Evaluation of the Dipeptide Approximation in Peptide Modeling by ab Initio Geometry Optimizations of Oligopeptides

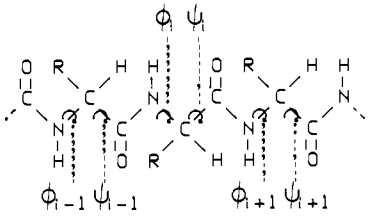
Lothar Schäfer,*,† Susan Q. Newton,† Ming Cao,† Anik Peeters,‡ C. Van Alsenoy,‡ Krzysztof Wolinski,†,§ and Frank A. Momany⊥

Contribution from the Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, Department of Chemistry, University of Antwerpen, B-2610 Wilrijk, Belgium, and Molecular Simulations Inc., 200 Fifth Avenue, Waltham, Massachusetts 02254. *Received June 17, 1992*

Abstract: HF/4-21G ab initio gradient geometry optimizations were performed on model tri- and hexapeptides to examine the validity and limitations of the widely used dipeptide approximation in empirical peptide conformational analyses. For the molecules *N*-formyl-Ala-Ala-amide, *N*-formylpentaglycine amide, and *N*-formylpentaalanine amide, several conformations were investigated, including repeated C7eq, repeated C5, helical, and bend forms. The results show that the order of low-energy regions obtained for dipeptides is significantly changed by long-range interactions in the hexapeptides. Helical forms are not energy minima in dipeptide space but are stable regions (nearly vanishing gradients) on the potential energy surface in the hexapeptides. β-Bends are stable conformations in the vacuum structures of extended chains, and interactions between two residues can significantly affect the torsional states of other residues in the bend. Compared with the binding energy of single residues in di- and tripeptides, cooperative energy effects are detected in extended chains. Significant variations with conformation are found in bond distances and angles, >7° in N-C(α)-C′, >4° in C′-C(α)-C(β), and up to 6° in C′-N-C(α). The C-N peptide bond length is shortened by multiple hydrogen bonding. This trend is important in view of an unexplained contrast in peptide structural chemistry, i.e., the observation that C-N peptide bonds in protein crystal structures are considerably shorter than C-N bonds in isolated amide units in the vapor phase. In general, conformationally dependent geometry changes are qualitatively similar in dipeptides and hexapeptides, but some parameter differences are enhanced by chain elongation and characteristic variations are observed where long-range interactions create special effects. Thus, the calculations confirm the importance of conformational geometry maps in describing peptide conformational properties, and they help to identify significant geometry trends that must be taken into account in the development of force field parameters for empirical peptide modeling.

Introduction

Peptide modeling first began with the work of Ramachandran and his group.[1] On the basis of these early studies, it was recognized that the flexibility of the backbone of peptide chains originates mainly from torsions about the N-C(α) and C(α)-C′ bonds. These torsional angles in peptides are customarily denoted as φ and ψ.



As the field evolved, peptide modeling proceeded from the hard sphere approximation[2] to molecular mechanics techniques with partitioned potential energy expressions[3] and to methods based on SCF procedures, such as extended Hückel,[4,5] CNDO/2,[5] PRDDO,[6] and PCILO.[7] A number of standard program packages, such as AMBER,[8] CHARMM,[9] GROMOS,[10] or INSIGHT,[11] are currently available for advanced empirical peptide conformational analyses.

Ab initio calculations of peptides first employed rigid (i.e., unoptimized) geometries for dipeptides[12] and oligopeptides.[13,14] In the early 1980s, the first full ab initio gradient geometry optimizations of peptides (*N*-acetyl-*N*′-methylamides of glycine and alanine) were performed.[15] These studies showed that conformational analyses with rigid geometries are insufficient, and conformational geometry maps[16] are needed to give a complete description of the conformational properties of such systems. A series of similar studies[17,18] has followed, including calculations

† University of Arkansas.
‡ University of Antwerpen.
§ Permanent address: Institute of Chemistry, Maria Curie-Sklodowska University, Lublin 20031, Poland.
⊥ Molecular Simulations, Inc.

(1) (a) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95. (b) Ramachandran, G. N. *Biopolymers* **1968**, *6*, 1494.
(2) (a) Ramakrishnan, C.; Ramachandran, G. N. *Biophys. J.* **1965**, *5*, 909. (b) Leach, S. J.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1966**, *4*, 369. (c) Sarathy, K. P.; Ramachandran, G. N. *Biopolymers* **1968**, *6*, 461. (d) Venkatachalam, C. M.; Ramachandran, G. N. *Annu. Rev. Biochem.* **1969**, *38*, 45. (e) Scheraga, H. A. *Adv. Phys. Org. Chem.* **1968**, *6*, 103. (f) Ramachandran, G. N. *Int. J. Prot. Chem.* **1969**, *1*, 5. (g) Scheraga, H. A. *Chem. Rev.* **1971**, *71*, 195.
(3) (a) Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 420. (b) Hagler, A.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319. (c) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361. (d) Zimmerman, S. S.; Pottle, M. S.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1. (e) Vasquez, M.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043. (f) Nemethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883. (g) Levitt, M. *J. Mol. Biol.* **1983**, *168*, 595. (h) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765. (i) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230. (j) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
(4) (a) Hoffmann, R.; Imamura, A. *Biopolymers* **1969**, *7*, 207. (b) Hopfinger, A. J.; Walton, A. G. *Biopolymers* **1970**, *9*, 29. (c) Govil, G. *J. Chem. Soc. A* **1970**, 2464.
(5) (a) Yan, J. F.; Momany, F. A.; Hoffmann, R.; Scheraga, H. A. *J. Phys. Chem.* **1970**, *74*, 420. (b) Momany, F. A.; McGuire, R. F.; Yan, J. F.; Scheraga, H. A. *J. Phys. Chem.* **1971**, *75*, 2286.
(6) Kleier, D. A.; Lipscomb, W. N. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1977**, *4*, 73.
(7) (a) Maigret, B.; Pullman, B.; Dreyfus, M. *J. Theor. Biol.* **1970**, *26*, 321. (b) Maigret, B.; Perahia, D.; Pullman, B. *J. Theor. Biol.* **1970**, *29*, 275. (c) Maigret, B.; Pullman, B.; Perahia, D. *J. Theor. Biol.* **1971**, *31*, 269. (d) Maigret, B.; Perahia, D.; Pullman, B. *Biopolymers* **1971**, *10*, 491. (e) Cabrol, D.; Broch, J.; Vasilescu, D. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1986**, *12*, 141.
(8) Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287.
(9) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(10) van Gunsteren, W. F.; Berendsen, H. J. C. *Biochem. Soc. Trans.* **1982**, *10*, 301.

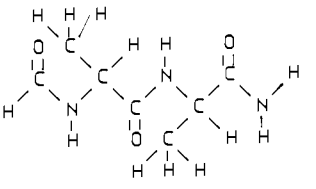
0002-7863/93/1515-272$04.00/0

with larger basis sets and MP2 single-point energy calculations[18] (MP2 energies calculated at SCF optimized, i.e., MP2 unoptimized geometries). As in the case of ab initio HF calculations, MP2-gradient optimizations[19] demonstrated that unoptimized MP2 energies are potentially inaccurate and not the "benchmark" results that they were originally taken[18] to be.

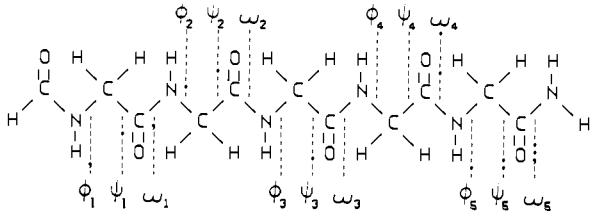
In developing parameters for empirical energy calculations, the "dipeptide approximation"[1,2,20] has often been applied. In this approximation it is assumed that the values of ϕ_i and ψ_i in the ith residue of a peptide chain depend mainly on one another and on the nature of the residue R_i but are largely independent of the neighboring pairs ϕ_{i-1},ψ_{i-1} and ϕ_{i+1},ψ_{i+1}. The model implies that essential conformational properties of polypeptides can be deduced from their isolated components. The approach has been quite successful in describing peptide conformational properties since short-range interactions are dominant in the folding of a polypeptide chain. At the same time, the model neglects cooperative phenomena in polymers and long-range interactions between groups that are remote from each other along the backbone of the peptide chain. Thus, characteristic differences in the properties of dipeptides and polypeptides were also found. For example, in contrast to many empirical potential energy studies, it is apparent from ab initio geometry optimizations of model dipeptides[15] that α-helical conformations, α_R, are not minima in dipeptide space, even though they are common in proteins. Furthermore, the C7eq form is the most stable dipeptide conformation in model calculations, but it is seldom found in proteins, probably as a result of competition between intramolecular hydrogen bonding and solvent effects.

In pursuing the main goal of peptide conformational analysis, i.e., computing the three-dimensional structures of polypeptides from their amino acid sequence, extensions beyond the dipeptide model are required. Thus, it is an interesting question how the properties of oligopeptides deviate from the sum of the properties of the component single residues. For some conformations, among them bends and helices, stabilization may arise from interactions between different single residues, and model calculations beyond the dipeptide approximation are needed for investigating such effects. In the past, such calculations were mainly performed with empirical procedures,[21] but there were also early attempts to identify cooperative effects in oligopeptides by using ab initio calculations.[14] By necessity, the latter were rigid geometry calculations. This was a disadvantage, recognized by their authors,[14] because such analyses can only be partially successful without geometry optimization.

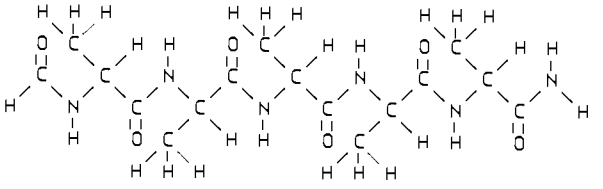
In the current paper we report the results of ab initio HF geometry optimizations of oligopeptides at a level (4-21G[22]) that allows for meaningful comparisons of calculated conformational properties, even though it is not as advanced as one might desire. For the alanine model tripeptide I (*N*-formyl-Ala-Ala-amide) and the model hexapeptides of glycine (II) and alanine (III) (C-terminal amide and N-terminal formyl derivatives of pentaglycine and pentaalanine), we have optimized repeating C5 and C7eq



I



II



III

conformations, as well as bend and helical conformers, in order to determine whether the conformational differences between dipeptides and polypeptides, as previously determined, are true features of the systems involved or artifacts of computation. Similarly, some time ago conformationally dependent geometry trends of dipeptides[15,16] were found to be of great significance for peptide modeling,[23] and it is important to determine to what extent such trends are accurate for larger systems.

Bend structures in proteins were first classified by Venkatachalam[24] and later by Lewis et al.[25] It is an interesting question whether such peptide turns are intrinsically stable or created by solvation or by any other environmental factors. In addition, to develop a thorough understanding of the stability of such systems, one must explore the variability of the conformational states of the residues in a bend. Since little experimental information is

(11) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Proteins: Struct., Funct., Genet. **1988**, *4*, 31.
(12) (a) Ryan, J. A.; Whitten, J. L. *J. Am. Chem. Soc.* **1972**, *94*, 2396. (b) Peters, D.; Peters, J. *J. Mol. Struct.* **1979**, *53*, 103. (c) Sawaryn, A.; Yadav, J. S. *Int. J. Quantum Chem.* **1982**, *22*, 547. (d) Peters, D.; Peters, J. *J. Mol. Struct.* **1982**, *88*, 137. (e) Peters, D.; Peters, J. *J. Mol. Struct.* **1982**, *90*, 305. (f) Stern, P. S.; Chorev, M.; Goodman, M.; Hagler, A. T. *Biopolymers* **1983**, *22*, 1901. (g) Peters, D.; Peters, J. *J. Mol. Struct.* **1984**, *109*, 137. (h) Peters, D.; Peters, J. *J. Mol. Struct.* **1984**, *109*, 149. (i) Williams, D. E. *Biopolymers* **1990**, *29*, 1367.
(13) (a) Peters, D.; Peters, J. *J. Mol. Struct.* **1980**, *68*, 243. (b) Peters, D.; Peters, J. *J. Mol. Struct.* **1980**, *69*, 249. (c) Wright, L. R.; Borkman, R. F. *J. Phys. Chem.* **1982**, *86*, 3956. (d) Mavri, J.; Avbelj, F.; Hadzi, D. *J. Mol. Struct.* **1989**, *187*, 307.
(14) van Duijnen, P. T.; Thole, B. T. *Biopolymers* **1982**, *21*, 1749.
(15) (a) Schäfer, L.; van Alsenoy, C.; Scarsdale, J. N. *J. Chem. Phys.* **1982**, *76*, 1439. (b) Scarsdale, J. N.; van Alsenoy, C.; Klimkowski, V. J.; Schäfer, L.; Momany, F. A. *J. Am. Chem. Soc.* **1983**, *105*, 3438. (c) Schäfer, L.; Klimkowski, V. J.; Momany, F. A.; Chuman, H.; van Alsenoy, C. *Biopolymers* **1984**, *23*, 2335. (d) Klimkowski, V. J.; Schäfer, L.; Momany, F. A.; van Alsenoy, C. *J. Mol. Struct.* **1985**, *124*, 143.
(16) (a) Siam, K.; Klimkowski, V. J.; van Alsenoy, C.; Ewbank, J. D.; Schäfer, L. *J. Mol. Struct.* **1987**, *152*, 261. (b) Siam, K.; Kulp, S. Q.; Ewbank, J. D.; Schäfer, L.; van Alsenoy, C. *J. Mol. Struct.* **1989**, *184*, 143.
(17) (a) Weiner, S. J.; Singh, U. C.; O'Donnell, T. J.; Kollman, P. A. *J. Am. Chem. Soc.* **1984**, *106*, 6243. (b) Sapse, A. M.; Fugler, L. M.; Cowburn, D. *Int. J. Quantum Chem.* **1986**, *29*, 1241. (c) Cheam, T. C.; Krimm, S. *J. Mol. Struct.* **1989**, *188*, 15. (d) Cheam, T. C.; Krimm, S. *J. Mol. Struct.* **1989**, *193*, 1. (e) Cheam, T. C.; Krimm, S. *J. Mol. Struct.* **1990**, *206*, 173. (f) Barone, V.; Fraternali, F.; Cristinziano, P. L. *Macromolecules* **1990**, *23*, 2038. (g) Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J. L.; Marcoccia, J. F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256. (h) Böhm, H. J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129.
(18) (a) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C.; Pople, J. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1989**, *16*, 311. (b) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989.
(19) Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schäfer, L. *J. Am. Chem. Soc.* **1992**, *114*, 5369.
(20) (a) Gibson, K. D.; Scheraga, H. A. *Biopolymers* **1966**, *4*, 709. (b) Ponnuswamy, P. K.; Sasisekharan, V. *Biopolymers* **1971**, *10*, 565. (c) Ramachandran, G. N.; Venkatachalam, C. M.; Krimm, S. *Biophys. J.* **1966**, *6*, 849. (d) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Isr. J. Chem.* **1973**, *11*, 121. (e) Pullman, B.; Pullman, A. *Adv. Protein Chem.* **1974**, *28*, 347.

(21) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797.
(22) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. *J. Am. Chem. Soc.* **1979**, *101*, 2550.
(23) Momany, F. A.; Klimkowski, V. J.; Schäfer, L. *J. Comput. Chem.* **1990**, *11*, 654.
(24) Venkatachalam, C. M. *Biopolymers* **1968**, *6*, 1425.
(25) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Biochim. Biophys. Acta* **1973**, *303*, 211.

known (outside the environment of a protein) that can be used to explore such questions, it is useful to investigate relevant model systems by computational techniques.

The results of the ab initio geometry optimizations of several structures of I–III, presented below, will show that (1) bend conformations can be intrinsically stable in vacuum systems, (2) in contrast to dipeptides, helical forms are stable regions on the potential energy surfaces of hexapeptides, and (3) repeating C7eq forms are not the most stable conformations of oligopeptides, even though C7eq is the global energy minimum in dipeptides. Furthermore, the geometry trends established for dipeptides are also displayed by the larger systems, with characteristic variations, and important trends are found for the peptide link as a function of chain length and hydrogen bonding.

Computational Procedures

All computations were executed with the program BRABO,[26] using the 4-21G basis set[22] and standard gradient procedures.[27] The ab initio gradients were relaxed in normal coordinate space[28] until all gradients on the atoms in Cartesian coordinates were <0.001 mdyn. Since a positive definite approximate Hessian was used to relax the normal coordinate forces, it is very likely that the geometries obtained have approached minima; they could be nonminimum stationary points only if, by accident, one normal coordinate force corresponding to a negative curvature was 0 in the initial geometry and in subsequent iterations. This is very unlikely in a nonsymmetric system, but positive proof is obtained only by calculating the second-order derivatives at the 4-21G level. Due to the size of the molecules investigated, this was not possible.

The structures selected for investigation are not the result of a complete conformational search. Rather, they were selected because their geometries represent characteristic regions of conformational space that are of interest for various reasons, as discussed below. It was specifically not the intent of this study to find the global energy minima of the investigated systems. Furthermore, if the potential energy surface is flat, then the first-order derivatives do not contain any information on how closely the optimized conformations have approached the exact locations of minimum energy.

Thus, the reported structures represent very likely, but not necessarily, regions (not points) around energy minima. In any case, these regions in conformational space are stable in the sense that the first-order derivatives are negligibly small. The primary structures, i.e., all bond distances and angles, are highly relaxed at the torsional angles listed. The exact nature of the results is emphasized because the neglect of similar limitations, defined in the early dipeptide work,[15] has at times led to some redundancy in subsequent analyses,[18] as pointed out before.[19]

All compounds are characterized by N-terminal formyl and C-terminal amide groups, the peptide bonds are taken to be in the trans configuration, and geometries were relaxed without any constraints. In the case of I (*N*-formyl-Ala-Ala-amide), two structures are included from a more extensive investigation of this system,[29] the repeating C7eq (I-C7) and C5 (I-C5) forms. In the case of II (*N*-formylpentaglycine amide), four structures are reported: the repeating C7eq (II-C7) form, the repeating C5 (II-C5) form, a β-bend structure (II-BET), and a helical structure obtained by refining an α-helical conformation (II-HEL). In the case of III (*N*-formylpentaalanine amide), the equivalent structures are III-C7, III-C5, III-BET, and III-HEL.

In the latter series, III-BET was not as fully optimized as the other conformations presented in this paper. After more than 100 cycles of geometry optimization, beginning with a structure close to that of II-BET, some of its Cartesian gradients are still on the order of 0.005 mdyn. The partially optimized structure is useful for our discussion because it shows in what direction pentaalanine moved away from the β-bend structure of pentaglycine. Furthermore, its bond distances and angles are sufficiently relaxed at the specified torsional angles to allow for meaningful comparisons with other systems. At the same time, additional refinements of this form are not of interest for the current investigation.

On the basis of recent MP2-gradient geometry optimizations of several forms of glycine[30] and the blocked alanine residue *N*-formylalanine amide,[19] the following error estimates can be given for the 4-21G results of this study: in the investigated systems, 4-21G energies were found to be closer to MP2 energies or experimental values than were results obtained with larger basis sets, with inaccuracies[19,30] of 0.3–1.5 kcal/mol for 4-21G and 1.4–2.4 kcal/mol for basis sets with polarization functions. This result may be surprising and is perhaps due to some fortuitous cancellation of errors. At the same time, it shows that the 4-21G basis set is well balanced for compounds of the kind considered here. Similarly, essential geometry trends (differences between comparable bond distances and angles in different conformations) are well determined at the 4-21G level, but the exact minima of torsional angles were found[19] to be inaccurate by up to 15°. Thus, future investigations of oligopeptides using correlation-gradient optimizations will yield improved results, but they are not possible now. In any case, the current 4-21G error estimates are small enough to support a number of important conclusions, as discussed below.

Table I. 4-21G Optimized Geometry Parameters for *N*-Formyl-Ala-Ala-Amide[a]

	C7	C5
Bond Lengths (Å)		
residue 1		
C=O	1.2262	1.2225
C—N	1.3466	1.3452
N—C(α)	1.4729	1.4556
C(α)–C′	1.5393	1.5276
C′=O	1.2329	1.2305
residue 2		
C—N	1.3410	1.3421
N—C(α)	1.4723	1.4558
C(α)—C′	1.5403	1.5279
C′=O	1.2211	1.2234
C′—N	1.3515	1.3515
Bond Angles (deg)		
residue 1		
O=C—N	123.98	124.41
C—N—C(α)	121.77	121.39
N—C(α)—C′	110.14	106.23
C(α)—C=O	121.56	121.62
C(α)—C′—N	114.85	115.62
O=C′—N	123.59	122.76
residue 2		
C—N—C(α)	122.21	121.47
N—C(α)—C′	110.35	106.04
C(α)—C=O	122.03	121.48
C(α)—C′—N	113.75	115.17
O=C′—N	124.22	123.34
Torsional Angles (deg)		
residue 1		
ϕ	−84.3	−166.7
ψ	66.4	170.5
ω	−174.3	177.1
residue 2		
ϕ	−84.8	−165.6
ψ	65.6	169.9

[a] Columns C5 and C7 give the parameters for I-C5 and I-C7. The total energy of I-C7 is −659.132887 au. I-C5 is 2.34 kcal/mol less stable than I-C7.

Results and Discussion

Calculated geometries and energies are presented in Tables I and II. The carboxyl-terminus ϕ and ψ values of calculated helical structures are compared with experimental results in Table III. The conformations of I–III considered in this study are shown in Figures 1–3. Some selected structural features are summarized in Figure 4,

Conformational Energy Trends. (A) The Helical Conformations. In 4-21G calculations of dipeptides (*N*-acetyl-*N*′-methylamides of glycine and alanine),[15] no energy minima were found in the vicinity of α_R. In contrast to this, 4-21G optimizations of II and III have led to stable regions (vanishing first-order gradients) in the helical part of the potential energy surface, as shown in Table II. The results show that computational procedures that do not yield helical energy minima for dipeptides can establish stable helical forms for larger systems in which stabilizing long-range interactions are possible.

(26) van Alsenoy, C. *J. Comput. Chem.* **1988**, *9*, 620.
(27) (a) Pulay, P. *Mol. Phys.* **1969**, *17*, 197. (b) Pulay, P. *Theor. Chim. Acta* **1979**, *50*, 299.
(28) Sellers, H. L.; Klimkowski, V. J.; Schäfer, L. *Chem. Phys. Lett.* **1978**, *58*, 541.
(29) Newton, S. Q.; Cao, M.; Schäfer, L.; Perczel, A.; Csizmadia, I. G.; Momany, F. A., to be published.
(30) Ramek, M.; Cheng, V. K. W.; Frey, R. F.; Newton, S. Q.; Schäfer, L. *J. Mol. Struct.* **1991**, *235*, 1.

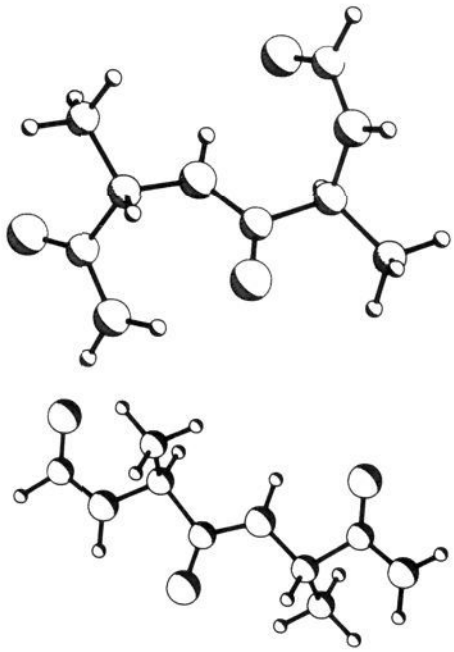


Figure 1. 4-21G optimized structures of *N*-formylalanine-alanine amide. Repeated C7eq conformation (top); repeated C5 conformation (bottom).

The long-range interactions that stabilize II-HEL and III-HEL involve a number of nonbonded H···O distances that are found in both systems in the 2.0–2.2-Å range. Such interactions exist between N—H in residue 4 and C═O in residue 1 (2.11 Å in II, 2.14 Å in III), the N-terminal formyl—oxygen and N—H in residue 3 (2.04 Å in II, 2.08 Å in III), C═O in residue 2, N—H in residue 5 (2.13 Å in II, 2.11 Å in III), and the C-terminal amide hydrogen and C═O in residue 3 (2.15 Å in II, 2.19 Å in III). Thus, the optimized helical structures of II and III contain H···O interactions between groups that are separated by two residues.

The helices obtained at the 4-21G level for II and III (Table II) are closer to type 3_{10}[31] than α_R. Considering a potential uncertainty of 15° in 4-21G ϕ and ψ angles, this aspect may be an artifact of the computational procedure. Nevertheless, the calculated structures fit experimental results very well. In particular, it is seen from Table III that in a number of nonhomologous proteins in which helical sections are found, their carboxyl-terminal ϕ and ψ angles (last two residues) show features that are in close agreement with those calculated here for the vacuum molecules. When the structure of III-HEL was optimized empirically using the QUANTA3.3/CHARMm parameters,[32] the ϕ,ψ torsional angles of residues 1–5 assumed the values (−50,−44), (−59,−42), (−69,−35), (−77,−29), (76,−31), i.e., residues are found both in the α_R and close to the 3_{10} regions.

(B) The β-Turn. The starting structure for the refinement of the β-turn of the glycine hexapeptide was a classic type II bend form[25] of the turn plus β-sheet conformations on each end. This form was chosen because it was obtained by geometry optimization using CHARMm.[32]

The structure resulting from 4-21G gradient relaxation, II-BET, did not retain the starting conformation. The dihedral angles (Table II) are quite unique in that residue 2 is puckered from the extended form ($\phi = -128, \psi = -149$), while residue 3 moved away from a typical turn into the C7eq conformation. Examination of the structure of II-BET (Figure 2) reveals that this unconventional turn very likely is the result of interactions between the C-terminal amide and N-terminal formyl groups. The hydrogen bond from the N-terminal formyl oxygen to the C-terminal amide hydrogen (2.1 Å) and of the N–H of residue 2 to the C-terminal amide carbonyl oxygen (1.85 Å) pulls the longer stretch of the β-sheet (i.e., residues 1 and 2) to the shorter stretch (residue 5 and end-group amide) to form the resulting favorable structure. Apparently the best way to establish these interactions is to pucker residue 2, because this pucker allows residue 1 to tilt toward the C-terminal amide.

Examples of dihedral angles for glycine in the regions described above are found in the X-ray structure of cyclo-(Gly-Ala-Gly-Ala-Ala-Gly),[33] in which the ϕ,ψ values of the Gly–Gly sequence are (139,158) and (84,−113). With the exception of the signs (which are not important for Gly), the dihedral angles of residues 2 and 3 of II-BET, (−128,−149) and (−78,77), are not too dissimilar. A turn similar to that of II-BET, but involving different residues, is found in Bovine ribonuclease S,[25] in which a type II bend is found involving the T G S S residues 87–90 with $i + 1$ and $i + 2$ ϕ,ψ values of (−90,100) and (150,−40). These values are within 25° of residues 3 and 4 of II-BET, i.e., (−78,77) and (149,−22).

In general, the β-turn II-BET is not common in proteins. It is important because it shows that the interaction between two residues, in this case the end groups, can significantly affect the torsional states of other residues in the bend. Thus, it illustrates the considerable flexibility of peptide chains in forming chain reversals. It is also an interesting aspect of this form that, without solvent interaction, it is more stable than the all-C7eq form II-C7. This feature, together with similar aspects of II-HEL, demonstrates that long-range interactions in oligopeptides can change the order of energy minima relative to that found for dipeptides. QUANTA3.3/CHARMm calculations yield similar results. All of these findings make it unlikely that the lack of stationary helical forms in dipeptides[15] is an artifact of the computational procedures applied.

When a structure for the alanine hexapeptide III is optimized starting with the dihedral angles of II-BET, residue 3 retains the C7eq conformation while residues 1 and 2 move toward the extended region. As a consequence the H···O interaction between the C-terminal amide and N-terminal formyl groups that is characteristic for II-BET is eliminated. It seems that this conformation is moving toward a totally different minimum. Its exploration is not part of the scope of the current study, but additional modeling calculations were performed to investigate this interesting difference between II-BET and the alanine analog.

When a methyl group is substituted at the α-carbon of residue 2 in II-BET, the dihedral angles of this residue (−128,−149) position the amide hydrogen of residue 3 close (approximately 1.9 Å) to the hydrogen atoms of the alanine side-chain methyl group. This destabilizing interaction is removed if the ϕ and ψ angles open toward the β-sheet (C5) conformational region. As residue 2 flattens, the long side (residues 1–3) of the β-bend conformation lengthens and the stabilizing hydrogen bond between the *N*-formyl carbonyl oxygen and the C-terminal amide hydrogen starts to lengthen and weaken. At residue 4, another change occurs upon adding the methyl side chain. In this case, the glycine dihedral angles (149,−22) are located in a conformational region that is energetically favorable for the glycine dipeptide but unfavorable in L-alanine dipeptide space. In this case, the carbonyl oxygen of residue 3 must move away from the methyl side-chain atoms of residue 4, which changes the ϕ value of residue 4 to −179°. This change, too, adds strain to the end-to-end interaction that characterizes II-BET. Similar small changes from other residues, all due to side-chain backbone interactions, contribute to this strain, each with a small but destabilizing contribution to a III-BET form that is similar to II-BET. The *accumulation* of these small energy changes, which are consistent with observed differences in dipeptide ϕ–ψ isoenergetic contour maps, is sufficient to cause large conformational differences in II-BET upon addition of the alanine methyl side chains. Substitution of a *single* methyl group, for example at residue 2, is not sufficient to break the end-to-end hydrogen bond.

(C) C5 and C7eq Forms. There are very few examples in the experimental protein crystal literature of residues in the C7eq

(31) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1968**, *201*, 601.
(32) QUANTA/CHARMm Release 3.3, 1992, Molecular Simulations Inc., 200 Fifth Ave., Waltham, MA 02254.
(33) Hossain, M. B.; van der Helm, D. *J. Am. Chem. Soc.* **1978**, *100*, 5191.

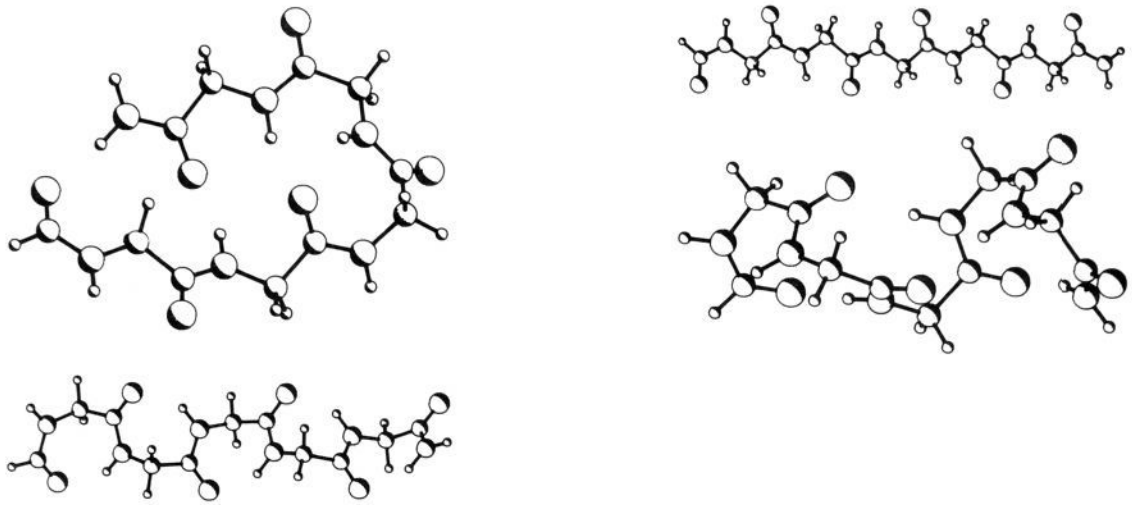


Figure 2. 4-21G optimized structures of *N*-formylpentaglycine amide. Repeated C7eq conformation (lower left); β-bend (upper left); repeated C5 conformation (upper right); helical conformation (lower right).

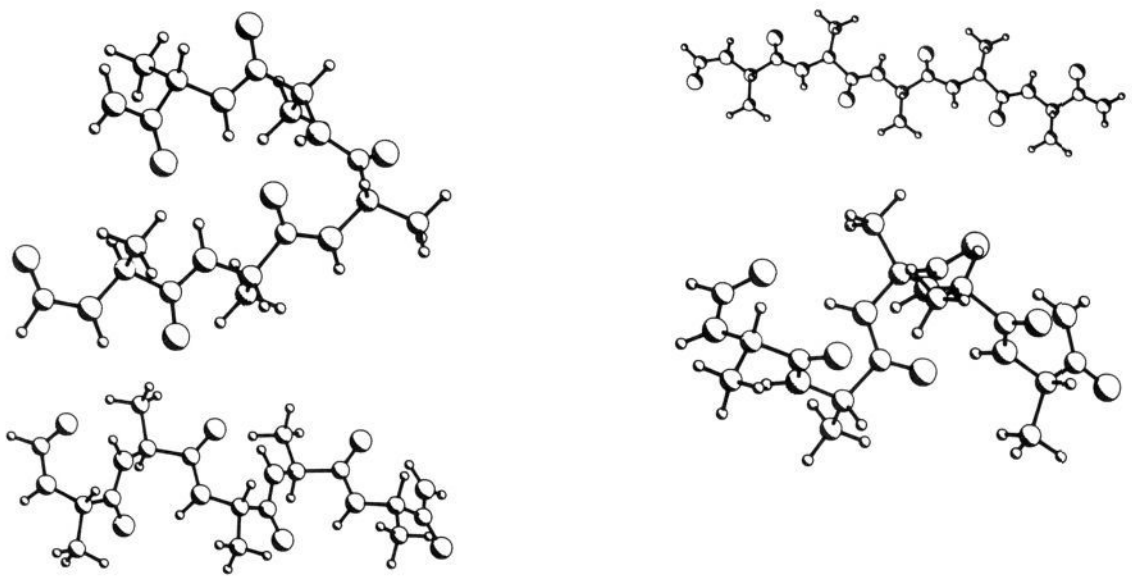


Figure 3. 4-21G optimized structure of *N*-formylpentaalanine amide. Repeated C7eq conformation (lower left); β-bend (upper left, partially optimized); repeated C5 conformation (upper right); helical conformation (lower right).

conformation, presumably because of a free energy preference for β-sheet, helix, or bend conformations. However, in all vacuum calculations of dipeptides, the C7eq form is low energy. Furthermore, it has been observed in solution studies of model dipeptides.[34]

In agreement with the dipeptide results, II-C7 and III-C7 represent stable regions on the potential energy surfaces of the hexapeptides, but they are not the lowest-energy forms. There are stabilizing hydrogen bonds, in each case involving the carbonyl oxygen in residue i and the amide hydrogen in residue $i + 2$, at approximately 1.96 Å in II-C7 and 1.98 Å in III-C7. The helical and bend forms of II are more stable than II-C7, and III-HEL is more stable than III-C7. For II-BET and II-C7, the difference is outside the expected error limits of 4-21G energies.

A recent study[19] in which the C5 and C7eq geometries of *N*-formylalanine amide were optimized at the MP2/6-311G** level revealed a peculiar handicap for HF/SCF calculations of peptides: at the important locations of the two most stable conformations, small basis sets like 3-21G and 4-21G yield energies relatively close to MP2 values (differences of about 0.4 kcal/mol) but torsional angles are poor (differences of up to 15°), whereas for large basis sets like 6-31+G* and 6-311G**, the locations of the minima are acceptable but the energies are poor (differences of about 1.5 kcal/mol). Thus, it is apparent that one must consider how the results of SCF calculations of peptides change when correlation effects are included.

The correlation results[19] were included in the refinement of the QUANTA3.3/CHARMm parameters.[32] When these parameters are used to calculate the geometry of III-C7, the resulting ϕ and ψ values range from −75 to −80° and from 73 to 84°. These values are closer to the MP2/6-311G** result of the blocked Ala residue, i.e., (−81,82), than those of the HF/4-21G III-C7 (Table II), in which the ϕ and ψ values are approximately −84° and 65°. The empirical calculations may be an indication of how the HF/4-21G geometries of II-C7 and III-C7 might change when electron correlation effects are included with larger basis sets in the ab initio calculations. Currently there is no other way to estimate this effect for the vacuum structures.

Table II shows that repeated C5 forms are considerably less stable than repeated C7eq forms. In contrast to the dipeptide energy pattern, they are the least stable forms of this series of

(34) (a) Avignon, M.; Houng, P. V.; Lascombe, J.; Marraud, M.; Neel, J. *Biopolymers* **1969**, *8*, 69. (b) Avignon, M.; Garrigou-Lagrange, C.; Bothorel, P. *J. Chim. Phys. Phys.-Chim. Biol.* **1972**, *69*, 62. (c) Avignon, M.; Lascombe, J. In *Conformation of Biological Molecules and Polymers*; Bergman, E. D., Pullman, B., Eds.; Academic Press: New York, 1973; p 97. (d) Zuk, W. M.; Freedman, T. B.; Nafie, L. A. *Biopolymers* **1989**, *28*, 2025.

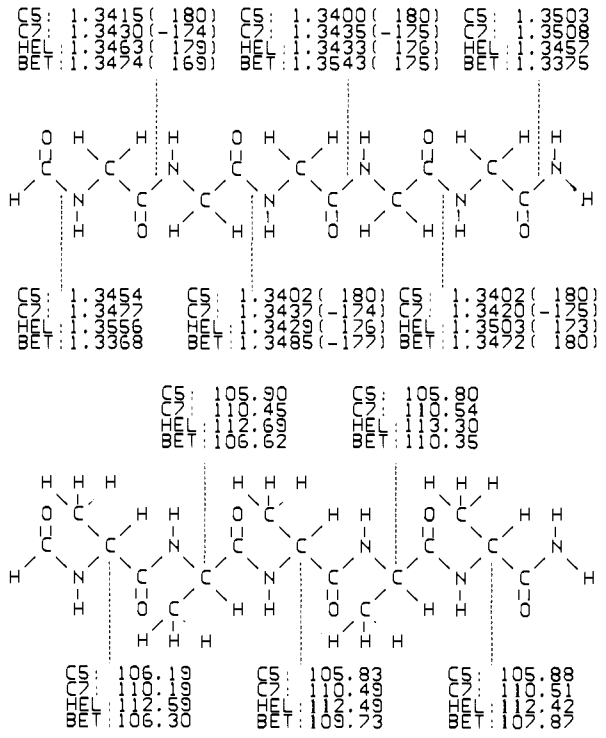


Figure 4. Local geometry trends: summary of the 4-21G optimized C–N peptide bond lengths in various conformations of *N*-formylpentaglycine amide and of the N–C(α)–C′ bond angles in *N*-formylpentaalanine amide.

II and III. This result suggests that the stability of β-sheets in proteins is a complicated tradeoff of inter- and intraresidue interactions,

(D) Cooperative Effects. Some time ago the strain energy of hydrocarbons was estimated from the energies of characteristic groups, such as CH_3 or CH_2, which had been derived from ab initio calculations of unstrained molecules.[35] Similarly, cooperative energy effects in hydrocarbons were recently derived by comparing the ab initio energies of some of their component conformational units.[36] In the case of peptides, the current study makes it possible to discuss cooperative energy effects in a similar way by considering the energy increments that can be associated with a single residue in characteristic conformational states.

The HF/4-21G total energy of the C7eq form of the blocked residue *N*-formylalanine amide is E(a) = −259 682.73 kcal/mol, and that of the I-C7 form of *N*-formyl-Ala-Ala-amide is E(aa) = −413 612,02 kcal/mol. The difference of these values, Δ, corresponds to the HF/4-21G energy increment that results when a single peptide fragment ((NH)C(α)(H,CH_3)(CO)) is inserted into a peptide chain; i.e., it is bonded to C′ (i − 1) on one end and to N (i + 1) on the other.

In the pentaalanine conformation III-C7, five peptide fragments, (NH)C(α)(H,CH_3)(CO), are linked in a chain. If the total energy of III-C7 were devoid of any cooperative effects, it should be equal to $3\Delta + E$(aa) = $4\Delta + E$(a) = −875 399.9 kcal/mol. However, the HF/4-21G energy of III-C7 is −875 402.2 kcal/mol (Table II). Thus, there is a difference of 2.3 kcal/mol that we ascribe to cooperative effects that are present in the hexapeptide but not in the dipeptide or tripeptide.

Similarly, for the C5 forms, the HF/4-21G energy of the dipeptide is E(a) = −259 681.34 kcal/mol, and that of the tripeptide is E(aa) = −413 609.68 kcal/mol; $3\Delta + E$(aa) = $4\Delta + E$(a) = −875 394.7 kcal/mol. The HF/4-21G energy of III-C5 (Table II) is −875 395.3 kcal/mol. Thus, there is a small cooperative energy difference of 0.6 kcal/mol. For helical forms, particularly large cooperative effects of this kind have been determined before[14] by standard geometry minimal basis ab initio calculations of oligopeptides of glycine and alanine. We cannot present a similar analysis for II-HEL or III-HEL because helical forms are not energy minima in di- and tripeptide space and HF/4-21G energies of constrained geometries of *N*-formylalanine amide and of I with the ψ and ϕ values of the hexapeptides are not available.

The quantitative aspects of the phenomena described in this section will undoubtedly change when correlation-gradient geometry optimizations of oligopeptides become possible. Nevertheless, it is likely that the HF/4-21G results describe reasonable trends. Such a performance was observed, for example, when more subtle cooperative energy increments associated with GG sequences in hydrocarbons were investigated at the HF/4-21G and MP4SDQ/6-31G*//HF/6-31G* levels.[36] Thus, on the basis of the 4-21G results discussed above, we expect that in the repeated C5 and C7eq forms of Ala, chain elongations will involve cooperative energy stabilization of a few tenths of a kcal/mol per residue.

Cooperative effects were also found in the dipole moments of oligopeptides of Gly and Ala.[14] The total 4-21G dipole moments for the C5 form of *N*-formylalanine amide, for I-C5 and III-C5, are 3.2, 6.3, and 10.8 D, respectively, and 3.4, 6.9, and 13.9 D for the corresponding C7eq forms. A quantitative analysis of the dependence of dipole moments on chain length, like that by van Duijnen et al.,[14] will require information on step-by-step elongated peptide chains that is currently not available for our calculations.

Conformational Geometry Trends. One of the most important general results arising out of the first ab initio gradient geometry optimizations of dipeptides[15] was the demonstration that the internal coordinates of a peptide can change significantly from one point of the conformational energy surface to another. This result was in contrast to a basic assumption adhered to for many years, i.e., that conformationally dependent local geometry changes can be neglected in empirical peptide conformational analyses. In general,[37] the early gradient optimizations of organic compounds revealed more clearly and more comprehensively than the common experimental techniques the diversity of structural details in organic molecules, illustrating differences in bond distances and angles that had often been taken to be equal.

A recent analysis[19] of the performance of different ab initio procedures that included MP2-gradient optimizations of the two most stable energy minima of *N*-formylalanine amide has shown that conformationally dependent geometry changes, i.e., differences between comparable bond distances and angles in different conformations of a system, to some extent depend on the computational method, but the fluctuations do not preclude the identification of some clear and useful structural trends. Thus, even at simple levels of theory such as HF/4-21G, conformational geometry maps of peptides capture structural trends that are not falsified by more advanced calculations.

One of the significant structural variations derived from dipeptide HF/4-21G geometries involves the backbone angle N–C(α)–C′.[15,16] In the HF/4-21G structures of the blocked single residues (*N*-acetyl-*N*′-methylamides) of Gly, Ala, and Ser,[15,16] this angle is characteristically small in the C5 forms (106–108°), intermediate in C7eq (110–112°), and large in the bridge region (114–116°). In view of the large number of proteins that have residues in the bridge region (ψ = 0),[3d] the ability to flex this angle to such an extent is important for the purpose of releasing steric strain. In fact, significant changes in the force field parameters of CHARMm[32] resulted when this and similar geometry variations were considered in the parameter development.

The current study for the first time makes it possible to test to what extent dipeptide conformational geometry trends are also applicable to larger systems. Several bond angles in the oligopeptides investigated for this paper display significant conformational changes. For example, C′–N–C(α) varies by nearly 6°,

(35) van Alsenoy, C.; Scarsdale, J. N.; Schäfer, L. *J. Comput. Chem.* **1982**, *3*, 53.

(36) Tsuzuki, S.; Schäfer, L.; Goto, H.; Jemmis, E. D.; Hosoya, H.; Siam, K.; Tanabe, K.; Osawa, E. *J. Am. Chem. Soc.* **1991**, *113*, 4665.

(37) (a) Schäfer, L. *J. Mol. Struct.* **1983**, *100*, 51. (b) Schäfer, L.; van Alsenoy, C.; Scarsdale, J. N. *J. Mol. Struct.* **1982**, *86*, 349.

Table II. 4-21G Optimized Structural Parameters and Conformational Energies for Various Conformations of *N*-Formylpentaglycine Amide and *N*-Formylpentaalanine Amide

	HEL		C5		C7		BET	
	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala
Bond Lengths (Å)								
residue 1								
C═O	1.2217	1.2228	1.2210	1.2228	1.2260	1.2267	1.2321	1.2243
C—N	1.3556	1.3540	1.3454	1.3446	1.3477	1.3467	1.3368	1.3424
N—C(α)	1.4625	1.4670	1.4490	1.4557	1.4655	1.4727	1.4479	1.4565
C(α)—C′	1.5263	1.5274	1.5235	1.5283	1.5309	1.5385	1.5154	1.5269
C′═O	1.2257	1.2284	1.2274	1.2306	1.2319	1.2328	1.2240	1.2326
C(α)—C(β)		1.5389		1.5440		1.5282		1.5440
residue 2								
C—N	1.3463	1.3449	1.3415	1.3410	1.3430	1.3430	1.3474	1.3390
N—C(α)	1.4562	1.4605	1.4503	1.4570	1.4622	1.4700	1.4425	1.4528
C(α)—C′	1.5294	1.5308	1.5243	1.5290	1.5309	1.5384	1.5294	1.5291
C′═O	1.2286	1.2323	1.2275	1.2309	1.2322	1.2330	1.2294	1.2358
C(α)—C(β)		1.5376		1.5436		1.5291		1.5444
residue 3								
C—N	1.3429	1.3414	1.3402	1.3399	1.3437	1.3441	1.3485	1.3441
N—C(α)	1.4546	1.4607	1.4508	1.4574	1.4615	1.4689	1.4667	1.4772
C(α)—C′	1.5295	1.5323	1.5244	1.5291	1.5310	1.5383	1.5372	1.5448
C′═O	1.2302	1.2326	1.2275	1.2308	1.2324	1.2332	1.2214	1.2232
C(α)—C(β)		1.5380		1.5435		1.5290		1.5277
residue 4								
C—N	1.3433	1.3427	1.3400	1.3397	1.3435	1.3438	1.3543	1.3496
N—C(α)	1.4534	1.4591	1.4511	1.4576	1.4615	1.4687	1.4564	1.4632
C(α)—C′	1.5326	1.5376	1.5241	1.5288	1.5318	1.5391	1.5243	1.5271
C′═O	1.2242	1.2267	1.2276	1.2309	1.2329	1.2339	1.2244	1.2255
C(α)—C(β)		1.5362		1.5435		1.5289		1.5465
residue 5								
C—N	1.3503	1.3475	1.3402	1.3401	1.3420	1.3422	1.3472	1.3545
N—C(α)	1.4518	1.4559	1.4499	1.4566	1.4635	1.4705	1.4511	1.4566
C(α)—C′	1.5307	1.5368	1.5247	1.5283	1.5340	1.5401	1.5237	1.5247
C′═O	1.2259	1.2273	1.2220	1.2236	1.2215	1.2220	1.2350	1.2298
C′—N	1.3457	1.3463	1.3503	1.3507	1.3508	1.3512	1.3375	1.3444
C(α)—C(β)		1.5362		1.5438		1.5282		1.5476
Bond Angles (deg)								
residue 1								
O═C—N	123.48	123.45	124.06	124.48	123.71	123.88	123.28	124.51
C—N—C(α)	120.17	120.46	120.73	121.50	121.39	121.78	117.61	121.52
N—C(α)—C′	113.77	112.59	107.55	106.19	113.01	110.19	111.50	106.30
C(α)—C═O	120.02	119.63	121.74	121.48	121.35	121.79	123.33	121.82
C(α)—C′—N	116.98	117.55	115.37	115.56	115.26	115.00	110.97	114.86
O═C′—N	123.00	122.80	122.90	122.95	123.37	123.21	125.69	123.30
N—C(α)—C(β)		110.32		111.59		110.13		111.80
C′—C(α)—C(β)		109.30		110.85		110.80		110.39
residue 2								
C—N—C(α)	120.21	120.53	120.93	121.71	121.57	121.88	123.36	120.88
N—C(α)—C′	113.83	112.69	107.23	105.90	113.40	110.45	108.60	106.62
C(α)—C═O	119.08	118.60	121.50	121.29	121.43	121.98	122.21	121.29
C(α)—C′—N	117.75	118.51	115.45	115.64	115.31	115.01	115.59	117.00
O═C′—N	123.16	122.89	123.06	123.06	123.23	123.00	122.20	121.71
N—C(α)—C(β)		110.58		111.64		109.85		111.08
C′—C(α)—C(β)		109.33		110.79		110.64		112.39
residue 3								
C—N—C(α)	119.31	119.53	120.97	121.73	121.56	121.83	121.23	121.98
N—C(α)—C′	113.70	112.49	107.13	105.83	113.39	110.49	112.71	109.73
C(α)—C═O	119.47	119.01	121.44	121.24	121.48	122.00	121.58	121.86
C(α)—C′—N	117.43	117.89	115.46	115.65	115.22	114.94	114.60	113.62
O═C′—N	123.10	123.10	123.10	123.10	123.28	123.05	123.82	124.52
N—C(α)—C(β)		110.42		111.64		109.91		109.75
C′—C(α)—C(β)		109.61		110.82		110.59		110.64
residue 4								
C—N—C(α)	119.78	120.30	120.97	121.73	121.56	121.83	120.87	122.79
N—C(α)—C′	114.74	113.30	107.10	105.80	113.44	110.54	113.69	110.35
C(α)—C═O	118.82	118.65	121.47	121.30	121.28	121.83	120.48	120.82
C(α)—C′—N	117.16	117.25	115.53	115.71	115.12	114.84	117.26	117.00
O═C′—N	124.03	124.08	123.00	122.98	123.57	123.32	122.26	122.03
N—C(α)—C(β)		110.59		111.66		109.89		112.22
C′—C(α)—C(β)		109.75		110.83		110.61		108.10
residue 5								
C—N—C(α)	120.54	121.66	120.81	121.58	121.85	122.20	119.58	119.88
N—C(α)—C′	114.46	112.42	107.26	105.88	113.27	110.51	108.28	107.87
C(α)—C═O	119.18	119.47	121.31	121.36	121.63	122.12	120.09	122.13
C(α)—C′—N	116.77	117.03	114.85	115.23	113.85	113.67	115.96	115.08
O═C′—N	124.04	123.49	123.84	123.41	124.50	124.20	123.95	122.75
N—C(α)—C(β)		111.01		111.53		109.77		111.40
C′—C(α)—C(β)		109.50		110.82		110.54		109.34

	Torsional Angle (deg)							
residue 1								
ϕ	−63.0	−63.9	180.0	−167.0	−81.9	−83.6	−179.6	(−167.7)
ψ	−25.2	−25.1	180.0	170.5	62.3	66.0	173.9	(169.6)
ω	178.8	−179.4	180.0	177.2	−174.2	−173.0	169.4	(177.0)
residue 2								
ϕ	−62.8	−64.0	180.0	−166.1	−81.9	−84.1	−128.3	(−160.7)
ψ	−17.3	−17.6	180.0	170.4	60.7	64.5	−149.3	(−174.0)
ω	176.4	177.9	180.0	177.0	−174.2	−173.2	−176.8	(−177.7)
residue 3								
ϕ	−64.1	−65.6	180.0	−166.2	−82.2	−84.4	−78.1	(−79.9)
ψ	−16.8	−17.9	180.0	170.5	60.9	64.5	77.2	(68.8)
ω	176.3	177.2	180.0	177.1	−174.6	−173.5	175.0	(−178.5)
residue 4								
ϕ	−69.0	−74.0	180.0	−166.1	−82.5	−84.7	149.3	(−179.1)
ψ	−9.1	−2.5	180.0	170.6	60.8	64.4	−22.3	(−39.6)
ω	173.3	172.9	180.0	177.0	−175.2	−174.0	179.9	(−175.4)
residue 5								
ϕ	−102.1	−104.7	180.0	−165.8	−83.3	−85.0	176.9	(−169.0)
ψ	11.3	11.0	180.0	170.1	62.2	65.1	162.5	(165.6)

	Relative Energies (kcal/mol)			
	HEL	C5	C7	BET
GLY	0.95	8.20	3.03	0.00
ALA	0.00	7.01	0.14	(4.34)
BET energy for GLY = 753 165.08 kcal/mol				
HEL energy for ALA = −875 402.35 kcal/mol				

Table III. Dihedral Angles (deg) of Two Residues Found at the Carboxy Termini of Helical Sections in Several Proteins and in II-HEL and III-HEL[a]

II-HEL[b]	III-HEL[b]	1HKG[c]	3WGA[d]	4ATC[e]
−69, −9	−74, −3	−68, −20	−81, 3	−67, −49
−102, 11	−104, 11	−100, −1	−123, 28	−103, 15
		−59, −16	−62, −20	−84, −13
		−119, 3	−97, 13	−107, −8
1CTS[f]	155C[g]	3DFR[h]	2C2C[i]	156B[j]
−66, −11	−64, −46	−67, −35	−71, −15	−83, −3
−108, −5	−93, 11	−107, 15	−101, 22	−113, 10
−64, −31		−60, −25		
−103, 31		−100, 0		
2MDH[k]	2CPP[l]	1CC5[m]	1SN3[n]	
−54, −33	−64, −20	−58, −33	−60, −50	
−127, 31	−101, 8	−121, 27	118, 0	
−66, 1	−67, −17			
−112, 45	−102, 21			

[a] Experimental structures were taken from the Protein Data Bank, Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Eur. J. Biochem.* **1977**, *80*, 319. [b] Results from this study, Table II. [c] Hexokinase A. [d] Wheat germ agglutinin. [e] Aspartate carbamoyl transferase. [f] Citrate synthase. [g] Cytochrome C550. [h] Dihydrofolate reductase. [i] Cytochrome C2. [j] Cytochrome B562. [k] Cytoplasmic malate dehydrogenase. [l] Cytochrome P450 CAM. [m] Cytochrome C5. [n] Scorpion neurotoxin 3.

C′–C(α)–C(β) by >4°, and N–C(α)–C′ by >7°. It is interesting to analyze to what extent these changes are affected by long-range interactions in a way that leads to deviations from the dipeptide results.

N–C(α)–C′ is approximately 107° in II-C5 and 106° in III-C5. In II-C7 and III-C7, it is approximately 113° and 110°. For the *N*-acetyl-*N*′-methylamides of Gly and Ala[15] this parameter is 108.0° and 106.4° in C5 and 111.9° and 109.5° in C7eq. In the C5 and C7eq forms of *N*-formylalanine amide,[19] it is 106.3° and 110.0°. In the helical forms II-HEL and III-HEL N–C(α)–C′ is close to 114° (II-HEL) and 113° (III-HEL). In the helical region of *N*-acetyl-*N*′-methylglycine amide, at slightly different torsional angles the value is 112.8°;[15] in the Ala homologue it is 113.3° (all data used are HF/4-21G; a summary of N–C(α)–C′ in II is given in Figure 4).

The data given in the last paragraph show that the variations of N–C(α)–C′ with conformation are qualitatively similar in all systems. At the same time, there is a noticeable chain-length effect. For example, the N–C(α)-C′ difference between C5 and C7eq in Gly is 4° for the dipeptide but 6° for the hexapeptide II (Table II). For Ala, this difference is 3° in the dipeptide, 4° in the tripeptide I (Table I), and nearly 5° in the hexapeptide III (Table II). Thus, in this series, chain elongation seems to enhance the conformational changes in N–C(α)–C′ by about 2°.

It is very interesting to inspect the variations of N–C(α)–C′ in II-BET and III-BET, in which residues populate different conformational states. For residues 1–5 of II-BET, the values are, with the associated ϕ and ψ torsions in parentheses, 111.5 (180,174), 108.6 (−128,−149), 112.7 (−78,77), 113.7 (149,−22), and 108.3 (177,163). In this sequence, residue 3 is C7, the ψ angle of residue 4 is in the helical region, and residue 5 is C5. In each of these cases, the N–C(α)–C′ angle is approximately in the expected range. In contrast to this, residue 1 is essentially C5, but its backbone angle is 4° larger than the characteristic value discussed above.

The prime cause for the apparent anomaly of residue 1 in II-BET can be seen in the hydrogen bond between the N-terminal formyl oxygen and the C-terminal amide hydrogen that pulls the formyl group in the direction of the amide group in such a way that the N–C(α)–C′ angle in residue 1 is opened up. In agreement with this trend, the C–N–C(α) angle in II-BET, residue 1, is compressed (117.6°) compared with values ranging from 119.6° to 123.4° in the other residues. The structure of III-BET is revealing in this context, because in it (Figure 3), the attractive interaction between the N-terminal formyl oxygen and the C-terminal amide hydrogen does not exist. Thus, there is no pressure on N–C(α)–C′ in residue 1 to open up nor on C–N–C(α) to contract. Indeed, the value is 106° for the former, i.e., typical for the conformational state (C5) of this residue, and, 122° for the latter, which is exactly in the range of this parameter in the other residues of III. These comparisons show that structures such as II-BET and III-BET, in which residues are in different conformational states, are a rich source of geometry trends that are important to consider in peptide modeling.

A significant trend is apparent in the calculated C–N peptide bonds. For many years it has been a puzzling problem that the C–N bond distances in small amides in the vapor phase are characteristically long compared to peptide bonds in proteins in the solid state. For example, the electron diffraction r_g C–N bond

lengths in formamide, acetamide, and *N*-methylformamide[38] are found in the region 1.366–1.380 Å, compared to typical trans peptide bonds at 1.34 Å[39] in the solid state. This difference is too large to be explained by thermal effects or by other operational differences of X-ray crystallography and gas electron diffraction.

Interestingly enough, HF/4-21G C–N bond distances in *N*-formylalanine amide[19] fall in the range 1.345–1.352 Å and are shorter by 0.01 Å than the corresponding MP2 values; i.e., compared to HF/4-21G, the shift in the presumably more accurate calculations is toward the experimental values. This is in contrast to HF calculations with polarization functions, such as HF/6-311G**,[19] in which the C–N bond lengths are shifted even farther away from the experiment. In the current case, we have the first theoretical indication that phenomena which can occur only in extended chains, specifically multiple hydrogen bonding, can lead to shorter peptide links.

In the helical forms of II and III, the C–N bond distances of residues 3 and 4 belong to multiply hydrogen-bonding peptide groups. It is seen from Table II that these bonds are shorter than helical C–N bonds in singly H-bonding peptide units. Similarly, multiply interacting internal peptide groups in the repeated C5 and C7eq conformations have shorter C–N bonds than the terminal groups. In II-BET, the shortest C–N bond distances are found (1.337 Å) for the multiply H-bonding terminal peptide groups. In III-BET, in which the terminal groups are not interacting with one another, the terminal C–N bonds are longer than in II-BET. A summary of C–N bonds of II is given in Figure 4. ω-Torsions are also listed, since peptide nonplanarity is an important factor for C–N bond extension.

The HF/4-21G calculated C–N bond lengths are still larger than peptide links obtained from crystallography. Nevertheless, compared with isolated peptide units in the vapor phase, the direction is clearly toward shorter peptide bonds in multiply hydrogen-bonded residues. The peptide link is an important gauge for measuring the success of a parameterization scheme in empirical peptide modeling. Thus, it is a significant finding that the contrast between peptide bond length in isolated molecules and crystal structures should be resolvable in this way.

Conclusions

The current study shows that ab initio gradient geometry optimizations of oligopeptides can be performed in a routine manner at a level at which useful information is obtained that is not readily available from other sources. Specifically, the calculations allow one to examine the validity of the dipeptide model in the development of force field parameters for empirical peptide conformational analyses.

The order of the low-energy regions of dipeptides obtained at the HF/4-21G level is significantly changed by long-range interactions in extended peptide chains. For example, helical forms are not energy minima in dipeptide space,[15] but they are stable regions (nearly vanishing gradients) for model hexapeptides (*N*-formylpentaglycine amide and *N*-formylpentaalanine amide). We conclude that the absence of such forms in ab initio dipeptide conformations is not an artifact of the calculations.

C7eq and C5 conformations are the most stable forms of dipeptides,[15-19] but bend and helical forms in hexapeptides, such as II-HEL, III-HEL, and II-BET, are more stable than the repeated C7eq and C5 conformations. In the tripeptide *N*-formyl-Ala-Ala-amide, the repeated C7eq form still is the most stable one.

Due to the interactions of the N-terminal formyl and C-terminal amide groups, the β-bend conformation of *N*-formylpentaglycine amide found in this study, II-BET, is unique. It confirms the stability of bends in vacuum structures and it is important because it shows that the interaction between two residues can significantly affect the torsional states of other residues in the bend. Thus, it illustrates the considerable flexibility of peptide chains in forming chain reversals.

Conformationally dependent geometry changes, previously determined for dipeptides[15,16] and incorporated in parameter developments for empirical peptide modeling,[23] are also found in the hexapeptides II and III. Specifically, variations of >7° are found in the important backbone angle, N–C(α)–C′, which is typically small in C5 forms, intermediate in C7eq, and large in conformations approaching the bridge region ($\psi = 0$). In view of the large number of proteins with residues in the bridge region,[3d] the ability of peptide chains to flex this angle is an important degree of freedom for releasing steric strain.

Other important angle changes are found for C′–N–C(α) (up to 6°) and C′–C(α)–C(β) (>4°). In general, the geometry trends observed for the smaller and larger peptides are qualitatively similar, but some of the parameter differences are enhanced by chain elongation. For example, differences between N–C(α)–C′ angles in C5 and C7eq are 2° larger in hexapeptides than in dipeptides. Furthermore, the effects of long-range interactions on structure, such as hydrogen bonding, are apparent in some cases. For example, the N–C(α)–C′ angle of residue 1 in II-BET is outside the range expected from dipeptide studies, apparently due to strain exerted by hydrogen bonding between the N-terminal formyl oxygen and the C-terminal amide hydrogen.

Hydrogen bonding is important for peptide links. C–N bonds in protein crystal structures are typically shorter by a few hundredths of an angstrom than those in isolated amides in the vapor phase. The calculations establish a connection between this phenomenon and hydrogen bonding; C–N bonds in multiply hydrogen-bonding peptide units are consistently shorter than those in noninteracting peptide groups.

It is an important aspect of HF/4-21G geometry trends of molecules of the kind considered here that they are very similar to MP2-optimized results[19] and experimental structures.[40] Thus, this study provides realistic estimates of the variations in geometry that occur within the same molecule in which individual residues are in different conformational states. The importance of such local geometry changes for empirical peptide modeling procedures has often been discounted, because it is argued that the molecular property of interest is energy and not geometry. However, if the primary structure of a model system is in error, calculated energy is also in error, because nonbonded interactions cannot be evaluated correctly. Thus, empirical peptide modeling procedures cannot be truly realistic if they do not reproduce, at least approximately, the most important conformational geometry changes that are characteristic for peptide chains.

Acknowledgment. The authors thank Prof. Peter Pulay for his support and helpful advice. Partial support by NSF Grant CHE 8814143 is gratefully acknowledged. A.P. thanks the Belgian National Science Foundation (NFWO) for appointment as Aspirant. C.V.A. thanks NFWO for appointment as Onderzoeksleider. Part of the research was supported by the Belgian DPWB-supercomputing program (contract IT/SC/23).

(38) (a) Kitano, M.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1973**, *46*, 3048. (b) Kitano, M.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1974**, *47*, 67. (c) Kitano, M.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1974**, *47*, 631.

(39) Benedetti, E. In *Peptides: Proceedings of the 5th American Peptide Symposium*, San Diego, CA, June 20–24, 1977; Goodman, M., Meienhofer, J., Eds.; Wiley: New York, 1977; p 257.

(40) de Smedt, J.; Vanhouteghem, F.; van Alsenoy, C.; Geise, H. J.; Schäfer, L. *J. Mol. Struct.* **1992**, *259*, 289.