J|A|C|S
ARTICLES
Published on Web 10/19/2006

Prediction of Folding Preference of 10 kDa Silk-like Proteins Using a Lego Approach and ab Initio Calculations

Gábor Pohl,[†] Tamás Beke,[†] János Borbély,[‡] and András Perczel*,[†]

Contribution from the Institute of Chemistry, Eötvös Loránd University, P.O. Box 32, 1518 Budapest 112, Hungary, and Department of Colloid and Environmental Chemistry, University of Debrecen, H-4010 Debrecen, Hungary

Received June 19, 2006; E-mail: perczel@chem.elte.hu

Abstract: Because of their great flexibility and strength resistance, both spider silks and silkworm silks are of increasing scientific and commercial interest. Despite numerous spectroscopic and theoretical studies, several structural properties at the atomic level have yet to be identified. The present theoretical investigation focuses on these issues by studying three silk-like model peptides: $(AG)_{64}$, $[(AG)_4EG]_{16}$, and $[(AG)_4PEG]_{16}$, using a Lego-type approach to construct these polypeptides. On the basis of these examples it is shown that thermoneutral isodesmic reactions and ab initio calculations provide a capable method to investigate structural properties of repetitive polypeptides. The most probable overall fold schema of these molecules with respect to the type of embedded hairpin structures were determined at the ab initio level of theory (RHF/6-311++G(d,p)//RHF/3-21G). Further on, analysis is carried out on the possible hairpin and turn regions and on their effect on the global fold. In the case of the $(AG)_{64}$ model peptide, the optimal β-sheet/turn ratio was also determined, which provided good support for experimental observations. In addition, lateral shearing of a hairpin "folding unit" was investigated at the quantum chemical level to explain the mechanical properties of spider silk. The unique mechanical characteristics of silk bio-compounds are now investigated at the atomic level.

Introduction

Biodegradable and biocompatible molecular systems composed of peptides or proteins are key issues when designing several types of new materials. The synthesis of macromolecules with equal or improved characteristics compared to silk would be of great importance to many applications. First, however, this kind of molecular "engineering" requires deciphering how these natural products show strong resistance toward stretching, combined with great elasticity and water insolubility at the atomic level.[1−3] Encoded by their diverse primary sequences, the outstanding physical properties of silk should be related to the 3D structure of their component proteins. Both the spider dragline silks and silkworm silks are therefore of scientific and commercial interest.[4−21] One of the main components of silks is silk fibroin proteins, which have amino acid sequences of great simplicity and repetition.[5,6] These fibers usually contain several multiple β-pleated sheets. Spider dragline silk, for instance, has two types of phases: a highly ordered crystalline phase consisting of multiple β-pleated sheets, and a partially aggregated phase with preformed β-sheet procrystals.[18] Thus, deciphering the folding type and preference of β-pleated sheets appearing in silk will provide useful information to better understand and improve the design of new composite biomaterial.

Previous studies showed that the primary constituents of various spider and silkworm silks are glycine (∼40%) and alanine (∼20%).[5,6] The remaining 30−40% of the fiber consists of bulkier amino acid residues such as glutamine, serine, and

* Web address: http://www.chem.elte.hu/departments/protnmr/index.html.

[†] Eötvös Loránd University.

[‡] University of Debrecen.

(1) Lewis, R. W. *Acc. Chem. Res.* **1992**, *25*, 392−398.
(2) Xu, M.; Lewis, R. W. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7120−7124.
(3) Vollrath, F. *Rev. Mol. Biotechnol.* **2000**, *74*, 67−83.
(4) Scheibel, T., Guest Ed. *Appl. Phys. A: Mater. Sci. Process*. **2006** (2), 191−376. Title: Special Issue: "Silk".
(5) Gührs, K. H.; Weisshart, K.; Grosse, F. *Rev. Mol. Biotechnol.* **2000**, *74*, 121−134.
(6) Bini, E.; Knight, D. P.; Kaplan, D. L. *J. Mol. Biol.* **2004**, *335*, 27−40.
(7) Tirrell, D. A. *Science* **1996**, *271*, 39.
(8) Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L. *Biomaterials* **2003**, *24*, 401−416.
(9) Jackson, C.; O'Brien, J. P. *Macromolecules* **1995**, *28*, 5975−5977.
(10) Elmorjani, K.; Thiévin, M.; Michon, T.; Popineau, Y.; Hallet, J.-N.; Guéguen, J. *Biochem. Biophys. Res. Commun.* **1997**, *239*, 240−246.
(11) Lewis, R. V.; Hinman, M.; Kothakota, S.; Fournier, M. J. *Protein Expression Purif.* **1996**, *7*, 400−406.
(12) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471−5475.
(13) Deguchi, Y.; Fournier, M. J.; Masoni, T. L.; Tirrell, D. A. *J. Macromol. Sci., Pure Appl. Chem. A31* **1994**, *11*, 1691−1700.
(14) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114* (2), 727−733.
(15) Marsh, R. E.; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1−33.
(16) Geddes, A. J.; Parker, K. D.; Atkins E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343−344.
(17) Termonia, Y. *Macromolecules* **1994**, *27*, 7378−7381.
(18) Simmons, A. H.; Michal, C. A.; Jelinski, L. W. *Science* **1996**, *271*, 84−87.
(19) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706.
(20) Panitch, A.; Matsuki, K.; Cantor, E. J.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1997**, *30*, 42−49.
(21) Krejchi, M. T.; Cooper, S. J.; Deguchi, Y.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1997**, *30*, 5012−5024.

10.1021/ja063933p CCC: $33.50

tyrosine.[7,8] The molecular weight of these large proteins can range from 70 up to 700 kDa.[8,9] The central and largest part of these proteins is composed of short and highly repetitive sequence units made from alanine and glycine residues, such as $(-GAGAGAS-)_n$, $(-GX-)_n$, $(-GGX-)_n$, $(-GGGX-)_n$, etc. ($1 \leq n \leq 15$ and "X" stands for G, L, Q, S, Y, or other, mostly hydrophobic residues), forming β-pleated sheets of currently unknown length and puckering.[22]

To investigate this further, numerous artificial repetitive polypeptides[10−14] were synthesized. McGrath et al.[14] studied alanylglycine-rich polypeptides composed of repeating $-(AG)_3PEG-$ units generated by recombinant DNA technology. Their basic rationale was as follows: (i) poly(glycyl-alanine) forms an antiparallel β-sheet "arrangement" in the solid state, (ii) since proline frequently initiates β-turns in proteins, it should be found at turns, (iii) furthermore, proline and glutamine are poor β-sheet formers; thus, they are involved in the reversal of the polypeptide chain, resulting in hairpin formation. Nevertheless, it was found[14] that the X-ray scattering pattern showed no direct evidence of β-sheets, and they concluded that these polymers form amorphous glasses at room temperature. They further suggested that when an odd number of amino acid residues form β-sheets with an alternating hydrogen bond pattern, strands could become incompatible with each other. Since "normal", two-residue long β-turns could be formed only at one end of the above sheets, further destabilization of the overall fold is expected, resulting in an amorphous glassy structure rather than a set of organized β-sheet structures.

The "next generation" of such model peptides[23−25] lacking proline and using the repetitive $-(AG)_kEG-$ sequences ($3 < k < 6$) were expected to have an even number of amino acid residues in their β-sheets. Thus, they could fold into a repeated antiparallel β-pleated sheet system with "regular" β-turns at both ends. Among these barely soluble polypeptides, $-(AG)_3EG-$ was subsequently analyzed and studied by mass, IR, and NMR spectroscopy, as well as by X-ray diffraction methods by Krejchi et al.[21] It was shown that the $-(AG)_3-EG-(AG)_3-EG-$ diad-like model system can form β-sheet layers. Interestingly, the absence of the diffraction signal in one direction made them rule out the presence of β-turns.[21] On the basis of modeling, they have proposed a mixture of inward and outward pointing γ-turns as backbone reversals instead of β-turns. The inward pointing γ-turn consists of a single residue, where the outward turns, also known as α-turns, are formed by three consecutive residues[26,27] Longer models such as $-(AG)_kEG-$ sequences with $k = 4$, 5, or 6 could also adopt a similar fold with longer central β-sheets. Glutamic acids must control the length of these sheets and are thus located in turns. However, there is still a lack of solid experimental evidence on the type of turn, the conformation of the glutamic acid, and the sequence of the stacking of the adjacent polypeptides in the crystal. In contrast, the structure of artificial polypeptides composed solely of L-alanylglycine, $(AG)_{64}$, was found to form several β-sheets intersected by γ-turns, which were possibly formed solely by alanine residues.[20] The latter observation suggests that the formation of an optimal number of hydrogen bonds between the strands is the main driving force of folding into multiple-stranded β-pleated sheets, which, in the case of $-(AG)_kEG-$ polypeptides, is affected by the sheet-breaking potential of glutamic acids.

It seems that the 3D structure of these models can be characterized by repeated antiparallel β-pleated sheets "interrupted" by turns of yet unknown type and nature. However, these turns play a key role in initiating the fold of silk proteins. We have investigated the turn formation of these sequences by using theoretical methods and attempted to delineate the structural consequences of turn preferences.

In the early 1990s, numerous theoretical calculations of polypeptides including model peptides with repetitive sequences were carried out to study the conformational properties of both the component β-pleated sheets and β-turns.[28] Recently, structural properties of β-turns and hairpins were studied using short model peptides.[29−31] It was shown that biologically relevant conclusions can be drawn on the basis of gas phase computations carried out on peptides. In addition, with the aid of thermoneutral isodesmic reactions,[32−34] suitable building units of silk fibroin[35] can be used to reconstruct both structure and energetic properties of macromolecules. Benson's additivity rules, one of the earliest in the field of additivity, details the enthalpies of formation, $\Delta H_f°$, and the entropy, $S°$, of organic molecules in the gas phase.[36,37] The additivity of electron density was described by Mezey et al.,[38−40] which was called the Lego approach. Furthermore, Bader et al.[41] and others[42−44] detailed the physics and energies of group additivity in several different molecules such as hydrocarbons and silanes. Recently, it was shown that such an "additivity-based" method can be used to estimate the energies of nanosized macromolcecules that are too large to be investigated directly by ab initio methods.[45−48]

(22) Winkler, S.; Kaplan, D. L. *Rev. Mol. Biotechnol.* **2000**, *74*, 85−93.
(23) Borbély, J.; Deguchi, Y. Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polym. Reprints* **1993**, *34* (1), 112.
(24) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427−1432.
(25) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471−5475.
(26) Pavone, V.; Gaeta, G.; Lombardi, A.; Nastri, F.; Maglio, O.; Isernia, C.; Saviano, M. *Biopolymers* **1996**, *38*, 705−721.
(27) Chou, K.-C. *Anal. Biochem.* **2000**, *286*, 1−16.
(28) Vásquez, M.; Némethy, G.; Scheraga, H. A. *Chem. Rev.* **1994**, *94*, 2183−2239, and the references therein.
(29) Perczel, A.; Gáspári, Z.; Csizmadia, I. G. *J. Comput. Chem.* **2005**, *26*, 1155−1168.
(30) Czinki, E.; Császár, A. G.; Perczel, A. *Chem.−Eur. J.* **2003**, *9* (5), 1182−1191.
(31) Perczel, A.; McAllister, M. A.; Császár, P.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1993**, *115*, 4849−4858.
(32) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4796−4801.
(33) George, P.; Bock, C. W.; Tracthman, M. *J. Chem. Educ.* **1988**, 61, 225−227.
(34) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986; *Pure Appl. Chem.* **1999**, *71* (10), 1919−1981.
(35) Zhou, P.; Li, G.; Shao, Z.; Pan, X.; Yu, T. *J. Phys. Chem. B* **2001**, *105*, 12469−12476.
(36) Benson, S. W.; Cruickshan, F. R.; Golden, D. M.; Hangen, G. R.; O'Neal, H. E.; Rodhers, A. S.; Shaw, R.; Wals, R. *Chem. Rev.* **1969**, *42*, 279−324.
(37) Benson, S. W. *Thermochemical Kinetics*; Wiley: New York, 1976; *J. Chem. Educ.* **1965**, *42* 502−516.
(38) Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 12423−12430.
(39) Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1994**, *116*, 12022−12032.
(40) Walker, P. D.; Mezey, P. G. *Can. J. Chem.* **1994**, *72*, 2531−2536.
(41) Bader, R. F. W.; Bayles, D. *J. Phys. Chem. A* **2000**, *104*, 5579−5589.
(42) Cortés-Guzmán, F. Bader, R. F. W. *Chem. Phys. Lett.* **2003**, *379*, 183−192.
(43) Wiberg, K. B.; Bader, R. F. W.; Lau, D. H. C. *J. Am. Chem. Soc.***1987**, *109*, 1001−10012.
(44) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.***2005**, *127*, 7215−7226.
(45) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2002**, *106*, 11791−11800.
(46) Exner, T. E.; Mezey, P. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061−4069.
(47) Szekeres, Zs.; Exner, T. E.; Mezey, P. G. *Int. J. Quantum Chem.* **2005**, *104*, 847−860.
(48) Beke, T.; Czajlik, A.; Csizmadia, I. G.; Perczel, A. *Phys. Biol.* **2006**, *3*, S26−S39.

Table 1. Backbone Conformation and Relative Stability of α-, β-, and γ-Turn-Containing Hairpin Structures of Different Model Peptides Obtained at the RHF/6-311++G(d,p)//RHF/3-21G Level of Theory

amino acid sequence	type of turn	hairpin conformation[b]	side-chain conf of Glu[c]	ΔE (kcal·mol^{-1})[d]
AGEGAGA	α (II-RS)[a]	$\beta_L\ \beta_L\ \underline{\epsilon_L\ \gamma_D}\ \alpha_L\ \gamma_D\ \beta_L$	g−, g+	0.00
	α (II-LS)	$\beta_L\ \epsilon_D\ \gamma_D\ \gamma_L\ \alpha_D\ \epsilon_L\ \beta_L$	g−, g+	8.25
GAGEGAG	α (II-LS)	$\beta_L\ \beta_L\ \underline{\epsilon_D\ \gamma_L}\ \alpha_D\ \gamma_L\ \beta_L$	g−, g+	0.00
	α (I-LS)	$\beta_L\ \beta_L\ \underline{\epsilon_L\ \alpha_D}\ \delta_D\ \epsilon_D\ \beta_L$	g+, g−	8.05
AGAGEGA	α (II-LU)	$\beta_L\ \beta_L\ \underline{\epsilon_L\ \alpha_D}\ \alpha_D\ \beta_L\ \beta_L$	g−, a	0.00
	α (II-RS)	$\beta_L\ \beta_L\ \underline{\alpha_L\ \gamma_L}\ \delta_D\ \epsilon_D\ \beta_L$	a, g−	8.97
AGEGAG	β (type II)	$\beta_L\ \beta_L\ \underline{\gamma_L\ \delta_D}\ \beta_L\ \beta_L$	g−, g+	0.00
	β (type II′)	$\beta_L\ \beta_L\ \underline{\gamma_D\ \delta_L}\ \beta_L\ \beta_L$	a, g−	10.0
GAGEGA	β (type II′)	$\beta_L\ \beta_L\ \underline{\gamma_D\ \delta_L}\ \beta_L\ \beta_L$	g+, g+	0.00
	β (type I′)	$\beta_L\ \beta_L\ \underline{\alpha_D\ \alpha_D}\ \epsilon_L\ \beta_L$	g+, g+	12.4
GEGAG	γ (classic)	$\beta_L\ \delta_L\ \underline{\gamma_D}\ \epsilon_L\ \beta_L$	g−, g−	0.00
	γ (inverse)	$\beta_L\ \delta_D\ \underline{\gamma_L}\ \epsilon_D\ \beta_L$	g−, g+	8.52
AGEGA	γ (inverse)	$\beta_L\ \gamma_D\ \underline{\gamma_L}\ \epsilon_D\ \beta_L$	g−, a	0.00
	γ (classic)	$\beta_L\ \delta_L\ \underline{\gamma_D}\ \epsilon_L\ \beta_L$	a, g−	7.20
GAGEG	γ (classic)	$\beta_L\ \delta_L\ \underline{\gamma_D}\ \epsilon_L\ \beta_L$	g−, a	0.00
	γ (inverse)	$\beta_L\ \delta_D\ \underline{\gamma_L}\ \epsilon_D\ \beta_L$	g+, a	6.78

[a] Based on nomenclature for α-turns applied by Pavone et al.[26] [b] Based on the conformation of the component subunit, cf. Perczel et al.,[50] with turns underlined. Only the highest and lowest energy conformations are shown herein. For more details see Supporting Information. [c] *anti*: a; *minus gauche*: g−; *plus gauche*: g+ [d] Energies are relative to the lowest energy structures of each sequence, $E = -2000.004015$, -1960.957296, -2000.005742, -1754.087511, -1754.083246, -1547.213490, -1508.163796, and -1508.167139 hartrees, respectively.

To provide "atomic level" insight into the structural properties of silk proteins, we use this novel method, which seems suitable for the estimation of the structural properties of repetitive polypeptides and proteins larger than 10 kDa. Based on the results, the following questions were addressed:

- Which conformational building units constitute the 3D structure of silk protein?
- Do the silk-like repetitive polypeptides such as $(AG)_{64}$ fold into a structure with a maximal number of hydrogen bonds? How does such a system select the ideal length of β-sheets and the optimal number of turns when folding? Is there a possible connection between the solvent accessibility of silk proteins and the length of the antiparallel β-sheets forming the 3D structure?
- What is the turn preference of polypeptides composed of $-(AG)_4EG-$ or $-(AG)_4PEG-$ repetitive sequences?
- What is the energy dependence of lateral shearing (opening up) of a $(GA)_3GE(GA)_3$ type model, when adopting antiparallel β-sheet and β-turn conformers?

Theoretical Basis

Computational Details and Considerations. All computations were carried out in a vacuum using the Gaussian 03 software package.[49] Model peptides with various sequences and hairpin conformations, as well as four different antiparallel β-pleated sheet assemblies, were subjected to full geometry optimization at the RHF/3-21G level of theory, followed by single-point energy calculations at the RHF/6-311++G(d,p)//RHF/3-21G level of theory (Table 1, Table 2, and Figure 1). For details see the Supporting Information.

All "byproducts", formamide, $HCO-NH_2$, and formamide dimer, $[HCO-NH_2]_2$, required by the isodesmic reactions were optimized at the RHF/3-21G level of theory with additional single-point energy calculations. To keep the structure required by the peptide model, the coordinates of all oxygen, nitrogen, and carbonyl carbon atoms were constrained in the latter dimer.

Finally, to estimate the energetic properties of lateral shearing of such polypeptides, geometry optimizations were performed on the $(GA)_3GE(GA)_3$ model peptide in an antiparallel β-sheet conformation connected through a β-turn, scanning the distance of the terminal H- - -O hydrogen bond from 1.93 Å in the relaxed state to 14.93 Å with a stepwise progression of 0.1 Å at the RHF/3-21G level of theory. Every tenth structure was extracted (e.g., structures where the distance increased by 1 Å), and single-point energy calculations were performed at the RHF/6-311++G(d,p) level of theory. To preserve an ideal lateral

Table 2. Energetic Properties of Different β-Pleated Sheet Models Obtained at the RHF/6-311++G(d,p)//RHF/3-21G Levels of Theory

type of sheet[a]	ΔE[b] (kcal·mol^{-1})
I	0.48
II	0.62
III	0.00
IV	0.41

[a] For more details see Figure 1 and Supporting Information. [b] Relative to the energy of the lowest energy β-sheet model (III) $E = -1243.580071$ hartrees.

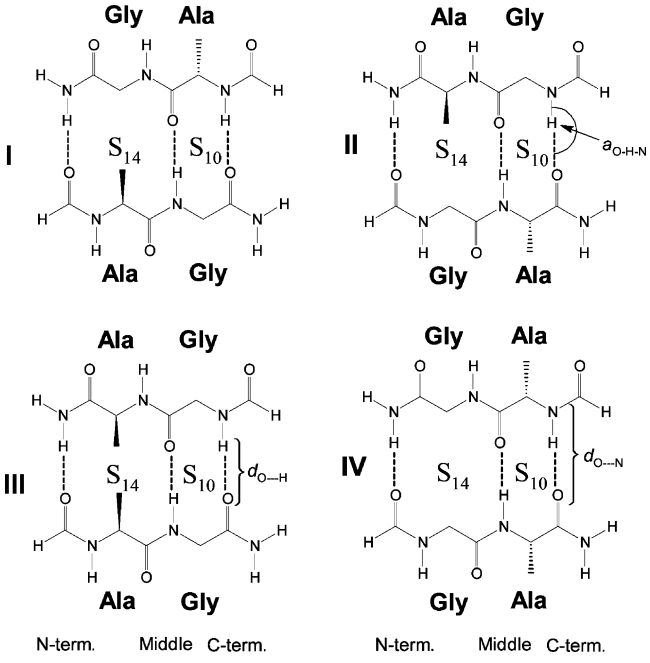


Figure 1. Dimers of peptide dimers are suitable antiparallel β-sheet building blocks for isodesmic reactions.

(49) Frisch, M. J.; et al. Gaussian, Inc.: Wallingford, CT, 2004.

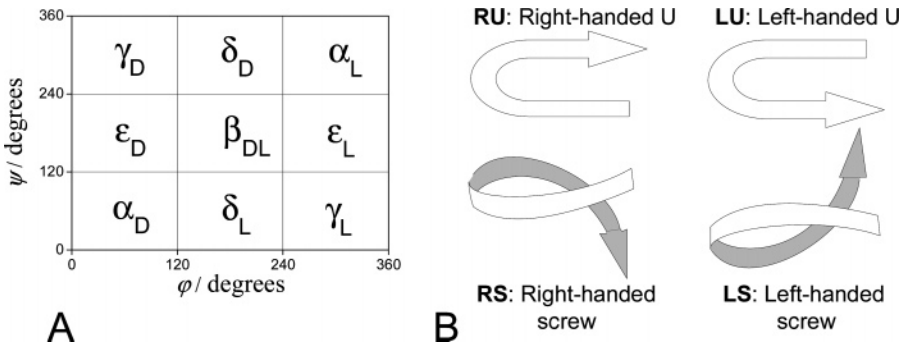
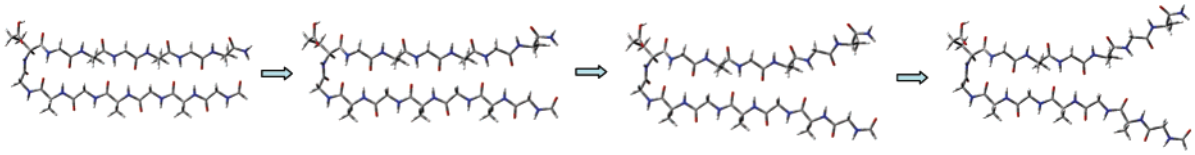


Figure 2. Nomenclature of the conformation of amino acid residues based on regions of the Ramachandran map[50] (A) and shape-based nomenclature of the different types of α-turns[26] (B).

Chart 1. Schematic View of the Lateral Shearing of $(GA)_3Ge(GA)_3$ by Increasing the Distance of the Terminal Hydrogen Bond from 2 Å to 15 Å



shearing, all backbone torsional angles of the residues in the β-sheets were constrained to the values obtained during the full optimization of the original structure (Chart 1).

Nomenclature. The nomenclatures applied in the case of antiparallel sheets S_{10} and S_{14} are based on the number of atoms involved in a pseudoring formed by two consecutive hydrogen bonds (Figure 1). The conformation of the amino acids is designated according to the regions of the Ramachandran map (Figure 2A). The symbols also represent the amino acid conformation characteristics in the most common secondary structures; for example, β_L stands for β-sheets, α_L for α-helix built from L-amino acids, etc.[50] The nomenclature of the four main types of α-turns is based on the shape of the turn (Figure 2B).[26]

Results and Discussion

Building Units of Silk-like Polypeptides. To understand both the folding properties and energetic driving force of the formation of the antiparallel β-pleated sheets of silk-like proteins, the following repetitive polypeptides were studied in detail at the ab initio level of theory: $(AG)_{64}$, $[(AG)_4EG]_{16}$, and $[(AG)_4PEG]_{16}$. The $-AG-$ and $-(AG)_kEG-$ $(3 \le k \le 6)$ repetitive sequences were characterized as antiparallel β-pleated sheets interconnected through turns,[20,21] while $-(AG)_kPEG-$ $(3 \le k \le 4)$ units were thought to form amorphous glasses.[14,23] Since the antiparallel β-sheet type backbone conformation was originally assumed for the latter sequence, its stability difference was also investigated.

To build these polypeptides, two main types of building units were obtained: various turn sequences and short antiparallel β-sheets. Since the properties of these units will determine those of the polypeptides, they are discussed in detail.

First, seven-, six-, and five-residue-long oligoglycines, G_7, G_6, and G_5, were optimized in various conformations incorporating different types of turns, namely, three-residue-long α-turns,[26,27] two-residue-long β-turns,[30,31] and one-residue-long γ-turns.[51] To mimic the hairpin-like secondary structure, both the first two and the last two amino acid residues of these oligoglycines were set to have extended-like backbone structure ($\varphi \sim 180°$ and $\psi \sim 180°$), or β_L for short,[50] with the typical interstrand H-bonds of antiparallel β-pleated sheets. The conformers that failed to form the latter antiparallel system were omitted from further analysis. Hydrogen atoms of selected prochiral centers were replaced by methyl groups to obtain the following sequences: AGAGAGA, GAGAGAG, AGAGAG, GAGAGA, AGAGA, GAGAG (Figure 3). All these models adopting multiple conformers were subsequently fully optimized. As the last step in building the model, glutamic acids and prolines were "placed into" all possible turn positions (e.g., in the case of α-turns, Glu was inserted into three different sequential positions, namely, AGEGAGA, GAGEGAG, and AGAGEGA). These final models were fully optimized (Chart 2). For glutamic acid, all backbone conformers were equipped with side-chain orientations that could form a backbone/side-chain hydrogen bond, hence lowering the total energy of the molecule. On the other hand, α- and β-turn conformations were obtained for the AGPEGAG sequence, though α-turn formation is possible only between adjacent β-sheets, where the "antiparallel" criterion is considered. Finally, single-point energy calculations were carried out on all the model conformers at the RHF/6-311++G(d,p)//RHF/3-21G level of theory (Tables 1, 3, and 4).

Although more accurate methods are already available, these are still rather expensive and would be required only for finer details beyond the goals of the present study, whereas the chosen computational approach seems sufficient for the level of the problems addressed.

Some of the conformations are slightly shifted (e.g., from ϵ_L to γ_L and from γ_D to δ_D in the case of II-RS type α-turns, etc.) from the previously determined torsion angle values,[26] due to the formation of side-chain/backbone or backbone/backbone type hydrogen bonds (Table 1). The propensity of the side-chain atoms of glutamic acid to form hydrogen bonds with carbonyl oxygen atoms is quite significant. In addition, oxygen−oxygen repulsion of the Glu side-chain and backbone carbonyls can

(50) Perczel, A.; McAllister, M. A.; Csaszar, P.; Csizmadia, I. G. *Can. J. Chem.* **1994**, *72*, 2050−2070.
(51) Vass, E.; Láng, E.; Samu, J.; Majer, Zs.; Kajtár-Peredy, M.; Mák, M.; Radics, L.; Hollósi, M. *J. Mol. Struct.* **1998**, *440*, 59−71, and references therein.

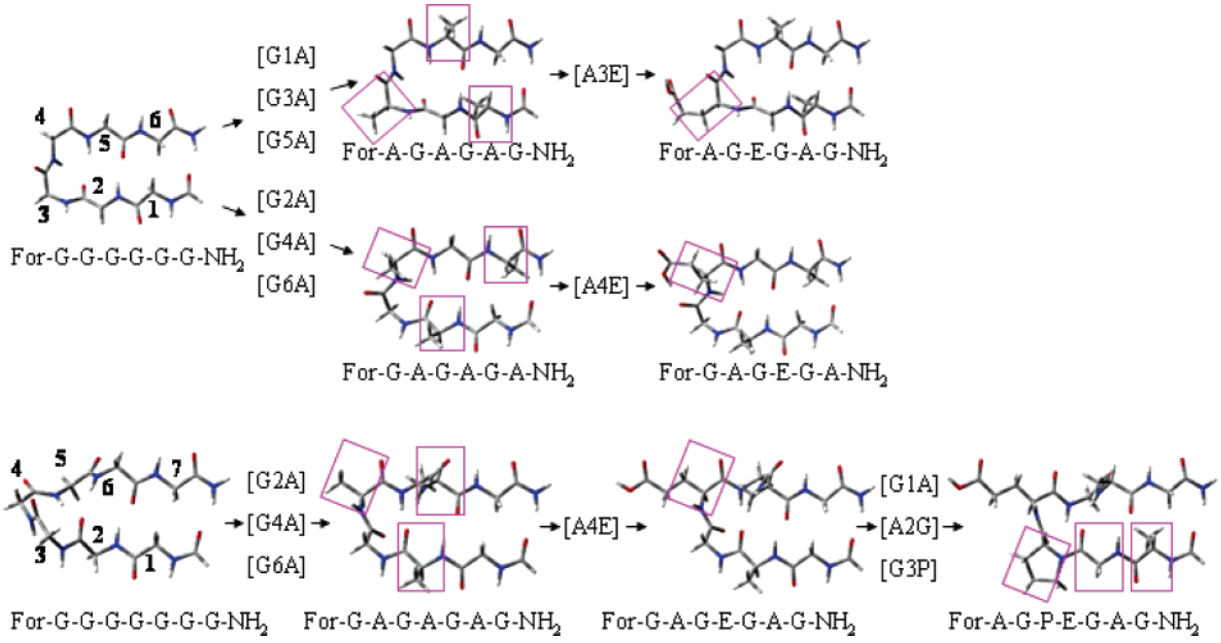
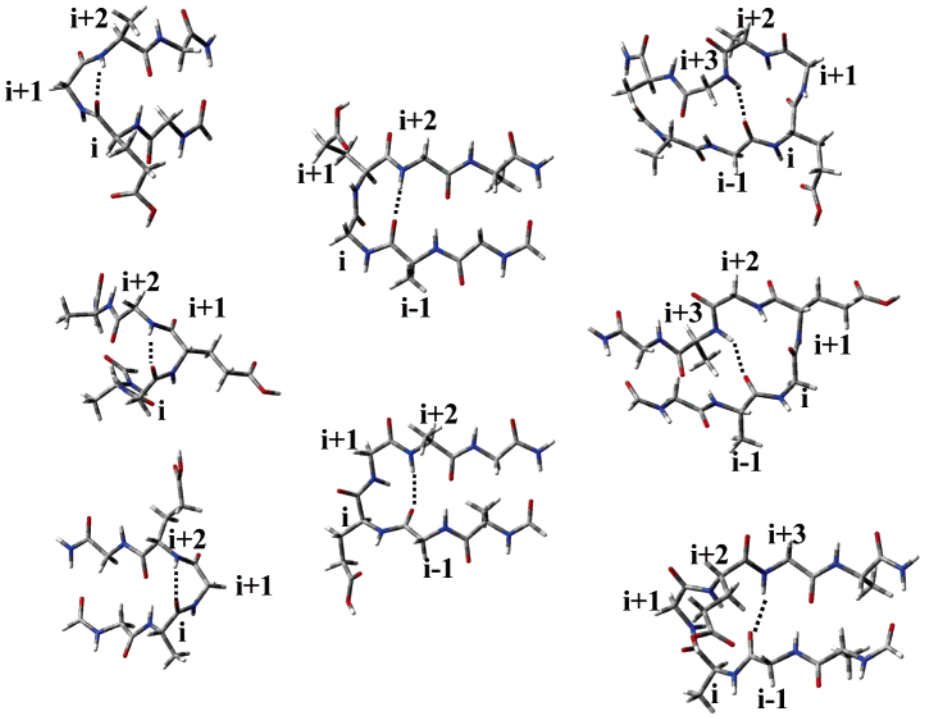


Figure 3. Schematic representation of the steps resulting in the final Glu- and Pro-containing hairpin models, beginning at the G_6 (above) and G_7 (below) model peptides.

Chart 2. Final Glu-Containing Hairpins within the Possible Types of Turns: γ-Turns (left), β-Turns (middle), and α-Turns (right) Columns[a]



[a] Note that as long as γ-turns have only one (i+1), β-turns have two (i and i+1) and α-turns have three amino acid residues (i, i+1, and i+2) in the turn region.

make selected conformers highly unfavorable. The latter effect can be as significant as an 8 kcal·mol^{-1} increase in total energy, as observed in in the case of β-turns such as GAGEGA (STable 2).

The second type of building unit is the antiparallel β-pleated sheets that are formed by covalently unattached segments of alanylglycine or glycylalanine (Figure 1). The different turn conformers can join in two distinct ways in the antiparallel β-pleated sheet subset; namely, interactions can be between two alanine residues or between alanine and glycine residues in the β-sheet. A total of four different structures can be constructed from Ala and Gly, using noncovalently attached tetrapeptides when forming an antiparallel β-pleated sheet with the latter (Table 2).

Interestingly, models that have alanine interacting with alanine and glycine with glycine (types III and IV) are slightly more stable than those of "mixed" types (alanine and glycine). Since the hydrogen bond parameters of the latter β-sheet models are

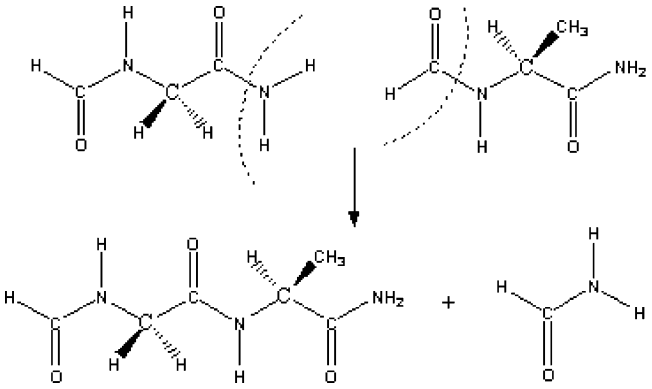


Figure 4. Thermoneutral isodesmic reaction, building up a HCO-Ala-Ala-NH_2 from two HCO-Ala-NH_2 and resulting in formamide, HCO−NH_2, as byproduct.

similar, the calculated ∼0.5 kcal·mol^{-1} stabilization energy difference as a function of the polypeptide sequence can be attributed to the van der Waals interactions between the methyl side chains. This small energetic difference suggests that the type of β-sheets formed has only a minor effect on the folding preference of the polypeptides in question. For more details on the building units, see the Supporting Information.

Thermoneutral Isodesmic Reactions. Since polypeptides larger than 11 kDa are far beyond the limits of current ab initio calculations, the stability of different selected backbone conformers of $(AG)_{64}$, $[(AG)_4EG]_{16}$, and $[(AG)_4PEG]_{16}$ type model peptides was investigated using suitable thermoneutral isodesmic reactions.

Isodesmic (ID) reactions involve chemical transformations with the condition that the same number of a given bond type must be present in both the reactants and products, as illustrated by the generalized equation (Figure 4)

$$\underset{\text{reactant}}{\text{A−a}} + \underset{\text{reactant}}{\text{b−B}} \rightarrow \underset{\text{product}}{\text{A−B}} + \underset{\text{byproduct}}{\text{a−b}} \qquad (1)$$

Thus, to obtain the total energy of the product, the following equation should be used:

$$E_{\text{A−B}} = \{E[\text{A−a}] + E[\text{b−B}]\} - \{E[\text{a−b}]\} \qquad (2)$$

As detailed earlier,[48] the stability properties of polypeptides with repetitive sequences can be estimated with relatively high confidence using thermoneutral isodesmic reactions and considering only a few different building units at a given time. This is in contrast to the energetic properties of systems with complex and multiple interactions, such as globular proteins where side-chain interactions have a major contribution to the overall fold of the protein. The reconstruction of these complex systems is thus much more difficult since many unique building units with various conformations need to be considered, and the protein cannot be built up from multiple similar small fragments.

The polypeptides under examination in this study consist of repetitive fragments. There are 16 fragments in both $[(AG)_4EG]_{16}$ and $[(AG)_4PEG]_{16}$, and their reconstruction will be discussed below.

The total energies of these polypeptides can be easily reconstructed if an $(AG)_3EG(AG)_4EGAG$ molecule is considered first. This consists of two six-residue hairpins (AGEGAG) and two antiparallel β-pleated sheet fragments ($[AG]_2$−$[AG]_2$) (Figure 1, Chart 2, and Figure 5A). When this molecule is a fragment used to build a longer structure, the hydrogen bonds

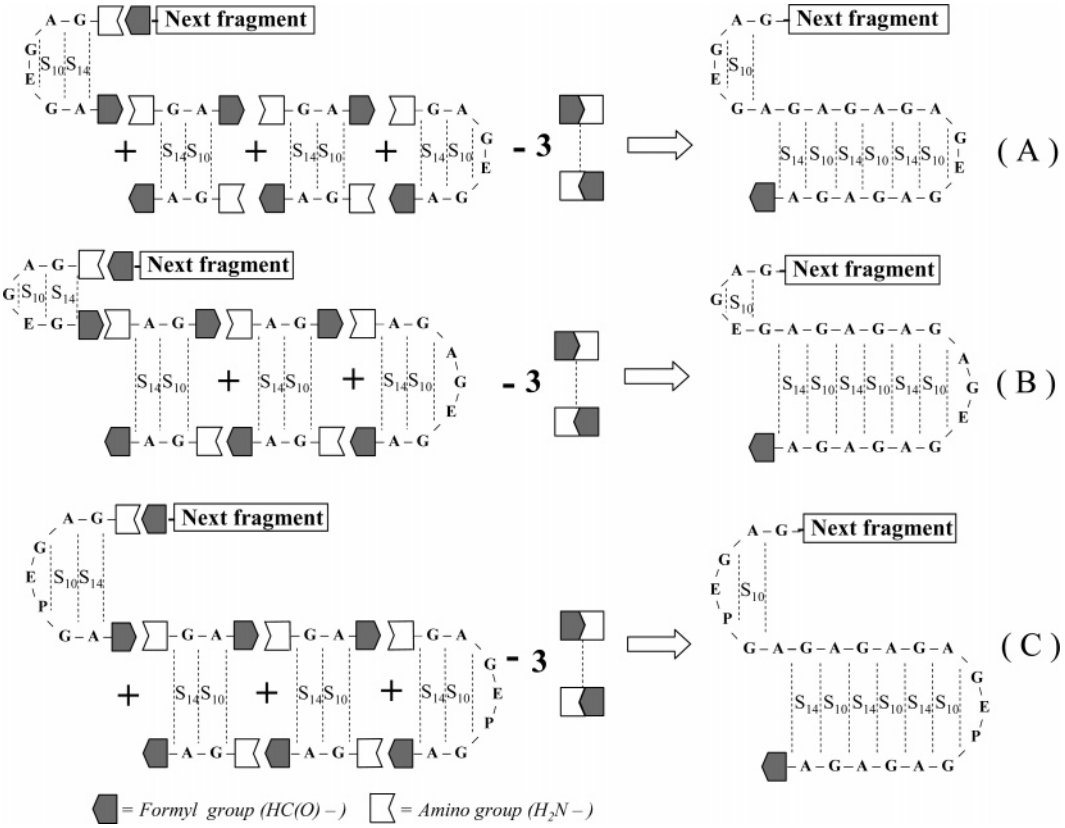


Figure 5. Building scheme for the $[(AG)_4EG]_{16}$ model peptide using hexapeptide turns (A); mixed hepta- and pentapeptide turns (B); model peptide $[(AG)_4PEG]_{16}$ using proline-containing turns (C). For more details on other sequences, see the Supporting Information.

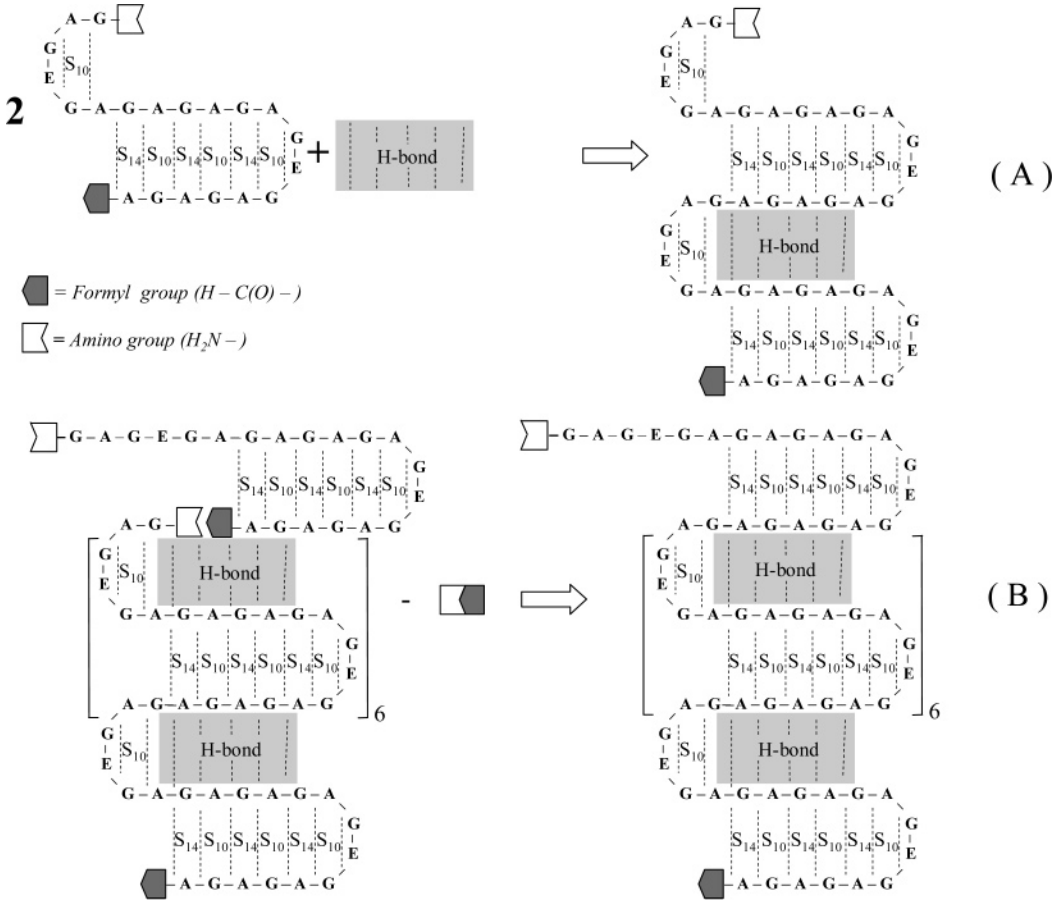


Figure 6. Building the whole polypeptide $[(AG)_4EG]_{16}$. (A) Addition of interactions between the "middle" fragments: the hydrogen bonds which appear between the sheets (H-bonds). (B) Multiplication of the middle fragment and addition of the final fragment in order to obtain the correct stoichiometry for the model peptide $[(AG)_4EG]_{16}$.

formed with the incoming fragment must be taken into account. Note that due to the solvent accessibility of the outward pointing amide hydrogens and carbonyl oxygens in the hairpins, as well as the unfavorable conformations of several hairpins, the two terminal hydrogen bonds that could appear between adjacent hairpins are not considered in our model (Figure 6A). Finally, these fragments must be multiplied and the final fragment that does not have a turn structure in the C-terminal must be added. This fragment still contains enough amino acid residues to model the original composition of the polypeptide (Figure 6B). For more details and equations of the applied thermoneutral isodesmic reactions see the Supporting Information.

Stability, Strength, and Solubility of Silk Proteins as a Function of the Pleated Sheet Length. As previously observed, most silk proteins form structures with antiparallel β-pleated sheets interconnected by hairpins. However, varying the number of amino acids involved in the sheet regions, i.e., the length of the repeating sheet units, could result in different numbers of intramolecular hydrogen bonds, which influences the stability and tensile strength of the polypeptide. Additinally, the number of hydrogen bond-forming donor and acceptor N−H and C═O groups that could interact with the solvent (H_2O) decreases, which would result in decreased solubility. Thus, these three properties are highly dependent on the length of the sheet regions. To demonstrate the relationship between these properties and the preferred structures, we chose the $(AG)_{64}$ polypeptide, which has been thoroughly investigated.[20] On the basis of X-ray diffraction data, the $(AG)_{64}$ polypeptide is reported to fold through similar structures as the $-(AG)_xEG-$ $(3 \le x \le 6)$ type polypeptides.[20] The driving force of this process could be the formation of an optimal number of hydrogen bonds that stabilize the structure.

Different folds of $(AG)_{64}$ were built using thermoneutral isodesmic reactions with the same type of β-turns in the hairpins, but with an increasing number of sheet-building units (denoted by n_S). The antiparallel β-sheet model I (Figure 1 and Figure 7A) was used as the sheet-building unit. Distribution of the relative energies of these different folds gives a parabolic curve, with a well-defined low-energy region ($4 \le n_S \le 7$). This suggests that the fold of polypeptides with repetitive sequences is highly coordinated (Figure 7B). According to X-ray data, these lamellae are 3−4 nm thick,[20] which would suggest that $n_S \approx 2$ or 3. Although our theoretical results do not quantitatively match the experimental data, there are several considerations that make this difference reasonable: (i) The n_S derived from experimental observations is only our rough approximation, based on the proposed length of the lamellae. (ii) The stabilizing effect of hydrogen bonds that could appear between the neighboring hairpins is not considered in our model. Turn regions are the most flexible parts of these artificial polypeptides and are the most accessible to solvent molecules. Thus, the hydrogen bond formation in these regions cannot be considered due to lack of factual information. (iii) It is possible that the application of higher level of theory or electron correlation methods could provide better energies for H-bonds, which would shift the n_S value toward the experimental observation. (iv) Solvent effects

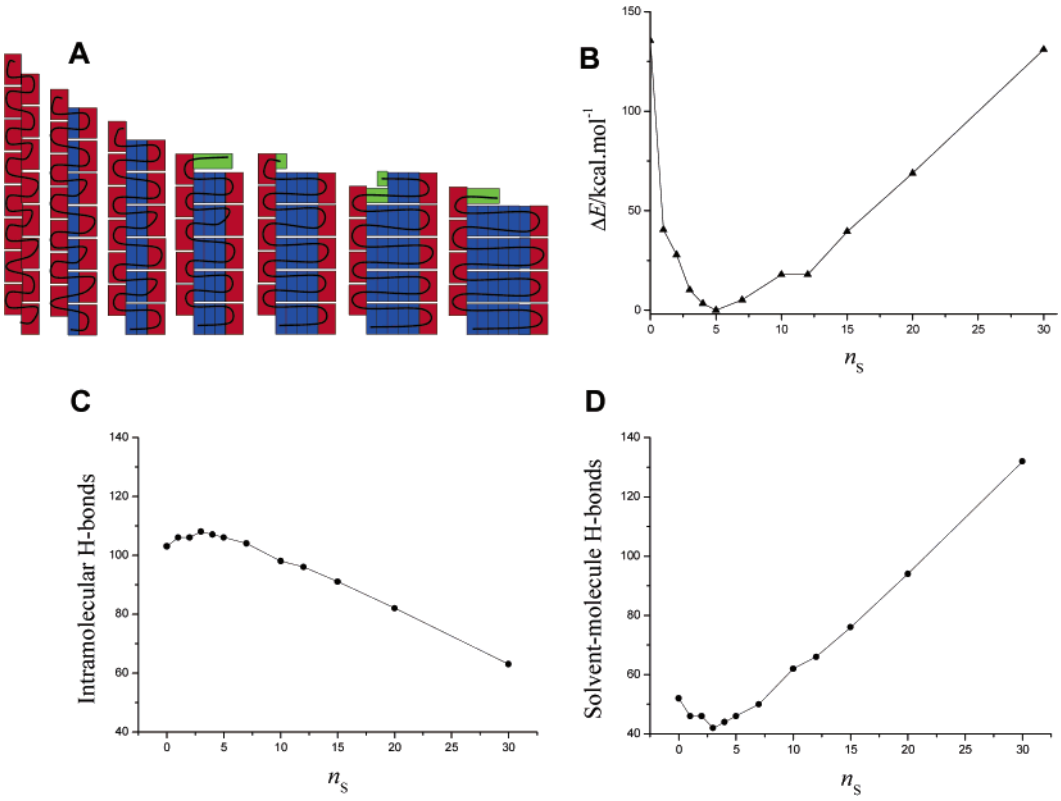


Figure 7. (A) Different $(AG)_{64}$ folds built up from hairpin units (red blocks), increasing number of antiparallel β-sheet units, n_S (blue blocks), and leftover extended strands (green blocks). (B) Relative energies as a function of the number of sheet-building units appearing between two sequential hairpins (n_S). (C) The number of intramolecular hydrogen bonds and (D) the number of supposed solvent−molecule hydrogen bonds as the function of n_s.

Table 3. Ab Initio (RHF/6-311++G(d,p)//RHF/3-21G) Stabilities of Selected Conformers of the $[(AG)_4EG]_{16}$ Polypeptide Obtained Using Isodesmic Reactions

type of turn (α, β, or γ)[a]	hairpin conformation[b]	Glu[c] side-chain conf	ΔE (kcal·mol^{-1})[d]
	Hairpin Sequence AGEGAG		
β (type II)	$\beta_L\ \beta_L\ \underline{\gamma_L\ \delta_D}\ \beta_L\ \beta_L$	g−, g+	−302.4
	Hairpin Sequence GAGEGA		
β (type II′)	$\beta_L\ \beta_L\ \underline{\gamma_D\ \delta_L}\ \beta_L\ \beta_L$	g+, g+	−259.0
	Hairpin Sequences AGEGA and GAGEGAG		
γ (inverse)	$\beta_L\ \gamma_D\ \underline{\gamma_L\ \epsilon_D}\ \beta_L$	g−, a	−283.69
α (II-LS)	$\beta_L\ \beta_L\ \underline{\epsilon_D\ \gamma_L\ \alpha_D}\ \gamma_L\ \beta_L$	g−, g+	
	Hairpin Sequences GAGEG and AGAGEGA		
γ (classical)	$\beta_L\ \delta_L\ \underline{\gamma_D\ \epsilon_L}\ \beta_L$	g−, a	−292.85
α (I-LU)	$\beta_L\ \beta_L\ \underline{\epsilon_L\ \alpha_D\ \alpha_D}\ \beta_L\ \beta_L$	g−, a	
	Hairpin Sequences GEGAG and AGEGAGA		
γ (classical) α (II-RS)	$\beta_L\ \delta_L\ \underline{\gamma_D\ \epsilon_L}\ \beta_L$	g−, g−	−299.09
	$\beta_L\ \beta_L\ \underline{\epsilon_L\ \gamma_D\ \alpha_L}\ \gamma_D\ \beta_L$	g−, g+	
	Unfolded (Reference) Structure		
extended	$\beta_L\ \beta_L\ \beta_L\ \beta_L\ \beta_L\ \beta_L$	a, a	0.0

[a] Subtypes of turns in parentheses. [b] Based on the conformation of the component subunit, cf. Perczel et al.,[50] with turns underlined. [c] *anti*: a; *minus gauche*: g−; *plus gauche*: g+ [d] Energies are relative to the extended structure of the sequence, $E = -40019.823393$ hartrees.

are not considered in current model, and explicit H-bonds to the solvent would definitely change the conclusions. Nevertheless, the observation of greatest importance is the trend in the β-sheet length preference. Finally, the same region ($1 \le n_S \le 5$) has the most intramolecular hydrogen bonds (Figure 7C), which suggests that the preferred folds are less soluble (Figure 7D).

Stability of the Differently Folded $[-(AG)_4EG-]_{16}$ Polypeptide. As discussed, the total energies of molecular systems that are too large for "direct" ab initio calculations can be reconstructed using suitable building units and thermoneutral isodesmic reactions (Figures 4−6 and Table 3). The accuracy of such isodesmic reactions was found to be acceptable[48] since each of the building units and isodesmic reactions were similar. Thus far, five main types of backbone folds were distinguished. Two contain hairpins with β-turns, AGEGAG and GAGEGA, with the glutamic acids at the (i+1) and (i+2) position connected to antiparallel β-sheets (form I and II), respectively (Figures 1,

5A and Chart 2). In the remaining three folds, the antiparallel β-pleated sheets are connected by either an α- or γ-turn. Three pairs of α- and γ-turns were considered as a function of the amino acid sequence of the polypeptide, namely, AGEGAGA−GEGAG; GAGEGAG−AGEGA; and AGAGEGA−GAGEG (Figure 5B). The relative positions of Glu were also different for structures with β-turns. Although several conformers exist where only the conformation of the Glu side-chain is different, only hairpins with the most stable α-, β-, and γ-turns are reported for simplicity. The fully extended structure of the polypeptide, i.e., with all backbone torsional angles *anti*, was used as a reference (Table 1).

When comparing the energies of the different foldamers relative to a fully extended or unfolded conformer, the energy difference gained by the formation of a typical hydrogen bond network is significant. Stabilization energy can be greater than 300 kcal·mol^{-1}, as in the case of the most stable structure composed of type II β-turns and β-sheets ($\Delta E^{\text{type II }\beta} = -302.4$ kcal·mol^{-1}). In accordance with the X-ray data,[21] each of the three structures with alternating turns have relative energies close to that of the most stable conformer: $\Delta E^{\gamma\ (\text{inverse})+\alpha(\text{II}-\text{LS})} = -283.69$ kcal·mol^{-1}, $\Delta E^{\gamma(\text{classical})+\alpha(\text{I}-\text{LU})} = -292.85$ kcal·mol^{-1}, $\Delta E^{\gamma(\text{classical})+\alpha(\text{II}-\text{RS})} = -299.09$ kcal·mol^{-1}. This suggests that alternating α- and γ-turn hairpins throughout the polypeptide chain result in a stable molecular fold.

The overall stability of these sequence-dependent conformers is due to preferred turn structure and stable β-sheet building units. The relative stability of structures with alternating turns (α-turn + γ-turn), such as hairpin GAGEG, AGAGEGA, GEGAG, and AGEGAGA, is mainly due to the antiparallel β-sheet regions (building unit III) (Table 2). Although these β-sheet structures are close in energy (Table 2), the gain in energy can be 12 kcal·mol^{-1} when permuted 30 times (see eq [S8] in the Supporting Information). The small ∼9 kcal·mol^{-1} stability difference between the appropriate two structures, $\Delta E^{\gamma(\text{inverse})+\alpha(\text{II}-\text{LS})} = -283.69$ and $\Delta E^{\gamma(\text{classical})+\alpha(\text{I}-\text{LU})} = -292.85$ kcal·mol^{-1}, is associated with the less stable central position of glutamic acids in hairpins AGEGA and GAGEGAG. The larger difference of ∼15 kcal·mol^{-1}, $\Delta\Delta E = (-283.69 - -299.09)$ kcal·mol^{-1}, is due to an additional intraresidual hydrogen bond appearing between the side-chain carbonyl oxygen atom and the amide hydrogen of the glutamic acid residue in the AGEGAGA hairpin (Table 3).

The same type of backbone/side-chain hydrogen bond is present in the most stable AGEGAG hairpin system, with a type II β-turn in its reversal. Since the latter type of stable turn structure is considered 15 times in eq [S8], as long as the alternating (α-turn + γ-turn) hairpins of AGEGAGA are considered only 8 times by removing the destabilizing effect of form I of the β-sheet, this structure will be the most stable ($\Delta E^{\text{type II }\beta} = -302.4$ kcal·mol^{-1}) (Tables 1 and 3). Note that the range in energy values of the four most stable structures of the 11.3 kDa polypeptide is less than 20 kcal·mol^{-1}; thus the differences are almost negligible if the size of the polypeptide is considered. This suggests that $[(AG)_xEG]_{16}$ ($3 \le x \le 6$) polypeptides form mainly alternating hairpins with some local β-turn-containing regions. Although this is in agreement with the previous structure analysis based on X-ray results,[21] our results also show that the main conformer should not be the one with the glutamic acids in the apex of the turns (e.g., the structure with AGEGA and GAGEGAG hairpins); rather, the glutamic acid is in the *i*th position in both cases in GEGAG and AGEGAGA hairpins to create a ∼16kcal·mol^{-1} more stable structure.

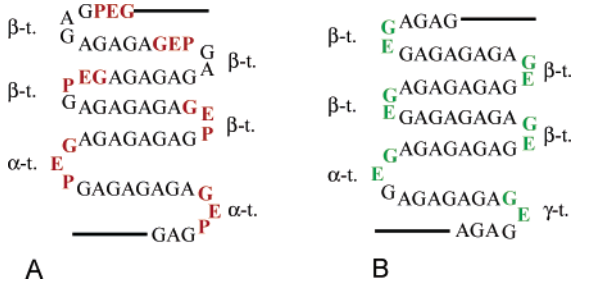


Figure 8. Position of the residues according to various turns. (A) Odd numbers in the repetitive sequence, e.g., $-(AG)_4PEG-$. Note that the PEG sequence shifts toward the β-sheet regions when there are β-turns in the hairpin, which leads to the disruption of the sheet structure. (B) Even numbers in the repetitive sequence, e.g., $-(AG)_4EG-$. The EG sequence always remains in the hairpins, irrespectively of the type of turn applied.

Stability of the Differently Folded $[(AG)_4PEG]_{16}$ Polypeptide. In the case of artificial proteins with $-(AG)_xEG-$ ($3 \le x \le 6$) sequences, previous results showed the antiparallel β-sheet arrangement of the structures.[21−25] Nevertheless, the first candidate that was predicted to form such a structure was a polypeptide with the $-(AG)_3PEG-$ repetitive sequence. In contrast, according to X-ray results, it formed amorphous glasses.[14] If the odd number of residues in the $-(AG)_3PEG-$ repetitive sequence is considered, one can draw the following conclusion:[14] it is highly unfavorable for antiparallel β-sheets to contain an odd number of amino acids, since the consecutive strands could not form the alternating S10−S14 type hydrogen bond patterns. However, antiparallel β-sheets containing an even number of residues could be easily formed for the latter repetitive sequence, if all the turn regions consist of an odd number of residues, such as α-turns. To see whether theoretical methods could provide additional reasons for the aggregation of the latter polypeptide, we investigated the energetic properties of a polypeptide with similar repeptitive sequence, $[(AG)_4PEG]_{16}$ (Figure 5C).

In light of the previous statement, if the hairpin regions are in β-turns, then the proline residues could not be preserved in the turn regions and would shift toward the middle of the strands, which would disrupt the sheet structure (Figure 8). The same shift would occur in the case of alternating α- and γ-turns. Therefore, only a structure formed from solely α-turns and antiparallel β-sheets could preserve the proline and glutamic acid residues in the hairpin regions. When analyzing energetic properties of these structures, the stabilizing energies of the most stable structures have the same order of magnitude compared to a fully extended structure as in the case of $[(AG)_4EG]_{16}$. This is seen in the case of $[(AG)_4PEG]_{16}$, e.g., $\Delta E = -292.9$ kcal·mol^{-1} for the structure with a II-LS type α-turn (Table 4). At first glance, this high stabilization energy would contradict the phenomena of the observed amorphous glasses.[14] However, the aggregation could be explained by the fact that the AGPEGA sequence is capable of forming stable β-turns, as calculated at the RHF/6-311++G(d,p)//RHF/3-21G level of theory (Table 5, Supporting Information). This means that both α- and β-turns could simultaneously form and be present in the $[(AG)_4PEG]_{16}$ polypeptide.

Table 4. Ab Initio (RHF/6-311++G(d,p)//RHF/3-21G) Stabilities of the Three Lowest Energy Conformers of the $[(AG)_4PEG]_{16}$ Polypeptide Obtained by Isodesmic Reactions

type of α-turn[a]	hairpin conformation[b]	side-chain conf of Glu[c]	ΔE (kcal·mol^{-1})[d]
α (I-LU)	$\beta_L\,\beta_L\,\underline{\epsilon_L\,\alpha_D\,\gamma_D}\,\epsilon_L\,\beta_L$	a, a	−271.8
α (II-LS)	$\beta_L\,\beta_L\,\underline{\epsilon_D\,\gamma_L\,\alpha_D}\,\gamma_L\,\beta_L$	g+, a	−254.0
	$\beta_L\,\beta_L\,\underline{\epsilon_D\,\gamma_L\,\alpha_D}\,\gamma_L\,\beta_L$	g−, a	−292.9
extended	$\beta_L\,\beta_L\,\underline{\delta_L\,\beta_L\,\beta_L}\,\beta_L$	a, a	0.0

[a] Based on nomenclature for α-turns applied by Pavone et al.[26] [b] Based on the conformation of the component subunit, cf. Perczel et al.,[50] with turns underlined. [c] *anti*: a; *minus gauche*: g−; *plus gauche*: g+. [d] Energies are relative to extended structure of the sequence, $E = -45185.105196$ hartrees.

On the basis of these results, the folding of these artificial polypeptides, $[(AG)_4EG]_{16}$ and $[(AG)_4PEG]_{16}$, could be presented as folding units of antiparallel β-sheet structures through various hairpins. In the case of $[(AG)_4EG]_{16}$, if a few strands interconnect through β-turns, the number of amino acids in the strands remains the same, the glutamic acids remain in the hairpin regions, and incoming strands could be interconnected through alternating turns (Figure 8B). However, in the case of $[(AG)_4PEG]_{16}$, if a β-turn is stabilized interconnecting two β-sheets, the position of the proline will shift toward the strands in the next turn, leading to the disruption of the antiparallel β-sheet motif, which results in an aggregated structure (Figure 8A). Thus, although in theory the $[(AG)_4PEG]_{16}$ polypeptides, similarly to $[(AG)_4EG]_{16}$ polypeptides, are capable of forming stable antiparallel β-sheets interconnected through α-turns, the local small structural changes in the turns can disrupt the sheet structures and consequently have a considerable impact on the macroscopic properties of the fiber.

Lateral Shearing of the $(GA)_3GE(GA)_3$ Model Peptide. Theoretical calculations provide a useful tool to understand the elasticity of silk-like polypeptides even at an atomic level. To observe the possible structural and energetic changes created by the destruction of hydrogen bonds in lateral shearing, we selected the $(GA)_3GE(GA)_3$ model peptide and performed constrained geometry optimizations when the distance between the (N)−H and O═(C) atoms of the terminal hydrogen bonds is gradually lengthened from 2 to 15 Å with a stepwise progression of 1 Å (Chart 1).

A monotonically increasing curve was obtained when the calculated relative energies were plotted as the function of the terminal (C)═O- - -H−(N) distance (Figure 9). However, the nonlinearity observed is rather unexpected and requires further structural explanation.

To understand the inflection point observed in the energy curve, we investigated the bond angles that would be affected by the destruction of hydrogen bonds. In the 14-membered hydrogen-bonded rings (S14), two types of bond angles are involved. Thus, the average of these two angles, ρ and σ, was considered (Chart 3 and Figure 10). The 10-membered hydrogen-bonded rings (S10) can be described by only one type of bond angle, τ, and were also investigated and plotted (Chart 3 and Figure 10).

As the length of the terminal hydrogen bond increases, there is at first a slow increase in bond angles, with a resultant increase in the tension in the molecule and a steeply increasing relative energy curve (Figures 9, 10A, and 10C). After reaching a certain distance of approximately 5 Å, the first hydrogen bond breaks.

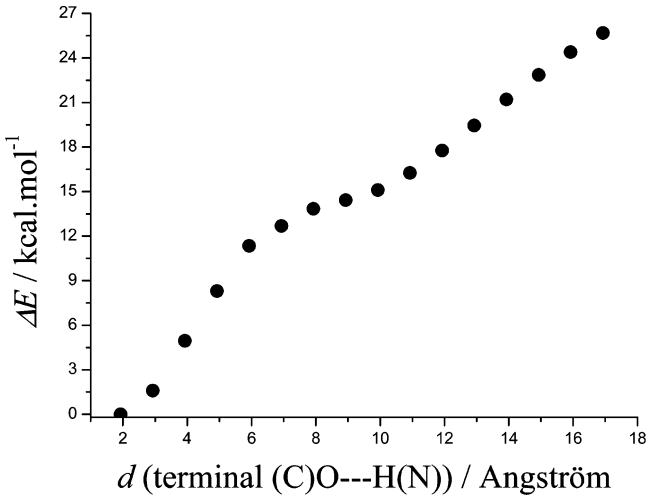
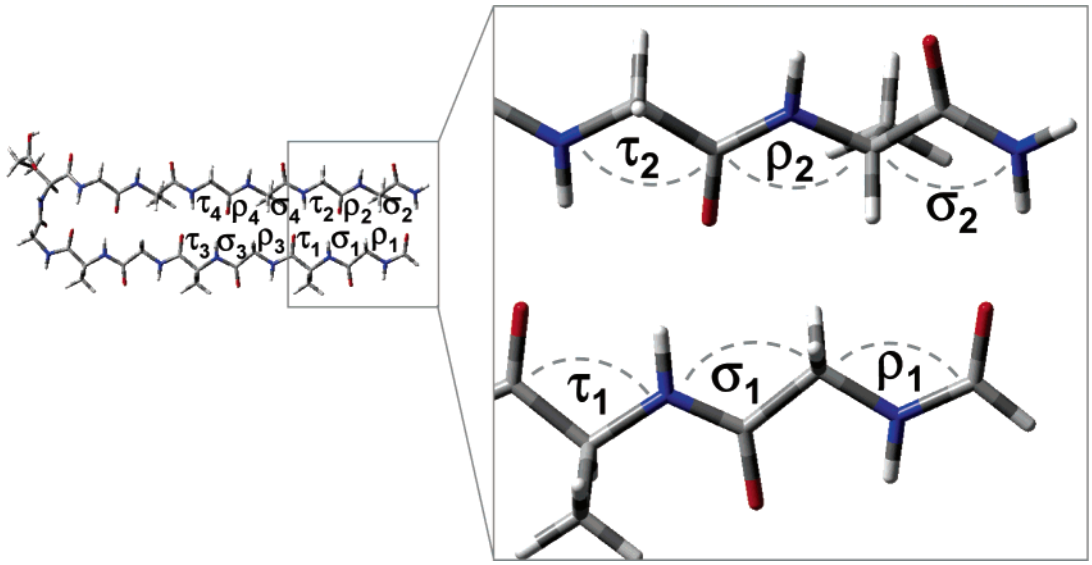


Figure 9. Relative energies of the $(GA)_3GE(GA)_3$ model as a function of the terminal (C)═O- - -H−(N) distance.

It can be verified by the sudden increase in the steepness of the energy curve, as well as the maximum angle values of the average ρ and average σ (Figure 10A). It seems that the break of an S14 hydrogen bond system is closely related and followed by the break of the following S10 hydrogen bond system at approximately 6 Å, as observed in the increase of τ_1 and τ_2 of the first S10 system (Figure 10C). After the break of the first S14−S10 hydrogen bond pattern, the relaxation of the free chain causes a decrease in the angle values (Figures 10A and 10C). The break of the S10 system is more rapid and can account for its rigidity and also its propensity to form an intraresidual hydrogen bond in an extended β_L conformation. The more relaxed bond angles appearing between the removal of the first S14−S10 hydrogen pattern and the second one result in a shallower increase in the relative energy between 7 and 9 Å. As the second S14 system begins to fall apart around 8−9 Å (Figure 10B), the steepness of the energy curve increases again (Figure 9). This is shortly followed by the radical increase in the bond angles of the second S10 system (Figure 10D). However, the second S14−S10 hydrogen bond pattern is less affected by the increase of the terminal (C)═O- - -H−(N) distance, and the breakup of this hydrogen bond pattern occurs less rapidly. We assume that another inflection point would occur in the increase of relative energy at around 18−20 Å, which would appear after the removal of the second S14−S10 hydrogen bond pattern. The high-energy demands of this lateral shearing system accounts for the strength of silk-like polypeptides, where numerous antiparallel β-sheets contribute to the excellent mechanical properties of silk fiber. Note that the curves describing the change in the bond angles throughout lateral shearing are very similar to the sawtooth pattern observed during forced extension of antiparallel sheet-containing proteins measured by atomic force microscopy.[52] The monotonically increasing energy curve that deviates from the sawtooth pattern could be explained by the fact that the effect of solvents is not considered in our model. The model system consequently lacks the stabilizing effect of solute−solvent hydrogen bonds that would form during the removal of one hydrogen bond in the polypeptide.

(52) Carrion-Vazquez, M.; Oberhauser, A. F.; Fisher, T. E.; Marszalek, P. E.; Li, H.; Fernandez, J. M. *Prog. Biophys. Mol. Biol.* **2000**, *74*, 63−91, and references therein.

Chart 3. Bond Angles That Were Plotted as a Function of Terminal Hydrogen Bond Distance[a]



[a] The C(O)−N−C (denoted by ρ) and N−C(O)−C (denoted by σ) bond angles describe the S14 hydrogen-bonding system, while N−C(O)−C (denoted by τ) bond angles describe the S10 hydrogen-bonding system.

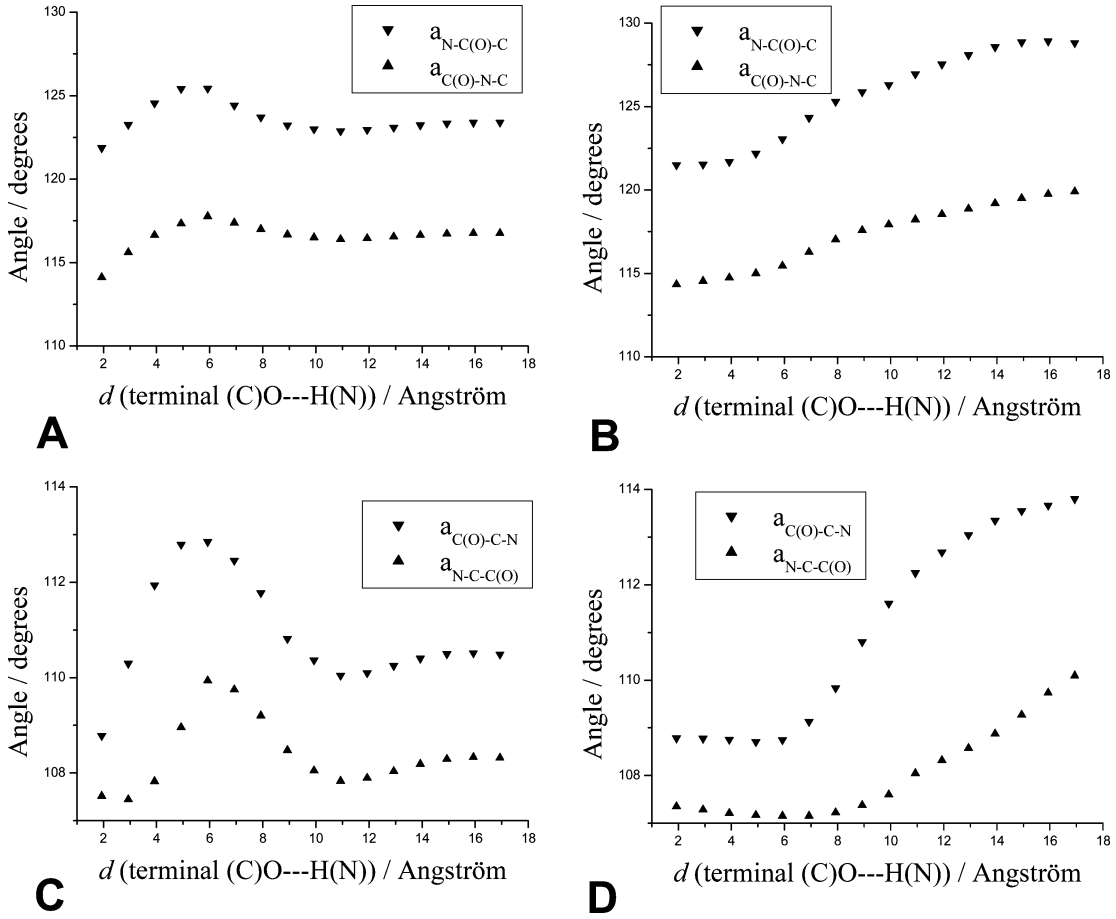


Figure 10. Bond angles of different hydrogen-bonding systems in the $(GA)_3GE(GA)_3$ model as a function of the terminal (C)═O- - -H−(N) distance. From the N-terminal: (A) bond angles of the first S14 system $a_{C(O)-N-C} = (\rho_1 + \rho_2)/2$, $a_{N-C(O)-C} = (\sigma_1 + \sigma_2)/2$; (B) bond angles of the second S14 system $a_{C(O)-N-C} = (\rho_3 + \rho_4)/2$, $a_{N-C(O)-C} = (\sigma_3 + \sigma_4)/2$; (C) bond angles of the first S10 system $a_{N-C-C(O)} = \tau_1$, $a_{C(O)-C-N} = \tau_2$; and (D) bond angles of the second S10 system $a_{N-C-C(O)} = \tau_3$, $a_{C(O)-C-N} = \tau_4$. For more details, see Chart 3.

Conclusions

Using thermoneutral isodesmic reactions and suitable models, the properties of silk proteins on silk-like model peptides have been examined. On the basis of our Lego-type approach, we have found that the stability of very large supermolecular systems can be estimated with acceptable accuracy, although it cannot be directly computed at the ab initio level of theory.

It is proposed that for simple polypeptides with repetitive sequences the stability, strength, and possibly solubility as well fall within a well-defined relationship relative to the number of sheets and turns. On the basis of our data, the most probable number of sheets (n_S) for ideal molecular packing is in the $4 \leq n_S \leq 7$ range. Note, that although the computed molecular topology is different from the one suggested by X-ray data[20] ($n_S \approx 2$ or 3), the observed trend is important in understanding the sheet-length preferences of such polypeptides.

Based on the stability properties of the individual hairpin conformers, the fold-dependent stability of different conformers of silk-like polypeptides was established. The most stable structure contains hairpins with the β(type II) turn conformation and the AGEGAG sequence (Table 3). When previously predicted structures of $[(AG)_4EG]_{16}$ based on X-ray diffraction spectroscopy[21] are compared with our results, it is found that the difference of stability between structures built from alternating γ- and α-turns and those built from β-turns is negligible if the size of the polypeptide is considered. On the basis of these results, we concluded that in the case of a local β-turn formation, the $[(AG)_4EG]_{16}$ polypeptide retains its lamellar structure. However, in the case of the proline-containing polypeptides, with $-(AG)_xPEG-$ ($x = 3, 4$) repetitive sequences, the formation of a stable local β-turn is also possible, but as a consequence prolines shift further "into" the sheet (Figure 8), which results in the breakup of the lamellar structure. This explanation supports previous X-ray diffraction results where $[(AG)_4EG]_{16}$ was found to contain antiparallel sheets, while $[(AG)_4PEG]_{16}$ had an amorphous glassy structure.

To study the strength and energetic properties of hydrogen bonds that stabilize silk proteins, the lateral shearing of the $(GA)_3GE(GA)_3$ hairpin structure was evaluated by gradually separating the N- from the C-terminal end. The relative energy increase as a function of the terminal (C)═O- - -H−(N) distance is far from linear, as one would have expected. In contrast, it has inflection points separating regions of different hydrogen bond characteristics. We found that the nonlinear increase of the energy is caused by two effects: the almost simultaneous breaking of the corresponding S14 and S10 hydrogen-bonded systems and the distortion in the bond angles of the peptide backbone. This distortion aims to preserve the remaining hydrogen bonds. On the basis of this phenomenon and the sudden relaxation of bond angles after breaking a hydrogen-bonded system, we assume that the breaking of hydrogen bonds is quantized instead of continuous. This observation is a useful step in understanding folding properties of proteins containing antiparallel β-pleated sheets. The phenomenon is in qualitative agreement with the sawtooth curves obtained by single-molecule AFM in protein unfolding experiments.[52]

Acknowledgment. This research was supported by grants from the Hungarian Scientific Research Foundation (OTKA T046994 and TS044730) and MediChem2. J.B. acknowledges the support of the US-Hungarian Joint Foundation (JFN 355) and D. A. Tirrell (California Institute of Technology) for artificial polypeptide samples. The authors would like to thank A. Pallo and S. Lau for helpful discussions.

Supporting Information Available: Computational details and consideration, building units of silk-like polypeptides, and thermoneutral isodesmic reactions, as well as the complete ref 49. This material is available free of charge via the Internet at http://pubs.acs.org.

JA063933P