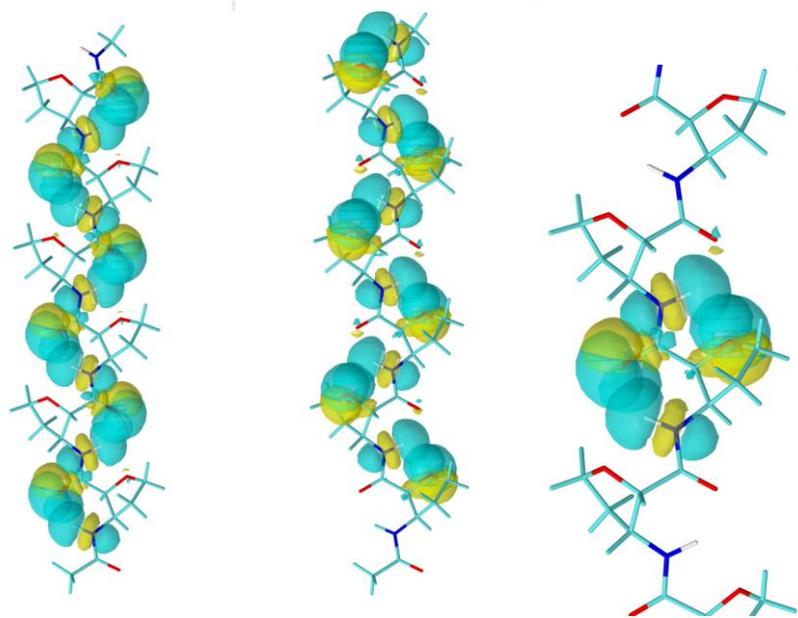


Figure 8 Selected NBOs mounted on the skeleton of $(H_{10}^P)_6$ (top row) and $(Z_6^M)_6$ (bottom row) conformers of Ac-(cATFC)₆-NHMe. The amide-amide (left columns), the amide-“THF-ring” H-bonds (middle columns), and those combined per residue (right columns) are depicted.

The *trans* isomer, just as seen in for ACPC, prefers the H_8^P conformation in the monomeric state (see **Table S13**) and the H_{12}^P conformation in hexameric state (see **Table S14**). In other words, the elongation of $tATFC \rightarrow (tATFC)_6$ doesn't refold the building block's structure, only a gentle shift is experienced as the backbone H-bond network evolves. A curious type of helical conformer was also found among stable (but rather unfavored) conformers of $(tATFC)_6$ which merits attention. Using the well-established descriptors based on the (ϕ, θ, ψ) torsion angles and nomenclature of β -peptides, conformer #9 (see **Table S14**, **Figure 9**) would be classified as an H_{10}^P helix. However, the structure is a helix only in the sense that it winds around a central pore, but it lacks the main-chain H-bond pattern that usually stabilizes such an arrangement. This structure is defined instead by the rigidity introduced by an apparent drive to sustain the $NH_i \dots (O_{\text{furanoid}})_i$ H-bonds of the building unit. Pseudo-Ramachandran backbone torsional angles, $(\phi_i, \psi_i) \sim (96^\circ, 140^\circ)$, indicate an ϵ -type (elongated) conformer.²⁵ It is interesting to note, that the two descriptors together give an accurate characterization; the fold is a helix (in line with the topology) and an elongated structure (indicating the absence of stabilizing backbone H-bonds) at the same time. This is also reflected in its chain-length of 18 Å, longer than that of a $(H_{12}^P)_6$ (13.4 Å) but shorter than a truly extended zigzag fold (e.g. $\zeta(Z_6^M)_6$ is 23.2 Å).

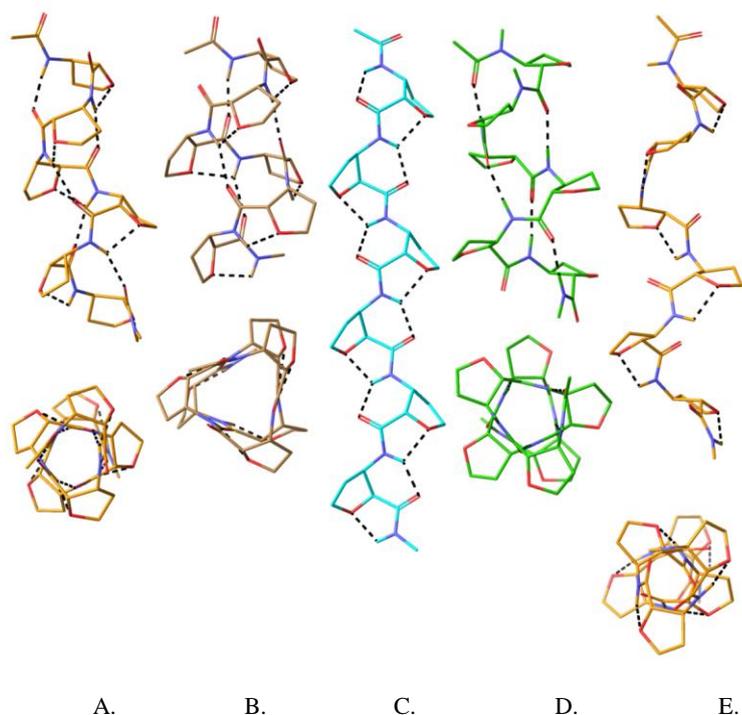


Figure 9 The calculated structure of H_{10}^P (A.), H_{14}^P (B.) and Z_6^{*M} (C.) conformations of $(cATFC)_6$ and the H_{12}^P (D.) and H_{10}^P (conformer #9) (E.) conformations of $(tATFC)_6$ homohexamers. Helices (and the helix-like conformation of H_{10}^P) are also shown from the top. Color scheme is the same as that used in the Tables of the Supporting Information.

The results thus indicate that configuration induced conformer selection is present in case of ATFC too and it is exclusive! Instead of being able to switch between a zigzagged and a helical conformation as seen for $(ACPC)_n$, the two alternatives in this case are the left-like and the right-handed helices: namely the *cis* configuration selects the former, while the *trans* the latter one. Both selections are exclusive leaving no room for conformational interchange at room temperature.

Fold-type preserved in oligomers: foldamers of $XylAFU^{\alpha/\beta}$ and $RibAFU^{\alpha/\beta}$ sugar amino acids

In the case of XylAFUs and RibAFUs building blocks, the two additional –OH groups (with respect to ATFC) stimulate the formation of additional intricate H-bond motifs, further influencing conformer selections. The maximum number of H-bonds formed in homohexamers of ACPC is 6, 12 in ATFC, while in case of the amino furanuronic acids discussed

below this number can be as high as 28, recalling the large number of H-bonds present in cellulose and starch.

The *cis* or (*S,R*)-stereoisomer of the α -D-Xylofuranuronic acid (XylAFU $^{\alpha}$)_n derivative, adopts exclusively the Z₆^{*M} conformation both as a monomer and as a hexamer (Tables S15 and S16). In the (Z₆^{*M})₆ foldamer 24 H-bonds stabilize the system (while the Z₆^{*M} monomer contains 4). The β -anomer, (XylAFU $^{\beta}$)_n, has very similar conformational preferences (Tables S20 and S21). Here also, the Z₆^{*M} conformer has the largest number of H-bonds, namely 3 in the monomer and 28 in the hexameric form (Figure 10). It is interesting to note, that the great number of H-bonds present in the (Z₆^{*M})₆ form de-selects the helical conformations in this case: the (H₁₀^P)₆ conformation – which was found to be the global minimum of c(ATFC)₆ - contains 7 and 16 (!) less H-bonds in case (XylAFU $^{\alpha}$)₆ and (XylAFU $^{\beta}$)₆ than the (Z₆^{*M})₆ arrangement.

On the other hand, for the α anomer of the *trans* (RibAFU $^{\alpha}$)_n (of (*S,S*) configuration), both the monomer and the hexamer was found to prefer the elongated-like Z₆^{*M} conformer. However, unlike for the monomer where population of Z₆^{*M} is dominant (~75%), the energy difference between the extended and helical forms is balanced for the homohexamer: ~42% of (Z₆^{*M})₆, ~41% of (H₁₂^P)₆ and 17% of (H₁₀^P)₆ were found, within an energy gap of < 3 kJ/mol with respect to the global energy minimum (Table S17, S18 and Figure 11). Because of the small energy differences, all 3 conformers were recalculated on a larger basis set (6-311++G(d,p), Table S19). The more accurate QM approach reordered the conformational states entering more structure discrimination: 88% (H₁₂^P)₆, 9% of (Z₆^{*M})₆ and 3% (H₁₀^P)₆ was found to constitute the equilibrium mixture and this was accepted as the presently determined best estimate of this state. It is worthwhile to note, that Z₆^{*M} conformer was found to be quite close in energy to that of the helical H₁₂P structure even at this higher level (at 5.6 kJ/mol). It is therefore quite possible that oligomers of RibAFU inserted into a protein might be stabilized in a zigzagged form too, by a couple of well-positioned interactions formed with the protein matrix - underlining that *trans* building blocks are not all necessarily pre-determined toward forming helical structures. This increase in structure variability seems to be coupled to the increased complexity and H-bonding capability of foldamers.

RibAFU $^{\alpha}$ affords the longest homohexamer of this study too, a truly extended aaa conformation (its (ϕ θ ψ) torsion angles approaching 150°), with 27.7Å chain length - conformer #8 at ΔE of 83.3 kJ/mol relative energy with respect to the global minimum (Table S18).

formázott: Betűtípus: Nem Félkövér

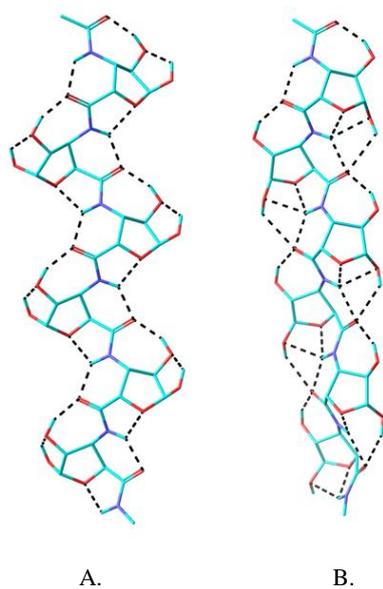


Figure 10 The calculated structure of Z_6^{*M} conformer of $(XylAFU^\alpha)_6$ (A.) and the Z_6^{*M} conformer of $(XylAFU^\beta)_6$ (B.) homo-hexamers. Color scheme goes with that of the Tables of the Supporting Information.

The *trans* (Ribo-)stereoisomer with β -anomeric orientation, $(RibAFU^\beta)$, just as seen previously for the *t*ACPC and *t*ATFC elements, prefers H_8^P (87%) as a monomer and exclusively (99%) $(H_{12}^P)_6$ as a hexamer. However, as described for the α -anomer, the Z_6^M conformer type occupies the close second place (at +10.6 kJ/mol, see **Tables S22, S23** and **Figure 11**).

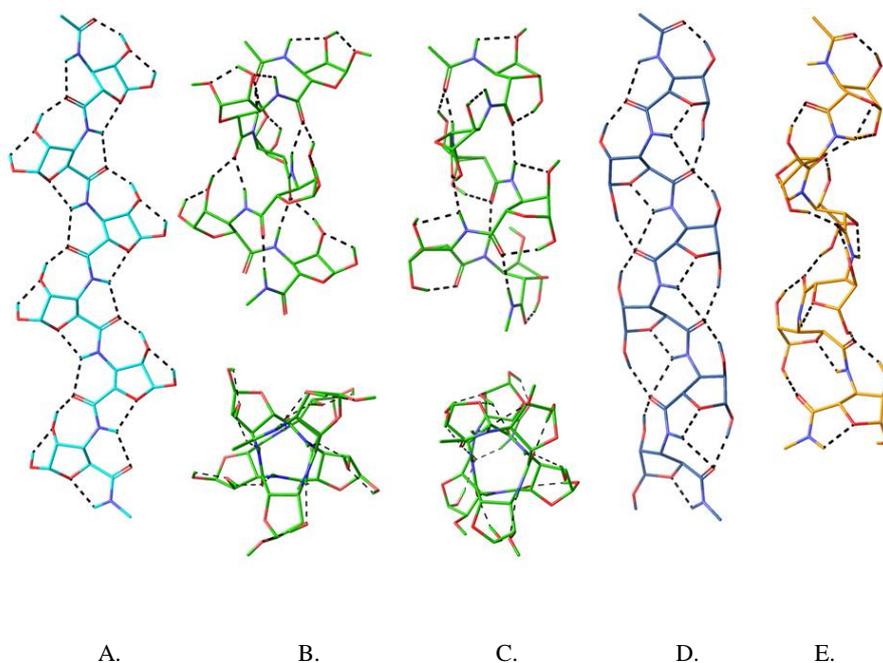


Figure 11 The calculated structure of Z_6^{*M} (A.) and H_{12}^P (B.) conformers of $(\text{RibAFU}^\alpha)_6$ and the H_{12}^P (C.), Z_6^M (D.) and H_{10}^P (E.) conformations of $(\text{RibAFU}^\beta)_6$ homo-hexamers. Color scheme is the same as that used in the Tables of the Supporting Information.

Due to mutarotation, the α - and the β -anomers of the furanoid ring of these sugar amino acid derivatives exchange in water and thus, the more favorable of the AFU^α and AFU^β anomers are to be considered only. For the *cis* stereoisomers, $(\text{XylAFU}^{\alpha/\beta})_6$, the α -anomer turns out to be the more stable one (17.9 kJ/mol), but for the other diastereomer, $(\text{RibAFU}^{\alpha/\beta})_6$, the β -anomer is the more stable one (9.9 kJ/mol). This variation however, does not affect the overall conformer selection, as the α and β form of both stereoisomers have a clear and common preference, namely $(Z_6^{*M})_6$ for *cis* or Xylo and $(H_{12}^P)_6$ for the *trans* or Ribo.

Fold-type preserved in oligomers: foldamers of *XylAFU^{α(ip)}* and *RibAFU^{α(ip)}* sugar amino acids

The 1,2-isopropylidene (*ip* for short) derivatives of both D-Xylo- and D-Ribo-sugar amino acids are key intermediates of AFU chemical synthesis as well as water soluble foldamer building blocks by their own right^(13, 20, 27). In the case of the *cis* stereoisomer, *XylAFU^{α(ip)}*, the H_{14}^P , a helix-prone conformer is found to be the most stable (70%, **Table S24**) and this same **left-handed-like helical fold** becomes exclusive for the homo-hexamer: $(XylAFU^{\alpha(ip)})_6 \rightarrow (H_{14}^P)_6$. (**Table S25** and **Figure 12**). The question of handedness and directionality of this helix is quite similar to those seen and discussed in case of *cATFC*, with its $NH_i \rightarrow C=O_{i+2}$ H-bond system pointing in the opposite direction as that of a regular α -helix (compare its pseudo position on the Ramachandran map **Figure 6**), but with a right handed N-term to C-term wind (**Figures 8, 9** and **12**). Note, that while $(XylAFU^{\alpha(ip)})_6$ adopts exclusively a left-handed-like helix (~10 Å in length) when the hydroxyl groups of the sugar are protected with an isopropylidene group (**Table S25**), it takes on the zigzagged (Z_6^{*M})₆ fold of very different shape and length (~23 Å) (**Table S16** and **S21**) when the –OHs are free.

The *trans* stereoisomer, *RibAFU^{α(ip)}*, favors the Z_8^{*M} conformation (67.5%, **Table S26**) as a monomer, while as a hexamer it selects the $(H_{12}^P)_6$ right-handed helical foldamer exclusively (100% **Table S27**). This is quite understandable, as $(H_{12}^P)_6$ of *RibAFU^{α(ip)}* contains 11 H-bonds (**Figure 12**): 5 of which are well positioned $-C=O_i \leftarrow NH_{i+3}$ backbone H-bonds as compared to the 6, quite strained $(O_{furanoid})_i \leftarrow NH_{i+1}$ type interactions within the $(Z_8^{*M})_6$ conformation.

formázott: Felső index

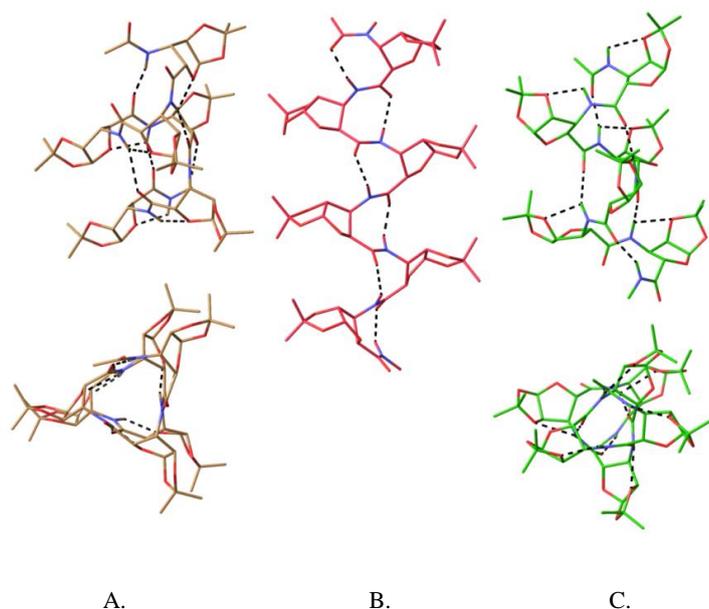


Figure 12 The calculated structure of H_{14}^P (A.) and the elongated Z_8^{*P} (B.) conformers of $(XylAFU^\alpha(ip))_6$ (Table S25) and the right-handed helical H_{12}^P (C.) conformer of $(RibAFU^\alpha(ip))_6$ homo-hexamers (Table S27). Helices shown from the top, coloring is in accordance with Tables of the Supporting Information.

Discussion

Ascertaining the secondary structure forming propensities of foldamer building blocks is important for two overlapping reasons. It allows the design of protein compatible and complementary premediated 3D structures and allows for selecting the ideal building block(s) for incorporation into a given molecular structure (e.g. β -turn, helix, sheet, coiled-coil).

In this study the *cis* and *trans* stereoisomers of ACPC, ATFC, two amino β -furanuronic acids (AFUs), and 1,2-*O*-isopropylidene protected derivatives of the latter building blocks were considered and calculated, all in their monomeric and homo-oligomeric form. There are a number of structural studies concerning such foldamer building blocks containing five membered ring systems. *Cis*-configuration results in elongated, strand-like structures in case of oligomers of homochiral *c*ACPC,^{3,16} while in helices in case of the more polar building units, like *c*ATFC,¹⁹ and the isopropylidene protected XylAFU(ip).¹⁴ Thus, the introduction of a ring O-atom (in place of the $-C^eH_2-$ segment) is seen to be an important

factor in determining the overall 3D-fold of these *cis* oligomers. On the other hand, the *trans* isomer of tetrahydrofuran β -amino acid, *t*ATFC, was found to form a 12-helix¹⁹, complying with that found for *t*ACPC too.^{16,18} The *trans* isomer was shown to preserve its oligomeric form even when a bulky thymine is attached to the tetrahydrofuranyl ring, folding into a H12²⁹ or H8³⁰ helical structure.

In line with the above we also found, in homohexamers of cyclic β -amino acids containing 5 membered rings, that the *trans* (*S,S*) configuration invokes almost exclusively the formation of (H12^P)₆ right-handed helices. These helices are remarkably stable and similar; their backbone structures, defined by the common H-bond network, are virtually unaltered by the different ring substituents. On the contrary, *cis* configuration results in either elongated (zigzag_z)₂ or \ast left-handed-like helical foldamers. More specifically, our comprehensive analysis revealed that homohexamers of *c*ACPC and the unprotected forms of the XylAFU ^{α} and XylAFU ^{β} prefer the lengthier zigzagged conformations ((Z₆ ^{\ast P})₆ and its mirror image, (Z₆ ^{\ast M})₆ fold), while *c*ATFC and the isopropanyl-protected XylAFU ^{α} (ip) will fold into helices: hexamers of *c*ATFC into a 10-helix, and XylAFU ^{α} (ip) into a 14-helix ((H10^P)₆ or (H14^P)₆ conformers, respectively) - in agreement with the (CD and NMR) experimental results^{14,19} (Figure 13).

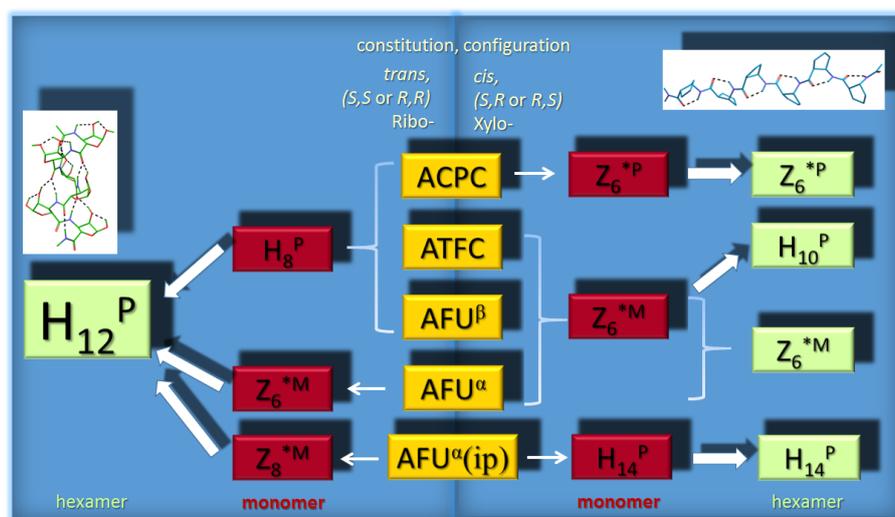


Figure 13 Constitution and configuration determined conformer types (Lego elements) of β -(sugar) amino acid derivatives. (Note that H12^P makes a right-handed, while both H10^P and H14^P form a left handed helix, but of different pitch.)

The energetic ordering of both the monomeric and hexameric forms correlate in general with the number of H-bonds available in a given conformer, attenuated by the unfavorable electrostatic interactions of the sidechain oxygen atoms and the carbonyl moieties in certain molecular packings. We found that in most – albeit not all - cases the

conformational preference of the monomer building blocks prevails in the homo-hexameric state too, if the additional stabilizing effect of inter-unit H-bonds - that cannot be present in monomers - is also taken into account. 10-, 12- 14- or even 16-membered main-chain H-bond loops are not only an additional possibility present in longer systems but also allow for much better oriented, thus stronger interactions than the 6- or 8-membered H-loops of the monomers. Z-type monomers with 6- or 8-membered H-bond loops in almost all cases assemble into zig-zagged hexamers, preserving their H-bond motifs in the oligomeric state too. Therefore, to adequately estimate the success of a helix-prone (H-type) monomer in the oligomeric state, an extra ~10 kJ/mol stabilization energy³¹ has to be added to the energy of the monomer to account for the H-bonds that will form upon hexamer formation. If this is done, the resultant energetic ordering of the monomer states will be a rather dependable estimate of the conformational preferences of the longer versions too, at least in predicting whether an elongated, zigzagged or helical type of conformation is to be expected.

To estimate the willingness of the monomers and oligomers of furanoid β -amino acids to participate in α -peptidic structures, we had to describe the extent of their similarity. Martinek and Fülöp proposed the use of inter-residue $\psi(n-1)$ vs. $\phi(n)$ shifted Ramachandran maps to distinguish between helical and extended conformations and simultaneously predict their compliance with α -peptidic structures.¹¹ Helices result in angle pairs of identical sign (eg. an α -helix is of (-,-)), strands and other extended motifs in opposing sign-pairs (eg. a β -strand is of (-,+)) (11). This works nicely for most structures calculated in this study: the right-handed H_8^P , H_{12}^P , H_{16}^P helices are (-,-), the left-handed-like H_{10}^P and H_{14}^P helices (+,+) (reflecting their non-conformity with α -helices), and the zigzag conformers of cACPC (Z_6^{*P} , Z_6^M) are (+,-) (indicating a strand-like structure that however will not match with the β -strands of α -peptides). On the other hand, most of the Z_6^{*M} structures calculated by us (appearing among low energy conformers of AFUs) would be characterized as a helix based on the (+,+) sign of their ϕ ψ angle-pairs (in case of homohexamers, shifting the ψ value to the previous residue has no real significance, since both ϕ s and ψ s are uniform along the backbone), although these structures are definitely elongated zigzag conformers. Thus, in most –albeit not all - cases, this method clearly distinguishes whether a helical or an extended structure is to be expected, but it does not describe how close the topology of that structure will be to those of their α -counterparts.

So, here a slightly different approach was applied: a dummy atom was introduced halfway between the C_α and C_β , which could be used to measure classical ϕ and ψ - like pseudo backbone angles (ϕ_d and ψ_d), similar to those of α -peptides and proteins (see **Figure 4**). While the extended Z_6^{*P} conformation of (cACPC)₆ –according to this measure - results in a backbone conformation that is rarely sampled by α -proteins (falls in the disallowed region when projected to the α -Ramachandran-map thus reflecting its strand-breaking nature²⁸), the Z_6^{*M} hexamers of XylAFU ^{α} and XylAFU ^{β} possess a backbone quite similar to that of the β -strand conformation of α -peptides (**Figure 14**), with ϕ_d , ψ_d values of -164.6°, 128.1° and -158.9°, 128.9°, corresponding to the β -strand region of the Ramachandran map (**Figure 6**). Also, even though the 10- and 14- helices formed by cATFC and XylAFU ^{α} (ip) hexamers fall

near the Ramachandran region corresponding to left-handed helices of α -peptides, due to the switched directionality of their hydrogen bond motif (NH \rightarrow C=O type H-bonds), these two conformers will also be incompatible with α -peptide structures (**Figure 87**). On the other hand, all H_{12}^P conformers of the *trans* β -amino acids studied here fall in the Ramachandran-region of the α -helix, and comply readily with it (see **Figure 6**. and **Figure 14**). Thus, of the studied systems, it is only the sugar amino acids (AFU^α and AFU^β) that are, in both their *cis* and *trans* forms, readily insertable into regular protein folds substituting segments of secondary structure elements (**Figure 14**).

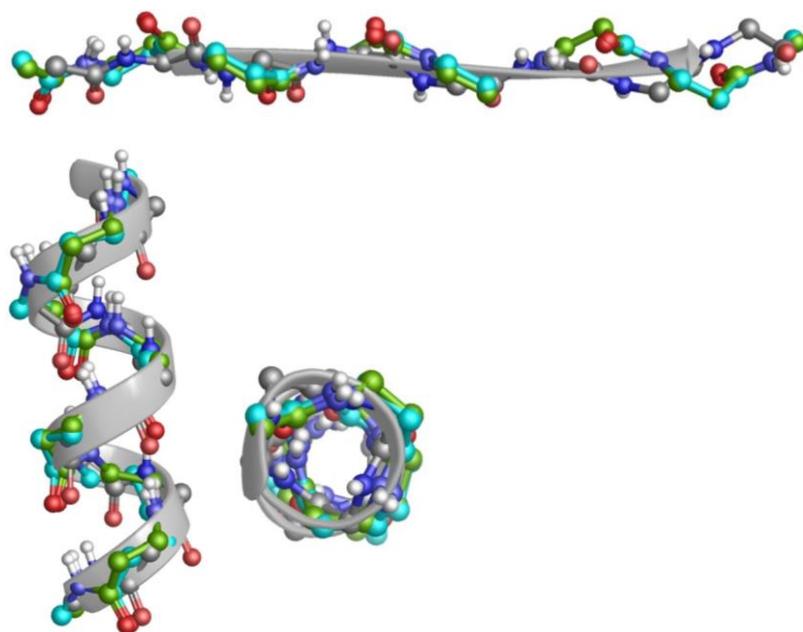


Figure 14 $Z6^{*M}$ backbone of $XylAFU^\alpha$ (C atoms in green) and $XylAFU^\beta$ (C atoms in cyan) fitted to the backbone of a β -strand of an α -protein (transferrin (pdb code: 1tta, residues: 114-121) (C atoms is grey) (top). H_{12}^P backbone of $RibAFU^\alpha$ (C atoms in green) and $RibAFU^\beta$ (C atoms in cyan) fitted to the backbone of a α -helix of an α -protein (glp-1 (pdb code: 3iol, residues: 38-45) (C atoms is grey) – 6 β residue to 8 α -residue to compensate for the difference in rise (bottom).

For successful design however, a few more factors should be considered as well. Even if the major geometrical properties match between the α - and β -peptides, sidechain position will always be shifted: “on- and off-register” regions appear when compared. For example, a hexamer of RibAFU has a backbone structure topologically analogous to that of an α -helix, but it will be longer than a six amino acid long α -helix. For an ideal fit, 6 monomers of

RibAFU would substitute 8 α -amino acid residues. Similarly, a hexameric stretch of XylAFU will fit to an 8 residue long β -strand, so for optimizing interactions, shifts in sidechain position have to be taken into account (as can be seen on **Figure 14**).

Substitution of an entire secondary structural elements should also be possible, where the inserted foldamer no longer has to comply with the backbone structure, simply serve as a well-defined stand-alone linker between protein regions, domains or motives. In this case the width and length of it will be most important, which we found also telling characteristics of any given homo-oligomer. In general, our findings indicate that in case of the studied homohexamers, fully elongated conformers reach over 25 Å in length, 6-zigzag conformers stretch to 22 - 25 Å, 8-zigzagged conformation results in an approximately 18 - 20 Å chain-length, while various helical arrangements lead to less than 21 Å chain-end separations. Thus, for example, the approximately 13 Å length of $(H_{12}^P)_6$ helices can also be bridged by a tetramer of H_8^P (a narrower helix) or by a heptamer of H_{16}^P (a wide and short type of helix) and even by a trimer of Z_6^{*M} – all well-defined, self-assembled structural motifs that can be built from AFU-derivatives of the appropriate constitution and configuration.

In this paper we focused our attention on homo-oligomers in an attempt to describe the standalone conformational preference of the studied building blocks, however, β -foldamers can be linked in heterogeneous (mixed) patterns too, or can be intermixed with various α or γ amino acids, which makes the versatility and plasticity of the attainable structures practically limitless¹¹. Having established that the conformational preferences of the monomers have a distinct influence over the evolving homo-oligomeric structures, it is a straightforward question whether the structure of alternating or mixed sequences can also be predicted using our methodology. A nice example of such ~~was presented by~~ [can be found in the results of Martinek et al.](#)¹⁶ who have shown using various experimental techniques that while joining six (R,R) - t ACPC units of uniform configuration leads to formation of a helix, using alternating (R,R) and (S,S) building blocks results in strand formation. Similarly, while the homo-oligomeric form of (R,S) - c ACPC₆ is a strand, using alternating (S,R) and (R,S) units, a H10/12 helix can be produced. When the configuration of a foldamer is thus switched, the resultant mirror image structure is expected to populate the enantiomeric pairs of the conformers established as favorable for the original molecule. For example, the (R,R) - t ACPC monomer should prefer the H_8^M conformation, which is the mirror image of H_8^P determined as the global minimum of (S,S) - t ACPC in this study (see Table S6). It is quite interesting to see, that by simply placing alternating (S,S) - and (R,R) - t ACPC monomers in H_8^P and H_8^M conformation next to each other –without carrying out any calculation – a strand structure stabilized by 8-membered H-loops appears, exactly the arrangement seen experimentally.¹⁶ On the other hand, the H10/12 helix of the alternating oligomer of (S,R) and (R,S) - c ACPCs, can be built from the Z_8^M and Z_8^P enantiomeric pair of low energy conformers (see **Table S1**, **S6** and **Figure 15**), allowing a slight (less than 30°) readjustment of its torsional angles. It is important to note, that while the alternation of mirror image building blocks changes the outcome topology (helix building blocks resulting in a strand, while Z-type monomers building a helix), the conformational preference of the monomer units prevails even within this non-uniform environment. The same behavior is expected when inserting β -monomers

into α - or various other backbones. It has also been shown that such single foldamer insertions may be useful for testing and modifying specificity of the parent α -peptide medications, as was seen, for example, in a recent study, that where by $\alpha \rightarrow \beta$ substitution in select positions of GLP-1 analogues, can introduce significant bias in the implicated GPCR-coupled signaling pathways could be introduced, thus suggesting that such single foldamer insertions may be useful for testing and modifying specificity of the parent α peptide medications.³³ According to our findings, the design of such insertions should also be possible relying on the conformational preferences of the monomeric β -foldamer unit, allowing for the selection of building blocks carefully matched to the local structural features of the α host.

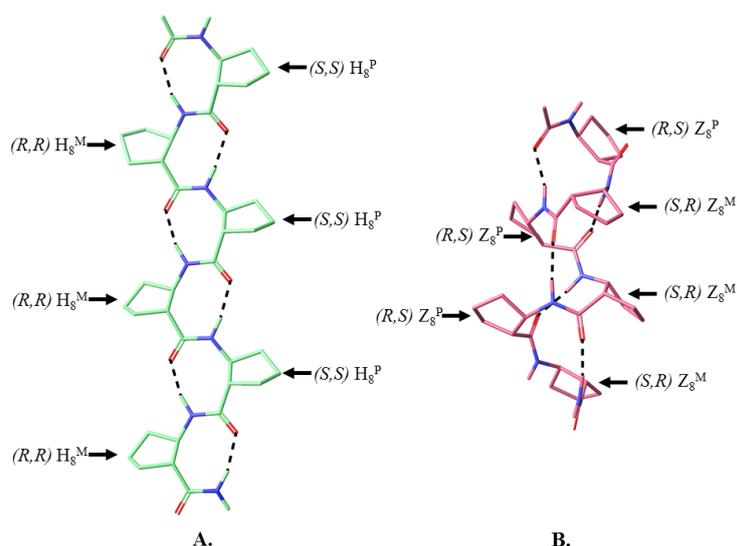


Figure 15 A. Strand and B. H10/12 helix formed from tACPC and cACPC building blocks of alternating configuration

Conclusion

We found that in case of β -furanoid -amino acids, due to the appearance of sidechain H-bond acceptors and donors, the configuration based uniformity of foldamers is lost in a sense that all *cis* monomers will not prefer elongated, zigzagged secondary structures in their oligomeric forms while *trans* monomers will not necessarily assemble into helices. However, we also found that the homo-oligomer forms of the studied systems have a definite **conformational preference** and those **structures are predictable** based on the conformational preferences of the monomeric units (corrected for the absence of inter-unit H-bonds (in case of helix-builder monomers)). The methodology described in this paper

correctly predicted the known structures of homo-oligomers of *c*ACPC, *t*ACPC, *c*ATFC, *t*ATFC and RibAFU(ip), demonstrating that the potential energy surface of monomers of β -amino acids can be sufficiently described by optimizing 27 representative structures of the dihedral space and based on the results, the preferred form of the homo-oligomer can be estimated. Thus, we propose that when designing new β -peptides or for an initial estimation of the propensity of forming α -compatible secondary structures in case of any new β -foldamer type, the described simple procedure might be followed.

In addition to the well-established ϕ , θ , ψ backbone triplets, Ramachandran-like pseudo backbone dihedrals (ϕ_i , ψ_i) and the chain-length of homo-multimers were found to be useful for describing the adaptability of β -peptides to α -protein structures. In case of β -furanoid sugar-amino acids, beside their bio-compatibility, their conformational compatibility with α -protein structures was also revealed, proposing them as ideal building blocks for incorporation not only in select single positions, but also in form of longer subunits, *e.g.* hexamers.

Supporting Information

Tables detailing results of the calculations (as described in the text) provided in Tables S1–S27.

ACKNOWLEDGEMENTS

This work was supported by grants from the Hungarian National Science Fund (OTKA, NK101072 and K116305).

References:

1. Appella, D. H.; Christianson L. A.; Klein, D. A.; Powell, D. R.; Huang, X.; Barchi Jr, J. J.; Gellman S. H. Residue-based Control of Helix Shape in β -peptide Oligomers. *Nature* **1997**, *387*, 381-384.
2. Appella, D. H.; Christianson L. A.; Klein, D. A.; Richards, M. R.; Powell, D. R.; Gellman S. H. Synthesis and Structural Characterization of Helix-Forming β -Peptides: *trans*-2-Aminocyclopentanecarboxylic Acid Oligomers. *J. Am. Chem. Soc.* **1999**, *121*, 7574–7581.
3. Martinek, T. A.; Tóth, G. K.; Vass, E.; Hollósi, M.; Fülöp, F. *Cis*-2-Aminocyclopentanecarboxylic Acid Oligomers Adopt a Sheetlike Structure: Switch from Helix to Nonpolar Strand. *Angew. Chem. Int. Ed.* **2002**, *41*, 1718-1721.

4. Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman S. H. Structural and Biological Mimicry of Protein Surface Recognition by Alpha/Beta-peptide Foldamers. *Proc. Natl. Acad. Sci. USA*. **2009**, *106*, 14751-14756.
5. Denton, E. V.; Craig, C. J.; Pongratz, R. L.; Appelbaum, J. S.; Doerner, A. E.; Narayanan A.; Shulman, G. I.; Cline G. W.; Schepartz A. A β -Peptide Agonist of the GLP-1 Receptor, a Class B GPCR. *Org. Lett.* **2013**, *15*, 5318-5321.
6. Johnson, L. M.; Barrick, S.; Hager, M. V.; McFedries, A.; Homan, E. A.; Rabaglia, M. E.; Keller, M. P.; Attie, A. D.; Saghatelian, A.; Bisello, A.; Gellman S. H. A Potent α/β -Peptide Analogue of GLP-1 with Prolonged Action in Vivo *J. Am. Chem. Soc.* **2014**, *136*, 12848-12851.
7. Checco, J. W.; Lee, E. F.; Evangelista, M.; Sleeb, N. J.; Rogers, K.; Pettikiriachchi, A.; Kershaw, N. J.; Eddinger, G. A.; Belair, D. G.; Wilson, J. L.; Eller, C. H.; Raines, R. T.; Murphy, W. L.; Smith, B. J.; Gellman, S. H.; Fairlie, D. α/β -Peptide Foldamers Targeting Intracellular Protein-Protein Interactions with Activity in Living Cells *J. Am. Chem. Soc.* **2015**, *137*, 11365-11375.
8. Checco, J. W.; Gellman, S. H. Targeting Recognition Surfaces on Natural Proteins with Peptidic Foldamers. *Curr. Opin. Struct. Biol.* **2016**, *39*, 96-105.
9. Fülöp, F.; Martinek, T. A.; Tóth, G. K. Application of Alicyclic β -amino Acids in Peptide Chemistry. *Chem. Soc. Rev.* **2006**, *35*, 323-334.
10. Goodman, C. M.; Choi, S.; Shandler, S.; DeGrado, W. F. Foldamers as Versatile Frameworks for the Design and Evolution of Function. *Nat. Chem. Biol.* **2007**, *3*, 252-262.
11. Martinek, T. A.; Fülöp, F. Peptidic Foldamers: Ramping up Diversity. *Chem. Soc. Rev.*, **2012**, *41*, 687-702.
12. Risseuw, M.; Overhand, M.; Fleet, G. W.; Simone, M. I. A Compendium of Cyclic Sugar Amino Acids and their Carbocyclic and Heterocyclic Nitrogen Analogues. *Amino Acids* **2013**, *45*, 613-689.
13. Gruner, S. A. W.; Locardi, E.; Lohof, E.; Kessler, H. Carbohydrate-based Mimetics in Drug Design: Sugar Amino Acids and Carbohydrate Scaffolds. *Chem Rev.* **2002**, *102*, 491-514.
14. Chandrasekhar, S.; Reddy, M. S.; Jagadeesh, B.; Prabhakar, A.; Ramana Rao M. H. V.; Jagannadh B. Formation of a Stable 14-Helix in Short Oligomers of Furanoid cis- β -Sugar-Amino Acid. *J. Am. Chem. Soc.* **2004**, *126*, 13586-13587.
15. Christianson, L. A.; Lucero, M. J.; Appella, D. H.; Klein, D. A.; Gellman, S. H. Improved Treatment of Cyclic β -amino Acids and Successful Prediction of β -peptide Secondary Structure Using a Modified Force Field: AMBER**C*. *J. Comput. Chem.* **2000**, *21*, 763-773.
16. Martinek, T. A.; Mándity, I. M.; Fülöp, L.; Tóth, G. K.; Vass, E.; Hollósi, M.; Forró, E.; Fülöp, F. Effects of the Alternating Backbone Configuration on the Secondary Structure and Self-Assembly of β -Peptides. *J. Am. Chem. Soc.* **2006**, *128*, 13539-13544.
17. Sussman, F.; Villaverde, M. C.; Estévez, J. C.; Estévez, R. J. Searching the Conformational Space of Cyclic β -amino Acid Peptides. *J. Phys. Chem. B* **2009**, *113*, 9669-9680.

18. Choi, S. H.; Guzei, I. A.; Spencer, L. C.; Gellman, S. H. Crystallographic Characterization of 12-Helical Secondary Structure in β -peptides Containing Side Chain Groups. *J. Am. Chem. Soc.* **2010**, *132*, 13879–13885.
19. Pandey, S. K.; Jogdand, G. F.; Oliveira, J. C. A.; Mata, R. A.; Rajamohanan, P. R.; Ramana, C. V. Synthesis and Structural Characterization of Homochiral Homooligomers of Parent cis- and trans-Furanoid- β -amino Acids. *Chem. Eur. J.* **2011**, *17*, 12946–12954.
20. Nagy, A.; Csordás, B.; Zsoldos-Mády, V.; Pintér, I.; Farkas V. and Perczel A. C-3 Epimers of Sugar Amino Acids as Foldameric Building Blocks: Improved Synthesis, Useful Derivatives, Coupling Strategies. *Amino Acids*, **2016**, *48* (in press). doi:10.1007/s00726-016-2346-5
21. Csordás, B.; Nagy, A.; Harmat, V.; Zsoldos-Mády, V.; Leveles, I.; Pintér, I., Farkas, V.; Perczel, A. Origin of Problems Related to Staudinger Reduction in Carbopeptoid Syntheses. *Amino Acids* **2016**, *48*, 2619-2633.
22. Beke, T.; Somlai, C.; Perczel, A. Toward a Rational Design of Beta-Peptide Structures. *J Comput Chem.* **2006**, *27*, 20-38.
23. Frisch, M. J., Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A. Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox D. J., *Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
24. Lovell, S. C., Davis, I. W., Arendall, W.B. 3rd, de Bakker, P. I., Word, J.M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by $C\alpha$ Geometry: ϕ, ψ and $C\beta$ deviation. *Proteins* **2003**, *50*, 437–450.
25. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J. L.; Marcocchia, J. F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256–6265.
26. Oberholser, K.; Sussman, J. L.; Hodis, E.; Decatur, W.; Livne, S.; Prilusky, J., Richardson, J. S., Berchansky, A. Ramachandran Plot *Proteopedia* **2013**. doi: 10.14576/381225.1772128
27. Foster, J. P.; Weinhold F., Natural Hybrid Orbitals, *J. Am. Chem. Soc.* **1980**, *102*, 7211-7218
28. Olajos, G.; Hetényi, A.; Wéber, E.; Németh, L. J.; Szakonyi, Z.; Fülöp, F.; Martinek, T. A. Induced Folding of Protein-Sized Foldameric β -Sandwich Models with Core β -Amino Acid Residues. *Chem. Eur. J.*, **2015**, *21*, 6173–6180.

29. Chandrasekhar, S.; Reddy, G. P. K.; Kiran, M. U.; Nagesh, C.; Jagadeesh, B. Nucleoside Derived Amino Acids (NDA) in Foldamer Chemistry: Synthesis and Conformational Studies of Homooligomers of Modified AZT. *Tetrahedron Lett.* **2008**, *49*, 2969-2973.
30. Threlfall, R.; Davies, A.; Howarth, N. M.; Fisher, J.; Cosstick, R. Peptides Derived from Nucleoside β -Amino Acids Form an Unusual 8-helix. *Chem. Commun.* **2008**, 585-587.
31. Wendler, K.; Thar, J.; Zahn, S.; Kirchner, B. Estimating the Hydrogen Bond Energy. *J. Phys. Chem. A* **2010**, *114*, 9529-9536.
32. Claridge, T. D. W.; Goodman, J. M.; Moreno, A.; Angus, D.; Barker, S. F.; Taillefumier, C; Watterson, M. P.; Fleet, G. W. J.; 10-Helical conformations in oxetane β -amino acid hexamers, *Tetrahedron Letters* **2001**, *42*, 4251-4255.
33. Hager, M. V.; Johnson, L. M.; Wootten, D.; Sexton, P. M.; Gellman, S. H. β -Arrestin-Biased Agonists of the GLP-1 Receptor from β -Amino Acid Residue Incorporation into GLP-1 Analogues. *J. Am. Chem. Soc.*, **2016**, *138*, 14970-14979.