Correlation Energy, Thermal Energy, and Entropy Effects in Stabilizing Different Secondary Structures of Peptides

Kerstin Möhle,[†] Martin Gussmann,[†] Andreas Rost,[||] Renzo Cimiraglia,[§] and Hans-Jörg Hofmann*[,‡]

Institute of Physical and Theoretical Chemistry, Institute of Organic Chemistry, Faculty of Chemistry and Mineralogy, University of Leipzig, Talstrasse 35, D-04103 Leipzig, Germany, Computational Center, University of Leipzig, Augustusplatz 10/11, D-04109 Leipzig, Germany, Department of Chemistry, University of Ferrara, Via Borsari 46, I-44100 Ferrara, Italy, and Institute of Biochemistry, Faculty of Biosciences, Pharmacy and Psychology, University of Leipzig, Talstrasse 33, D-04103 Leipzig, Germany

Received: July 7, 1997; In Final Form: July 23, 1997[⊗]

Quantum chemical calculations on some typical elements of secondary structure in peptides and proteins (β sheets, β and γ turns) at the Hartree−Fock and MP2 correlation energy levels show considerable differences in the stability orders of alternative structures. The correlation energy data indicate an overestimation of hydrogen-bonded structures. Thus, correlation energy data may be misleading when comparing peptide structures of different type, as for instance, conformations with and without hydrogen bonds or with a different number of hydrogen bonds. This effect is corrected at the Gibbs free energy level when including thermal energy and entropy contributions. Considerable compensation of correlation energy and entropy contributions is mainly responsible for the relatively good correspondence of Hartree−Fock energy differences obtained with more extended basis sets and the free enthalpy data at the correlation energy level.

Introduction

The three-dimensional structure of peptides and proteins is controlled by the variation of the simplest elements of secondary structure: helices, β sheets, and reverse turns. The correct description of the sometimes very small stability differences between the various structure alternatives that are possible for a given amino acid sequence represents a very delicate matter. It is obvious that even relatively small errors when examining smaller peptide units may lead to incorrect results on larger peptides and proteins with consequences for the general understanding of structure formation in these molecules. Moreover, when considering that molecular mechanics will continue to be the basis for the description of the structure and dynamics of biological macromolecules in the future, special care is necessary in establishing empirical force fields. It is a promising and generally accepted procedure to develop force fields on the basis of ab initio molecular orbital (MO) theory. However, it is by no means clear which level could be sufficient for this. Most applications of ab initio MO theory on peptides concern geometry optimizations of some diamides, e.g. the *N*-formylglycinamides and *N*-formylalaninamides (For-Gly-NH_2 and For-L-Ala-NH_2) or the corresponding *N*-acetyl-*N*′-methylamide derivatives.[1a−l] Only a few consider triamides, e.g. selected β turn structures.[2a−d] The influence of correlation effects is mostly neglected in these optimizations, and only for the simplest peptide models For-Gly-NH_2 and For-L-Ala-NH_2 are optimized structures available at various basis set levels considering correlation effects.[1h,k] The tedious procedure of geometry optimization including correlation energy even for smaller peptide units makes it understandable that estimations of the influence of zero-point vibration energies, thermal energy, and entropy contributions to the Gibbs free energy, for which knowledge of the vibration frequencies is necessary, demand still greater efforts. Thus, these contributions were generally neglected at the higher levels of ab initio MO theory and their role in stabilizing different secondary structure elements in peptides and proteins is not well understood until now. It is the aim of this paper to characterize the influence of these contributions on the energetic relations between selected diamide and triamide conformations representing typical elements of secondary structure in peptides and proteins.

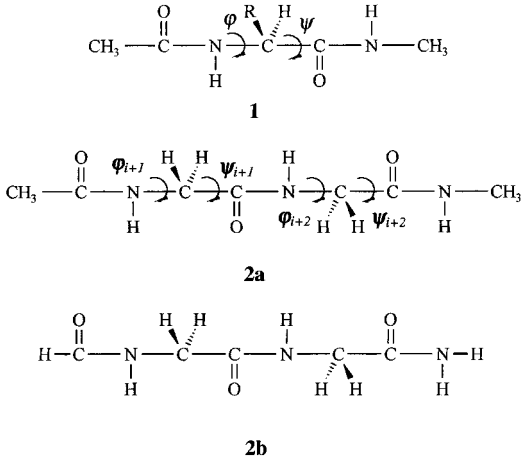

Figure 1. Sketch of the model compounds that the calculations are based on.

Methodology

All calculations were performed employing the Gaussian 94 program package.[3] As model compounds, conformers of *N*-acetylalanine-*N*′-methylamide (**1**) (Ac-L-Ala-NHMe), *N*-acetylglycylglycine-*N*′-methylamide (**2a**) (Ac-Gly-Gly-NHMe), and *N*-formylglycylglycinamide (**2b**) (For-Gly-Gly-NH_2) were selected (Figure 1). All geometries were completely optimized at the HF/6-31G* and MP2/6-31G* levels of ab initio MO

* Corresponding author. FAX: ++49-(0)341-9736749. E-mail: hofmann@rz.uni-leipzig.de.
[†] Faculty of Chemistry, University Leipzig.
[||] Computational Center, University Leipzig.
[§] University of Ferrara.
[‡] Faculty of Biosciences, University Leipzig.
[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1997.

S1089-5639(97)02192-0 CCC: $14.00

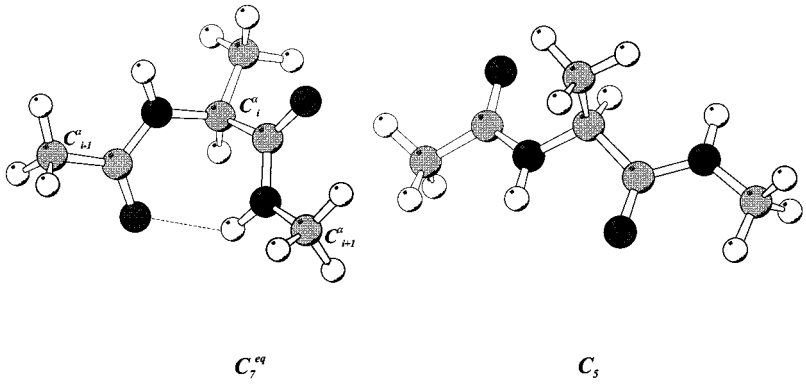

Figure 2. C_{7eq} and C_5 conformers of Ac-L-Ala-NHMe.

theory and characterized by the eigenvalues of the force constants matrix. The optimizations were considered to be finished when the four default criteria of Gaussian 94 were fulfilled. In all cases the maximum residual force was below 0.001 mdyn. The frequencies obtained at the two approximation levels were used for the estimation of the zero-point vibration energies, the thermal energy, and entropy contributions. To test whether the conclusions might be influenced by the overestimation of the frequencies at the HF and MP2 levels, the calculations of the thermodynamic properties were repeated with scaled frequencies according to the recommendations of Pople et al.[4] for the 6-31G* basis set.

Results and Discussion

The conformations of Ac-L-Ala-NHMe (**1**) and the simpler For-Gly-NH_2 and For-L-Ala-NH_2 compounds have frequently been the subject of investigations at the Hartree−Fock (HF) level of ab initio MO theory employing various basis set levels.[1a−k] At least for the latter two molecules the influence of the correlation energy on the stability relations between various conformers was examined.[1h,k] However, thermodynamic data including zero-point vibration energies are not available. Regardless of the basis set level, the most important conformers for all three compounds are the so-called C_{7eq} and C_5 conformations (Figure 2). The C_{7eq} form is the simplest model structure for a γ turn in peptides realizing a complete change of the direction of a peptide sequence via three amino acids supported by a hydrogen bond between the peptide bonds of the first ($i - 1$) and third ($i + 1$) amino acid (Figure 2). The C_5 form corresponds to a nearly extended conformation and reflects the basic conformation of β sheet structures. To illustrate some energetic and structural aspects, Table 1 summarizes data from different sources for For-L-Ala-NH_2 which were obtained both at the HF and the MP2 correlation energy levels employing various basis sets. The following conclusions could be drawn from these data:

(i) The MP2 correlation energy results show the pseudocyclic C_{7eq} form distinctly more stable than the extended C_5 conformation, whereas both structures are energetically nearly equivalent with only a small preference of the γ turn at the HF level when employing more extended basis sets. On the basis of the correlation energy data it is sometimes concluded that more compact peptide structures such as hydrogen-bonded conformations are generally favored over extended ones.[1l]

(ii) Within the same series, HF and MP2, respectively, the energy differences agree rather well when going from the 6-31G* to the 6-311++G(d,p) basis set, whereas the low-level 3-21G basis set shows the C_{7eq} conformation distinctly more stable in the two series. Thus, the 6-31G* or 6-31G** basis set levels may be judged to be of sufficient quality to describe the energetic relations in the two approximations of ab initio MO theory.

TABLE 1: Conformation Characteristics for the C_{7eq} and C_5 Forms of For-L-Ala-NH_2 at Various Levels of ab Initio MO Theory from Different Sources[a]

	C_{7eq}[b]		C_5[b]		
method	φ[d]	ψ[d]	φ[d]	ψ[d]	ΔE[c]
HF/3-21G	−84.5	67.3	−168.3	170.6	5.2
HF/4-21G	−84.7	66.8	−166.6	169.9	5.8
HF/6-31G*	−85.2	76.4	−158.0	161.7	1.5
HF/6-31G**	−85.3	76.0	−157.9	162.2	1.3
HF/6-311G(d,p)	−85.5	78.3	−156.8	162.2	1.0
HF/6-311++G(d,p)	−86.2	78.8	−161.0	161.0	0.5
MP2/3-21G	−81.9	68.4	−170.2	174.1	10.0
MP2/4-21G	−81.6	68.2	−169.5	174.6	10.3
MP2/6-31G*	−82.7	77.7	−158.7	167.5	6.0
MP2/6-31G**	−83.0	77.6	−159.8	166.6	6.2
MP2/6-311G(d,p)	−81.8	81.8	−158.8	168.3	5.9
MP2/6-311++G(d,p)	−82.8	80.6	−157.1	163.2	5.1

[a] From refs 1e−k and this work. [b] See Figure 1. [c] In kJ/mol; $E(C_5) - E(C_{7eq})$. [d] In degrees.

(iii) The agreement of the torsion angle values for φ and ψ calculated at the various basis set levels is rather satisfactory both within the HF and MP2 series, respectively, and between the two approximation series. The largest variations of φ and ψ are about 10°.

(iv) Surprisingly, the energetic data obtained with the 3-21G basis set at the Hartree−Fock level agree rather well with the MP2 correlation energy data for the high-level 6-311++G(d,p) basis set. Thus, some authors conclude that basis set extension and correlation energy effects are compensating.[1k]

In fact, the results of our HF/6-31G* and MP2/6-31G* calculations on the larger Ac-L-Ala-NHMe compound **1** confirm the above-mentioned stability aspects (Table 2).

There are more examples where correlation energy favors cyclic structures over extended or open ones. Thus, the higher stability of nonclassical carbocations over their classical counterparts is only found when the correlation energy is considered.[5a−c] The same is true for the cyclic structure of trithiapentalene and its derivatives which can only be reproduced after inclusion of correlation energy, whereas the open structure alternatives are still the preferred minimum conformations on the Hartree−Fock energy hypersurface and even disappear at the correlation energy level.[6] Similar effects seem to appear in pseudocyclic peptide structures formed by hydrogen bonds. However, the situation reverses when additionally calculating the thermal energy and entropy contributions. At the Gibbs free energy level (Table 2), the C_{7eq} and C_5 conformers are again of comparable stability as already predicted at the Hartree−Fock level employing larger basis sets. A detailed analysis

TABLE 2: Torsion Angles, Total Energies, Zero-Point Vibration Energies, and Free Enthalpies and Entropies for the C_{7eq} and C_5 Conformation of Ac-L-Ala-NHMe (1) at the Hartree−Fock and Correlation Energy Levels and the Energy and Free Enthalpy Differences between the Two Conformers

	MP2/6-31G*		HF/6-31G*	
	C_{7eq}[a]	C_5[a]	C_{7eq}[a]	C_5[a]
φ[b]	−82.9	−158.6	−85.4	−157.4
ψ[b]	77.9	161.1	79.4	158.8
E_T[c]	−494.310 898	−494.308 149	−492.861 542	−492.860 889
ΔE[d,e]	0.0	7.2	0.0	1.7
ZPVE[c]	0.190 620	0.190 136	0.200 765	0.200 229
G[c]	−494.159 913	−494.158 569	−492.700 161	−492.700 521
ΔG[d,e]	0.0	3.5	0.0	−0.9
S[f]	463.6	473.9	457.9	464.3

[a] See Figure 1. [b] In degrees. [c] In au. [d] In kJ/mol. [e] Related to C_{7eq}. [f] In J/mol K.

shows the considerable compensation of correlation and entropy effects mainly responsible for this behavior and a small destabilizing effect of the thermal energies on the pseudocyclic structures (Table 2). The hydrogen-bonded ring system of C_{7eq} has a higher order than the extended C_5 form as indicated by an entropy decrease. This destabilizing effect is compensated by the correlation energy which favors the pseudocyclic structures over the extended ones. Remembering the above-mentioned good agreement of the low-level HF/3-21G energy differences between C_{7eq} and C_5 and those from high-level MP2/6-31G* or even MP2/6-311++G(d,p) ab initio calculations, it can be concluded that the MP2 correlation energy results may be misleading in calculations on peptide structures when conformations of different type are compared as is the case for the pseudocyclic C_{7eq} and the extended C_5 form.

A second example may illustrate these aspects. The compounds **2a**, and **2b** are simple models to describe β turn conformations.[7a−c] β turns reverse the direction of peptide chains via four amino acids. Two important β turn conformations are the βI (common) and βII (glycine) turns, which are frequently found in peptides and proteins. Both turns are characterized by a hydrogen bond between the peptide bonds of the first and fourth amino acid (Figure 3). The various β turns are defined by the torsion angles φ_{i+1}, ψ_{i+1}, φ_{i+2}, and ψ_{i+2} of the second and third amino acids (Figures 1 and 3). The βII turn is preferred, when the amino acid glycine enters the position $i + 2$, whereas the βI conformation predominates for most L-amino acids. The energetic relations between these important pseudocyclic elements of secondary structure and the extended conformation might characterize the folding tendency of the open peptide chain. It is obvious, that the above-mentioned problems may essentially influence the conclusions drawn at the various approximation levels. In Table 3, the results of the calculations on **2a** and **2b** at the HF level show the βII more stable than the βI turn in agreement with experimental data. However, the extended form is still distinctly more stable than both turns.

When including correlation energy, the situation changes. Now, the two β turn structures are more stable than the extended form, whereas the energy relation between the two turns is not influenced since they are of comparable structure type. The preference of the extended conformation over the turn structures is again established at the Gibbs free energy level when considering correlation energy, thermal energy, and entropy contributions. When repeating the calculations of the thermodynamic properties with scaled frequencies at both approximation levels, nearly the same free enthalpy differences between the corresponding conformations were obtained without correction. It should be noticed that the characteristic conformation angles agree rather well again for all structures at both approximation levels (Tables 2 and 3).

Conclusions

The quantum chemical results for some typical elements of secondary structure in peptides and proteins show considerable differences between the stability orders obtained at the Hartree−Fock and the correlation energy levels. Surprisingly, the energy differences at the higher correlation energy level might be misleading when comparing peptide conformations of different type, e.g. structures with or without hydrogen bonds or structures with a different number of hydrogen bonds. Consideration of solely the correlation energy overestimates the importance of hydrogen-bonded structures, which is corrected by the thermal energy and entropy contributions at the Gibbs free energy level. Compensation of entropy and correlation energy effects is mainly responsible for the relatively good correspondence between Hartree−Fock energy differences obtained at a sufficiently high basis set level, e.g. HF/6-31G* and higher, and the free enthalpy data at the correlation energy level. Establish-

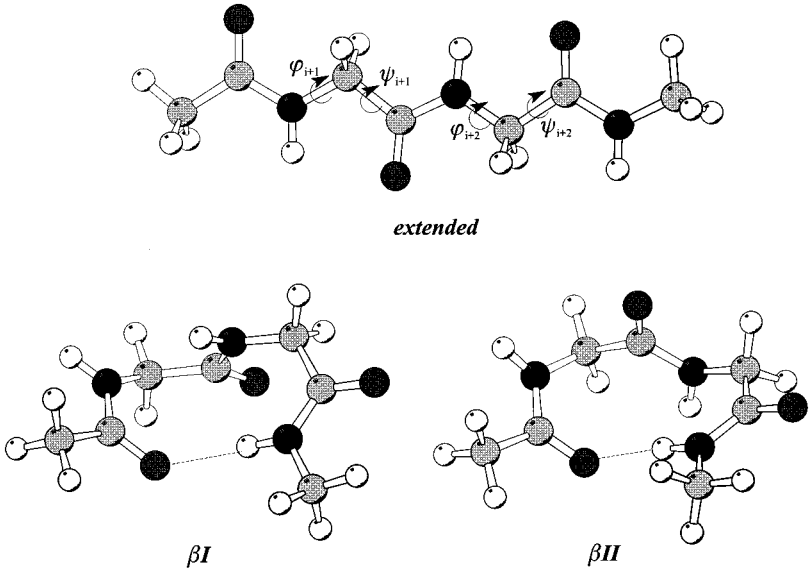

Figure 3. Extended, βI, and βII turn conformations for Ac-Gly-Gly-NHMe.

TABLE 3: Torsion Angles, Total Energies, Zero-Point Vibration Energies, and Free Enthalpies and Entropies for the Extended Form and the βI and βII Turns of the Model Compounds Ac-Gly-Gly-NHMe (2a)[a] and For-Gly-Gly-NH_2 (2b)[b] at the Hartree−Fock and Correlation Energy Levels and the Energy and Free Enthalpy Differences between the Conformers

	MP2/6-31G*			HF/6-31G*		
	extended[c]	βI[c]	βII[c]	extended[c]	βI[c]	βII[c]
φ_{i+1}[d]	−171.2	−72.1	−58.6	−179.9	−73.3	−60.9
	−180.0	−71.1	−59.7	180.0	−72.2	−61.8
ψ_{i+1}[d]	−176.9	−21.2	139.8	−179.7	−17.7	136.4
	−180.0	−20.9	137.9	180.0	−18.1	134.7
φ_{i+2}[d]	−179.8	−99.6	92.7	−179.7	−101.9	95.5
	−177.5	−101.8	96.4	−179.9	−103.5	97.6
ψ_{i+2}[d]	−179.8	15.3	−14.0	−179.7	11.9	−11.7
	−179.8	12.8	−11.4	180.0	10.3	−10.2
E_T[e]	−662.542 161	−662.546 443	−662.547 719	−660.641 395	−660.639 135	−660.640 946
	−584.206 460	−584.208 500	−584.209 562	−582.565 953	−582.562 801	−582.564 254
ΔE[f,g]	0.0	−11.2	−14.6	0.0	5.9	1.2
	0.0	−5.4	−8.1	0.0	8.3	4.5
ZPVE[e]	0.217 266	0.218 868	0.219 081	0.229 991	0.231 450	0.231 671
	0.159 130	0.161 235	0.161 578	0.170 076	0.171 679	0.171 928
G[e]	−662.374 123	−662.371 191	−662.372 070	−660.458 731	−660.451 053	−660.452 395
	−584.090 736	−584.085 813	−584.086 375	−582.436 383	−582.429 458	−582.430 556
ΔG[f,g]	0.0	7.7	5.4	0.0	20.2	16.6
	0.0	12.9	11.4	0.0	18.2	15.3
S[h]	580.4	522.5	520.7	558.6	516.2	513.5
	502.0	448.4	446.5	469.7	443.1	441.8

[a] First values. [b] Second values. [c] See Figure 3. [d] In degrees. [e] In au. [f] In kJ/mol. [g] Related to the extended form. [h] In J/mol K.

ing empirical force fields for peptides and proteins on the basis of correlation energy data without consideration of the thermodynamic contributions cannot be recommended.

Acknowledgment. Support of this work by the Deutsche Forschungsgemeinschaft (Innovationskolleg "Chemisches Signal und biologische Antwort"), the German Fonds der Chemischen Industrie, and the Italian M.U.R.S.T is gratefully acknowledged.

References and Notes

(1) (a) Schäfer, L.; Van Alsenoy, C.; Scarsdale, J. N. *J. Chem. Phys.* **1982**, *76*, 1439−1444. (b) Schäfer, L.; Klimkowski, V. J.; Momany, F. A.; Chuman, H.; Van Alsenoy, C. *Biopolymers* **1984**, *23*, 2335−2347. (c) Scarsdale, J. N.; Van Alsenoy, C.; Klimkowski, V. J.; Schäfer, L.; Momany, F. A. *J. Am. Chem. Soc.* **1983**, *105*, 3438−3445. (d) Klimkowski, V. J.; Schäfer, L.; Momany, F. A.; Van Alsenoy, C. *J. Mol. Struct. (THEOCHEM)* **1985**, *124*, 143−153. (e) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks C. L., III; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989−5997. (f) Böhm, H.-J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129−7135. (g) Gould, I. R.; Kollman, P. A. *J. Phys. Chem.* **1992**, *96*, 9255−9258. (h) 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*, 5369−5377. (i) Rommel-Möhle, K.; Hofmann, H.-J. *J. Mol. Struct. (THEOCHEM)* **1993**, *285*, 211−219. (j) Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc.* **1994**, *116*, 9250−9256. (k) Endredi, G.; Perczel, A.; Farkas, O.; McAllister, M. A.; Csonka, G. I.; Ladik, J.; Csizmadia, I. G. *J. Mol. Struct. (THEOCHEM)* **1997**, *391*, 15−26. (l) Brooks, C. L., III; Case, D. A. *Chem. Rev. (Washington, DC)* **1993**, *93*, 2487−2502.

(2) (a) Perczel, A.; McAllister, M. A.; Csaszar, P.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1993**, *115*, 4849−4858. (b) Böhm, H.-J. *J. Am. Chem. Soc.* **1993**, *115*, 6152−6158. (c) Schäfer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Van Alsenoy, C.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.* **1993**, *115*, 272−280. (d) Möhle, K.; Gussmann, M.; Hofmann, H.-J. *J. Comput. Chem.* **1997**, *18*, 1415−1430.

(3) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Peterson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challcombe, V.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94 (Revision D.4)*; Gaussian, Inc.: Pittsburgh, PA, 1996.

(4) Pople, J. A.; Scott, A. P.; Wong, M. W.; Radom, L. *Isr. J. Chem.* **1993**, *33*, 345−350.

(5) (a) Koch, W.; Liu, B.; DeFrees, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 1527−1528. (b) Raghavachari, K.; Haddon, R. C.; Schleyer, P. v. R.; Schaefer, H. F., III *J. Am. Chem. Soc.* **1983**, *105*, 5915−5917. (c) Yoshimine, M.; McLean, A. D.; Liu, B.; DeFrees, D. J.; Binkley, J. S. *J. Am. Chem. Soc.* **1983**, *105*, 6185−6186.

(6) Cimiraglia, R.; Hofmann, H.-J. *J. Am. Chem. Soc.* **1991**, *113*, 6449−6451.

(7) (a) Venkatachalam, C. M. *Biopolymers* **1968**, *6*, 1425−1436. (b) Richardson, J. *Adv. Protein Chem.* **1981**, *34*, 167−339. (c) Chou, P. Y.; Fasman, G. D. *J. Mol. Biol.* **1977**, *115*, 135−175.