(OiBu)-p-U(OBz)$_2$ (6900 A_{260}, 0.26 mmol) was allowed to react with pyridinium MMTr-C^{Bz}(OBz)-p (1 mmol) and DCC (2 g) in pyridine (5 ml) for 72 hr. The reaction mixture was treated with aqueous pyridine and extracted with *n*-pentane. Cyclohexylurea was removed by filtration and the solution was evaporated with pyridine. The pyridine solution was precipitated with ether and the precipitate was washed with ether. The powder was treated with 80% acetic acid for 1 hr at 25°. Acetic acid was evaporated with added aqueous ethanol and the residue was dissolved in pyridine (5 ml) and 70% ethanol. The solution was applied to a column (3 × 45 cm) of TEAE-cellulose (acetate) preequilibrated with 70% ethanol. The trinucleotide **14** was eluted at salt concentration of about 0.11 *M*. The yield was 2690 A_{280} units, 0.06 mmol, 26%.

G^{iBu}(OiBu)-p-C^{Bz}(OBz)-p-G^{iBu}(OiBu)-p-U(OBz)$_2$ (8). The pyridinium salt of trinucleotide **14** (1260 A_{280} units, 0.03 mmol) was allowed to react with pyridinium MMTr-G^{iBu}(OiBu)-p (0.25 mmol) using DCC (2 mmol) in pyridine (1 ml) for 81 hr. Water (1 ml) was added and the mixture was kept at room temperature for 12 hr and concentrated to *ca.* 0.5 ml. The solution was applied to a column (1.5 × 25) of Sephadex LH-20 preequilibrated with 90% ethanol. The fractions of 2 ml were collected and fractions (51–57) which contained the product (1080 A_{280} units) were combined. The purity of the tetranucleotide was checked by ppc and found to be 42%. The yield was 24%. The tetranucleotide was further purified by ion-exchange chromatography on a TEAE-cellulose column (1.7 × 30 cm). The elution pattern is shown in Figure 6 and the identification of some of the peaks is given in Table III. Peak IV contained the tetranucleotide 292 A_{260} units (4.7 μmol), 15%.

Condensation of R-G^{iBu}(OiBu)-p-G^{iBu}(OiBu)-p (7b) with G^{iBu}-(OiBu)-p-C^{Bz}(OBz)-p-G^{iBu}(OiBu)-p-U(OBz)$_2$ (8). The tetranucleotide **8** (230 A_{260}, 3.7 μmol) and the dinucleotide **7** (600 A_{260}, 19 μmol) were treated with TPS (67 μmol) in pyridine (0.2 ml) at 22° for 6 hr. Aqueous pyridine (50%, 0.5 ml) and triethylamine (0.13 mmol) were added at 0° and the mixture was kept at room temperature for 16 hr. The solution was evaporated with pyridine and the dry residue was treated with 15 *M* methanolic ammonia (15 ml) for 18 hr at 26°. Volatile materials were evaporated and the residue was dissolved in 7 *M* urea (10 ml). The solution was adjusted to pH 7.5 with 2 *N* ammonium hydroxide and applied to a column (0.7 × 80 cm) of DEAE-cellulose equilibrated with 7 *M* urea and 0.02 *M* Tris–HCl, pH 7.5. The elution pattern and conditions are shown in Figure 7. Peak VIII contained the hexanucleotide, GpGpGpCpGpU (34 A_{260} units, 0.57 μmol). Fractions were combined and desalted by gel filtration[10] using a column (2 × 60 cm) of Biogel P-2 (100–200 mesh). The desalted oligonucleotide was subjected to paper electrophoresis and found to be homogeneous. In paper chromatography, however, two spots were detected besides the compound at the origin. An aliquot (8 A_{260} units) of the desalted product was subjected to rechromatography on a column (0.7 × 90 cm) of DEAE-cellulose equilibrated with 7 *M* urea and 0.02 *M* Tris–HCl. Elution was performed with a gradient of sodium chloride, 0.1 *M* (80 ml) in 7 *M* urea and 0.02 *M* Tris–HCl. Fractions of the main peak (2.7 A_{260} units) were desalted as described above. The hexanucleotide was characterized by two-dimensional tlc after RNase St[24] and venom phosphodiesterase digestion as shown in Figure 8. Base analysis by digestion with pancreatic RNase and RNase St gave Gp and Cp in a ratio of 4.3:1.0, theoretical being 4:1. Venom phosphodiesterase digestion gave the ratio pG:pU:pC = 2.8:1.0:1.0 (theoretical, 3:1:1).

Acknowledgment. The authors are indebted to Dr. Masachika Irie of Kyoto University for a gift of RNase M, to Dr. Nobuo Yoshida of Shionogi Research Institute for RNase St, and to Kohjin Co. Ltd. for ribonucleotides. We thank Dr. Dieter Söll for reading the manuscript. A part of this work was supported by grants from the Ministry of Education of Japan and Toray Science Foundation.

(24) N. Yoshida, H. Inoue, A. Sasaki, and H. Otsuka, *Biochim. Biophys. Acta*, 228, 636 (1971).

Ab Initio Calculations on Large Molecules Using Molecular Fragments. Polypeptides of Glycine[1a]

Lester L. Shipman[1b] and Ralph E. Christoffersen[1c]

Contribution from the Department of Chemistry, University of Kansas, Lawrence, Kansas 66044. *Received September 5, 1972*

Abstract: The *ab initio* SCF molecular fragment approach is applied to a study of mono-, di-, tri-, tetra-, and pentapeptides of glycine. Fully extended, anti-parallel-chain pleated sheet, parallel-chain pleated sheet, and α-helix conformations are studied. Trends in total energy, orbital energies, populations, and bond orders are discussed, and intramolecular hydrogen bonding is identified as an important stabilizing force in the α-helix conformation of the tetra- and pentapeptides of glycine.

The use of X-ray diffraction techniques[2] and other physical and chemical methods of analysis[3] has led to major advances in the knowledge of the geometric structure of polypeptides. However, even though such techniques may, in optimum cases, reveal the detailed coordinates of virtually every nonhydrogen nucleus in a polypeptide, the elucidation of the forces giving rise to these structures (*i.e.*, the reasons for the stability of such conformations) and how stable they will be relative to external perturbations are extremely difficult to extract. Consequently, there is a need for additional analytical techniques that do not need to introduce, *via ad hoc* postulates, the kinds of forces that may be acting, but rather, that will allow the nature of the interactions to be extracted from the results of studies utilizing the analytical techniques. Such a possibility exists in principle through the use of the techniques of molecular quantum mechanics. This work is a report of the application of an *ab initio* quantum mechanical technique to a study of polyglycine, and will show how the nature of the forces

(1) (a) This work was supported in part by the National Science Foundation, the University of Kansas, and the Upjohn Company, Kalamazoo, Mich. 49001. (b) NSF Trainee, 1969–1972; (c) Alfred P. Sloan Research Fellow, 1971–1973.

(2) See, for example, W. N. Lipscomb, *Accounts Chem. Res.*, 3, 81 (1970).

(3) For a discussion of some of the available techniques, see S. Bernhard, "The Structure and Function of Enzymes," W. A. Benjamin, New York, N. Y., 1968.

that are acting can be extracted directly from the calculations.

There have been a number of quantum mechanical studies of peptides as mono- and dipeptides. Some of these studies have been *ab initio*,[4-14] while the majority have utilized other methods, mostly semiempirical.[15-48] In other quantum mechanical studies, the electronic band structure of extended hydrogen bonded peptide systems has been examined.[49-52] Quantum mechanical studies of single-chain polypeptides longer than a dipeptide are very limited in number,[53-56] and none have been carried out using *ab initio* techniques.

In a previous study,[14] *ab initio* molecular fragment SCF calculations were carried out on formamide, *N*-methylacetamide, and 2-formamidoacetamide, in order to examine the kinds of interactions expected along a single polypeptide chain, and the suitability of the molecular fragment approach[57-66] to describe them. In that study, the molecular fragment approach was found to provide a description of these model peptides that is in both qualitative and quantitative agreement with experimental and other theoretical studies. In particular, the valence molecular orbital structure, dipole moments, and energetic interactions between a given amide unit and attached hydrocarbon groups as well as with neighboring amide units were found to be in excellent agreement with available data from experimental and/or large basis set theoretical studies. The current studies represent application of this approach to single-stranded polypeptides of glycine, where the interactions are expected to be similar to those characterized in the prototype studies.

Specifically, the current studies use the molecular fragment approach to carry out *ab initio* SCF calculations on the di-, tri-, tetra-, and pentapeptides of glycine (see Figure 1). Calculations have been carried out at the fully extended (FE), anti-parallel-chain pleated-sheet (APC-PS), parallel-chain pleated sheet (PC-PS), and α-helix (α) conformations for the dipeptide (DP), tripeptide (TRP), and tetrapeptide (TEP). Calculations were also carried out at the fully extended and α-helix conformations for the pentapeptide (PP) and at a planar geometry for the monopeptide (MP). The monopeptide was chosen as formamide. The nuclear geometry used in these calculations is the "standard" geometry given by Momany, *et al.*, in Table I in ref 47, which was based upon a search of the recent structural literature. The C–H bond length for the "carbonyl end" of the peptide chain was taken as 1.102 Å. This value is the C–H bond length found

(4) H. Basch, M. B. Robin, and N. A. Kuebler, *J. Chem. Phys.*, 47, 1201 (1967).
(5) M. A. Robb and I. G. Csizmadia, *Theor. Chim. Acta*, 10, 269 (1968).
(6) M. A. Robb and I. G. Csizmadia, *J. Chem. Phys.*, **50**, 1819 (1969).
(7) J. B. Moffat, *J. Theor. Biol.*, 26, 437 (1970).
(8) (a) M. Dreyfus, B. Maigret, and A. Pullman, *Theor. Chim. Acta*, 17, 109 (1970); (b) M. Dreyfus and A. Pullman, *ibid.*, 19, 20 (1970).
(9) R. Bonaccorsi, A. Pullman, E. Scrocco, and J. Tomasi, *Chem. Phys. Lett.*, 12, 622 (1972).
(10) D. H. Christensen, R. N. Kortzeborn, B. Bak, and J. J. Led, *J. Chem. Phys.*, 53, 3912 (1970).
(11) C. R. Brundle, D. W. Truner, M. B. Robin, and H. Basch, *Chem. Phys. Lett.*, 3, 292 (1969).
(12) R. Ditchfield, J. E. Del Bene, and J. A. Pople, *J. Amer. Chem. Soc.*, 94, 703 (1972).
(13) J. A. Ryan and L. L. Whitten, *J. Amer. Chem. Soc.*, **94**, 2396 (1972).
(14) L. L. Shipman and R. E. Christoffersen, *J. Amer. Chem. Soc.*, **95**, 1408 (1973).
(15) R. Rein, N. Fukuda, H. Win, and G. A. Clarke, *J. Chem. Phys.*, **45**, 4743 (1966).
(16) A. Julg and P. Carles, *Theor. Chim. Acta*, 7, 103 (1967).
(17) J. A. Pople and M. Gordon, *J. Amer. Chem. Soc.*, 89, 4253 (1967).
(18) J. E. Bloor, B. R. Gibson, and F. A. Billingsly, II, *Theor. Chim. Acta*, 12, 360 (1968).
(19) J. F. Yan, F. A. Momany, R. Hoffmann, and H. A. Scheraga, *J. Phys. Chem.*, 74, 420 (1970).
(20) A. Pullman and H. Berthod, *Theor. Chim. Acta*, 10, 461 (1968).
(21) F. A. Momany, R. F. McGuire, J. F. Yan, and H. A. Scheraga, *J. Phys. Chem.*, **74**, 2424 (1970).
(22) K. N. Shaw and L. W. Reeves, *Chem. Phys. Lett.*, 10, 89 (1971).
(23) S. Nagakura, *Mol. Phys.*, 3, 105 (1960).
(24) A. Julg and P. Carles, *Theor. Chim. Acta*, 1, 140 (1963).
(25) K. Kaya and S. Nagakura, *Theor. Chim. Acta*, 7, 117 (1967).
(26) D. Poland and H. A. Scheraga, *Biochemistry*, 6, 3791 (1967).
(27) R. Hoffmann and I. Imamura, *Biopolymers*, 7, 207 (1969).
(28) L. B. Kier and J. M. George, *Theor. Chim. Acta*, 14, 258 (1969).
(29) B. Maigret, B. Pullman, and M. Dreyfus, *J. Theor. Biol.*, 26, 321 (1970).
(30) C. H. Bushweller, P. E. Stevenson, J. Golini, and J. W. O'Neil, *J. Phys. Chem.*, 74, 1155 (1970).
(31) A. S. N. Murthy, K. G. Rao, and C. N. R. Rao, *J. Amer. Chem. Soc.*, 92, 3544 (1970).
(32) A. S. N. Murthy, S. N. Bhat, and C. N. R. Rao, *J. Chem. Soc. A*, 1251 (1970).
(33) B. Maigret, B. Pullman, and D. Perahia, *Biopolymers*, 10, 107 (1971).
(34) B. Pullman, B. Maigret, and D. Perahia, *Theor. Chim. Acta*, 18, 44 (1970).
(35) B. Pullman, B. Maigret, and D. Perahia, *C. R. Acad. Sci.*, 270, 1396 (1970).
(36) B. Maigret, D. Perahia, and B. Pullman, *J. Theor. Biol.*, 29, 275 (1970).
(37) B. Maigret, D. Perahia, and B. Pullman, *Biopolymers*, 10, 491 (1971).
(38) G. Govil, *J. Chem. Soc. A*, 386 (1971).
(39) G. Govil, *J. Indian Chem. Soc.*, 48, 731 (1971).
(40) G. Govil, *J. Chem. Soc. A*, 2464 (1970).
(41) D. Perahia, B. Maigret, and B. Pullman, *Theor. Chim. Acta*, 19, 121 (1970).
(42) B. Maigret, B. Pullman, and J. Caillet, *Biochem. Biophys. Res. Commun.*, 40, 808 (1970).
(43) G. D. Zeiss and M. A. Whitehead, *J. Chem. Soc. A*, 1727 (1971).
(44) B. Maigret, B. Pullman, and D. Perahia, *J. Theor. Biol.*, 31, 269 (1971).
(45) J. Almlöf and O. Mårtensson, *Acta Chem. Scand.*, 25, 355 (1971).
(46) B. Pullman, J. L. Coubeils, P. Courriere, and D. Perahia, *Theor. Chim. Acta*, 22, 11 (1971).
(47) F. A. Momany, R. F. McGuire, J. F. Yan, and H. A. Scheraga, *J. Phys. Chem.*, 75, 2286 (1971).
(48) M. G. Evans and J. Gergeley, *Biochim. Biophys. Acta*, 3, 188 (1949).
(49) M. Suard, G. Berthier, and B. Pullman, *Biochim. Biophys. Acta*, 52, 254 (1961).
(50) M. Suard, *Biochim. Biophys. Acta*, 59, 227 (1962).
(51) M. Suard, *Biochim. Biophys. Acta*, 64, 400 (1962).
(52) S. Yomosa, *Biopolym. Symp.*, 1, 1 (1964).
(53) A. Rossi and C. W. David, *Theor. Chim. Acta*, 14, 429 (1969).
(54) A. R. Rossi, C. W. David, and R. Schor, *J. Phys. Chem.*, 74, 4551 (1970).
(55) R. Schor, H. Stimne, G. Wettermark, and C. W. David, *J. Phys. Chem.*, 76, 670 (1972).
(56) G. Govil and A. Saran, *J. Chem. Soc. A*, 3624 (1971).
(57) R. E. Christoffersen and G. M. Maggiora, *Chem. Phys. Lett.*, 3, 419 (1969).
(58) R. E. Christoffersen, D. W. Genson, and G. M. Maggiora, *J. Chem. Phys.*, 54, 239 (1971).
(59) G. M. Maggiora, D. W. Genson, R. E. Christoffersen, and B. V. Cheney, *Theor. Chim. Acta*, 22, 337 (1971).
(60) R. E. Christoffersen, *J. Amer. Chem. Soc.*, 93, 4104 (1971).
(61) R. E. Christoffersen, *Advan. Quantum Chem.*, 6, 333 (1972).
(62) R. E. Christoffersen, L. L. Shipman, and G. M. Maggiora, *Int. J. Quantum Chem.*, 5S, 143 (1971).
(63) B. V. Cheney and R. E. Christoffersen, *J. Chem. Phys.*, 56, 3503 (1972).
(64) D. W. Genson and R. E. Christoffersen, *J. Amer. Chem. Soc.*, 94, 6904 (1972).
(65) L. L. Shipman and R. E. Christoffersen, *Chem. Phys. Lett.*, 15, 469 (1972).
(66) D. W. Genson and R. E. Christoffersen, *J. Amer. Chem. Soc.*, 95, 362 (1973).

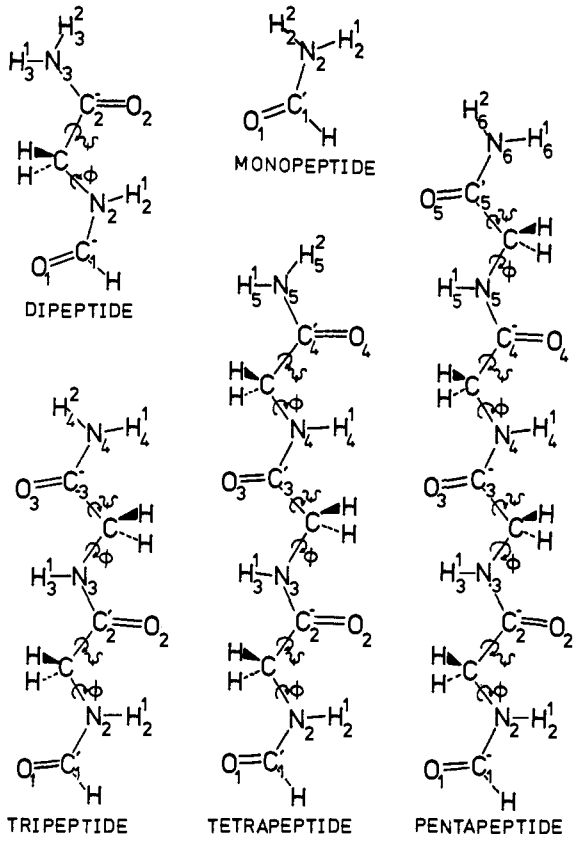

Figure 1. Polypeptides studied. Each molecule has been depicted in its fully extended conformation, *i.e.*, with all $(\phi, \psi) = (180, 180°)$.

Table I. Four Standard Peptide Conformations[a,b]

Conformation	ϕ	ψ
FE	180	180
APC-PS	−142	145
PC-PS	−119	113
α	−48	−57

[a] ϕ and ψ angles are given in degrees. [b] See ref 68.

in the most thorough microwave structural determination of formamide.[67] Standard values[68] of ϕ and ψ are used for the four conformations studied (see Table I). ϕ and ψ are defined according to the tentative rules (1969) established by the IUPAC–IUB Commision of Biochemical Nomenclature.[69]

Since the conceptual basis of the molecular fragment procedure, as well as the advantages and limitations of the procedure as revealed by applications to prototype systems, have been described elsewhere[14,57–66,68] they will not be repeated here. The molecular fragment floating spherical Gaussian orbital (FSGO) basis set[70] that has been taken from earlier studies and

(67) C. C. Costain and J. M. Dowling, *J. Chem. Phys.*, 32, 158 (1960).

(68) J. T. Edsall, P. J. Flory, J. C, Kendrew, A. M. Liquori, G. Nemethy, G. N. Ramachandran, and H. A. Scheraga, *Biopolymers*, 4, 121, 1149 (1966); *J. Biol. Chem.*, 241, 1004, 4167 (1966); *J. Mol. Biol.*, 15, 339 (1966); 20, 589 (1966).

(69) IUPAC–IUB Commission on Biochemical Nomenclature, *Biochemistry*, 9, 3471 (1970).

(70) These basis orbitals were introduced and utilized for small molecules by A. A. Frost and coworkers [*J. Chem. Phys.*, 54, 764 (1971), and in earlier references contained therein]. Formulation and characterization of these orbitals for large molecule calculations are given in ref 14 and 57–66.

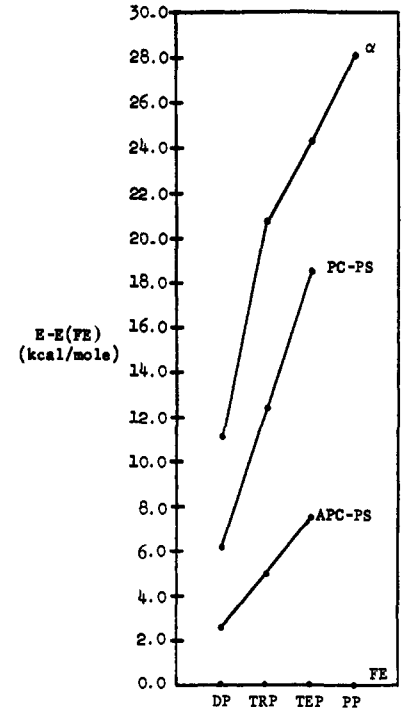

Figure 2. Energy of various conformations as a function of chain length, relative to the fully extended form: DP = dipeptide, TRP = tripeptide, TEP = tetrapeptide, and PP = pentapeptide.

Table II. Molecular Fragment Data[a]

Fragment type	FSGO type	FSGO distance from "heavy" atom	FSGO radii (ρ)
CH_4 (T_d)	C–H	1.23379402	1.67251562
R(C,H) = 2.05982176	C inner shell	0.0	0.32784375
$\cdot CH_3$ (planar)	C–H	1.13093139	1.51399487
R(C,H) = 1.78562447	C–π	±0.1	1.80394801
	C inner shell	0.0	0.32682735
·OH (sp hybrid)	O–H	0.76467773	1.23671871
R(O,H) = 1.54774058	O–LP (σ)	0.21614258	1.28753780
	O–LP (ρ)	±0.1	1.19741696
	O–π	±0.1	1.12242182
	O inner shell	0.00057129[b]	0.24028227
$:NH_3$ (planar)	N–H	0.75201903	1.39424495
R(N,H) = 1.93131910	N–π	±0.1	1.50625972
	N inner shell	0.0	0.27684894

[a] See ref 82. [b] This is the distance from the oxygen nucleus along the OH bond axis, toward the H nucleus.

utilized throughout the current calculations is described in Table II. The molecular fragments utilized were a planar $\cdot CH_3$ fragment, a planar $:NH_3$ fragment, a planar ·OH fragment, and a tetrahedral CH_4 fragment. The implications of these calculations, both chemical and quantum mechanical, are discussed in the following sections. However, even though considerable effort has been expended in order to create a "balanced basis set" for polypeptide description, it should be remembered that these basis sets will provide a description that is far (energetically) from the Hartree–Fock limit. Consequently, the inadequacies of such a basis set as well as the difficulties of interpreta-

tion that are inherent in methods of analysis such as the Mulliken analysis should be kept in mind in the following discussions.

Total Energies

In Figure 2 are given conformational energies relative to the fully extended conformations (whose energies are arbitrarily zeroed) as a function of chain length. The anti-parallel-chain pleated sheet and parallel-chain pleated sheet conformational energy curves are linear, suggesting that the total energy increase in going from the fully extended conformation to either of these conformations is composed of approximately equal and independent contributions from each pair of adjacent ϕ and ψ rotations. Further evidence for this suggestion will be seen in the analysis of the charge distributions and their changes as a function of conformation that is given in later sections. Also, as the chain length increases, it is seen that the anti-parallel-chain pleated sheet and parallel-chain pleated sheet conformations become less stable relative to the fully extended conformation. This result is not necessarily in conflict with the known occurrence of anti-parallel-chain pleated sheet and parallel-chain pleated sheet peptide conformations in nature,[71] since an important contribution to the stability of these conformations comes from interchain hydrogen bonding. Indeed, a rough estimate of the role which interchain hydrogen bonding may play can be obtained if additional assumptions are made concerning the possibility of interchain hydrogen bonding in the fully extended form. In particular, if it is assumed that interchain hydrogen bonding is not important in the fully extended conformation[72] of polyglycine, then examination of Figure 2 indicates that, on the average, hydrogen bonds must be stronger than 2.6 and 6.2 kcal/mol, for the anti-parallel-chain pleated sheet and parallel-chain pleated sheet conformations, respectively, if these forms are to be more stable than the fully extended form. Also, these results are independent of chain length.

Of particular interest is the α-helix conformational energy curve, which is seen to have a discontinuity in slope. As the following analysis indicates, this behavior can be rationalized in terms of hydrogen bonding interactions, that first become apparent at the tetrapeptide length.

Expressing this more quantitatively, let us assign the total dipeptide energy difference between the fully extended and α-helix form ($\Delta E_2^{\text{FE-}\alpha}$) to the energy of interaction between two adjacent amide units (labeled n and $n + 1$), *i. e.*

$$\Delta E_2^{\text{FE-}\alpha} = \Delta E_{n,n+1}^{\text{FE-}\alpha} \quad (1)$$

For the tripeptide, the total $\Delta E_3^{\text{FE-}\alpha}$ will be partitioned as

$$\Delta E_3^{\text{FE-}\alpha} = 2\Delta E_{n,n+1}^{\text{FE-}\alpha} + \Delta E_{n,n+2}^{\text{FE-}\alpha} \quad (2)$$

where $\Delta E_{n,n+2}^{\text{FE-}\alpha}$ is the interaction between a given amide unit (n) and another amide unit ($n + 2$) located two amide units from amide unit n.[73] Analogously, the total $\Delta E_4^{\text{FE-}\alpha}$ for the tetrapeptide and total $\Delta E_5^{\text{FE-}\alpha}$ for the pentapeptide can be partitioned as

$$\Delta E_4^{\text{FE-}\alpha} = 3\Delta E_{n,n+1}^{\text{FE-}\alpha} + 2\Delta E_{n,n+2}^{\text{FE-}\alpha} + \Delta E_{n,n+3}^{\text{FE-}\alpha} \quad (3)$$

and

$$\Delta E_5^{\text{FE-}\alpha} = 4\Delta E_{n,n+1}^{\text{FE-}\alpha} + 3\Delta E_{n,n+2}^{\text{FE-}\alpha} + 2\Delta E_{n,n+3}^{\text{FE-}\alpha} + \Delta E_{n,n+4}^{\text{FE-}\alpha} \quad (4)$$

The generalization of this type of analysis for comparison of any two conformations of a polypeptide of arbitrary length is obvious. Application to the results of the current studies yields

$$\Delta E_{n,n+1}^{\text{FE-}\alpha} = +11.2 \text{ kcal/mol} \quad (5)$$

$$\Delta E_{n,n+2}^{\text{FE-}\alpha} = -1.6 \text{ kcal/mol} \quad (6)$$

$$\Delta E_{n,n+3}^{\text{FE-}\alpha} = -6.1 \text{ kcal/mol} \quad (7)$$

$$\Delta E_{n,n+4}^{\text{FE-}\alpha} = +0.3 \text{ kcal/mol} \quad (8)$$

$\Delta E_{n,n+3}^{\text{FE-}\alpha}$ is expected, due to the geometric arrangement of the α-helix, to reflect primarily the stabilization of the α-helix due to hydrogen bond formation, if it exists. As is seen from the magnitude of $\Delta E_{n,n+3}^{\text{FE-}\alpha}$ the assignment of it as a direct measure of the strength of a hydrogen bond formed between amide unit n and amide unit $n + 3$ gives rise to a hydrogen bond strength that is significant, but well within the range of hydrogen bond energies found in other quantum mechanical studies (see Table III) on peptides. Additional evidence for the identification of this interaction as a hydrogen bonding effect will be given in the section on populations and bond orders, where the charge distribution in the various conformations is discussed.

From distance considerations alone, the absolute value of the interaction energy between amide unit n and amide units more distant than $n + 4$ would be expected to be less than $|\Delta E_{n,n+4}|$. Thus, for a polyglycine chain containing N amide units ($N \geq 4$), if $\Delta E_{n,n+i}$ ($i > 4$) terms are assumed to be negligible, the α-helix conformation should be energetically less favorable than the fully extended conformation by $\sim[24.3 + 3.8(N - 4)]$ kcal/mol. In other words, for any chain length, polyglycine should prefer the fully extended conformation over the α-helix conformation.

This conclusion is supported by several different experimental observations.[74-76] For example, McDiarmid[74] has found, in a far-ultraviolet spectroscopic study of oligoglycines (n = 2–5), that a random structure is preferred. In addition, Fraser, *et al.*,[75] have studied ordered sequences of glycyl and γ-ethyl-L-glutamyl residues using X-ray diffraction, infrared,

(71) J. A. Schellman and C. Schellman, *Proteins*, 2, 1 (1964).

(72) While this assumption is probably justified for general fully extended polypeptides, where bulky R groups in side chains are available for interference with interchain hydrogen bond formation, the validity of such an assumption here for R = H is less obvious. Consequently, the discussion of interchain hydrogen bonding effects given here must be considered within the context of this possible complication.

(73) This type of analysis implicitly assumes that the observed effects can be described by additive interactions. Thus, for example, in eq 2 it is assumed that $\Delta E_{n,n+1}^{\text{FE-}\alpha}$ is unchanged by the addition of a new interaction, $\Delta E_{n,n+2}^{\text{FE-}\alpha}$. While this is not the only type of analysis that can be carried out for these data, it will be seen to form both a convenient and consistent description of the observed results.

(74) R. S. McDiarmid, Dissertation, Harvard University, 1965, and private communication.

(75) R. D. B. Fraser, B. S. Harrap, T. P. MacRae, F. H. C. Stewart, and E. Suzuki, *Biopolymers*, 5, 251 (1967).

(76) (a) C. H. Bamford, L. Brown, E. M. Cant, A. Elliot, W. E. Hanby, and B. R. Malcolm, *Nature (London)*, 176, 396 (1955); (b) A. Rich and F. H. C. Crick, *ibid.*, 176, 915 (1955).

Table III. Hydrogen Bond Energies from Quantum Mechanical Studies[a]

Method	Entity studied	Hydrogen bond energy, kcal/mol	Ref
CNDO/2	Formamide planar parallel linear dimer	4.9	20
		9.30	21
		5.0	31
	Formamide planar antiparallel linear dimer	4.6	20
		9.21	21
		5.0	31
Ab initio		7.95	8b
CNDO/2	Formamide planar cyclic dimer	5.8	20
		11	21
		6.5	31
		6.5	32
Ab initio		7–9.5	8a
CNDO/2	Formamide planar linear trimer	10	21
		6	31
	Formamide planar forked trimer	8.4	21
	N-Methylacetamide linear dimer	9.4	21
	N-Methylacetamide trans dimer	5	31
	N-Methylacetamide dimer	3.5	32
CNDO	Acetamide dimer in a PC-PS arrangement	4.1	56
	Acetamide dimer in an APC-PS arrangement	5.1	56
	Acetamide and *N*-methylformamide in an α-helical arrangement	3.8	56

[a] While all of the studies in this table deal with peptide interactions, direct comparisons with the current studies may be complicated by the fact that many of the studies included here deal with intermolecular hydrogen bonds (instead of intramolecular hydrogen bonds, as studied here), where additional geometric constraints may affect the strength of the hydrogen bonds.

and optical rotatory dispersion techniques, and concluded that the presence of glycyl residues in polypeptide chains caused a marked reduction in the stability of the α-helix, relative to a random conformation in solution and to the β conformation in the crystalline state. Finally, in the crystalline state,[77] polyglycine is found in two forms, neither of which is α-helical. Polyglycine I is a β form in which the polyglycine chains are fully extended or nearly fully extended,[76a] and polyglycine II consists of helices with a threefold screw axis,[76b] having $(\phi, \psi) = (-80, 150°)$.

In contrast to the α-helix studies, application of the previous analysis to the other conformers gives

$$\Delta E_{n,n+1}^{\text{FE-APC-PS}} = +2.6 \text{ kcal/mol} \quad (9)$$

$$\Delta E_{n,n+2}^{\text{FE-APC-PS}} = -0.1 \text{ kcal/mol} \quad (10)$$

$$\Delta E_{n,n+3}^{\text{FE-APC-PS}} = 0.0 \text{ kcal/mol} \quad (11)$$

and

$$\Delta E_{n,n+1}^{\text{FE-PC-PS}} = +6.2 \text{ kcal/mol} \quad (12)$$

$$\Delta E_{n,n+2}^{\text{FE-PC-PS}} = 0.0 \text{ kcal/mol} \quad (13)$$

$$\Delta E_{n,n+3}^{\text{FE-PC-PS}} = 0.0 \text{ kcal/mol} \quad (14)$$

which indicates in perhaps a more striking fashion the lack of interactions beyond the nearest neighbors in going from the fully extended to either the anti-parallel-chain pleated sheet or parallel-chain pleated sheet conformers.

The results of the current *ab initio* conformational energy study can be compared with the results of EHT[53,54] and CNDO/2[55] conformational studies on a tetrapeptide of glycine by Rossi, *et al.*, and Schor, *et al.*, respectively. The peptide geometry used in all three of these studies was the geometry of Leach, *et al.*[78] The actual tetrapeptide used in all three studies was $CH_3CONHCH_2CONHCH_2CONHCH_2CONHCH_3$.

In the first EHT study, Rossi, *et al.*,[53] found a relative minimum at $(\phi, \psi) = (-43, -52°)$ in the α-helix region, which was 5.4 eV (124 kcal/mol) above the energy of the fully extended conformation. In this EHT calculation, the Cusachs approximation[79] was used. In a second EHT study,[54] the same authors found the α-helix and fully extended regions comparable in energy, with the fully extended region being slightly lower in energy. Instead of using the Cusachs approximation, the Wolfsberg–Helmholtz approximation[80] was used in the latter calculations.

In the CNDO/2 study, Schor, *et al.*,[55] found three nonequivalent minima; an absolute minimum at $(\phi, \psi) = (-20, -60°)$, a second minimum at $(\phi, \psi) = (-60, 60°)$ which was 12 kcal/mol above the absolute minimum, and a third minimum at $(\phi, \psi) = (180, 180°)$ which was 15 kcal/mol above the absolute minimum. The first minimum was near the 3_{10} helix, $(\phi, \psi) = (-49, -26°)$, and the α-helix regions, the second minimum corresponded to a seven-membered ring, and the third minimum corresponded to the fully extended conformation. In addition, the same authors found that the energy of the α-helix conformation was 1.0 kcal/mol below that of the fully extended conformation.[81]

From calculations on small peptides, it has been established that EHT fails to account adequately for hydrogen bond formation, while CNDO and CNDO/2 do account for hydrogen bond formation.[21,32,47,56] In addition, the appearance of a hydrogen bonded seven-membered ring as an important minimum in the CNDO/2 calculation, and its failure to appear in the EHT calculation, are consistent with previous dipeptide results.[47]

In EHT and CNDO studies of the tetraglycine, $CH_3CONHCH_2CONHCH_2CONHCH_2CONH_2$, Govil and Saran[56] have found that the α-helix conformation is less stable than the fully extended conformation by 16.2 kcal/mol using EHT and 2.6 kcal/mol using CNDO.

In addition to the information concerning energetic interactions that arise when the conformation of polyglycine is changed from the fully extended to other conformations, it is also of interest to examine the nature of the contributions to a given conformer. Of particular interest in this case is an examination of the fully extended conformer, to see if there are delocalization effects that are already present within this particular conformation, before any rotations occur. To

(77) Comparisons here are complicated by solid-state effects, *e.g.*, crystal packing effects but, as indicated, still appear consistent with the results of the current studies.

(78) S. J. Leach, G. Nemethy, and H. A. Scheraga, *Biopolymers*, 4, 369 (1966).

(79) L. C. Cusachs, *J. Chem. Phys.*, 38, 1607 (1963).

(80) M. Wolfsberg and L. Helmholtz, *J. Chem. Phys.*, 20, 837 (1952).

(81) Private communication: C. W. David, Department of Chemistry, the University of Connecticut, Storrs, Conn.

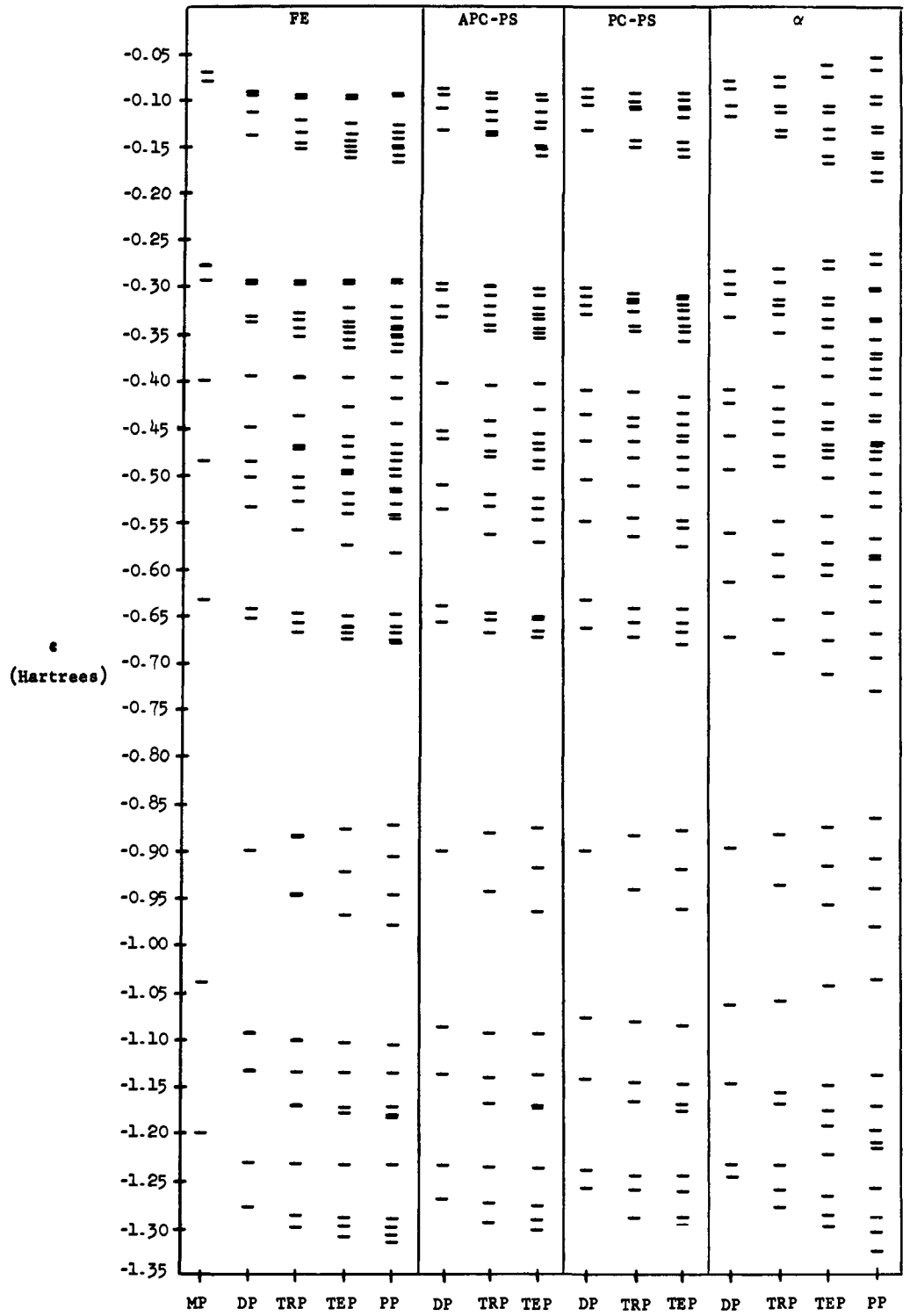

Figure 3. Valence molecular orbital energies *vs.* chain length for various conformations of polyglycine: DP = dipeptide, TRP = tripeptide, TEP = tetrapeptide, and PP = pentapeptide.

do this, we write the polypeptides studied here using the formula

$$HCONH(CH_2CONH)_nCH_2CONH_2,\ n = 0, 1, 2, 3 \quad (15)$$

The total energy of $HCONHCH_2CONH_2$ ($n = 0$) is just the total energy of the dipeptide (−320.1198 hartrees). Next, assuming that the energy of each $-CH_2-CONH-$ unit has the same energy, the energy of this unit may be calculated algebraically from the total energies of the tri-, tetra-, and pentapeptides. The results, in hartrees, are −176.2986, −176.2986, and −176.2985 for the tri-, tetra-, and pentapeptides, respectively. The difference in the fourth decimal place may be due only to numerical round-off. Thus, there is no evidence of significant stabilization due to delocalized effects, since such stabilization should increase with increasing chain length. Therefore, the total energy,[82] E_n, of the fully extended polyglycine

(82) All energies and distances, unless otherwise specified, are reported in Hartree atomic units. See H. Shull and G. G. Hall, *Nature (London)*, 184, 1559 (1959).

Table IV. Distribution of Electrons in the HOMO-1, HOMO, and LUMO[a] among the Amide Units and CH_2's for Various Polypeptides of Glycine in the FE, APC-PS, PC-PS, and α Conformations[b]

Conformation	MO	Molecule	All CH_2's combined	Amide unit no. 1	2	3	4	5
	HOMO-1	MP		2.000 (σ_n)				
	HOMO			2.000 (π_n)				
	LUMO			2.000 (π_a)				
FE	HOMO-1	DP	0.015	1.983 (σ_n)	0.002			
		TRP	0.070	1.917 (π_n)	0.013	0		
		TEP	0.070	1.920 (π_n)	0.010	0	0	
		PP	0.070	1.919 (π_n)	0.011	0	0	0
	HOMO	DP	0.067	1.885 (π_n)	0.048			
		TRP	0.016	1.983 (σ_n)	0.001	0		
		TEP	0.015	1.983 (σ_n)	0.002	0	0	
		PP	0.015	1.983 (σ_n)	0.002	0	0	0
	LUMO	DP	0.031	0.021	1.948 (π_a)			
		TRP	0.039	0.008	1.671 (π_a)	0.282 (π_a)		
		TEP	0.042	0	0.201 (π_a)	1.661 (π_a)	0.096	
		PP	0.045	0.033	0.634 (π_a)	1.241 (π_a)	0.047	0
APC-PS	HOMO-1	DP	0.019	1.948 (σ_n)	0.033			
		TRP	0.027	1.936 (σ_n)	0.035	0.002		
		TEP	0.028	1.934 (σ_n)	0.038	0	0	
	HOMO	DP	0.075	1.661 (π_n)	0.264 (σ_n)			
		TRP	0.071	1.763 (π_n)	0.146	0.020		
		TEP	0.069	1.798 (π_n)	0.125	0.008	0	
	LUMO	DP	0.028	0.177	1.795 (π_a)			
		TRP	0.033	0.057	1.539 (π_a)	0.371 (π_a)		
		TEP	0.040	0.009	0.485 (π_a)	1.335 (π_a)	0.131	
PC-PS	HOMO-1	DP	0.028	0.352 (σ_n)	1.620 (π_n)			
		TRP	0.037	0.517 (π_n)	0.203 (σ_n)	1.243 (π_n)		
		TEP	0.044	0.362 (π_n)	0.077	0.247 (π_n)	1.270 (π_n)	
	HOMO	DP	0.087	1.192 (π_n)	0.721 (σ_n)			
		TRP	0.082	0.914 (π_n)	0.532 (π_n)	0.472 (π_n)		
		TEP	0.085	1.080 (π_n)	0.388 (π_n)	0.187	0.260 (π_n)	
	LUMO	DP	0.019	0.574 (π_a)	1.407 (π_a)			
		TRP	0.027	0.168	1.478 (π_a)	0.327 (π_a)		
		TEP	0.032	0.056	0.799 (π_a)	1.010 (π_a)	0.103	
α	HOMO-1	DP	0.106	0.639 (π_n)	1.255 (σ_n)			
		TRP	0.113	0.019	0.377 (π_n)	1.491 (σ_n)		
		TEP	0.117	0	0.006	0.316 (π_n)	1.561 (σ_n)	
		PP	0.116	0	0	0.007	0.413 (π_n)	1.464 (σ_n)
	HOMO	DP	0.004	0.274 (σ_n)	1.722 (π_n)			
		TRP	0.008	0.004	0.075	1.913 (π_n)		
		TEP	0.005	0	0.001	0.044	1.950 (π_n)	
		PP	0.008	0	0	0.001	0.055	1.936 (π_n)
	LUMO	DP	0.005	1.492 (π_a)	0.500 (π_a)			
		TRP	0.011	0.971 (π_a)	0.973 (π_a)	0.045		
		TEP	0.005	1.665 (π_a)	0.311 (π_a)	0.019	0	
		PP	0.022	0.465 (π_a)	1.463 (π_a)	0.048	0.002	0

[a] The electron distribution in the LUMO is a "virtual" distribution since the LUMO is unoccupied. [b] Where the number of electrons in an amide unit is greater than 0.2, the symbol denoting the local symmetry (see Figure 4) making the largest contribution is enclosed in parentheses.

chain defined in eq 15 may be estimated by the formula

$$E_n = -[320.1198 + n(176.2986)] \quad (16)$$

Of course, this conclusion does not imply that the shape and energy of individual molecular orbitals must also be additive functions of the peptide chain length.

Govil and Saran[56] have found that EHT and CNDO also predict that the total energy of polyglycine in the fully extended conformation is a linear function of the number of amide units in the chain.

Molecular Orbital Structure

The orbital energies for the valence molecular orbitals of polyglycine in the four conformations studied are plotted *vs.* chain length in Figure 3. The monopeptide orbital energies have been included in the FE section for reference purposes. Note that a band structure is discernible, although there is considerable mixing of the bands between $\epsilon = -0.60$ and $\epsilon = -0.25$ hartrees, and also between $\epsilon = -0.20$ and $\epsilon = -0.05$ hartrees. The latter region in which mixing occurs is the region of the higher occupied molecular orbitals. These molecular orbitals are formed with local σ_n and π_n symmetries (see Figure 4).

For brevity, the following discussion of the approximate Hartree–Fock molecular orbitals will be limited to the "chemically most interesting" molecular orbitals, *i.e.*, the higher occupied and lower unoccupied molecular orbitals. In each category is a group of molecular orbitals whose orbital energies are isolated in a cluster from other orbital energies, *i.e.*, there is a significant energy gap between the orbital energies in the cluster and the nearest orbital energies not in the cluster.

It should be noted that (although not depicted in Table IV), in all four conformations studied, the energy of the lowest unoccupied molecular orbital (LUMO) decreased with increasing chain length. Also, in the fully extended conformation, the energy of the highest occupied molecular orbital apparently

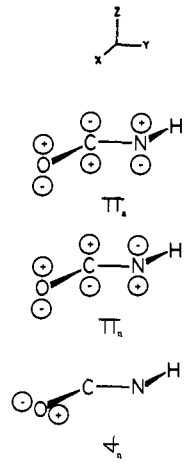

Figure 4. Amide unit orbital symmetries.

rapidly approaches an asymptotic value with increasing chain length. The same trend is found for the anti-parallel-chain pleated sheet and parallel-chain pleated sheet conformations, although the convergence to an asymptotic value is not as rapid. On the other hand, the energy of the HOMO in the α-helix conformation behaves quite differently in that, as the chain length increases, the energy of the HOMO *increases*.

As observed in previous studies,[14,57–66] the magnitudes of the orbital energies in Figure 3 are not the values expected from near Hartree-Fock wave functions. However, the previous studies have shown that, to a good approximation, the orbital energies of all valence molecular orbitals as computed from near Hartree–Fock wave functions are a linear function of the corresponding molecular orbital energies calculated using the molecular fragment approach. For example, the result of a least-squares fit of the data for formamide to a straight line is[14,82]

$$\epsilon_i(\text{near Hartree–Fock}) = 0.8511\epsilon_i(\text{molecular fragment}) - 0.354 \pm 0.015 \quad (17)$$

Thus, although the absolute magnitude of the various orbital energies will not be directly comparable to ionization potentials through Koopmans theorem,[83] the spacing between the molecular orbital energies is expected to be adequately represented.

Table IV shows how pairs of electrons in each HOMO-1, HOMO, and LUMO would be distributed among the amide units and CH_2's of polyglycine in various conformations and for various chain lengths. Where a single amide unit contributes more than 10% (0.2 electron) to a given molecular orbital, the symbol denoting the local symmetry (see Figure 4) making the largest contribution is entered in the table. The number of electrons in each amide unit has been calculated by summing up the number of electrons in each symmetric orthonormalized basis function residing in the amide unit. Further details of this procedure are given in the next section.

Although not necessarily expected from the SCF procedure, the HOMO-1 and HOMO are consistently localized in the first amide unit in the fully extended conformation, *i.e.*, they are localized at the "carbonyl end" of the molecule for all chain lengths. The LUMO, on the other hand, is localized in the second amide unit of the dipeptide and the second and third amide units for the tri-, tetra-, and pentapeptides. As the results of the conformational energy study in the previous section indicated, polyglycine molecules might be expected to be in a fully extended or nearly fully extended conformation in solution. Thus, a consideration of the molecular orbitals of this conformation may be helpful in rationalizing the reactivity of polyglycine. Under the simple assumptions of frontier orbital theory[84] that the electron densities defined by the LUMO and HOMO can be used to estimate the location of nucleophilic and electrophilic attack, respectively, it is seen that nucleophilic attack would be expected to occur at the second or third amide unit and electrophilic attack would be expected to occur at the first amide unit. However, it should be emphasized that such considerations are, at best, only rough approximations, and should be treated as such. Also, it is interesting to note that the nature of HOMO is π_n for the dipeptide and σ_n for the tri-, tetra-, and pentapeptides.

In the anti-parallel-chain pleated sheet conformation, the SCF procedure results in the HOMO-1 being localized in the first amide unit, while the HOMO is again localized in the first two amide units, although the degree of HOMO localization is not as great as in the fully extended conformation. The LUMO is localized in the second amide unit of the dipeptide and the second and third amide units of the tri- and tetrapeptides. The nature of the HOMO is π_n in the first amide unit of the dipeptide and σ_n in the second amide unit. The nature of the HOMO in the tri- and tetrapeptides is π_n. In the parallel-chain pleated sheet conformation, the HOMO-1, HOMO, and LUMO appear quite delocalized. In the α-helix conformation, the HOMO-1 and HOMO are localized in the last two amide units, and the LUMO is localized in the first two amide units. The nature of the HOMO in the dipeptide is σ_n in the first amide unit and π_n in the second amide unit. The nature of the HOMO in the tri-, tetra-, and pentapeptides is π_n.

Populations and Bond Orders

In order to examine these charge distributions in greater detail, variations in charges and bond orders have been followed as a function of conformation, chain length, and amide unit number. The particular type of population analysis utilized consists of a consideration of the **P** matrix in the basis of the symmetric orthonormalized molecular fragment FSGO,[65] which can be considered to be a modification of the Mulliken analysis[85] in a manner suitable for application to molecular fragment FSGO basis sets.

Consider the original FSGO basis, ϕ, and the overlap matrix in this basis, **S**. The **P** matrix in the original FSGO basis, $\mathbf{P}^{(\varphi)}$, is calculated as

$$P_{ij}^{(\varphi)} = 2\sum_{k}^{\text{occ}} C_{ik}^{(\varphi)} C_{jk}^{(\varphi)} \quad (18)$$

(83) T. A. Koopmans, *Physica*, 1, 104 (1933).

(84) See, for example, L. Salem, "Molecular Orbital Theory of Conjugated Systems," W. A. Benjamin, New York, N. Y., 1966, pp 326–332. See also K. Fukui, T. Yonezawa, and H. Shingu, *J. Chem. Phys.*, 20, 722 (1952); K. Fukui, T. Yonezawa, and C. Nagata, *Bull. Chem. Soc. Jap.*, 27, 423 (1954).

(85) R. S. Mulliken, *Phys. Rev.*, 41, 66 (1932); *J. Chem. Phys.*, 3, 573 (1935); 23, 1833, 1841, 2238, 2343 (1955); 36, 3428 (1962); *J. Chim. Phys. Physicochim. Biol.*, 46, 497, 675 (1949).

Table V. Selected Orbital Populations and Bond Orders for σ-Type Orbitals[a]

Atom	Fully extended					Anti-parallel-chain pleated sheet				Parallel-chain pleated sheet				α-Helix				
	MP	DP	TRP	TEP	PP	MP	DP	TRP	TEP	MP	DP	TRP	TEP	MP	DP	TRP	TEP	PP
									σ-Type Orbital Populations[b,c]									
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
O_1(LP-σ)	1.996	1.995	1.995	1.995	1.995	1.996	1.995	1.995	1.995	1.996	1.995	1.995	1.995	1.996	1.995	1.995	1.995	1.995
							(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)		(0.000)	(−0.001)	(−0.001)	(−0.001)
O_2(LP-σ)		1.994	1.994	1.994	1.994		1.994	1.994	1.994		1.994	1.994	1.994		1.994	1.993	1.993	1.993
								(0.000)	(0.000)			(0.000)	(0.000)			(0.000)	(−0.001)	(−0.001)
O_3(LP-σ)			1.994	1.994	1.994			1.994	1.994			1.994	1.994			1.994	1.994	1.993
									(0.000)				(0.000)				(0.000)	(−0.001)
O_4(LP-σ)				1.994	1.994				1.994				1.994				1.994	1.993
																		(0.000)
O_5(LP-σ)					1.994													1.994
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)	(−0.000)	(−0.010)	(−0.009)	(−0.010)
O_1(LP-π)	1.976	1.972	1.972	1.972	1.972	1.976	1.972	1.972	1.972	1.976	1.972	1.972	1.972	1.976	1.962	1.962	1.963	1.962
							(+0.002)	(+0.002)	(+0.002)		(+0.005)	(+0.006)	(+0.005)		(+0.006)	(−0.004)	(−0.004)	(−0.002)
O_2(LP-π)		1.953	1.950	1.950	1.950		1.955	1.952	1.952		1.958	1.956	1.955		1.959	1.946	1.946	1.948
								(+0.002)	(+0.003)			(+0.006)	(+0.006)			(+0.007)	(−0.003)	(−0.003)
O_3(LP-π)			1.952	1.949	1.949			1.954	1.952			1.958	1.955			1.959	1.946	1.946
									(+0.002)				(+0.006)				(+0.007)	(−0.003)
O_4(LP-π)				1.952	1.949				1.954				1.958				1.959	1.946
																		(+0.007)
O_5(LP-π)					1.952													1.959
						(0.000)	(+0.001)	(+0.001)	(+0.001)	(0.000)	(+0.002)	(+0.002)	(+0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(−0.001)
N_2–H_2'	1.995	1.992	1.992	1.992	1.992	1.995	1.993	1.993	1.993	1.995	1.994	1.994	1.994	1.995	1.992	1.992	1.992	1.991
							(0.000)	(+0.001)	(+0.001)		(0.000)	(+0.003)	(+0.003)		(0.000)	(+0.001)	(0.000)	(0.000)
N_3–H_3'		1.994	1.991	1.991	1.991		1.994	1.992	1.992		1.994	1.994	1.994		1.994	1.992	1.991	1.991
								(0.000)	(+0.001)			(0.000)	(+0.003)			(0.000)	(−0.003)	(0.000)
N_4–H_4'			1.994	1.994	1.991			1.994	1.992			1.994	1.994			1.994	1.991	1.991
									(0.000)				(+0.003)				(0.000)	(0.000)
N_5–H_5'				1.994	1.991				1.994				1.994				1.994	1.991
																		(0.000)
N_6–H_6'					1.994													1.994
									σ-Bond Orders[d,e]									
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(−0.001)	(−0.001)
C_1'–O_1	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.960	0.959	0.959
							(0.000)	(0.000)	(0.000)		(0.000)	(+0.001)	(+0.001)		(0.000)	(0.000)	(0.000)	(0.000)
C_2'–O_2		0.959	0.959	0.959	0.959		0.959	0.959	0.959		0.959	0.960	0.960		0.959	0.959	0.959	0.959
								(−0.001)	(+0.001)			(−0.001)	(+0.001)			(+0.001)	(0.000)	(0.000)
C_3'–O_3			0.960	0.959	0.959			0.959	0.960			0.959	0.960			0.959	0.959	0.959
									(0.000)				(0.000)				(0.000)	(0.000)
C_4'–O_4				0.959	0.959				0.959				0.959				0.959	0.959
																		(0.000)
C_5'–O_5					0.959													0.959
						(0.000)	(0.000)	(+0.001)	(+0.001)	(0.000)	(0.000)	(+0.001)	(+0.001)	(0.000)	(0.000)	(+0.001)	(+0.002)	(+0.002)
C_1'–N_2	0.946	0.943	0.942	0.942	0.942	0.960	0.943	0.943	0.943	0.960	0.943	0.943	0.943	0.946	0.943	0.943	0.944	0.944
							(0.000)	(0.000)	(0.000)		(0.000)	(+0.001)	(0.000)		(+0.001)	(+0.001)	(+0.001)	(+0.001)
C_2'–N_3		0.943	0.941	0.941	0.941		0.943	0.941	0.941		0.943	0.942	0.941		0.944	0.942	0.942	0.943
								(0.000)	(0.000)			(0.000)	(+0.001)			(+0.001)	(+0.001)	(+0.001)
C_3'–N_4			0.943	0.941	0.941			0.943	0.941			0.943	0.942			0.944	0.942	0.942
									(0.000)				(0.000)				(+0.001)	(+0.002)
C_4'–N_5				0.943	0.941				0.943				0.943				0.944	0.943
																		(+0.001)
C_5'–N_6					0.943													0.944

[a] The identification of atoms is given in Figure 1. [b] In parentheses is given: Pop. (i) − Pop. (fully extended), where i refers to the conformation to which rotation occurred. [c] Although one of the oxygen lone-pair orbitals is a π-type orbital, it contributes density to the same (local) plane as the other σ orbitals, and its inclusion in this table is appropriate. [d] In parentheses is given: bond order (i) − bond order (fully extended), where i refers to the conformation to which rotation occurred. [e] These bond orders refer to the bond order between the two symmetrically orthonormalized FSGO in the bonding region between the nuclei involved.

Table VI. Orbital Populations and Bond Orders for π-Type Orbitals[a]

Atom	Fully extended					Anti-parallel-chain pleated sheet				Parallel-chain pleated sheet				α-Helix				
	MP	DP	TRP	TEP	PP	MP	DP	TRP	TEP	MP	DP	TRP	TEP	MP	DP	TRP	TEP	PP
									π-Type Orbital Populations[b]									
						(0.000)	(+0.001)	(+0.001)	(+0.001)	(0.000)	(+0.004)	(+0.003)	(+0.003)	(0.000)	(−0.017)	(−0.019)	(+0.003)	(+0.026)
O_1	1.170	1.139	1.336	1.136	1.136	1.170	1.140	1.137	1.137	1.170	1.143	1.139	1.139	1.170	1.122	1.117	1.159	1.162
							(−0.002)	(+0.001)	(0.000)		(−0.006)	(0.000)	(−0.009)		(−0.031)	(−0.046)	(−0.050)	(−0.008)
O_2		1.211	1.182	1.180	1.180		1.209	1.183	1.180		1.205	1.182	1.171		1.180	1.136	1.130	1.172
								(−0.003)	(0.000)			(−0.007)	(−0.001)			(−0.023)	(−0.039)	(−0.042)
O_3			1.206	1.177	1.175			1.203	1.177			1.199	1.176			1.183	1.138	1.133
									(−0.003)				(−0.008)				(−0.017)	(−0.033)
O_4				1.207	1.178				1.204				1.199				1.190	1.145
																		(−0.014)
O_5					1.206													1.192
						(0.000)	(+0.001)	(+0.001)	(+0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(+0.005)	(+0.010)	(−0.017)	(−0.021)
C_1'	1.069	1.074	1.074	1.074	1.074	1.069	1.075	1.075	1.075	1.069	1.074	1.074	1.074	1.069	1.079	1.084	1.057	1.053
							(+0.001)	(+0.002)	(+0.001)		(+0.005)	(+0.006)	(+0.004)		(+0.016)	(+0.022)	(+0.027)	(+0.002)
C_2'		1.078	1.079	1.080	1.079		1.079	1.081	1.081		1.083	1.085	1.084		1.094	1.010	1.107	1.081
								(+0.001)	(+0.002)			(+0.004)	(+0.005)			(+0.013)	(+0.019)	(+0.024)
C_3'			1.083	1.084	1.084			1.084	1.086			1.087	1.089			1.096	1.103	1.108
									(+0.001)				(+0.005)				(+0.012)	(+0.017)
C_4'				1.083	1.084				1.084				1.088				1.095	1.101
																		(+0.009)
C_5'					1.084													1.093
						(0.000)	(−0.006)	(−0.006)	(−0.006)	(0.000)	(−0.015)	(−0.014)	(−0.014)	(0.000)	(+0.006)	(+0.003)	(−0.010)	(−0.010)
N_2	1.761	1.790	1.792	1.792	1.792	1.761	1.784	1.786	1.786	1.761	1.775	1.778	1.778	1.761	1.796	1.795	1.782	1.782
							(−0.001)	(−0.008)	(−0.007)		(+0.001)	(−0.015)	(−0.013)		(+0.020)	(+0.023)	(+0.021)	(+0.008)
N_3		1.739	1.771	1.773	1.773		1.738	1.763	1.766		1.740	1.756	1.760		1.759	1.794	1.794	1.781
								(−0.001)	(−0.007)			(+0.001)	(−0.014)			(+0.014)	(+0.018)	(+0.016)
N_4			1.739	1.770	1.772			1.738	1.763			1.740	1.756			1.753	1.788	1.788
									(−0.001)				(+0.001)				(+0.008)	(+0.014)
N_5				1.738	1.769				1.737				1.739				1.746	1.783
																		(+0.009)
N_6					1.738													1.747
									π-Bond Orders[c,d]									
						(0.000)	(−0.001)	(−0.001)	(−0.001)	(0.000)	(−0.003)	(−0.001)	(−0.001)	(0.000)	(+0.005)	(+0.004)	(−0.004)	(−0.004)
C_1'–O_1	0.879	0.893	0.894	0.894	0.894	0.879	0.892	0.893	0.893	0.879	0.890	0.893	0.893	0.879	0.898	0.898	0.890	0.890
							(0.000)	(−0.002)	(−0.002)		(−0.001)	(−0.005)	(−0.003)		(+0.008)	(+0.011)	(+0.010)	(+0.002)
C_2'–O_2		0.853	0.868	0.869	0.869		0.853	0.866	0.867		0.852	0.863	0.866		0.861	0.879	0.879	0.871
								(0.000)	(−0.002)			(0.000)	(−0.003)			(+0.005)	(+0.011)	(+0.008)
C_2'–O_3			0.853	0.868	0.869			0.853	0.866			0.853	0.865			0.858	0.877	0.877
									(0.000)				(0.000)				(+0.002)	(+0.007)
C_4'–O_4				0.853	0.867				0.853				0.853				0.855	0.874
																		(+0.002)
C_5'–O_5					0.853													0.855
						(0.000)	(+0.001)	(0.000)	(0.001)	(0.000)	(+0.002)	(0.000)	(+0.001)	(0.000)	(−0.015)	(−0.014)	(+0.008)	(+0.009)
C_1'–N_2	0.472	0.441	0.439	0.438	0.438	0.472	0.442	0.439	0.439	0.472	0.443	0.439	0.439	0.472	0.426	0.425	0.446	0.447
							(+0.001)	(+0.002)	(+0.001)		(−0.002)	(0.000)	(−0.002)		(−0.024)	(−0.037)	(−0.036)	(−0.016)
C_2'–N_3		0.490	0.459	0.457	0.457		0.491	0.461	0.458		0.488	0.459	0.455		0.466	0.422	0.421	0.441
								(+0.001)	(+0.001)			(−0.003)	(−0.001)			(−0.016)	(−0.031)	(−0.031)
C_3'–N_4			0.489	0.459	0.457			0.490	0.460			0.486	0.458			0.471	0.428	0.426
									(0.000)				(−0.003)				(−0.012)	(−0.027)
C_4'–N_5				0.490	0.460				0.490				0.487				0.478	0.433
																		(−0.012)
C_5'–N_6					0.490													0.478

[a] The identification of atoms is given in Figure 1. [b] In parentheses is given: Pop. (i) − Pop. (fully extended), where i refers to the conformation to which rotation occurred. [c] In parentheses is given: bond order (i) − bond order (fully extended), where i refers to the conformation to which rotation occurred. [d] These bond orders refer to the bond orders between the two symmetrically orthogonalized orbitals on the atoms involved in the π bond.

where $C^{(\varphi)}$ is the coefficient matrix that transforms the φ basis into the approximate canonical Hartree–Fock molecular orbitals, ψ ($\psi = \varphi\mathbf{C}^{(\varphi)}$). The $\mathbf{P}$ matrix in the symmetric orthonormalized basis, χ, is calculated as

$$P_{ij}^{(\chi)} = 2\sum_{k}^{\text{occ}} C_{ik}^{(\chi)}C_{jk}^{(\chi)}, \text{ where } \mathbf{C}^{(\chi)} = \mathbf{S}^{1/2}\mathbf{C}^{(\varphi)} \quad (19)$$

The χ basis has the important property that the χ_i's are the orthonormal molecular orbitals that are the closest (in a least-squares sense) to the φ_i's. Thus, χ_i is localized in predominantly the same region of space in which φ_i is found, and therefore χ_i may be thought of as the orthonormalized version of φ_i. The diagonal elements, $P_{ii}^{(\chi)}$, of $\mathbf{P}^{(\chi)}$ are then interpreted as orbital populations, and the off-diagonal elements, $P_{ij}^{(\chi)}$, of $\mathbf{P}^{(\chi)}$ are interpreted as bond orders. In the preceding section, use was made of the population of each of the χ_i's in particular molecular orbitals. The population of χ_i in the kth molecular orbital is simply $2[C_{ik}^{(\chi)}]^2$ (see ref 65 for further discussion of the relationships among ϕ, χ, and ψ).

The trends in populations and bond orders that are observed provide additional insight in several ways into the nature of the charge redistribution that occurs upon changes in conformation. For example, we note from Table V that, in general, the populations and bond orders in the σ system vary only slightly with conformation and chain length. For the one case where greater charge redistribution is observed (the oxygen LP-π population), it is seen that the trends remain the same as a function of the chain length. However, it is seen from Table VI that the populations and bond orders of the amide π systems show substantially greater changes on the whole. These π-system changes, as discussed below, appear to be the primary changes that occur on rotation from the fully extended to other conformations using the molecular fragment technique. They will be seen to quantify further the identification of hydrogen bonding effects mentioned earlier, by relating them to changes in orbital populations and bond orders in the π systems.

Considering the π-type orbital populations and bond orders in greater detail, we note that, for the anti-parallel-chain pleated sheet and parallel-chain pleated sheet conformers, the variations in population and bond order in going from the fully extended form to either of these are generally quite small (except for the N population in the parallel-chain pleated sheet conformer), and in all cases the trends remain the same as a function of chain length. Also, there appears to be no evidence of hydrogen bonding effects in these conformers. Thus, it appears that the rotations from the fully extended form to the anti-parallel-chain pleated sheet or parallel-chain pleated sheet forms do cause some charge redistribution to take place, but the nature and magnitude of these (generally small) changes can be estimated for longer chain lengths merely by extension of the cases treated here.

Of considerable interest are the changes observed in going from the fully extended to the α-helix conformer, which will be seen to show clearly the effect of hydrogen bond formation in the tetra- and pentapeptide cases. In particular, we note a marked change in the trends of the π-type orbital populations in O_1, C_1', and N_2 (in the tetra- and pentapeptide cases), and in O_2, C_2', and N_3 (in the pentapeptide case). Similar but smaller effects are noted for the corresponding C_1'–O_1, C_2'–O_2, C_1'–N_2, and C_2'–N_3 π-type bond orders. These changes indicate that O_1 and O_2 become more negative than the previous trends would have predicted, while C_1', C_2', N_2, and N_3 become more positive. In terms of bond orders, the C_1'–O_1 and C_2'–O_2 π bonds appear to be weakened, compared to the previous trends, while the C_1'–N_2 and C_2'–N_3 π bonds appear to be strengthened. These are just the effects that would be expected if hydrogen bonds are formed, and indicate that the primary effects of hydrogen bonding can be described (using a valence bond description) by a change in emphasis from resonance structure I to structure II (see Figure 5) in the amide unit when the hydrogen bond is formed.

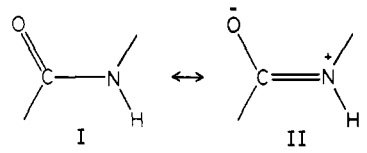

Figure 5. Contributing resonance structures in amides.

It is also of interest to compare the lack of reorganization of the σ structure found here in the *intramolecular* hydrogen bonds between pairs of adjacent π systems with the results that have been obtained for other kinds of hydrogen bonding environments. In particular, when there are not adjacent π systems (*e.g.*, the $H_2CO\cdots H_2O$ system[86] or the $ROH\cdots OHR'$ systems[87]), the major rearrangement that is found is *within* the σ system. On the other hand, for cases involving *intermolecular* hydrogen bonds in adjacent π systems within amides,[7,8] the effects have been found to involve primarily a redistribution of the π system, although rearrangement of the σ system is also observed. These differences in σ-orbital charge rearrangement can be rationalized by further examination of the FSGO basis set in the following manner.

If the FSGO are assumed to give strictly localized contributions to the molecular orbital description, then the molecular fragment description of the σ-electron density in the region where the intramolecular hydrogen bond is formed can be thought of as consisting of contributions from only a few orbitals. In particular, each of the oxygen lone pairs is described by a single orbital, and the pair of electrons in the N–H bond is also described by a single orbital. (The oxygen and nitrogen inner shell orbitals, that are approximately in the region of interest, are also described by single orbitals that are doubly occupied.) Since each of these orbitals is doubly occupied, and no other orbitals are present in this region of space, the SCF procedure will only transform these doubly occupied orbitals among themselves, leaving the total charge density invariant.[88] Thus, the particular choice of basis orbitals for the σ-electron description is not flexible enough to allow the σ-charge redistribution effect

(86) For the case of $H_2CO\cdots H_2O$ hydrogen bonding, see K. Morokuma, *J. Chem. Phys.*, **55**, 1236 (1971).

(87) For the case of hydrogen bonds of the type $ROH\cdots OHR'$ (R,R' = H, CH_3, NH_2, OH, and F), see J. E. Del Bene, *J. Chem. Phys.*, **57**, 1899 (1972).

(88) F. L. Pilar, "Elementary Quantum Chemistry," McGraw-Hill, New York, N. Y., 1968, p 345.

to be seen. However, this more subtle deficiency of the basis set does not affect the ability to identify the presence of hydrogen bonding, since the energetics and π-orbital reorganization have been shown to be sufficient in that regard in the previous discussion.

On the other hand, this kind of analysis indicates the substantial advantages that result from the use of localized orbitals in an *ab initio* framework, where the characteristics (both good and bad) of calculated results can be related directly and in a reasonably straightforward manner to the basis set. In addition, improvement of the model is frequently possible without major alterations in the basic approach. For example, σ-orbital charge redistribution can be allowed for in several ways, such as inserting a virtual orbital in the region of space of interest (*e.g.*, placing an additional FSGO on the hydrogen atom involved in the hydrogen bond), or allowing adjustment of the nonlinear parameters that are present in the original FSGO basis. Of course, the advantages of improving the description *via* techniques such as these must be weighed against the computational difficulties that are introduced and the effect on properties of interest. Such studies are currently under investigation and will be reported at a later date.

Summary and Conclusions

The current studies on polyglycine[89] have shown several results of interest, both in general, and specifically to polyglycine. In general, the applicability of the molecular fragment approach as a viable tool for the study of large molecular systems has been clearly demonstrated, since the calculations on the pentapeptide of glycine (144 electrons) represent the study of a system of considerable size and complexity.[90] In addition, the relative stability of the various conformations studied and the charge distributions in these conformations appear to be adequately described. Perhaps most significant, evidence for hydrogen bonding effects in glycine polypeptides of reasonably large size as important stabilizing forces has, for the first time, been extracted directly from the calculated energetic results and charge distribution analysis, without having to postulate their existence, or having concern that integral approximations may have biased the results.

The hydrogen bond strength in the α-helix conformation has been calculated to be 6.1 kcal/mol, but it has been shown that the α-helix conformation is not expected to be the most stable conformer for single-stranded polyglycine of any chain length. The fully extended conformation is predicted to be the most stable conformation (of those examined) for a single polyglycine chain, although the anti-parallel-chain pleated sheet and parallel-chain pleated sheet conformations may become more stable than the fully extended conformation for polystranded polyglycine, if interchain hydrogen bonds in the pleated sheet are stronger than 2.6 and 6.2 kcal/mol, respectively.

The HOMO-1, HOMO, and LUMO have been characterized for all chain lengths and conformations studied, and their tendency to be localized in the fully extended, anti-parallel-chain pleated sheet, and α-helix conformations has been noted.

Acknowledgments. The authors express their appreciation to the University of Kansas for partial support of the computing required for this study. In addition, the authors would like to thank the referee for particularly constructive comments.

(89) For a preliminary report concerning the evidence for hydrogen bonding effects in single-stranded polyglycine, see L. L. Shipman and R. E. Christoffersen, *Proc. Nat. Acad. Sci. U. S.*, 69, 3301 (1972).

(90) For other examples of the development and application of *ab initio* techniques for large molecules, see E. Clementi, J. Mehl, and W. von Neissen, *J. Chem. Phys.*, 54, 508 (1971), and references therein.