***Ab Initio* Calculations on Large Molecules Using Molecular Fragments. Evidence of Hydrogen Bonding in Polypeptides of Glycine**

(hydrogen bonding/poly(glycine)/self-consistent field calculations)

LESTER L. SHIPMAN AND RALPH E. CHRISTOFFERSEN*

Department of Chemistry, University of Kansas, Lawrence, Kans. 66044

Communicated by William N. Lipscomb, August 25, 1972

ABSTRACT *Ab initio* calculations on di-, tri-, tetra-, and pentapeptides of glycine in various conformations are reported. Hydrogen bonding is observed as an important stabilizing force in α -helical conformations. These studies on polypeptides of glycine of reasonable size show for the first time that the forces acting to stabilize polypeptide conformations can be extracted directly from theoretical studies, without prior postulation of their existence or need for concern that neglect or approximation of various integrals may have biased the results.

Understanding of the mechanism of fundamental biological processes has deepened substantially recently, due to major strides made possible by various physical and chemical methods (1). For example, in optimum cases, the detailed coordinates of virtually every nucleus, other than hydrogen, in a protein can be obtained in the solid state by use of high resolution x-ray diffraction techniques (2). However, extraction of the nature and magnitude of the forces that contribute to the stability of various structures is frequently extremely difficult, and many of the previous investigations used *ad hoc* postulates regarding their nature (3) in order that progress could be made.

In this communication, evidence for hydrogen bonding forces in polypeptides has been extracted directly from *ab initio* quantum mechanical calculations. In particular, calculations on poly(glycine), representing the largest molecular system treated by *ab initio* techniques to date (144 electrons in the largest case[†]), show evidence from *ab initio* theoretical studies of these forces on actual polypeptides of reasonable size for the first time. There is no necessity for prior postulation of the existence of these forces, or need for concern that neglect or approximation of integrals may have biased the results.

DESCRIPTION OF CALCULATIONS

The method used in these studies (4-13) is an *ab initio* self-consistent field procedure, based upon the use of investigations of molecular fragments in order to obtain basis orbitals for the construction of large molecules. Since the nature of the procedure and the basis set parameters for the various molecular fragments needed for these calculations have been de-

Abbreviations: SCF, self-consistent field; MP, mono-peptide; DP, dipeptide; TRP, tripeptide; TEP, tetrapeptide; PP, pentapeptide; FE, fully extended; APC-PS, antiparallel chain pleated sheet; PC-PS, parallel chain pleated sheet; α , alpha helix.

* To whom correspondence should be addressed.

† For other recent *ab initio* calculations on large molecules, see E. Clementi, J. Mehl, & W. von Niessen (1971) *J. Chem. Phys.* **54**, 508-520.

TABLE 1. Four standard peptide conformations*.[†]

Conformation	ϕ	ψ
Fully extended	180	180
Anti-parallel-chain pleated sheet	-142	145
Parallel-chain pleated sheet	-119	113
α -Helix	-48	-57

* ϕ and ψ angles are given in degrees.

† See ref. 15 and footnote[†].

scribed (9, 13) and have been shown to describe the various interactions in peptides adequately, they will not be repeated here.

The molecules investigated include di-, tri-, tetra-, and pentapeptides of glycine, and are depicted in Fig. 1. The nuclear geometries were chosen to be the "standard" geometries recommended by Momany *et al.* (14), augmented by a choice of 1.102 Å for the C—H bond length at the "carbonyl end" of the molecules. Four conformations were studied for the di-, tri-, and tetrapeptides, including the fully extended, anti-parallel-chain pleated sheet, parallel-chain pleated sheet, and α -helix conformers. For the pentapeptide of glycine, the fully extended and α -helix forms were investigated. Standard values (15) for the dihedral angles (ϕ and ψ) representing these conformers were used, and are given in Table 1 and shown pictorially in Fig. 1.

In Fig. 2 are given energy differences of various conformers of poly(glycine), relative to the fully extended form, as a function of polypeptide length. The energy has been arbitrarily zeroed for the fully extended conformers for convenience in this figure. The calculations for each conformer have been performed by double-precision arithmetic (72 bits) on a Honeywell 635 computer to full convergence of the self-consistent field procedure, with a typical root mean-square difference between charge and bond order matrix elements of successive interactions of $<10^{-5}$ at convergence.

DISCUSSION

As indicated in Fig. 2, the energy difference plots for the anti-parallel-chain pleated sheet and parallel-chain pleated sheet structures are linear functions of the polypeptide chain length over the range studied. This is not an unexpected result, and

† These dihedral angles are defined according to the tentative rules established by the IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry* **9**, 3471-3479.

FIG. 1. Polypeptides of glycine studied.

indicates that the energy required to rotate from the fully extended conformer to the anti-parallel-chain pleated sheet and parallel-chain pleated sheet conformers can be viewed as being composed of approximately equal and independent contributions from each pair of adjacent (ϕ , ψ) rotations as additional glycine subunits are added. Also, the decreased stability of the anti-parallel-chain pleated sheet and parallel-chain pleated sheet structures (relative to the fully extended structure) is not inconsistent with the occurrence of pleated-sheet structures in nature (16), since interchain hydrogen bonding is an important source of stabilization. Thus, these results indicate that *single* strands of poly(glycine) in the anti-parallel-chain pleated sheet or parallel-chain pleated sheet forms are not expected to be stable, relative to the fully extended structure, for poly(glycine) of any chain length.

Furthermore, if all interchain stabilization forces are assumed to arise from interchain hydrogen bonding, then a more quantitative lower limit estimate of the magnitude of the interchain hydrogen bonding forces can be obtained. In particular, the graphs of Fig. 2 indicate that the interchain hydrogen bonds of polystranded poly(glycine) must be stronger than 2.6 kcal/mol per glycine subunit for the anti-parallel-chain pleated sheet, and stronger than 6.2 kcal/mol per glycine subunit for the parallel-chain pleated sheet structures, if they are to be more stable than the fully extended form.

Of particular interest is the α -helix conformational curve in Fig. 2, which is seen to be nonlinear. Furthermore, the nonlinearity becomes apparent at the tetrapeptide chain length, which is the minimum chain length at which intramolecular hydrogen bonding is a possibility. It is believed

FIG. 2. Energies of glycine polypeptides relative to the fully extended form, as a function of polypeptide length. (DP = dipeptide; TRP = tripeptide; TEP = tetrapeptide; PP = pentapeptide; FE = fully extended; APC-PS = anti-parallel-chain pleated sheet; PC-PS = parallel-chain pleated sheet; α = α -helix.) $\Delta E = E_i - E_{FE}$, where $n = 2, 3, 4, 5$, and $i =$ APC-PS, PC-PS, and α .

that the nonlinearity is associated with intramolecular hydrogen-bonding effects, as the following analysis will show.

To quantify the magnitude of the various interaction energies, let us partition the energy differences as follows. For the dipeptide, the entire ΔE_{DP} will be assigned to the energy of interaction between two adjacent amide units (e.g., labeled n and $n + 1$), i.e.,

$$\Delta E_{DP} = \Delta E_{n,n+1} \quad [1]$$

For the tripeptide, the total ΔE_{TRP} will be partitioned as

$$\Delta E_{TRP} = 2\Delta E_{n,n+1} + \Delta E_{n,n+2} \quad [2]$$

where $\Delta E_{n,n+2}$ is the energy of interaction between a given subunit (n) and another subunit ($n + 2$) located two subunits from the n th subunit. Analogously, the total ΔE_{TEP} for the tetrapeptide and ΔE_{PP} for the pentapeptide can be partitioned as

$$\Delta E_{TEP} = 3\Delta E_{n,n+1} + 2\Delta E_{n,n+2} + \Delta E_{n,n+3} \quad [3]$$

and

$$\Delta E_{PP} = 4\Delta E_{n,n+1} + 3\Delta E_{n,n+2} + 2\Delta E_{n,n+3} + \Delta E_{n,n+4} \quad [4]$$

The generalization of this type of partitioning procedure to a polypeptide of arbitrary length is obvious.

Application of this procedure to the current studies results in

$$\Delta E_{n,n+1} = +11.2 \text{ kcal/mol}, \quad [5]$$

$$\Delta E_{n,n+2} = -1.6 \text{ kcal/mol}, \quad [6]$$

TABLE 2. Selected total π -orbital populations in poly(glycine)*

Atom	Fully extended conformer					α -Helix conformer				
	MP†	DP	TRP	TEP	PP	MP	DP	TRP	TEP	PP
O ₁	1.170	1.139	1.136	1.136	1.136	1.170	1.122	1.117	1.159	1.162
O ₂		1.211	1.182	1.180	1.180		1.180	1.136	1.130	1.172
O ₃			1.206	1.177	1.175			1.183	1.138	1.133
O ₄				1.207	1.178				1.190	1.145
O ₅					1.206					1.192
C ₁ '	1.069	1.074	1.074	1.074	1.074	1.069	1.079	1.084	1.057	1.053
C ₂ '		1.078	1.079	1.080	1.079		1.094	1.101	1.107	1.081
C ₃ '			1.083	1.084	1.084			1.096	1.103	1.108
C ₄ '				1.083	1.084				1.095	1.101
C ₅ '					1.084					1.093

Abbreviations are as in legend to Fig. 2.

* See Fig. 1 for the definition of the atom labels.

† The mono-peptide that has been included for reference purposes is formamide.

$$\Delta E_{n,n+3} = -6.1 \text{ kcal/mol}, \quad [7]$$

and

$$\Delta E_{n,n+4} = +0.3 \text{ kcal/mol}. \quad [8]$$

If $\Delta E_{n,n+3}$ is identified as representing the energy of a hydrogen bond that is formed between the carbonyl oxygen of subunit n and the amide hydrogen of the $(n+3)$ rd subunit, the quite reasonable estimate of 6.1 kcal/mol is obtained for the strength of the hydrogen bond (§, 17–23). Of course, such an identification assumes that all of the $\Delta E_{n,n+3}$ interaction can be assigned to a “through-space” hydrogen-bond formation, and that it does not contain any terms resulting from transmission of energetic effects through the σ -framework of the polypeptide backbone.

Further evidence for the identification of the nonlinearity of the α -helix plots in Fig. 2 with hydrogen bonding results from the use of the Mulliken population analysis (24), suitably modified for use with molecular fragment basis orbitals that are not necessarily localized on nuclei. Table 2 presents a portion of that analysis that is relevant to the discussion here. Two observations are of interest from this table. First, the π -type orbital on oxygen atom O₁ in the tetrapeptide and oxygen atoms O₁ and O₂ in the pentapeptide (the ones that would be involved in hydrogen bonding) have different trends from the other oxygen atoms as the fully extended conformer is changed to the α -helix form. In particular, these orbital populations *increase* (the O₂ population in the pentapeptide stays approximately constant) in the change from the fully extended to α -helix form, while all of the other oxygen π -type orbital populations decrease during the same process. Second, the carbon atom π -type orbital populations generally increase in going from the fully extended to the α -helix form, except for the populations of the π -type orbitals on C₁' in the tetrapeptide and on C₁' and C₂' in the pentapeptide, which generally *decrease* (or stay approximately constant). These trends indicate that, although actual charge transfer from the carbonyl group to the amide (three subunits away) does not take place, a change in the carbonyl charge distribution does

take place that would be expected if hydrogen bonds were formed. Namely, the carbonyl charge distribution becomes more polar, thus increasing the relative emphasis on resonance structure II (see Fig. 3) for those carbonyl groups involved in the hydrogen bond, and facilitating the formation of those hydrogen bonds.

Although these studies indicate the presence of hydrogen bonding as a stabilizing force in single strands of poly(glycine), they do not necessarily imply that the α -helix form of poly(glycine) should be most stable. In fact, for the di-, tri-, tetra-, and pentapeptides of glycine, the results of these studies (Fig. 2) indicate the clear preference for the fully extended form over that of the α -helix form. Furthermore, if $\Delta E_{n,n+j}$ is assumed to be negligibly small for $j > 4$ (as would be expected from distance considerations), then the energy difference between the α -helix and fully extended form of poly(glycine) having N subunits ($N \geq 4$) is seen from Eqs. [1–4] to be given by

$$\Delta E_N = (N-1)\Delta E_{n,n+1} + (N-2)\Delta E_{n,n+2} + (N-3)\Delta E_{n,n+3} + (N-4)\Delta E_{n,n+4}. \quad [9]$$

Substitution of the values obtained in these studies (Eqs. [5–8]) yields

$$\Delta E_N = 9.1 + 3.8N, \quad [10]$$

which is obviously positive for all N . Thus, single-stranded poly(glycine) will not prefer the α -helix form for *any* chain length.

Available experimental evidence also supports the observation that the α -helix is not preferred in poly(glycine), relative to other conformations. For example, in the crystal state, poly(glycine) is found in two forms. Poly(glycine) I is a fully extended or nearly fully extended form (25), and poly(glycine) II contains helices with a three-fold screw axis (26), but they

FIG. 3. Contributing resonance structures in amides.

§ The value of 6.1 kcal/mol falls well within the range of hydrogen bond energies (3.5–11 kcal/mol) that have been found in other quantum mechanical studies. See refs. 17–23.

are not α -helices. In addition, infrared, optical rotatory dispersion, and x-ray diffraction studies of ordered sequences of glycyl and γ -ethyl L-glutamyl residues (27) indicate that the presence of glycyl residues in polypeptide chains caused a marked reduction in the stability in solution of the α -helix conformation, relative to a random conformation. Finally, a far-ultraviolet study of a series of oligoglycines (28) ($n = 2$ to 5) indicated that the various peptide bonds are randomly oriented.

Further discussion of these calculations, including analysis of the molecular orbital structure, the effect of the small basis set, and comparison with semiempirical calculations (23, 29–32) on similar systems will be given elsewhere.

One of us (REC) would like to express his appreciation to W. N. Lipscomb for gracious hospitality and stimulating conversations during the preparation of this work. Also, we would like to express our appreciation to the University of Kansas for partial support of the computing time required for this work. This work was supported in part by the National Science Foundation, the University of Kansas, and the Upjohn Co., Kalamazoo, Mich. 49001. L. L. S. is a National Science Foundation trainee (1969–1972). R. E. C. is an Alfred P. Sloan research fellow (1971–1973).

1. For a review of some of the physical techniques that are currently in use, see (1970) "Spectroscopic Approaches to Biomolecular Conformation," ed. Urry, D. W. (American Medical Association, Chicago, Ill.).
2. For a recent report on the status of protein crystallography, see (1971) "Structure and Function of Proteins at the Three-Dimensional Level," *Cold Spring Harbor Symp. Quant. Biol.* **36**, 1–644.
3. See, for example, Scheraga, H. A. (1971) *Chem. Rev.* **71**, 195–217; see also, Scheraga, H. A. (1968) *Advan. Phys. Org. Chem.* **6**, 103–184.
4. Christoffersen, R. E. & Maggiora, G. M. (1969) *Chem. Phys. Lett.* **3**, 419–423.
5. Christoffersen, R. E., Genson, D. W. & Maggiora, G. M. (1971) *J. Chem. Phys.* **54**, 239–252.
6. Maggiora, G. M., Genson, D. W., Christoffersen, R. E. & Cheney, B. V. (1971) *Theor. Chim. Acta* **22**, 337–352.
7. Christoffersen, R. E. (1971) *J. Amer. Chem. Soc.* **93**, 4104–4111.
8. Christoffersen, R. E., Shipman, L. L. & Maggiora, G. M. (1971) *Int. J. Quantum Chem.* **5S**, 143–149.
9. Christoffersen, R. E. (1972) *Advan. Quantum Chem.* **6**, 333–393.
10. Cheney, B. V. & Christoffersen, R. E. (1972) *J. Chem. Phys.* **56**, 3503–3518.
11. Genson, D. W. & Christoffersen, R. E. (1972) *J. Amer. Chem. Soc.*, in press.
12. Shipman, L. L. & Christoffersen, R. E. (1972) *Chem. Phys. Lett.* **15**, 469–474.
13. Shipman, L. L. & Christoffersen, R. E. (1972) *J. Amer. Chem. Soc.*, in press.
14. Momany, F. A., McGuire, R. F., Yan, I. F. & Scheraga, H. A. (1971) *J. Phys. Chem.* **75**, 2286–2297.
15. Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Nemethy, G., Ramachandran, G. N. & Scheraga, H. A. (1966) *Biopolymers* **4**, 121–130, 1149–1150.
16. Schellman, J. A. & Schellman, C. (1964) in *The Proteins*, ed. Neurath, H. (Academic Press, New York), Vol. 2, pp. 1–137.
17. Dreyfus, M., Maigret, B. & Pullman, A. (1970) *Theor. Chim. Acta* **17**, 109–119.
18. Dreyfus, M. & Pullman, A. (1970) *Theor. Chim. Acta* **19**, 20–37.
19. Pullman, A. & Berthod, H. (1968) *Theor. Chim. Acta* **10**, 461–464.
20. Momany, F. A., McGuire, R. F., Yan, J. F. & Scheraga, H. A. (1970) *J. Phys. Chem.* **74**, 2424–2438.
21. Murthy, A. S. N., Rao, K. G. & Rao, C. N. R. (1970) *J. Amer. Chem. Soc.* **92**, 3544–3548.
22. Murthy, A. S. N., Bhat, S. N. & Rao, C. N. R. (1970) *J. Chem. Soc. (A)* **1970**, 1251–1256.
23. Govil, G. & Saran, A. (1971) *J. Chem. Soc. (A)* **1971**, 3624–3627.
24. Mulliken, R. S. (1962) *J. Chem. Phys.* **36**, 3428–3439, and earlier references therein.
25. Bamford, C. H., Brown, L., Cant, E. M., Elliot, A., Handy, W. E. & Malcolm, B. R. (1955) *Nature* **176**, 396–397.
26. Rich, A. & Crick, F. H. C. (1955) *Nature* **176**, 915–916.
27. Fraser, R. D. B., Harrap, B. S., MacRae, T. P., Stewart, F. H. C. & Suzuki, E. (1967) *Biopolymers* **5**, 251–257.
28. McDiarmid, R. S. (1965) Ph.D. dissertation, Harvard University.
29. Rossi, A. R. & David, C. W. (1969) *Theor. Chim. Acta* **14**, 429–431.
30. Rossi, A. R., David, C. W. & Schor, R. (1970) *J. Phys. Chem.* **74**, 4551–4555.
31. Schor, R., Stimme, H., Wettermark, G. & David, G. W. (1972) *J. Phys. Chem.* **76**, 670–672.
32. McGuire, R. F., Momany, F. A. & Scheraga, H. A. (1972) *J. Phys. Chem.* **76**, 375–393, and earlier references given therein.