JOC*Article*

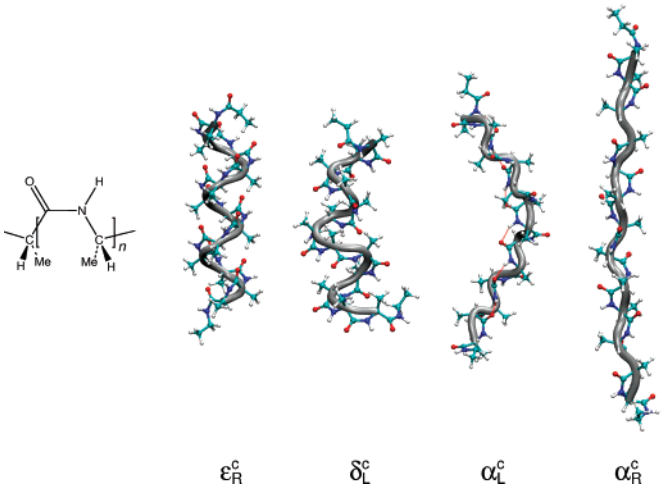
All-*cis* Helical Polypeptides

Romuald Poteau and Georges Trinquier*

Laboratoire de Physique et Chimie des Nano-Objets (UMR5215, CNRS-UPS-INSA), IRSAMC, Institut National des Sciences Appliquées, 31077 Toulouse Cedex 4, France

trinquier@irsamc.ups-tlse.fr

Received June 6, 2007



The possibility of all-*cis* open-chain polypeptides is rarely addressed, owing to three main reasons, namely, (i) the extreme scarcity of cis peptide bonds in naturally occurring proteins and peptides, (ii) the lesser thermodynamic stability (by about 2.5 kcal/mol) of cis amide bonds with respect to their trans counterparts, and (iii) widely held preconceptions about the so-called "steric clash" between lateral chains borne by two successive α carbons. Quantum-chemistry calculations performed on alanine tridecamers show how the latter constraints can be efficiently relieved through proper ϕ/ψ adjustments along the backbone, leading to several helical arrangements—presumably the only permitted regular structures. Four more-or-less regular helices were thus characterized, one of them, a superhelix, exhibiting intramolecular hydrogen bonds. Understanding and anticipating all-*cis* open-chain structures not only make use of the classical Ramachandran maps at each $C_{\alpha i}$, relating to $E = f(\phi_i,\psi_i)$, but also require the profile of a new kind of conformational dependence, the *plaque* maps, relating to $E = f(\phi_i,\psi_{i-1})$. The obvious coupling between two such maps enforces conformational dependence between two consecutive C_α's, somewhat questioning in this context the customary "local effects", and presumably reducing the whole chain plasticity. Whereas cis thermodynamic penalty cannot be abolished locally, energy clues indicate that assembling cis-prepared building units is an exothermic process. Besides, once built up, the all-*cis* backbone should be difficult to unlock, thus affording reasonable kinetic stability.

Introduction

Helical structures seem to represent a quasiuniversal option for spatial organization of matter at both macroscopic and microscopic echelons. At the molecular level, inorganic[1] and organometallic[2] materials, as well as organic polymers and biopolymers frequently elicit such kind of arrangements. Except for lipids, all classes of biopolymers provide examples of helical structures: nucleic acids with DNA, carbohydrates with helical starch and its inclusion complexes, proteins with the well-known

(1) Giraldo, O.; Brock, S. L.; Marquez, M.; Suib, S. L.; Hillhouse, H.; Tsapatsis, M. *Nature* **2000**, *405*, 38.

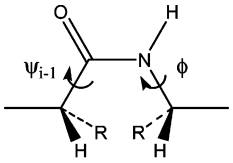
(2) Piguet, C.; Borkovec, M.; Hamacek, J.; Zeckert, K. *Coord. Chem. Rev.* **2005**, *249*, 705.

10.1021/jo0711790 CCC: $37.00
Published on Web 10/04/2007

α-helix secondary structure, not to mention the nested multiple-stranded helices of frozen water predicted to form within carbon nanotube at high pressure.[3] In the protein realm—here the matter in hand—besides the familiar right-handed α-helix, a few helical alternatives are also encountered, although less frequently. Some are likewise stabilized by intramolecular hydrogen bonding, such as 3_{10} or π helices,[4] while others can be entirely devoid of it, like polyproline-type helices PPII, the involvement of which in denatured states or protein elasticity has been abundantly discussed.[5] Most peptide foldamers[6] or peptidomimetic variants also elicit helices. Remarkably, here again this concerns not only foldamers committing their N−H parts into intramolecular hydrogen bonds,[7,8] but also those peptidomimetics which are devoid of it, like peptoids, who bear tertiary amide functions.[9]

In a previous work, we proposed potential cyclic arrangements for polypeptides built from cis-only peptide bonds. These more-or-less planar structures happen to be kinetically and thermodynamically stable for sizes around eight peptide units and beyond.[10] Although non-prolyl cis-peptide bonds are extremely scarce in natural peptides and proteins (<0.03%),[11] we propose in this work to explore theoretically the prospective structures of open chains of α-aminoacids linked by cis-peptide bonds only, seeking in particular the existence, in such polymers, of regular secondary structures like helices, as envisaged in fact by Linus Pauling more than half a century ago.[12] We are confirmed in this approach by noting, first, that any planar chain obliged to unidimensional growing with roughly constant curvature—typically the constraint sustained at cis peptide links—quite naturally adopts an helical layout, as do helicenes, oligoarylamides, aminophthalimides, or pyridine-pyrimidine oligomers.[13] On another hand, one instance is known where a cis-type peptide-link environment can produce helical structure: this happens in proline oligopeptides with all peptide bonds in the cis configuration, a structure called "polyproline helix of type I" (PPI).[14]

Let us tackle from scratch the challenge of making open chains of cis-peptide links. According to commonly accepted views, the main impediment in the building of an all-*cis* backbone stems from the so-called *steric clashes* between the lateral groups R attached to the α carbons, **1**.



1

This is, however, a rather naïve and restrictive picture, as the two edging α-carbons here assume the standard extended conformation encountered in all-*trans* β-strands ($\phi \approx \psi \approx 180°$). While such arrangements are actually quite favored in normal all-*trans* sequences, as evidenced from the "β zone" in theoretical Ramachandran maps, this should be no longer justified, a priori, for all-*cis* links. Obviously, any small variation of the $[\psi_{i-1},\phi_i]$ set should efficiently relieve the *syn*−*syn* constraint, hence tempering the contact between neighboring lateral chains,

(3) Bai, J.; Wang, J.; Zeng, X. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19664.

(4) Weaver, T. M. *Protein Sci.* **2000**, *9*, 201.

(5) (a) Stapley, B. J.; Creamer, T. P. *Protein Sci.* **1999**, *8*, 587. (b) Rucker, A. L.; Creamer, T. P. *Protein Sci.* **2002**, *11*, 980.(c) Pappu, R. V.; Rose, G. D. *Protein Sci.* **2002**, *11*, 2437. (d) Ma, K.; Wang, K. *Biochem. J.* **2003**, *374*, 687. (e) Ferreon, J. C.; Hilser, V. J. *Protein Sci.* **2003**, *12*, 447. (f) Whittington, S. J.; Creamer, T. P. *Biochemistry* **2003**, *42*, 14690. (g) Chellgren, B. W.; Creamer, T. P. *J. Am. Chem. Soc.* **2004**, *126*, 14734. (h) Chen, K.; Liu, Z.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15352. (i) Hamburger, J. B.; Ferreon, J. C.; Whitten, S. T.; Hilser, V. J. *Biochemistry* **2004**, *43*, 9790. (j) Bochicchio, B.; Aït-Ali, A.; Tamburro, A. M.; Alix, A. J. P. *Biopolymers* **2004**, *73*, 484. (k) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins*: *Struct., Funct., Bioinf.* **2004**, *55*, 502. (l) Vila, J. A.; Baldoni, H. A.; Ripoll, D. R.; Ghosh, A.; Scheraga, H. A. *Biophys. J.* **2004**, *86*, 731. (m) Fleming, P. J.; Fitzkee, N. C.; Mezei, M.; Srinivasan, R.; Rose, G. D. *Protein Sci.* **2005**, *14*, 111. (n) Zagrovic, B.; Lipfert, J.; Sorin, E. J.; Millett, I. S.; van Gunsteren, W. F.; Doniach, S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11698. (o) Cubellis, M. V.; Caillez, F.; Blundell, T. L.; Lovell, S. C. *Proteins*: *Struct., Funct., Bioinf.* **2005**, *58*, 880. (p) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. *Biochemistry* **2005**, *44*, 6269.

(6) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. *Chem. Rev.* **2001**, *101*, 3893.

(7) (a) Seebach, D.; Matthews, J. L. *J. Chem. Soc., Chem. Commun.* **1997**, 2015. (b) Hanessian, S.; Luo, X.; Schaum, R.; Michnick, S. *J. Am. Chem. Soc.* **1998**, *120*, 8569. (c) Andrews, M. J. I.; Tabor, A. B. *Tetrahedron* **1999**, *55*, 11711. (d) Baldauf, C.; Günther, R.; Hofmann, H.-J. *J. Org. Chem.* **2006**, *71*, 1200.

(8) (a) Semetey, V.; Hemmerlin, C.; Didierjean, C.; Schaffner, A. P.; Gimenez-Giner, A.; Aubry, A.; Briand, J. P.; Marraud, M.; Guichard, G. *Org. Lett.* **2001**, *3*, 3843. (b) Semetey, V.; Rognan, D.; Hemmerlin, C.; Graff, R.; Briand, J. P.; Marraud, M.; Guichard, G. *Angew. Chem.* **2002**, *115*, 1973; *Angew. Chem., Int. Ed.* **2002**, *41*, 1893. (c) Semetey, V.; Didierjean, C.; Briand, J. P.; Aubry, A.; Guichard, G. *Angew. Chem.* **2002**, *115*, 1975; *Angew. Chem., Int. Ed.* **2002**, *41*, 1895. (d) Hemmerlin, C.; Marraud, M.; Rognan, D.; Graff, R.; Semetey, V.; Briand, J. P.; Guichard, G. *Helv. Chim. Acta* **2002**, *85*, 3692. (e) Violette, A.; Averlant-Petit, M. C.; Semetey, V.; Hemmerlin, C.; Casimir, R.; Graff, R.; Marraud, M.; Briand, J. P.; Rognan, D.; Guichard, G. *J. Am. Chem. Soc.* **2005**, *127*, 2156. (f) Violette, A.; Fournel, S.; Lamour, K.; Chaloin, O.; Frisch, B.; Briand, J.-P.; Monteil, H.; Guichard, G. *Chem. Biol.* **2006**, *13*, 531. (g) Oakley, M. T.; Guichard, G.; Hirst, J. D. *J. Phys. Chem. B* **2007**, *111*, 3274.

(9) (a) Moehle, K.; Hofmann, H.-J. *Biopolymers* **1996**, *38*, 781. (b) Armand, P.; Kirshenbaum, K.; Falicov, A.; Dunbrack, R. L., Jr.; Dill, K. A.; Zuckermann, R. N.; Cohen, F. E. *Folding Des.* **1997**, *2*, 369. (c) Möhle, K.; Hofmann, H.-J. *J. Pept. Res.* **1998**, *51*, 19. (d) Wu, C. W.; Sanborn, T. J.; Zuckermann, R. N.; Barron, A. E. *J. Am. Chem. Soc.* **2001**, *123*, 2958. (e) Wu, C. W.; Sanborn, T. J.; Huang, K.; Zuckermann, R. N.; Barron, A. E. *J. Am. Chem. Soc.* **2001**, *123*, 6778. (f) Sanborn, T. J.; Wu, C. W.; Zuckermann, R. N.; Barron, A. E. *Biopolymers* **2002**, *63*, 12. (g) Burkoth, T. S.; Beausoleil, E.; Kaur, S.; Tang, D.; Cohen, F. E.; Zuckermann, R. N. *Chem. Biol.* **2002**, *9*, 647. (h) Wu, C. W.; Kirshenbaum, K.; Sanborn, T. J.; Patch, J. A.; Huang, K.; Dill, K. A.; Zuckermann, R. N.; Barron, A. E. *J. Am. Chem. Soc.* **2003**, *125*, 13525. (i) Lee, B. C.; Zuckermann, R. N.; Dill, K. A. *J. Am. Chem.Soc.* **2005**, *127*, 10999. (j) Baldauf, C.; Günther, R.; Hofmann, H.-J. *Phys. Biol.* **2006**, *3*, S1.

(10) Poteau, R.; Trinquier, G. *J. Am. Chem. Soc.* **2005**, *127*, 13875.

(11) A comprehensive bibliography on non-proline cis peptide bonds in proteins and peptides is given in ref 10; see also: Bairaktari, E.; Mierke, D. F.; Mammi, S.; Peggion, E. *J. Am. Chem. Soc.* **1990**, *112*, 5383. Scherer, G.; Kramer, M. L.; Schutkowski, M.; Reimer, U.; Fischer, G. *J. Am. Chem. Soc.* **1998**, *120*, 5568. Forbes, C. C.; Beatty, A. M.; Smith, B. D. *Org. Lett.* **2001**, *3*, 3595.

(12) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1952**, *38*, 86.

(13) (a) Gong, B. *Chem. Eur. J.* **2001**, *7*, 4337. (b) Haldar, D.; Jiang, H.; Léger, J.-M.; Huc, I. *Tetrahedron* **2007**, *63*, 6322. (c) Hu, Z.-Q.; Hu, H.-Y.; Chen, C.-F. *J. Org. Chem.* **2006**, *71*, 1131. (d) Appella, D. H.; Christianson, L. A.; Karle, I. L.; Powell, D. R.; Gellman, S. H. *J. Am. Chem. Soc.* **1999**, *121*, 6206. (e) Delsuc, N.; Léger, J.-M.; Massip, S.; Huc, I. *Angew. Chem., Int. Ed.* **2007**, *46*, 214. (f) Katayama, H.; de Greef, T. F. A.; Kooijman, H.; Spek, A. L.; Vekemans, J. A.; Meijer, E. W. *Tetrahedron* **2007**, *63*, 6642. (g) Ohkita, M.; Lehn, J.-M.; Baum, G.; Fenske, D. *Chem. Eur. J.* **1999**, *5*, 3471.

(14) (a) Krimm, S.; Venkatachalam, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 2468. (b) Beausoleil, E.; Lubell, W. D. *Biopolymers* **2000**, *53*, 249. (c) Sandvoss, L. M.; Carlson, H. A. *J. Am. Chem. Soc.* **2003**, *125*, 15855. (d) Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **2004**, *108*, 4885. (e) Kakinoki, S.; Hirano, Y.; Oka, M. *Polym. Bull.* **2005**, *53*, 109. (f) Kang, Y. K.; Park, H. S. *Biophys. Chem.* **2005**, *113*, 93. (g) Horng, J. C.; Raines, R. T. *Protein Sci.* **2006**, *15*, 74. (h) Zhong, H.; Carlson, H. A. *J. Chem. Theory Comput.* **2006**, *2*, 342.

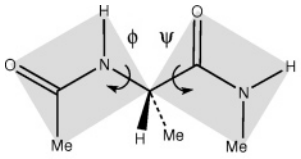
and opening the way for a practicable open chain—in other words, crucial degrees of freedom are neglected in this crude depiction.

The present study will focus on a prime example of an all-*cis* L-homochiral α-aminoacid open chain, namely all-*cis*-tridecaalanine. We will try to find its best possible arrangements using quantum-chemistry methods at the DFT level in in vacuo context. The paper is organized as follows. In the first section, some prerequisite spadework on local effects is made from conformational-map analyses. The next section reports and comments on the energies and geometries obtained for some regular structures. After a subsequent discussion section, the various computational and technical details are gathered in the Computational Section.

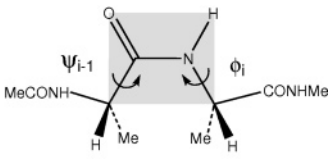
Results and Discussion

Scouting from Conformational Maps. Just as exploration of new territories is made easier by guides leading straightforwardly to the most interesting areas, the search for the most stable three-dimensional structures of cis polypeptides will be facilitated by preliminary inspection of theoretical Ramachandran maps. In α-aminoacid natural proteins, such conformational-energy maps proved to be efficient tools for analyzing the observed $[\phi,\psi]$ sets. Calculated on dipeptide models made of a single junction of two peptide plaques, the theoretical maps $E = f(\phi,\psi)$ essentially account for most of the fully or partly allowed regions as well as the forbidden regions, were latent steric clashes can be identified precisely.[15]

A sequence of cis plaques is actually ruled by *two* kinds of conformational maps. The first one is the classical Ramachandran map, corresponding to the two rotations at a given C_α pivot, $E = f(\phi_i,\psi_i)$, **2**. The second one, here called the *plaque* conformational map, corresponds to the two rotations about the terminal hinges of a given plaque, $E = f(\phi_i,\psi_{i-1})$, **3**. As mentioned, most of the steric interaction between neighboring lateral chains can be efficiently relieved by fine-tuning these dihedral angles ψ_{i-1} and ϕ_i. This second degree of freedom is a specific feature of all-*cis* peptide chains. It stems from the strong coupling occurring in cis plaques between two consecutive C_αHR groups, whereas, from intrinsic geometrical grounds, this interaction is by far negligible in classical trans peptide arrangements.



2



3

Both kinds of conformational maps have been calculated, and are reported below as 2-D potential-energy landscapes $E(\phi,\psi)$ (see the Computational Section for details). If many theoretical calculations of Ramachandran maps are available, occasionally for cis backbones,[16] to date none of them concerned *plaque* maps—the cornerstone governing cis-peptide conformational accessibility.

(15) Ho, B. K.; Thomas, A.; Brasseur, R. *Protein Sci.* **2003**, *12*, 2508.
(16) Baldoni, H. A.; Zamarbide, G. N.; Enriz, R. D.; Jauregui, E. A.; Farkas, Ö.; Perczel, A.; Salpietro, S. J.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **2000**, *500*, 97.

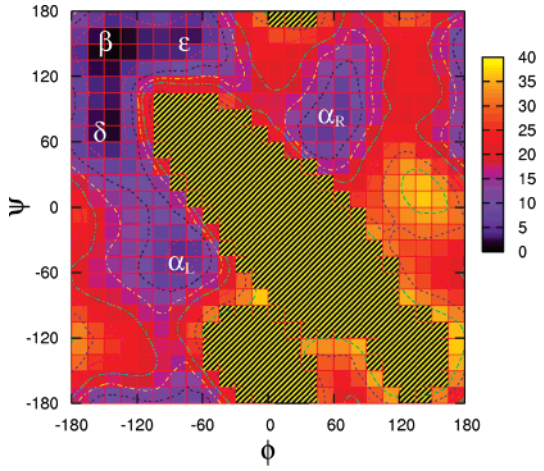
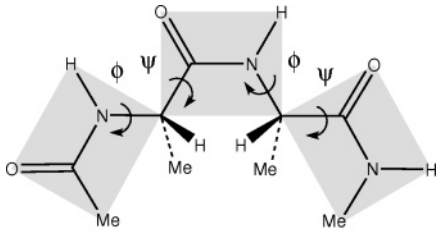


FIGURE 1. Global conformational map $E = f(\phi,\psi)$ for an all-*cis* alanine tripeptide model in which both $[\phi,\psi]$ sets are treated equivalently and both ω dihedral angles are constrained to zero (RHF/3-21G level, contour lines separated by 5 kcal/mol, dashed areas for $E > 40$ kcal/mol).

Prior to specific discussions on the two categories of maps, let us examine the particular case of a *regular* all-*cis* polypeptide. A convenient and natural pattern for obtaining structural hints is here a tripeptide model carrying two C_α's and three peptide plaques, with the constraint of full equivalency of all ϕ's on one hand, and all ψ's on the other hand, **4**.



4

An overall conformational map $E = f(\phi,\psi)$ will ensue, which in a way concentrates all the information pertaining to both the two residue maps and the single plaque map associated with this tripeptide. The corresponding energy plot is drawn in Figure 1.[17] A large forbidden region appears along a diagonal, from $[\phi,\psi] \approx [-90°,+90°]$ to $[\phi,\psi] \approx [+180°,-180°]$. Five favorable low-energy zones are revealed. The deepest one matches the β basin around $[\phi,\psi] \approx [-150°,150°]$, surrounded by a δ zone around $[-150°,+60°]$, and an ϵ zone around $[-90°,+150°]$, connected to β through a tiny barrier.[18] The map further suggests that regular cis-chains could also explore higher energy uplands within α regions, here located around $[\phi,\psi] \approx [-90°,-90°]$ for α_L and $[+60°,+90°]$ for α_R.[19]

(17) For this preliminary exploration, the RHF/3-21G level is used, the reliability of which in such a context has been otherwise established (see: Perczel, A.; Farkas, Ö.; Jakli, I.; Topol, I.A.; Csizmadia, I. G. *J. Comput. Chem.* **2003**, *24*, 1026).
(18) We use here a convenient labeling proposed for trans-peptide maps (see ref 17 and: Perczel, A.; Csizmadia, I. G. In *The amide linkage*; Greenberg A., Breneman, C. M., Liebman, J. F., Eds.; Wiley−Interscience: New York, 2003; Chapter 13).
(19) Unlike in classical trans-dipeptide models, in a cis ϕ/ψ map, the $[-60°,-60°]$ zone corresponds to local left-handed helicity (α_L), while the $[60°,60°]$ zone corresponds to local right-handed helicity (α_R).

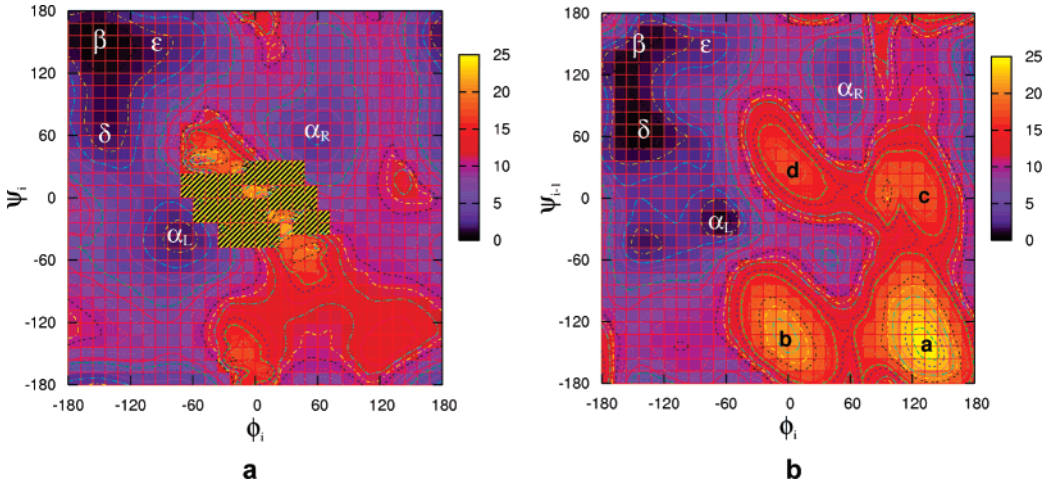


FIGURE 2. (a) Residue conformational map $E = f(\phi,\psi)$ for an all-*cis* alanine dipeptide model. (b) Plaque conformational map $E = f(\phi_i,\psi_{i-1})$ for an all-*cis* alanine tripeptide model including two C_α and three plaques, the two varied parameters being the dihedral anges ψ_{i-1} and ϕ_i at the middle plaque (see **3**). In both maps, all dihedral angles ω were kept to zero, the contour lines are separated by 1 kcal/mol, and the dashed areas are for $E > 25$ kcal/mol.

The above constraints $\phi_i = \phi_{i-1}$ and $\psi_i = \psi_{i-1}$ compel the cis-peptide chain to a regular shape. In the general case, however, these two degrees of freedom that are the dihedral torsions (ϕ_i,ψ_i) at a given C_α and the dihedral torsions (ψ_{i-1},ϕ_i) at the adjacent plaque should be regarded, a priori, as independent. To anticipate cis-peptides with nonregular shapes, we must therefore address the classical residue map and the plaque map separately. These are reported in Figure 2.

Both maps exhibit low-energy basins located rather similarly to one another, and, not surprisingly, similar to the above-discussed constrained synthetic map. In particular no new minimum occurs. Although located somewhat differently, the main zones β, ϵ, δ, α_L, and α_R are clearly identifiable on the two maps. These differ essentially by the forbidden regions. In the classical residue map (Figure 2a), there is a mountainous area around [0°,0°] corresponding to the steric clash between the two end methyl groups.[20] This zone extends along the diagonal while a second high-energy knoll occurs around [−30°,−150°], corresponding to the interaction between the left-hand terminal methyl group with the oxygen of the next plaque. In the plaque map (Figure 2b), four conspicuous forbidden zones are generated whenever two non-hydrogen groups are facing each other in the plaque plane. In decreasing order of hindering, these correspond to the sets R···R, R···C′, N···R, and N···C′, located ideally at [+120°,−120°], **a**, [0°,−120°], **b**, [+120°,0°], **c**, and [0°,0°], **d**, respectively, although the last maximum is significantly distorted at [0°,+40°].

The R···R or more generally R···R′ interaction may not be always that disadvantageous, however. Obviously, this contact will be favored when the two lateral chains come from charged residues carrying alternating charges, or from residues bearing an aromatic component open to hydrophobic stacking. The possibly resulting salt bridges or aromatic hydrophobic bridges would no doubt lower both the maximum **a** at [+120°,−120°], corresponding to R···R′ facing proximity, and the ideal starting β region at [−180°,+180°], corresponding to R···R′ lateral proximity responsible for the initial steric clash. Consequently, plaque maps for couples of residues such as EK, DR, or FF could somewhat differ from Figure 2b.

This illustrates how the plaque map is different from a classical residue map in that it is associated to a couple—*stricto sensu* an oriented couple—of residues, not to mention the cases of heterochirality. Where there were 20 Ramachandran maps for evaluating the local effect at a given residue R in classical trans peptides, there are in cis peptides as many as $20^2 = 400$ plaque maps to document the local effects induced by a given binomial RR′ of lateral chains.

Last, we would emphasize a crucial difference between the trans and cis worlds. In the former, due to the trans position of the two C_α atoms on the peptide plaque, these are little coupled so that the consecutive (ϕ,ψ) sets can be considered as roughly independent, resulting in a total number of possible conformations scaling as A^n (n is the chain length, A is the number of basic relative orientations of two adjacent plaques, which can be taken at 2 or 3, or is set as an effective parameter). This entails the well-known conformational diversity of classical trans polypeptide chains—the "proteiformity" of proteins. This is no longer the case in all-*cis* chains, where each C_α is coupled with its neighbors via the above-discussed plaque maps. Because two successive (ϕ,ψ) sets are now no longer independent, the total number of potential conformations for the whole chain can no longer be assessed, here, from a simple product, as previously. Consequently, reduced conformational diversity—hence less plasticity—is expected in all-*cis* chains. As a corollary, the "local effect" often discussed in the trans world should assume in the cis world, if any, a somewhat different status.

With this outline of cartography in hand, we can now set off exploring our all-*cis terra incognita*.

Regular Helical Structures. The choice of our model, tridecaalanine, emerges as a compromise between computational feasibility and lessening of finite size effects. The molecule is amidified at both ends so that the 13 α carbons or residues in fact convey 14 peptide plaques. For structural convenience, each terminal amide function is further ethylated, the studied entity finally being written as Et-CO-A^c_{13}-NH-Et, which will be labeled hereafter A^c_{13} for the sake of conciseness.

The systematic conformational search for this tridecaalanine model is started from numerous C_α conformational sets belong-

(20) The main features of the Ramachandran map for the cis-dipeptide model formylglycinamide have been discussed in ref 16. In that model, both the lateral and the two end methyl groups are lacking, which account for some differences with the present model.

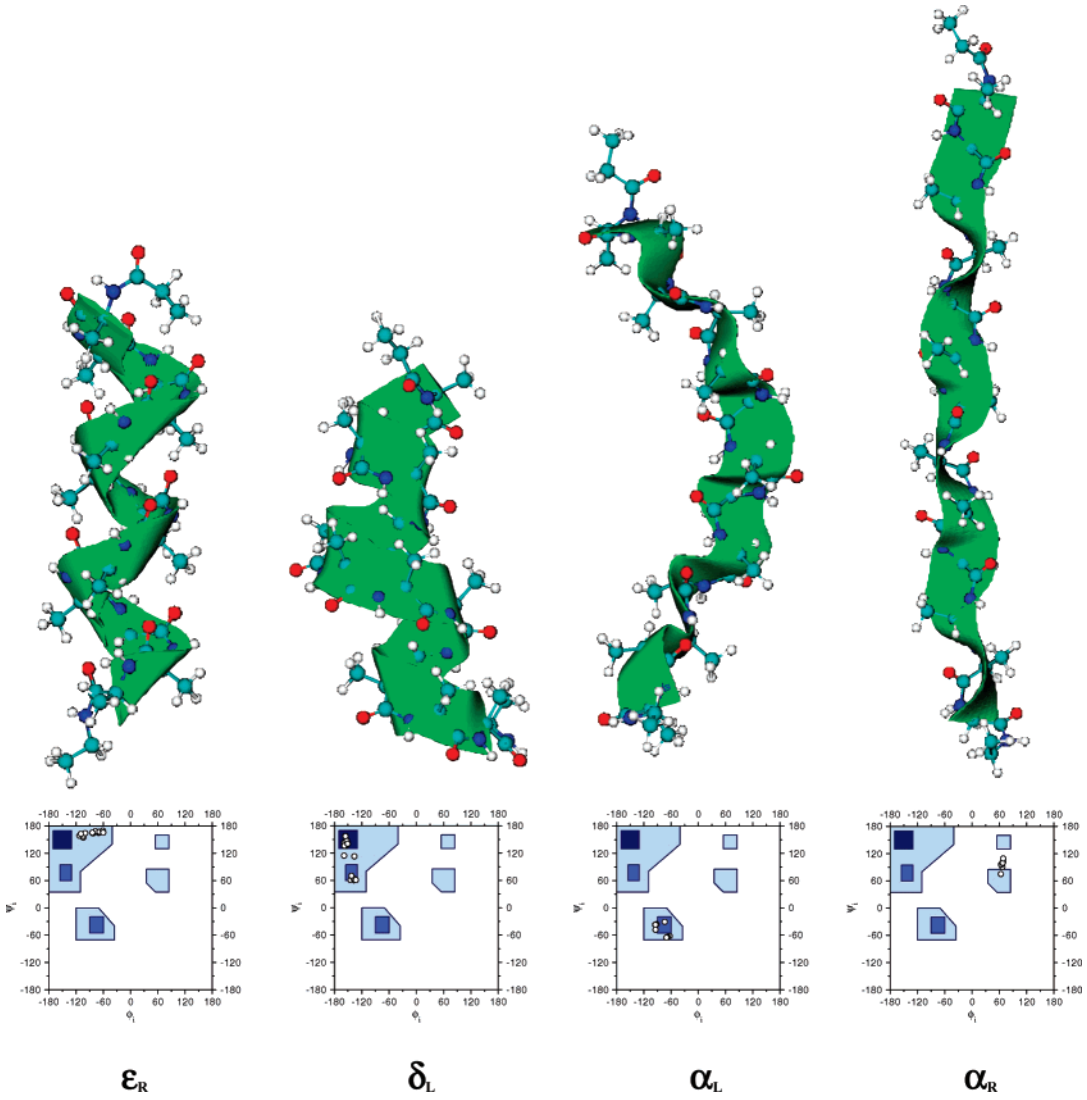


FIGURE 3. Some regular helical structures found for all-*cis*-tridecaalanine A^{c}_{13}, with corresponding Ramachandran maps.

ing or not to the four basins associated to the PES of the above-discussed dialanine models. This careful exploration led to only four kinds of essentially regular periodic structures of helical shapes (Figure 3), matching reasonably the valleys of both residue map and plaque map. For convenience, each helical structure is labeled according to the zone in the Ramachandran map which best gathers its $[\phi,\psi]$ set.

The thermodynamically more stable structure, a right-handed helix ϵ_R-A^{c}_{13}, does not strictly correspond to the β-cis bassin. It cannot be so because a cis dipeptide model ignores the impossibility for an open chain to fold approximately within the same plane. Only the quasiplanar cyclic peptides would match strictly the true β area, indeed.[10] The structure ϵ_R-A^{c}_{13} therefore explores a compromise between the β-cis basin ($[\phi,\psi] \approx [-150°,150°]$) and a neighboring ϵ-cis zone ($[\phi,\psi] \approx [-60°,-150°]$), which, not surprisingly, corresponds to the conformational parameters of cis polyproline helices PPI generally defined around $[\phi,\psi] \approx [-75°,160°]$. If ψ keeps here a rather constant value around 150°, ϕ actually spans the interval −60° to −120°. In such a curled arrangement, the steric clashes between neighboring lateral methyl groups have fallen down, as these point outside the helix (Figure 3), while remaining slightly more buried here than in a classical the α-helix.

Since, at our level of calculation, the cis form of the parent *N*-methylacetamide (NMA) lies 2.5 kcal/mol above its trans form, our cis tridecamers, bearing 14 peptide plaques, are expected to lie at least 14 × 2.5 ≈ 35 kcal/mol above an all-trans counterpart free of intramolecular hydrogen bonding, like a β strand of normal polyalanine. The actual energy difference between such a strand β-A^{t}_{13} and the helix ϵ-A^{c}_{13} is calculated at 69 kcal/mol, which would correspond to a mean plaque increment of 5 kcal/mol instead of the 2.5 kcal/mol in the isolated plaque. This extra penalty of 2.5 kcal/mol (≈5−2.5) appears reasonably low when compared to the *syn*−*syn* repulsion between methyl groups in the primitive plaque epitome **1,** estimated at around 10 kcal/mol by convergent procedures.[21] Simple $[\psi_{i-1},\phi_i]$ adjustements thus prove sufficient to bring what was a 10-kcal/mol *steric clash* in the ill-designed cis-plaque prototype back to a more affordable 2.5-kcal/mol *steric discomfort* in the fine-tuned chain—an energetic clue we consider quite encouraging. Incidentally, this 5-kcal/mol per-plaque energy difference between β-A^{t}_{13} and ϵ_R-A^{c}_{13} open chains is near that calculated between all-*cis*- and all-*trans*-cyclohexaalanine rings (4.7 kcal/mol).

(21) Mathieu, S.; Poteau, R.; Trinquier, G. Manuscript in preparation.

TABLE 1. Calculated Energetics and Estimated Helix Parameters for Tridecaalanines

		energies[a]					helix parameters		
		ΔE total	ΔE per plaque	MPJE[b]	MJCE[c]	NH···OC[d]	res per turn	pitch[e]	diameter[e]
all-*cis*	α_R^c	136.0	9.7	2.4	2.7		6.0	11.4	1.8
	α_L^c	123.2	8.8	1.4	1.7	6	14.0	30.0	7.1
	δ_L^c	99.1	7.1	−0.4	−0.3		4.3	6.4	4.1
	ϵ_R^c	92.7	6.6	−0.9	−0.9		4.0	6.0	3.0
all-*trans*	β^t	23.7	1.7	−3.4	−0.2				
	α_R^t	0.	0.	−5.2	−2.2	11	3.8	5.6	4.8
	exptl[f]						3.6	5.4	5.0

[a] In kcal/mol. [b] Mean plaque junction energy (see text for definition). [c] Mean junction coupling energy (see text for definition). [d] Total number of intramolecular hydrogen bonds. [e] In Å. [f] Standard values for mean trans α-helix.

In a previous work we proposed to use indexes based on isodesmic reactions to estimate the energy required for joining peptide plaques into polypeptides.[10] One of them, the *mean plaque junction energy* (MPJE), suggested that beyond six peptide units, an all-*cis* plaque association is almost as favorable as an all-*trans* one. For the presently studied open chains, the MPJE refers to the reaction

14 Et-CONH-Et →
Et-(CONH-CHMe)$_{13}$-CONH-Et + 13 EtH

The MPJE index, corresponding to the energy of the above reaction brought to a single junction, is therefore calculated as the energy of the formal equation

(14/13) Et-CONH-Et →
(1/13) Et-(CONH-CHMe)$_{13}$-CONH-Et + Me-Me

In agreement with the weak steric interaction of lateral methyl groups, we find for ϵ_R-A_{13}^c a negative MPJE value of −0.9 kcal/mol. As expected, this value is weaker than that obtained for β-A_{13}^t, calculated at −3.4 kcal/mol, but it clearly shows that cis-plaque association is still an exothermic process. Nevertheless, it appears to be less competitive in this open chain than in cyclic cis-peptides, for which MPJE indexes were estimated for instance at −1.4 kcal/mol in cA_6^c. Free of lateral chains, *cis*-cyclopolyglycines are of course even more favored (typically MPJE = −1.7 kcal/mol in cG_6^c).

In this context, the classical trans α-helix, α_R-A_{13}^t, is worth a comparison. At the presently used computational level, it lies at 92.6 kcal/mol *below* ϵ_R-A_{13}^c, corresponding to a difference of 6.6 kcal/mol per plaque. The 92.6-kcal/mol energy difference encompasses the 14 intrinsic cis−trans plaque differentials and the 11 intramolecular hydrogen bonds occurring in the all-*trans* α-helix α_R-A_{13}^t. The latter can be estimated to contribute 5−6 kcal/mol each, at the same level of description. After deduction of the 11 hydrogen bonds, the difference between helices ϵ_R-A_{13}^c and α_R-A_{13}^t therefore reduces to 27−38 kcal/mol, corresponding to increments of 1.9−2.7 kcal/mol per plaque. This happens to be close to the 2.5-kcal/mol/plaque intrinsic trans preference, suggesting a backbone strain of about the same magnitude in both helices, or only slightly unfavored in the all-cis case. Note that by performing a similar arithmetic from all-trans β-A_{13}^t to α_R-A_{13}^t, one gets a destabilizing contribution of the backbone from strand to helix of about 2−3 kcal/mol per residue—an increment in line with the relative energy of α and β zones in a calculated Ramachandran map of alanine.

The three other calculated obvious secondary structures correspond to two left-handed helices, δ_L-A_{13}^c and α_L-A_{13}^c, and a right-handed one, α_R-A_{13}^c. Only δ_L-A_{13}^c affords stability in vacuum according to MPJE criterion (see Table 1). Despite several attempts, this structure does not correspond to a regular helix as it does not remain within the δ zone only ($[\phi,\psi] \approx [-150°,60°]$), but mixes in with the β zone (see Figure 3), all the more when direct transits between the two bassins are associated with low barriers, as calculated above in the alanine dipeptide models.

α_L-A_{13}^c, the left-handed helix corresponding to the α zone in the Ramachandran map ($[\phi,\psi] \approx [-60°,-60°]$),[19] is the only structure exhibiting intramolecular hydrogen bonds. Alternating horizontal and vertical peptide plaques along the mean helical axis result in a superhelix, in which the vertical plaques are connected to each other by hydrogen bonds roughly parallel to the helix axis. In terms of plaques, every other plaque, the NH group is linked to a CO counterpart *two plaques ahead* in the sequence.[22] In terms of residues, every other residue, the NH group is linked to the CO counterpart of the *next residue* in the sequence. A total of six such hydrogen bonds occur in the present tridecamer. In the median part of the chain, their typical geometry corresponds to d(H···O) = 2.05 Å and θ(N−H···O) = 148°, suggesting hydrogen bonds slightly weaker here than in classical trans α-helices, for which the present treatment on α_R-A_{13}^t gives average values of d = 2.08 Å and θ = 160°. For the record, in the relaxed *cis*-NMA dimer, typifying face-to-face coupling of two cis plaques—a connection scheme quite different from that in α_L-A_{13}^c—the corresponding hydrogen-bond parameters are calculated at d = 1.86 Å and θ = 179°. The energy stabilization brought by the six hydrogen bonds in the α_L-A_{13}^c helix can be estimated at about 6(4−5) ≈ 25−30 kcal/mol, scalling down to 2−2.5 kcal/mol per residue, which is less than half the increment available in a classical trans α-helix. Unlike this case, the hydrogen-bond bonus appears not sufficient here to surmount the energy unstability caracterizing the cis α_L zones in both the *residue* ϕ/ψ map and the *plaque* ϕ/ψ_{i-1} map discussed above.

Figure 4 illustrates an interesting feature of the α_L-A_{13}^c layout. Viewed from the top, the vertical plaques point to the interior of the helix, while the horizontal plaques are oriented outside. One can wonder about the possibility of building a

(22) Unlike in classical trans α-helix where the C=O groups point onwards along the helix direction, and the NH groups point backwards, in the vertical plaques of cis α_L-A_{13}^c, NH points onwards and CO points backwards, even if these bonds are here less parallel to the mean helical axis than in a the trans α-helix.

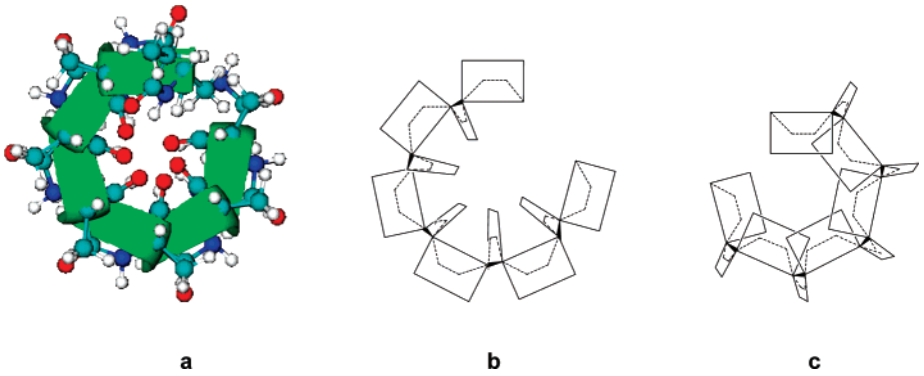


FIGURE 4. α_L^chelix: (a) top view; (b) schematic representation of peptide plaques, showing the outward horizontal plaques; (c) an upturned and unworkable alternative with inward horizontal plaques.

reverse layout, in which the vertical plaques would be outwards from the helix and the horizontal plaques inwards, as schematized in Figure 4c. Were such an orientation possible, it would open the way for self-association into a double helix, like DNA, with face-to-face paired cis plaques playing the role of planar Watson−Crick base pairing. Actually, such a construction is not geometrically feasible with α-polypeptides, but presumably β- or γ-polypeptides would be flexibile enough to make it achievable.

The right-handed helix α_R-A_{13}^c corresponds to a rather regular [ϕ,ψ] set around [60°,90°], roughly matching the α_L zone in a trans Ramachandran map. The resulting structure is quite elongated, with large pitch and weak diameter, so that one could be tempted to see it as a strand—α_R-A_{13}^c would be to ϵ_R-A_{13}^c what the β-strand is to the α-helix in the trans world. This is not the case, actually, as the only secondary structures allowed in the cis world should necessarily be helical in nature.

An alternative energy index has been proposed, still defined from an isodesmic reaction, but now relating a polypeptide chain to dipeptide units: the *mean junction coupling energy* (MJCE).[10] In the present polyalanine chain with ethylamide ends, the corresponding equation is written straightforwardly from the polycondensation of dipeptide-model units into the tridecamer with release of the consequent *N*-ethyl propionamide molecules:

13 EtCONH-CHMe-CONHEt →
EtCO-(NH-CHMe-CO−)$_{13}$-NHEt + 12 Et-CONH-Et

Brought to each pair of consecutive C_α residues—a junction of plaques—this formal equation becomes

(13/12) EtCONH-CHMe-CONHEt →
(1/12) EtCO-(NH-CHMe-CO)$_{13}$-NHEt + Et-CONH-Et

the energy of which defines the MJCE index. It translates the energy brought by the stabilization of the merging of dipeptides or the juxtaposition of residues—specifically residue lateral chains. Table 1 shows that both indexes happen to be nearly identical for the cis helices. Interestingly and encouragingly, MJCE, which can be viewed as chemically more meaningfull in producing oligopeptides from aminoacids, tends to relatively favor the cis helices over the trans structures even more than MPJE does. Arising from formal and conventional definitions, these indicators should not be over interpreted, but they consistently point to clear exothermicity in two cases of the cis-helix creation process.

Because the torsional coordinate around a cis-peptide bond, ω, is associated with an energy curve flatter than that around a trans-peptide bond,[23] this parameter could be called upon as an incidental degree of freedom in the fully relaxed all-*cis* sequences—in other words, an adjustment variable along the (ϕ,ψ,ω) backbone dihedral-angle continuum. This happens to some extent, actually, as significant deviations (up to 27° in the most stable helix ϵ_R^c) are observed.[24] Hence, it is the whole backbone that finds its way to alleviate the initial steric clash between lateral chains. To further evaluate the robustness of such backbone solutions, we have compared the converged geometries of ϵ_R^c-type minima for ethyl-amidified hexadecamers of alanine A_6^c and phenylalanine F_6^c. Regardless of extra adjustment parameters χ_1 and χ_2 in the latter, both backbones are alike, as reflected by the proper similarity of the 19 dihedral angles (ϕ,ψ,ω) constituting the whole backbone ($r^2 = 0.994$).

The helix parameters recapitulated in Table 1 show few differences from those of trans helices, except for the superhelix α_L^c, a 14_8 helix obviously shaped at larger scale. If ϵ_R^c logically reminds PPI, it is remarkable that this preferred form corresponds to the helical arrangement Pauling sensed as the most viable and interesting.[12]

Discussion

Irregular, nonhomogeneous structures have also been searched, starting from [ϕ,ψ] sets belonging to different zones on the above-discussed maps. Although most attempts led to high-energy arrangements, an interesting case is encountered with the structure shown in Figure 5, built from ϵ_R and α_L basins. Inspection of its backbone parameters suggests it can foreshadow either a large-diameter regular helix (the bias shown in Figure 5) or a more or less pronounced bend, which would open the way for tertiary organization. Remarkably, this structure has an MPJE index of −0.1 kcal/mol and an MJCE index of +0.1 kcal/mol, indicating fair stability as these numbers are in the range of those associated with the more regular helices. Further evidence for some structural diversity is given by variants to

(23) At the B3LYP/6-311G** level of calculation, a 25°-deviation of ω with respect to its ideal value ($\omega = 0°$) in *cis*-NMA costs only +0.3 kcal/mol, while the same deviation from ideal value ($\omega = 180°$) in *trans*-NMA costs +1.4 kcal/mol.

(24) The standard deviations for ω in the 14-peptide-bonds alanine tridecamers are calculated at 1.8° in β^t, 3.9° in α_R^t, 6.9° in ϵ_R^c, 13.7° in δ_L^c, 6.2° in α_L^c, 5.2 in α_R^c, and 8.7° in the A_{13}^c mixed form. For further discussion on these themes, see: MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1996**, *264*, 1180.

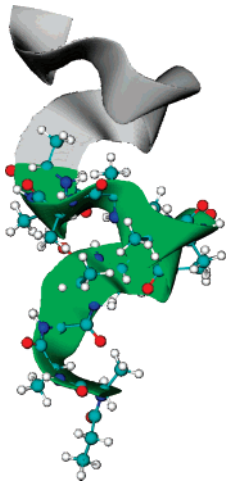


FIGURE 5. An example of a calculated nonregular helix, with an extrapolated upper part.

δ_L^c obtained when optimizing from different β/δ combinations of $[\phi,\psi]$ sets. Typically, one of these structures has the same energy as the above ϵ/α mixed form. Such favored "oddities" tend to point to some structural versatility and call for further careful explorations.

If all-*cis* polypeptide chains seem to hold a single class of *regular* secondary structures, helices, the present results tend to suggest these afford a certain polymorphism. However, little can be stated definitely on the extent of structural diversity all-*cis* proteins would sustain insofar as (i) the $\alpha/\beta/\delta/\epsilon$ combinatorics might be more wide-ranging than the present sampling suggests and (ii) many tertiary structures can be constructed with helices only, provided loops and bends are reachable. At this point of our explorations, we do have not enough results to answer these questions and more exploration of the energy hypersurfaces is needed. Nevertheless, one would be tempted to suggest overall reduced structural diversity, not because of steeper low-energy regions in the Ramachandran maps, but because of the interdependence between residues via the plaque maps. This reasonably suggests a reduced plasticity and flexibility in cis proteins with respect to the trans proteins.

Conclusions

In parallel with previous works on all-*cis* cyclic peptides, this study aimed to preliminarily explore, by theoretical means, the feasibility of all-*cis* open-chain polypeptides. Its key result lies in that the a priori dreaded steric clashes between lateral chains can be efficiently released through proper adjustment of dihedral angles along the backbone. This leads to helical chains that are more or less regular, and possibly further structured by hydrogen bonds. Again, once the cis/trans local irreducible penalty is satisfied, the association of such cis-prepared plaques is thermodynamically favorable in various instances.

Inevitably, this work also calls for several complementary calculations, refinements, and checks as many questions remain open. (i) How would heterochiral sequences modify the presently found energy and structural trends? Remember heterochirality prevents the formation of or widely alters the properties of β sheets in trans chains. (ii) What about β-, γ-, or δ-peptides? In these cases, more structural variety is expected, offering opportunities for further intramolecular structuring interactions and intermolecular self-assembling. (iii) Is globular folding possible from helix components separated by proper bends? How much flexibility would such compact arrangements possess? (iv) How could solvation influence all the above structures and relative energies? In most of them, the cis amide groups are oriented outside, therefore more available to solvent molecules than trans amide groups in classical secondary structures. (v) How would these helices interact, or pack together? The classical rules for fitting ridges into grooves applying in α-helices are no longer effective here, where plaque pairing should predominate instead.

While some of these points are currently being studied, reasonable conjectures about kinetic stability are permitted for now. It is obvious that once built, all-*cis* chains should remain efficiently locked, as isomerization in an unzipping way along the chain seems improbable due to the impractical steric upset that would ensue. Yet this could be fulfilled in vivo through elaborate and specific enzymatic machinery—still unknown.[25] Granted with sufficient kinetic stability the all-*cis* molecules could thus be viable species. Grounding on both thermodynamic and kinetic arguments, the prospect of such peptidomimetic alternatives should therefore promote active research for suitable polymerization processes.

Computational Section

All calculations were performed with the Gaussian 03 quantum chemistry package, at the DFT level of theory, using the B3LYP hybrid functional.[26] Geometry optimizations were achieved in the gas phase, without symmetry constraints, using the 6-31G(d) basis set. The quality of the B3LYP/6-31G(d) procedure has been assessed as a good cost/accuracy compromise for addressing structural and energy issues in the present class of molecules.[10] 2-D energy plots were calculated as fully relaxed (ϕ,ψ) maps, with ω constrained to its ideal cis value in order to avoid unwanted cis → trans inversions in high-energy regions. $[\phi,\psi]$ grids with 12° spacing were generated at the MPW1K/6-31G(d) level for both conformational maps of Figure 2, whereas the regular-helicity preliminary map of Figure 1 was generated at the Hartree−Fock/3-21G level of theory only, with 15° spacing. The MPW1K functional has been shown to be more reliable than B3LYP for kinetic applications.[27] Because the converged "regular" helical structures are nonideal helices, mostly due to finite-size effects, the helix parameters given in Table 1 are rounded coarse ruler-gauged values.

Acknowledgment. The authors thank the CALMIP and CINES computing centers for allocating computational ressources.

Supporting Information Available: Complete list of authors for ref 26, and total energies and Cartesian coordinates for all tridecamers. This material is available free of charge via the Internet at http://pubs.acs.org.

JO0711790

(25) If peptidyl-prolyl cis/trans isomerases are well documented (for a review, see: Fischer, G.; Aumüller, T. *Rev. Physiol. Biochem. Pharmacol.* **2003**, *148*, 105), few occurrences are reported for the non-prolyl secondary-amide cases; see for instance: Schiene-Fischer, C.; Habazettl, J.; Schmid, F. X.; Fischer, G. *Nat. Struct. Biol.* **2002**, *9*, 419.

(26) Frisch, M. J.; et al. *Gaussian 03*, revision B.05; Gaussian, Inc.: Wallingford, CT, 2004.

(27) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 48.