Helix compactness and stability: Electron structure calculations of conformer dependent thermodynamic functions

This Letter is dedicated to Professor Imre G. Csizmadia on the occasion of his 80th birthday by the co-authors and friends in appreciation of his contribution to the advancement of computational chemistry.

Imre Jákli, Imre G. Csizmadia, Szilard N. Fejer, Ödön Farkas, Bela Viskolcz, Svend J. Knak Jensen', Andras Perczel*

1. Introduction

Polypeptides can adopt a number of different conformers, but helices and β-strands are the most abundant structural motifs of proteins. Among helices the α- or 3.6_{13}-helix (sometimes it is called 4_{13}), with the characteristic $i \leftarrow (i + 4)$ type H-bonds is the most common [1]. Less frequently 3_{10}- and sporadically π-helical turns can also be assigned in proteins, all of right handedness. Regardless of their length α-helices often start and terminate by shorter 3_{10}-helical turn(s), where the intramolecular H-bonds are of $i \leftarrow (i + 3)$ type. This results in a narrower fold typically found by computational methods. In vacuum, the latter helical structure is the more stable backbone fold [2]. Furthermore, the mixing and interconversion of 3_{10}- and α-helices is allowed and was successfully obtained by quantum mechanical (QM) calculations in various solvents [2]. Helices have noticeable dipole moments, as each of the component homo-conformers (α_L for short [3]) has their own dipole moment. The constructive summation is due to the similar orientation of the adjacent amide planes around and along the central axis of a helix. This results in a significant macro dipole moment, with a positive end at the N- and a negative end at the C-terminus.

Helices can be involved in a variety of motifs, such as all-α, βαβ, Rossman-fold and TIM barrel [4]. Even in*i*ntrinsically *d*ynamic *p*roteins (IDP), where the backbone of the protein presents an unexpectedly large structure fluctuation, residual helices could be assigned [5], as temporarily existing structural motifs [6]. Helix packing is an important issue, as there are only a limited number of modes how helices can self organize. In the most common four-helix bundle arrangement, the topology is 'up-and-down' allowing the side-chains formed ridges to ideally pack into each other's grooves. Different type of membrane proteins were already characterized; those equipped with a single α-helical tail anchoring the globular part to the membrane and those passing through the membrane several times (e.g., bacteriorhodopsin) resulting in seven transmembrane (7TM) helices. The longest helical regions are in keratine, the protein of hair, in myosine and tropomyosine [7] and [8], which form muscular fibers, winding around each other by forming the coiled-coil superstructure.

The length of an α-helix varies as function of several factors, but in globular proteins most helices comprise around 10–15 amino acid residues [9]. By analyzing over 150 globular proteins the average length was found to be ~18 ± 8, ranging between 10 and 50 [10]. In coiled-coils [11], the average length is ~17, based on a recent survey by using PDB select 2008 [12].

However, the helices forming coiled-coils can be rather long, comprising well over hundred residues (e.g., myosine). The analysis of a total of 160 transmembrane helices of 15 non-homologous proteins resulted in an average length of 17(18) ± 2(3) [10]. Finally, in a single α-helical motif, the length of this secondary structural element can be significantly longer,

exceeding 60 helical residues [13]. Thus, elucidation of the backbone structural features and the inherent cooperativity of helices of various lengths are of significance. Long helices could have additional structural features, they can be curved or kinked [14].

Today computational quantum chemistry (QM) is suitable to study even longer biopolymers at an acceptable level of accuracy. The great feature of a QM approach is that beside structural information (comparable to experiments), it can provide stability data on molecular conformers. Varieties of QM methods have been shown to give reliable information on the stability of foldamers and transition state structures. It has been shown previously [2], in the case of $(Ala)_8$, that the greater stability of the 4_{13}- or α-helix is due to solvation. Direct hydration with the inclusion of explicit water molecule(s) has also been reported [15]. Consequently the 3_{10}-helix must exhibit the intrinsic properties of the right-handed helical structure, since it is the most stable form when environmental effects are excluded. Several QM computations have been published on 3_{10}-helical structures [16], [17], [18] and [19]. However, these studies reported electronic energies rather than thermodynamic functions. Here we have used electron structure calculations to obtain thermodynamic properties of oligopeptides, namely $(Gly)_n$ and $(L\text{-}Ala)_n$ in their extended- and 3_{10}-helical conformers, where *n*ranges between 1 and 34. These data allow us to study the build-up of the internal H-bond network and the associated folding process as a function of the number of residues within the peptide. The thermodynamics of the folding process, particularly the associated entropy change, is expected to contain a great deal of useful information [20]. In helices, L-Ala is regarded as the most preferred (most stabilizing) residue, while Gly (beside Pro) is the one to destabilize helical motifs the most [21]. Although, the structural difference between the above two residues is extremely small (–H versus $–CH_3$), the propensity difference is very significant. The difference in helix stabilization is typically explained by hydrophobic and folding entropy effects. We will focus here on the entropy term with respect to the overall stabilization.

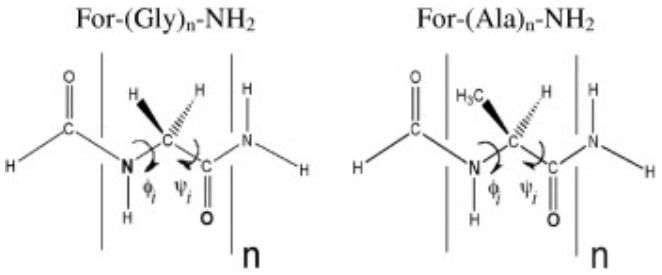
Our goal is to decipher the entropy and stability measures of a helix of various lengths. This Letter has to precede studies on more complex bundled helices (e.g., ion channels, 7TM systems) of great biochemical significance. In addition, we wanted to separate stability arising from backbone/backbone interaction, rather than from more specific backbone/side-chain interactions or from hydration.

Thus, the following specific questions were asked:

(i)Is a 3_{10}-helix more stable and compact than a single β-strand?
(ii)How do side chains and their chirality affect molecular packing and helix stability?
(iii)How accurately can the stability of a longer helix be predicted from its fragments?
(iv)What is the magnitude and location preference of subunit cooperativity in a helix?

2. Methods and computational details

The oligo- and poly(Gly) and Ala peptides used here have the following chemical structures:



where *n* stands for the number of residues.

Peptide geometries, both β-strand and 3_{10}-helical forms were determined by full optimization, using the GAUSSIAN 09 software [22] following the strict conformer selection rules published earlier [23] and [24]. The chosen levels of theory are *ab initio* Hartree–Fock and B3LYP implementation of the density functional theory. In both cases the chosen basis set was 6-31G(d). Selected structure calculations were also completed at the larger basis set 6-311+G(d,p), to assess the quality of the data. The optimized structures allow calculation of the harmonic vibration frequencies, which in turn allow determination of the thermodynamic functions (*H*°, *G*°, and *S*°). We note that the entropy is sensitive to low frequency modes which are found for long peptides. For a selected *n* the lowest frequency is found in the β-strand – both for L-Ala and for Gly.

For describing the 'folding process', thermodynamic functions and their normalized forms were calculated as follows:

equation(1)
$$\Delta H^\circ_{\beta\rightarrow\alpha} = H^\circ[(\alpha_L)_n] - H^\circ[(\beta_L)_n] \text{ and } \Delta H^\circ[n]_{\beta\rightarrow\alpha}/n$$

equation(2)
$$\Delta G^\circ_{\beta\rightarrow\alpha} = G^\circ[(\alpha_L)_n] - G^\circ[(\beta_L)_n] \text{ and } \Delta G^\circ[n]_{\beta\rightarrow\alpha}/n$$

equation(3)
$$\Delta S^\circ_{\beta\rightarrow\alpha} = S^\circ[(\alpha_L)_n] - S^\circ[(\beta_L)_n] \text{ and } \Delta S^\circ[n]_{\beta\rightarrow\alpha}/n$$

for $n \leq 34$. The subscript $\beta \rightarrow \alpha$ indicates conformational transition from the β-strand to the 3_{10}-helical conformer. We have completed geometry optimizations up to $n = 34$, where the overall length of the folded peptide is about 65 Å in its helical structure, similar to recent estimates [25]. For practical reasons the vibration frequencies were only determined for $n = 34$ in the case of Gly.

The calculations were done for the gas phase, as reliable methods for calculation of entropies in solution are computationally very demanding [26] and [27]. This restriction may not be too important if data are to mimic structure and stability of secondary structural elements shielded from the solvent (e.g., inside of a protein or a membrane).

In the world of the experimentalists the reference-state for the experimental observations favoring Ala as a helix maker is not a β-strand, which is a relatively uncommon secondary structural motif, but rather a random coil. In the context of the present calculations, this means that the reference-state should consists of a large number, *x*, of 'coil' conformations in thermal equilibrium, the large number itself adding to the entropy as $R \cdot ln\ x$. Since in the present case a single conformation, the β-strand, was used as reference-state the computed

internal entropy cannot be compared quantitatively to the experimentalists' entropy, but qualitative comparisons can indicate trends.

3. Results

3.1. Molecular structure I: torsion angles and twisting

Molecular structures of both the β-stranded and helical structures of Ala_{10} (Figure 1) are very similar to those calculated previously [1], [16], [17], [18], [19] and [28]. The backbone folds of helices in membrane proteins [$\varphi \sim -60°$, $\psi \sim -45°$ [29] and [30] and in water soluble globular proteins [$\varphi \sim -65°$, $\psi \sim -40°$ [31] and [32] are close to each other, and are somewhat different from those of 3_{10}-helices [$\varphi \sim -68°$, $\psi \sim -18°$ [31] and [33]. As expected, the dihedral angles of the QM optimized structures (Figure S1) are similar to those of 3_{10}-helices retrieved from PDB. Values near the C-terminal end tend to deviate from the 'average' values [$\varphi < -68°$, $\psi > -18°$] (Figure S1), characteristic of the central end of the N-terminal segment of a 3_{10}-helix. Dihedral angles of the extended β-form (Figure S1) show a weak oscillation (1–2°) along the backbone, pronounced especially for B3LYP calculations. The average values of the φ and ψ angles agree well with reported experimental data.

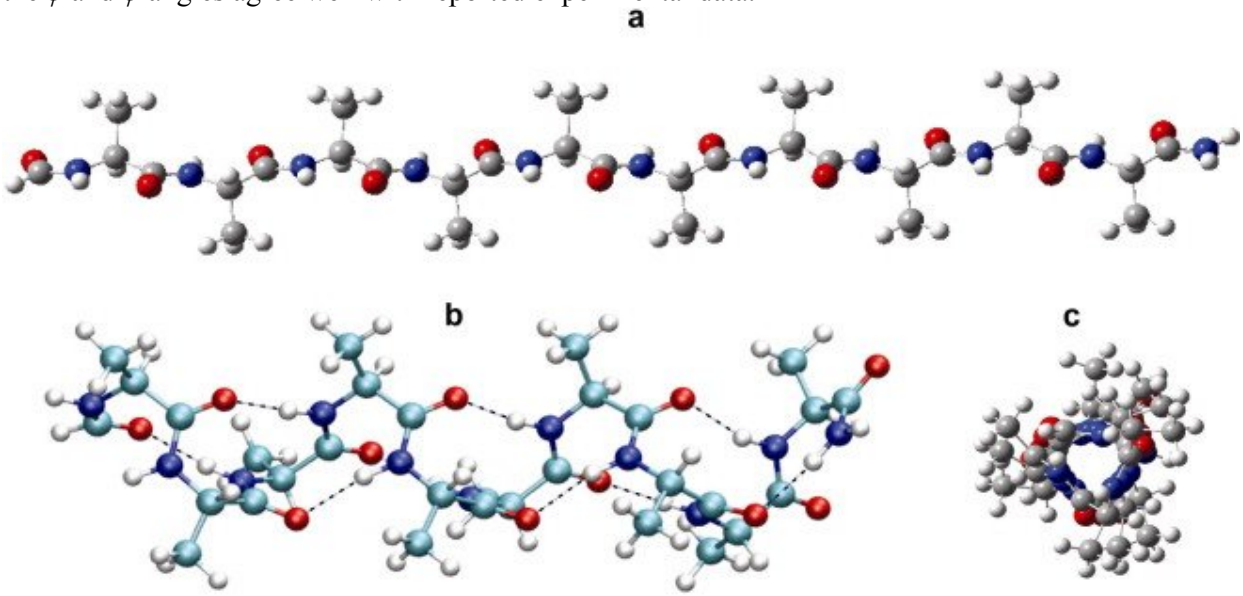


Figure 1.
The extended- (a) and the 3_{10}-helical (b and c) conformers of For-(L-Ala)$_{10}$-NH_2. Internal –CO)$_i$···(HN)$_{i+3}$– hydrogen bonds are indicated by dashed lines.

The extended conformers of both $(Ala)_n$ and $(Gly)_n$ do present an observable backbone twisting and bending not discussed here. However, the absence of a perfect flatness for $(\beta_L)_n$ due to segment cooperativity has a minor effect on the overall molecular stability and is therefore ignored for the moment. For example, at the B3LYP/6-31G(d) level of theory the difference between the electronic energy of the optimized conformer of $–Gly_{10}–$ and that of an imposed flat structure is only 0.038 kcal/mol.

The results from the preliminary Hartree–Fock calculations were included in order to show that the effect of electron correlation is relatively small.

3.2. Molecular structure II: intramolecular hydrogen bonds

The intramolecular H-bond lengths (–CO···HN–) of the 3_{10}-helices are reported here for $(Gly)_n$ and $(Ala)_n$(n = 10, 16, and 34, respectively) (Figure 2). The H-bond length varies along the backbone for both types of molecules, starting from the N-terminus. The longer H-bonds at both ends of the helices get shorter at the middle region by ~0.1 Å regardless of the theory applied. The H-bond network self-optimizes for a longer polypeptide chain: the distance, $d_{CO\cdots HN}$, gets shorter as the length of the polypeptide chain increases. The optimum value is reached for quite long systems: for n = 34 (B3LYP/6-31G(d)) $d_{CO\cdots HN}$ = 2.003 Å and 1.968 Å for $(Ala)_{34}$ and for $(Gly)_{34}$, respectively. In general, the H-bond distances at the central segment of the polypeptide chain are up to 5% shorter than those located at the N- or C-terminus. The H-bond length variation along the backbone is consistent with the red shift of the vibrational normal modes with a dominant contribution of C═O stretch relative to C═O stretch in the amino acid itself [34]. These normal modes usually have a contribution from several C═O groups that perform symmetric or asymmetric stretch vibrations. It turns out that the modes with groups close to the terminals have a smaller red shift than the other modes with a dominant C═O stretch. Clearly, a typical H-bond length is shorter in $(Gly)_n$ than it is in $(Ala)_n$, which is likely related to steric effects from the methyl group, resulting in a tighter molecular packing for $(Gly)_n$. To check if the oscillations observed for B3LYP calculations in the bond length are likely to be related to the basis set 6-31G(d), data were recalculated with a larger basis set 6-311+G(d,p) (Figure 2). This does not appear to be the case.

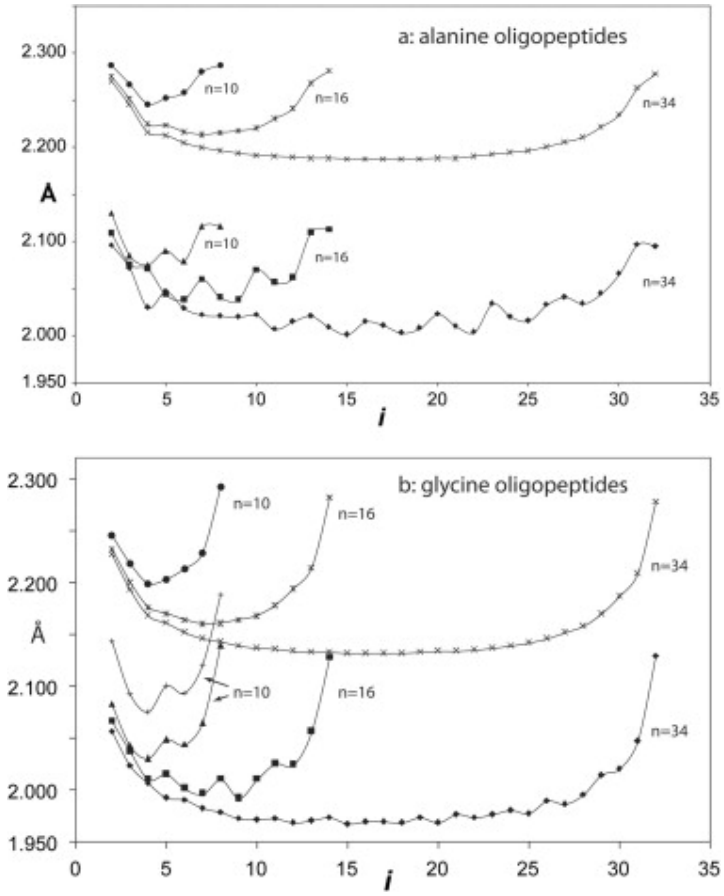


Figure 2.

$-(CO)_i\cdots(HN)_{i+3}-$ hydrogen bonds lengths in 3_{10}-helices formed by Ala (a) and Gly (b) residues at increasing residue number; *n*. Parameter *i* indicates the position of the H-bond relative to the N-terminus of the polypeptide. ♦: B3LYP/6-31G(d), $n = 34$; ■: B3LYP/6-31G(d), $n = 16$; ▲: B3LYP/6-31G(d), $n = 10$; **x**: HF/6-31G(d), $n = 34$; × HF/6-31G(d), $n = 16$; ●: HF/6-31G(d), $n = 10$; (+): B3LYP/6-311 + G(d,p)), $n = 10$.

Calculated dipole moments for $(Ala)_n$ and $(Gly)_n$ are shown in Figure 3 (B3LYP/6-31G(d)) and in Figure S2. It would be desirable to have experimental data for the dipole moment of a peptide related to the chains in our study to assess the accuracy of the calculated values. We have not been able to find such a system, so as a modest benchmark we use formamide – the building block of peptides. The experimental value of the dipole moment of formamide is 3.73 ± 0.07 D [35], while the calculated values are 3.82 D (B3LYP/6-31G(d)) and 4.10 D (HF/6-31G(d)), respectively. In the case of helical structures the difference between the HF and B3LYP data for long peptides is about 1%, whereas it is about 20% in the case of extended structures. The larger difference in the latter case is likely to be related to the differences in dihedral angles obtained by the two methods. Figure 3 compares the dipole moments of polyglycine and polyalanine. It appears that the difference between the dipole moments of the two peptides is around 1% and 5% for helical and extended structures, respectively. The biochemical significance of this observation is that side chain effects on the electron distribution are larger on the extended structure than on the helix. In other words, if this is generally true, the helix structures are almost the same no matter what the side chains

are. This may foreshadow the possibility that, in the biochemical sense, all helices respond in an analogous way.

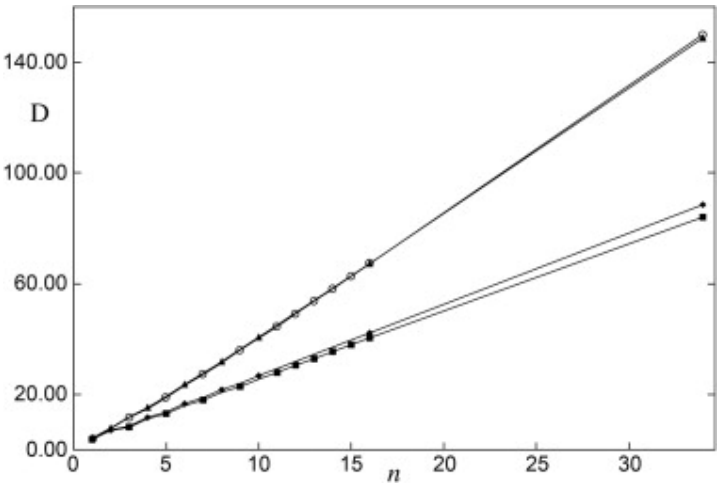


Figure 3.
Dipole moments of the polypeptides For-$(Ala)_n$-NH_2 and For-$(Gly)_n$-NH_2 as a function of the number of residues, n. The data are calculated at the B3LYP/6-31G(d) level. ○: 3_{10}-helical $(Ala)_n$; ▲: 3_{10}-helical $(Gly)_n$; ♦: extended $(Gly)_n$; ■: extended $(Ala)_n$.

4. Molecular stability

In terms of electronic energy, ΔE, Wieczorek and Dannenberg [18] and others [2], [16], [17] and [19] have shown that isolated helices seem to be more stable over β-strands; $\Delta E_{\beta\to\alpha} < 0$. Both for $(Ala)_n$ or $(Gly)_n$ at$n$ = 10–15, but also for peptides composed of residues of more complex side chains (e.g., Vla, Leu, Ile), a helical secondary structure acquires a considerable stability ($\Delta E_{\beta\to\alpha} < 25$–$35$ kcal/mol). The overall stability is proportional to the number of intramolecular H-bonds and thus to n, $\Delta E_{\beta\to\alpha}(n)$, the number of residues forming the polypeptide chain. Based on $\Delta E_{\beta\to\alpha} < 0$, a helix is an intrinsically stable foldamer, with respect to the $(\beta_L)_n$ backbone structure. As the stability difference between the folded and unfolded forms of a globular protein of average size is about 10–20 kcal/mol, the seemingly too high energy difference of the QM calculated unfolded and folded forms of an α-helix is perturbing. This discrepancy is often claimed as QM calculations neglect solvent effect. This argument holds however, some misleading perspective, as the folded protein's backbone is generally only weakly hydrated, or totally un-hydrated. In our view, the source of the problem is that instead of ΔE the free energy difference, $\Delta G°$, should be used (T – room temperature and$P$ – constant). Therefore, we use here the backbone conformation dependent thermodynamic functions ($H°$,$G°$ and $S°$), as well as their single residue normalized forms to explain the above challenge. We have characterized the folding process in terms of $\Delta X_{\beta\to\alpha}$ ($X = E$, $H°$, $S°$ and $G°$) (1)–(3) as function of the number of amino acid residues, n, both for $(Gly)_n$ and $(L\text{-}Ala)_n$ (Table 1 and Figure 4 and S3.)

Table 1.
Calculated standard state thermodynamic functions associated with the (β → α) conformational interconversion of For-$(Gly)_n$-NH_2 and For-$(L\text{-}Ala)_n$-NH_2 along with the change in electronic energy, $\Delta E_{\beta\to\alpha}$. Data were calculated at $T = 25$ °C using

B3LYP/6-31G(d) and HF/6-31G(d) levels of theory with the latter data given in parenthesis.

	n	$\Delta E_{\beta\rightarrow\alpha}$ (kcal/mol)	$\Delta H°_{\beta\rightarrow\alpha}$ (kcal/mol)	$\Delta S°_{\beta\rightarrow\alpha}$ (cal/mol K)	$\Delta G°_{\beta\rightarrow\alpha}$ (kcal/mol)
For-(Gly)$_n$-NH$_2$	1	2.01 (0.45)	2.22 (0.87)	−2.80 (−2.73)	3.05 (1.69)
	2	1.29 (1.98)	1.80 (2.45)	−9.06 (−6.39)	4.50 (4.35)
	3	1.08 (1.81)	1.78 (2.45)	−14.21 (−10.45)	6.02 (5.57)
	4	−0.31 (0.56)	0.69 (1.44)	−21.35 (−15.40)	7.05 (6.03)
	5	−2.01 (−1.06)	−0.79 (0.07)	−28.94 (−20.61)	7.84 (6.22)
	6	−4.02 (−2.96)	−2.52 (−1.58)	−32.92 (−25.92)	7.29 (6.15)
	7	−6.39 (−5.13)	−4.57 (−3.48)	−39.19 (−31.52)	7.12 (5.92)
	8	−8.88 (−7.47)	−6.78 (−5.55)	−45.00 (−37.16)	6.64 (5.53)
	9	−11.49 (−9.92)	−9.16 (−7.74)	−51.22 (−42.86)	6.12 (5.04)
	10	−14.25 (−12.49)	−11.63 (−10.03)	−57.73 (−48.63)	5.58 (4.47)
	11	−17.02 (−15.14)	−14.26 (−12.41)	−61.94 (−54.39)	4.21 (3.81)
	12	−19.99 (−17.84)	−16.85 (−14.84)	−70.98 (−60.19)	4.32 (3.11)
	16	−32.02 (−29.10)	−27.89 (−24.99)	−97.05 (−83.63)	1.05 (−0.06)
	34	−89.57 (−82.58)	−81.44 (−73.47)	−209.49 (−189.80)	−18.99 (−16.88)
For-(L-Ala)$_n$-NH$_2$	1	1.80 (2.11)	1.88 (2.21)	0.31 (1.25)	1.79 (1.84)
	2	0.86 (1.52)	1.10 (1.79)	−3.08 (−2.58)	2.02 (2.55)
	3	0.20 (0.87)	0.49 (1.24)	−6.45 (−6.26)	2.41 (3.11)
	4	−1.40 (−0.67)	−0.83 (−0.16)	−12.08 (−9.62)	2.77 (2.71)
	5	−3.27 (−2.59)	−2.60 (−1.93)	−15.01 (−13.25)	1.88 (2.02)
	6	−5.38 (−4.75)	−4.53 (−3.93)	−19.57 (−17.01)	1.31 (1.14)
	7	−7.77 (−7.17)	−6.86 (−6.17)	−21.48 (−20.94)	−0.45 (0.07)
	8	−10.28 (−9.76)	−9.20 (−8.60)	−27.35 (−24.92)	−1.05 (−1.16)
	9	−13.01 (−12.45)	−11.72 (−11.11)	−31.06 (−28.95)	−2.46 (−2.48)
	10	−15.81 (−15.26)	−14.38 (−13.74)	−33.74 (−33.05)	−4.32 (−3.88)
	11	−18.69 (−18.15)	−17.09 (−16.45)	−38.44 (−37.17)	−5.64 (−5.37)
	12	−21.64 (−21.09)	−19.82 (−19.21)	−43.90 (−41.32)	−6.73 (−6.89)
	16	−33.92 (−33.30)	−31.43 (−30.68)	−60.60 (−58.06)	−13.36 (−13.37)
	34	−92.28 (−91.14)			

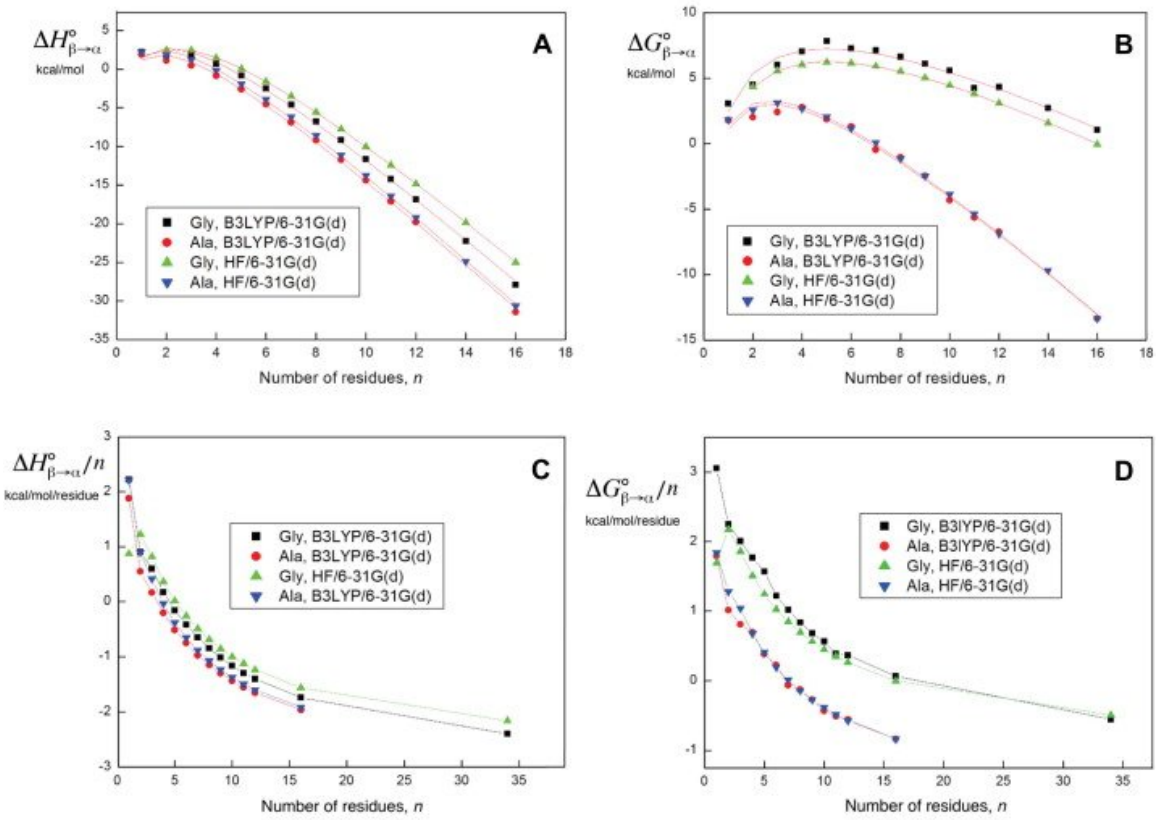


Figure 4.
Thermodynamic functions for the folding process (1) and (2) for For-$(Gly)_n$-NH_2 and For-$(Ala)_n$-NH_2 as function of the number of residues, n, at two levels of theory. (A) The enthalpy function $\Delta H°_{\beta\to\alpha}$ (1). (B) The free energy function, $\Delta G°_{\beta\to\alpha}$(2). (C) The enthalpy function per residue $\Delta H°_{\beta\to\alpha}/n$. (D) The free energy function per residue, $\Delta G°_{\beta\to\alpha}/n$.

A preliminary study on $\Delta S°_{\beta\to\alpha}$ has been reported earlier [36] along with the fitting function with three constants (c_1, c_2 and c_3), as follows:
equation(4)

$$c_1 + c_2 \ln(n) + c_3 n.$$

The justification for the above analytical form used for entropy is based on the simplifying features of the ideal gas model. It turns out that (4) also provides a good fit to $\Delta H°_{\beta\to\alpha}$, $\Delta G°_{\beta\to\alpha}$ and $\Delta E_{\beta\to\alpha}$ despite the absence of an obvious theoretical justification. The c_1, c_2, and c_3 parameters were obtained by fitting (4) to the data (Table 2). The constant c_1 is the intercept on the vertical axis associated with the thermodynamic function at hand. The logarithmic term c_2 reflects to the first phase of foldamer stabilization (Figure 4), namely the addition of a new residue to a helix still largely increases the normalized stability $\Delta G°[n]_{\beta\to\alpha}/n$. This signals a structural rearrangement and a significant amount of cooperativity between adjacent helical residues. In contrast, the linear term c_3 reflects to the second phase of foldamer stabilization, where the addition of an extra residue to the formed helix changes only marginally the normalized stability of it, and thus the foldamer is long

enough so that no major rearrangement is expected between residues. Therefore, cooperativity between residues seems marginal at this stage.

(i)Both for $(Ala)_n$ and $(Gly)_n$ the $\Delta E_{\beta\to\alpha}(n)$ and $\Delta H°_{\beta\to\alpha}(n)$ functions are positive below $n < 3$ (and 4), respectively (Figures S3 and S4). Beyond the 'zero- or crossing-points' their values are negative, indicating that at both levels of theory, the helical conformer is favored over $(\beta_L)_n$ for peptides longer than five residues. By analyzing the normalized enthalpy function, $\Delta H°_{\beta\to\alpha}(n)/n$, for smaller helices ($n < 8$) the above described cooperativity between residues is prevalent. However, for longer oligopeptides ($n > 12$) a 'linear phase' of the helical build-up is reached [17]. Although, the elongation of the polypeptide chain further increases the helix stability, this happens virtually in a 'linear way', easy to predict.

(ii)The analysis of the folding entropy function, $\Delta S°(n)_{\beta\to\alpha}$, is more informative via its normalized form: $\Delta S°_{\beta\to\alpha}(n)/n$. At small n a sign of cooperativity is seen. Thus, for shorter polypetides, in the 'build up phase' of a 3_{10}-helix ($n < 5$) cooperativity is conspicuous, as when the backbone folds into a helix its complexity increases. The shortest 3_{10}-helix is characterized by at least two consecutive hydrogen bonds formed between the main-chain CO of residue i and NH of residue $(i + 3)$ [33]. However, passing beyond this length cooperativity drops, entropy and complexity changes quasi linearly. The uptake of a new residue to the helix results in just the elongation of the foldamer. In this latter phase, complexity stays almost constant and at the DFT level of theory we find $\Delta S°_{\beta\to\alpha}/n \sim -6.37 \pm 0.07$ cal/(mol K) for Gly and -4.26 ± 0.16 cal/(mol K) for Ala. Clearly, molecular packing measured by entropy is characteristic of the amino acid composition of the polypeptide, as for $(Ala)_n$ it is smaller, $-T(\Delta S°_{\beta\to\alpha})/n \sim 1.27 \pm 0.05$ kcal/mol, but 50% larger for $(Gly)_n$; $-T(\Delta S°_{\beta\to\alpha})/n \sim 1.90 \pm 0.02$ kcal/mol. Between these two small side chains, comprising Gly and Ala the difference is considerably large; $\sim 0.63 \pm 0.07$ kcal/mol per residue. At the HF level the difference is 0.55 ± 0.03 kcal/mol per residue. This signals that the helix without a side chain is packed tighter and its loosening arises due to the presence of space requiring side chain atoms.

(iii)The function $\Delta G°(n)_{\beta\to\alpha}$ initially increases both for $(Ala)_n$ and $(Gly)_n$. This shows that in contrast to $\Delta E_{\beta\to\alpha}$ (and $\Delta H°_{\beta\to\alpha}$), the likelihood of a helical fold (partial propensity) decreases. The $\Delta G°(n)_{\beta\to\alpha}$function has a maximum at around $3 \le n \le 5$ and drops afterwards (cf. Table 1). Eventually, it crosses zero, but at two very different values of n. For Ala crossing occurs at $n = 7$, but at $n = 16$ for Gly. Thus, the $\Delta G°(n)_{\beta\to\alpha}$ function gives a more realistic explanation for helix stability than $\Delta E(n)_{\beta\to\alpha}$. Helix and β-strand are in *quasi* thermal-equilibrium for shorter and medium range oligopeptides. However, for longer polypeptides the helical folds become more and more prevalent. Thus, unlike for very specific helix forming residues, shorter polypeptides ($n \le 10$) have no clear structural preferences [28]. Such a quasi thermo-neutrality has the advantage that by altering the primary sequence, the conformer dependent thermodynamic equilibrium can self-tune. The large difference between the 'zero- or crossing-point' for $(Gly)_n$ and $(Ala)_n$, (16 and 7, respectively), is also in line with the above self-tuning process. The lack of a side chain group for Gly (a unique feature between the 20 natural amino acids) allows a higher degree of self packing, and thus a larger entropy term (Figure 5) of $(Gly)_n$, which makes it less easy to reach a negative free energy change for adopting a helical backbone. In contrast, the introduction of the smallest side chain group (a-CH_3 group with the right steric configuration for Ala) makes the helical structure for $(L\text{-}Ala)_n$ less compact. Thus, a $\Delta H°$ term of about the same magnitude for both (Figure 5), but

with a considerably smaller $T\Delta S°$ term for $(Ala)_n$ results in a $\Delta G°(n)_{\beta\rightarrow\alpha}$ function with a 'zero- or crossing-point' occurring at a significantly smaller n. This explanation, on the one hand, clearly underlines the importance of the use of $\Delta G°(n)_{\beta\rightarrow\alpha}$ rather then $\Delta E(n)_{\beta\rightarrow\alpha}$. On the other hand, it explains at least partly how side chains contribute to secondary structure preferences. In conclusion, besides electrostatic and dispersive forces operating between side chains, foldamer (e.g., helix) stability is largely determined by molecular packing and/or compactness.

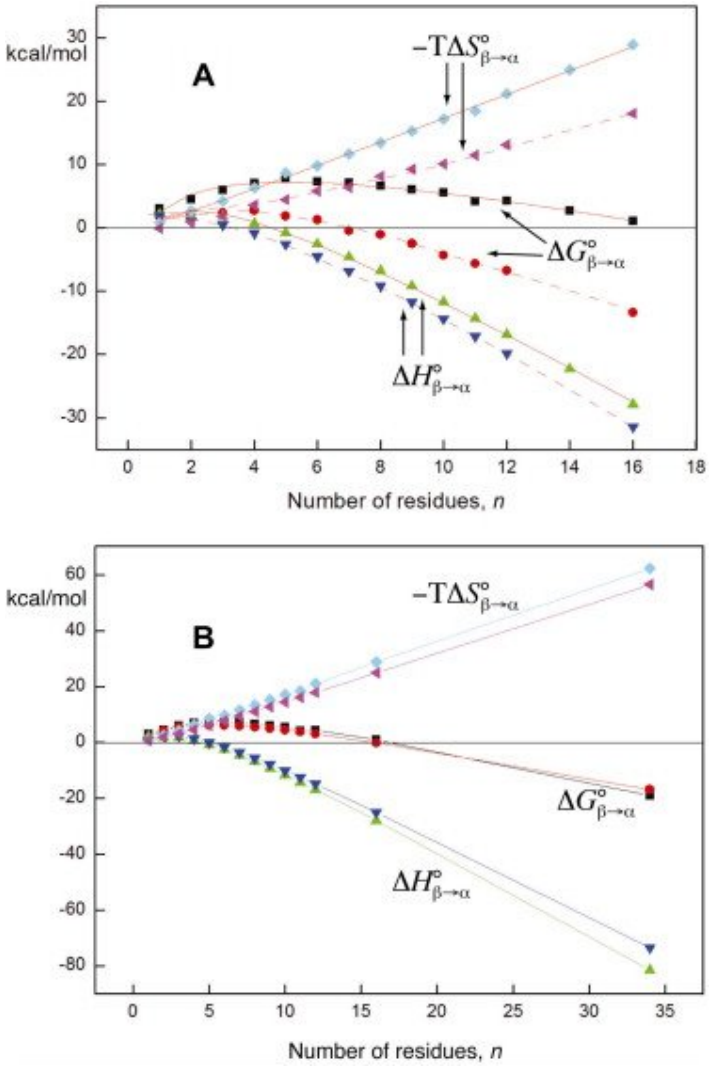


Figure 5.
(A) Thermodynamic functions, $\Delta H°_{\beta\rightarrow\alpha}$, $-T\Delta S°_{\beta\rightarrow\alpha}$ and $\Delta G°_{\beta\rightarrow\alpha}$ for For-$(Gly)_n$-$NH_2$and for For-$(L\text{-}Ala)_n$-NH_2 as function of the number of residues, n. The functions are calculated at the B3LYP/6-31G(d) level of theory. (B) Data for For-$(Gly)_n$-NH_2 obtained at the B3LYP/6-31G(d) and HF/6-31G(d) levels of theory.

Table 2.
Parameters c_1, c_2 and c_3 obtained by fitting the function (4) to the data for $\Delta E_{\beta\rightarrow\alpha}$, $\Delta H°_{\beta\rightarrow\alpha}$, $\Delta G°_{\beta\rightarrow\alpha}$ and $\Delta S°_{\beta\rightarrow\alpha}$. Data for B3LYP/6-31G(d) and HF/6-31G(d) with the latter data in parenthesis.

Function	Amino acid type	Fitted parameters (average value ± standard deviation)			
		c_1	c_2	c_3	χ^2
$\Delta E_{\beta\to\alpha}$(kcal/mol)	Gly	4.5 ± 0.4(3.33 ± 0.05)	6.9 ± 0.3(7.79 ± 0.05)	−3.47 ± 0.03(−3.38 ± 0.01)	0.24(0.001)
	L-Ala	4.5 ± 0.3(4.9 ± 0.3)	5.6 ± 0.4(5.9 ± 0.3)	−3.35 ± 0.07(−3.39 ± 0.06)	0.17(0.12)
$\Delta H°_{\beta\to\alpha}$(kcal/mol)	Gly	4.4 ± 0.4(3.35 ± 0.05)	7.0 ± 0.3(7.56 ± 0.06)	−3.24 ± 0.03(−3.08 ± 0.01)	0.28(0.001)
	L-Ala	4.4 ± 0.3(4.9 ± 0.2)	5.5 ± 0.4(5.7 ± 0.3)	−3.16 ± 0.08(−3.19 ± 0.05)	0.18(0.10)
$\Delta G°_{\beta\to\alpha}$(kcal/mol)	Gly	3.4 ± 0.4(2.47 ± 0.07)	6.7 ± 0.3(6.16 ± 0.08)	−1.35 ± 0.03(−1.22 ± 0.01)	0.30(0.002)
	L-Ala	3.1 ± 0.4(3.4 ± 0.2)	5.1 ± 0.5(5.1 ± 0.2)	−1.89 ± 0.08(−1.91 ± 0.04)	0.21(0.06)
$\Delta S°_{\beta\to\alpha}$(cal/mol K)	Gly	3.1 ± 0.8(3.4 ± 0.2)	1.1 ± 0.6(3.9 ± 0.1)	−6.37 ± 0.07(−6.09 ± 0.01)	1.15(0.04)
	L-Ala	4.4 ± 0.7(5.0 ± 0.2)	1.4 ± 0.9(1.8 ± 0.3)	−4.26 ± 0.16(−4.23 ± 0.06)	0.7(0.10)

The B3LYP method is lacking in the dealing with dispersion forces. An estimate of the effect of dispersion forces on the electronic energy can be obtained by applying Grimme's uncomplicated and highly useful correction, DFT-D3 [37] to the optimized structures at the B3LYP/6-31G(d) level. We find for a selected *n*that (i): the correction is larger (numerically) for L-Ala than for Gly and (ii): the correction is larger for the 3_{10}-helix than for the β-strand. The latter implies that $\Delta H°_{\beta\to\alpha}$ will be changed and it becomes important to investigate if the variation of $\Delta G°_{\beta\to\alpha}(n)$ with a positive region and a negative region is preserved when dispersion forces are considered. To explore this issue we did additional optimizations using the functional wB97xd (part of the G09 package) which incorporates dispersion effects. Using the basis set 6-31G(d) we find that the variation is preserved but the crossing from positive to negative values now takes place at lower values of *n*, between 3 and 4 and between 4 and 5, for L-Ala and Gly, respectively.

The possibility of extrapolating relative stability as well as the source of it for very long secondary structural elements (e.g., helices) is of significance. By deciphering (i) the contribution of cooperativity between residues, (ii) the changes of it as function of the polypeptide chain length, as well as (iii) the compactness of it fine tuned by side-chain size, configuration and chemical nature give a better insight into foldamer stability.

5. Discussion and summary

We have shown that shorter helices [$1 < n \leq 6(8)$] are nascent and show characteristics of a build up phase, while medium size or longer helices ($6(8) \leq n \leq \infty$) are matured secondary

structural elements. During the first phase, a larger cooperativity is operative between the residues, where the $\Delta G°(n)_{\beta\rightarrow\alpha}$ function goes through a maximum as the raising intramolecular H-bonds induce significant structural reorganization. On the contrary, the elongation phase can be characterized by low or insignificant cooperativity between the folding residues, $\Delta G°(n)_{\beta\rightarrow\alpha}$ decreases and approaches a phase where stability changes becomes linear as function of n. The dipole moment for neutral side chains (e.g., $–CH_3$ of Ala) affects electron distribution such that it has a larger effect on an extended than on a helical fold.

A longer 3_{10}-helix is a stable and compact secondary structural element, compared to a β-strand. However, even for the shortest helix its molecular complexity is higher than that of the appropriate $(\beta_L)_n$ structure. Therefore, conclusions derived from $\Delta E(n)_{\beta\rightarrow\alpha}$ may be misleading. Although, it is computationally demanding to calculate vibration frequencies, needed for free energies, it is recommended to use $\Delta G°(n)_{\beta\rightarrow\alpha}$, when stability issues are in focus. The C-terminus of a 3_{10}-helix is less tightly packed, compared to its N-terminus, as a larger structural distortion is typically present at the C-terminus of a helical foldamer (Figure S1). This is simple to explain as here the closing structural unit is the less compact δ_L-rather than the more 'packed' α_L-subunit [2], [16], [18] and [21]. This presumption is now fully confirmed by the residue specific $\Delta S°_{\beta\rightarrow\alpha}$ values (Table 1). The introduction of chirality requires the asymmetric substitution at the C^{α}-atoms, replacing the appropriate H^{α} by a bulkier *R*-group. In fact not the appearance of the chirality, but the insertion of a larger space-requiring *R*-groups (even a $–CH_3$) makes the helical fold less compact. The decrease of the molecular packing is significant even for the smallest Ala it reaches ~50 –70%. However, with the appearance of even bulkier side chains (e.g., Phe, Trp, Glu, Lys) no further decrease in helical packing is expected to occur, as in all proteogenic amino acid residues a $–C^{\beta}H_2–$ spacer is almost always present positioning side-chain groups at an optimum distance from the backbone atoms. This argument on entropy is in line with the experimental fact that Ala stabilizes a helical fold (Ala is preferred in helices) relative to Gly, as its entropy term is smaller (Table 1 and Figure 5) [21]. Luque et al. [38] have suggested that the backbone entropy difference upon folding into helices accounts for almost all the difference between Ala and Gly. The magnitude of this difference was quantified as –0.72 kcal/mol at 298 K, as the energy effect caused by the difference in backbone entropy changes. The numerical agreement between experimental and herein computed entropy term favoring Ala (~−0.6 ± 0.1 kcal/mol per residue) is excellent.

One can predict quite accurately the relative stability of longer secondary structural elements, via their fragments, as shown here for helices. The error is small and it is in the chemical range (~ ± 0.5 kcal/mol). Cooperativity between residues forming a single β-strand is negligible and does not change with the lengthening of this secondary structural element. In contrast, the C-terminus of a helix shows significant cooperativity, unlike residues located at the middle or at the N-terminus of the foldamer. This is clearly related to the irregular molecular packing (Figures S1 and S2) and to the above mentioned molecular distortions localized at the very same region of the computed helical structures: $–(\alpha_L)_{n\text{-}4}–(\alpha_L)3\delta L–$. The magnitude of such cooperativity is well reported by suitable structural measures (e.g., torsion angles and H-bond lengths) and by the entropy term.

The conclusion herein, however, raises a challenging problem. If $\Delta G°(n)_{\beta\rightarrow\alpha}$ monotonically decreases beyond the 'zero- or crossing-point', and thus the stability of a helix monotonically increases with n (compared to $(\beta_L)_n$), then what prevents the entire polypeptide (or protein) to

refold into a single α-helical motif? Both a charged single α-helical motif [13] and a coiled-coil structure are typically composed of very long helices (n > 50-75). If the stability of a helix increases constantly with the growing number of intramolecular H-bonds, then once a helical template is formed, how can such polypeptide escape from this thermodynamic pitfall? Or, in other words, how and at what free energy can such an ultrastabilized helical structure re- or unfold? How is it that helices in globular proteins have an average length ~18 ± 8, and they are not longer [11]? Are helix-breaker amino acid residues (Pro, Hyp, etc.) inserted along the primary sequence of helical preference for this purpose? Actually, the challenging problem mentioned above is only a puzzle in the gas phase. In aqueous solution a peptide makes strong hydrogen bonds to water and other residues in β-sheets and coil segments in tertiary structures in addition to the internal hydrogen bonds. The actual structure of the peptide is determined from all of these contributions. The present investigation only addresses the properties of a peptide model without environmental effects.

Acknowledgements

This work was supported by grants from the Hungarian Scientific Research Fund (OTKA K72973, NK101072) and TÁMOP-4.2.1.B-09/1/KMR.

References

[1]
A. Liljas, L. Liljas, J. Piskur, G. Lindblom, P. Nissen, M. Kjeldgaard
Textbook of Structural Biology, World Scientific, Publishing Co. Pte. Ltd., Singapore (2009)

[2] I.A. Topol, S.K. Burt, E. Deretey, T.-H. Tang, A. Perczel, A. Rashin, I.G. Csizmadia
J. Am. Chem. Soc., 123 (2001), p. 6054

[4] C.A. Oregano, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, J.M. Thornton
Structure, 5 (1997), p. 1093

[5] P. Tompa: **Structure and Function of Intrinsically Disordered Proteins**, Chapman and Hall/CRC (2009)

[6] R. Kiss, D. Kovács, P. Tompa, A. Perczel Biochemistry, 47 (2008), p. 6936

[14] S. Kumar, M. Bansal Biophys. J., 71 (1996), p. 1574

[15] A. Perczel, Ö. Farkas, I.G. Csizmadia J. Am. Chem. Soc., 117 (1995), p. 1653

[17] T. Beke, A. Czajlik, I.G. Csizmadia, A. Perczel Phys. Biol., 3 (2006), p. S26

[19] A. Jagielska, J. Skolnick J. Comput. Chem., 28 (2007), p. 164

[20] B. Viskolcz, S.N. Fejer, S.J. Knak Jensen, A. Perczel, I.G. Csizmadia Chem. Phys. Lett., 450 (2007), p. 123

[21] J. Lopez-Llano, L.A. Campos, J. Sancho Proteins, 64 (2006), p. 769

[24] J.M.S. Law, M. Szőri, R. Izsak, B. Penke, I.G. Csizmadia, B. Viskolcz J. Phys. Chem. A, 110 (2006), p. 6100

[26]S. Wan, R.H. Stote, M. Karplus J. Chem. Phys., 121 (2004), p. 9539

[29] J. Wang, S. Kim, F. Kovacs, T.A. Cross Protein Sci., 10 (2001), p. 2241

[30] S. Kim, T.A. Cross Biophys. J., 83 (2002), p. 2084

[31] R.C. Paga, S. Kim, T.A. Cross Structure, 16 (2008), p. 787

[32] L.J. Smith, K.A. Bolin, H. Schwalbe, M.W. MacArthur, J.M. Thornton, C.M. Dobson J. Mol. Biol., 255 (1996), p. 494

[33] P. Enkhbayar, K. Hikichi, M. Osaki, R.H. Kretsinger, N. Matsushima Proteins, 64 (2006), p. 691

[35] Handbook of Chemistry and Physics, 85th edn., CRC Press, Boca Raton, Fl, USA, 2004, pp. 9–47.

[36] B. Viskolcz, I.G. Csizmadia, S.J. Knak Jensen, A. Perczel Chem. Phys. Lett., 501 (2010), p. 30

[37] S. Grimme, J. Antony, S. Ehrlich, H. Krieg J. Chem. Phys., 132 (2010), p. 154104