Peptide Models. 1. Topology of Selected Peptide Conformational Potential Energy Surfaces (Glycine and Alanine Derivatives)

András Perczel,[†] János G. Angyán,[†] Márton Kajtar,[†] Wladia Viviani,[‡] Jean-Louis Rivail,[‡] John-Frank Marcoccia,[§] and Imre G. Csizmadia*,[§]

Contribution from the Institute of Organic Chemistry, Eötvös University, Budapest, Hungary, Laboratoire de Chimie Théorique, Université de Nancy I, Vandoeuvre-les-Nancy, France, and Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A1.
Received July 16, 1990

Abstract: *N*-Formylglycinamide (For-Gly-NH_2), *N*-formyl-L-alaninamide (For-L-Ala-NH_2), and *N*-formyl-D-alaninamide (For-D-Ala-NH_2) were subjected to conformational studies by molecular mechanics (MM) and ab initio SCF methods. Both methods predicted the corresponding γ (usually labeled as C_{7eq}) conformations to be the global minimum on the Ramachandran map, $E(\phi,\psi)$. The number of minima and their approximate location obtained by MM corresponded to those that one might have expected on the basis of multidimensional conformational analysis. However, only the ab initio SCF study was capable of properly describing the effects associated with excessive repulsive or attractive interactions that lead to the annihilation and creation of certain critical points. Consequently, the topology of the MM and SCF conformational potential energy surfaces, $E(\phi,\psi)$, were found to be remarkably different for the three diamide systems investigated in this work.

Introduction

The shapes of peptides and proteins have been a central theme of scientific investigation for decades. Throughout the years it became customary to classify primary, secondary, and tertiary structures of proteins. To the best of our knowledge, the DNA molecule, which codes proteins, can predetermine only the sequence of the amino acids, which is the primary structure. The secondary and the tertiary structures are not coded *directly* by the DNA; therefore, they must be a consequence of the primary structure. Following the hard-sphere approximation, the secondary and tertiary structures can be defined in terms of a set of conformational angles $[\phi_i,\psi_i]$, associated with the amino acid units that make up the polypeptide chain.

For this reason, the conformation of diamides such as $CH_3CONHCHRCONHCH_3$ or $HCONHCHRCONH_2$ that can be derived from the naturally occurring amino acids are of great importance as they can mimic a segment of proteins. It is not surprising, therefore, that numerous papers have already been published, using a variety of theoretical methods, aimed at studying the $E(\phi,\psi)$ potential energy surface of the naturally occurring amino acid diamides.

However, no paper has been reported on the topological description of the potential energy surfaces of these diamides. The 19 common amino acid side chains (the proline has no flexible side chain) that necessarily influence the $E(\phi,\psi)$ surface can be divided into four groups, depending on the different degree of substitution at the β-carbon (C^β) atom. With the lack of a C^β atom (i.e., R = H), glycine forms the sole example of the achiral family. For the chiral family of diamides, several examples exist containing an sp^3 carbon as the first atom of the substituent. For example, the substituent in alanine is methyl (CH_3), in serine it is CH_2OH, and for valine it is $CH(CH_3)_2$.

family type	side chair[a]	example
1	H	glycine
2	CH_3	alanine
3	CH_2R	serine
4	CHR'R''	valine

[a] R = OH; R'' = R' = CH_3.

For each of the four types of amino acids, a grid search in terms of $[\phi,\psi]$ was performed, but the features of the potential energy surfaces, based on MM calculations, show a strong resemblance between the four groups of amino acids. The topology of MM and ab initio surfaces generated for the first two types (glycine and alanine) are compared in the present paper.

During the past 40 years, protein chemists have simplified their approach to the study of protein folding by separating, at least mentally, the problem of backbone conformation from the problems of nearest-neighbor and long-range interactions. This philosophy implied the conviction that first we have to understand the problem of backbone conformation in the absence of nearest-neighbor and long-range interactions (such as electrostatic attraction or repulsion, dipole–dipole stabilization, or destabilization and H bondings) before we can gain a full comprehension of the overall problem. Of course, even the problem of backbone conformation includes local side-chain/backbone as well as intrabackbone interactions. However, even such a basic notion as the concept of the Ramachandran's map has the above approach at its very core.

According to this approach, the backbone conformational problem of a protein might be viewed in terms of a corresponding conformational potential energy hypersurface in which nearest-neighbor and long-range interactions are eliminated. As a mathematical description of this traditional idea, let us consider the conformational potential energy hypersurface (PEHS) of a protein $E = E(x)$, where the variables are torsional angles, $x = x(\phi_1,\psi_1, ..., \phi_n,\psi_n)$, defined according to the IUPAC–IUB convention for peptides and proteins. If one retains in addition to the intrabackbone interaction one of the major effects, operative in determining the backbone conformational PEHS, namely the local side-chain/backbone interaction, and ignores, at least initially, the nearest-neighbor interaction, then the conformation of a single amino acid residue becomes very important. In view of that, the overall expression for the potential energy hypersurface can now be subdivided into n potential energy surfaces (PES) of the type $E(\phi_i,\psi_i)$

$$E(\phi_1,\psi_1, ..., \phi_i,\psi_i, ..., \phi_n,\psi_n) \Rightarrow \{E(\phi_1,\psi_1), ..., E(\phi_i,\psi_i), ..., E(\phi_n,\psi_n)\} \quad (1)$$

where n is the number of amino acid residues in the peptide chain. As a result of the partitioning of the $2n$-dimensional space to n two-dimensional subspaces

$$\{\phi_1,\psi_1, ..., \phi_i,\psi_i, ..., \phi_n,\psi_n\} \rightarrow \begin{cases} \{\phi_1,\psi_1\} \\ \vdots \\ \{\phi_i,\psi_i\} \\ \vdots \\ \{\phi_n,\psi_n\} \end{cases} \quad (2)$$

[†] Eötvös University.
[‡] Université de Nancy I.
[§] University of Toronto.

0002-7863/91/1513-6256$02.50/0

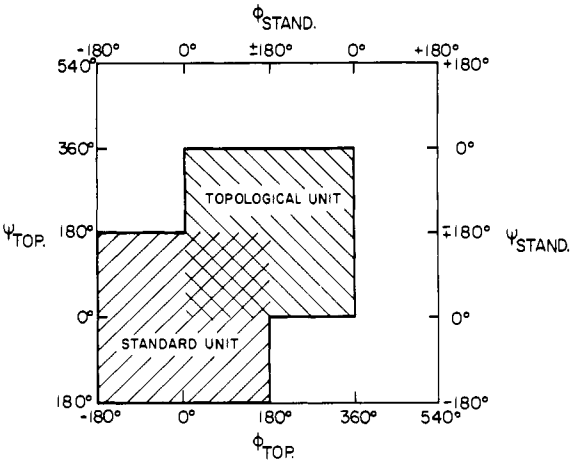


Figure 1. Domain of the original (standard unit) and the transformed (topological unit) Ramachandran map, $E(\phi,\psi)$, for an amino acid residue in peptides. If $0° \leq \phi_{STAND} \leq 180°$, then $\phi_{TOP} = \phi_{STAND}$. If $-180° \leq \phi_{STAND} \leq 0°$, then $\phi_{TOP} = \phi_{STAND} + 360°$. If $0° \leq \psi_{STAND} \leq 180°$, then $\psi_{TOP} = \psi_{STAND}$. If $-180° \leq \psi_{STAND} \leq 0°$, then $\psi_{TOP} = \psi_{STAND} + 360°$.

one can replace the study of the overall protein PEHS, $E = E(x)$, by n regular and more manageable PES, $E = E(\phi_i,\psi_i)$, where $1 < i < n$. The PES of $E(\phi_i,\psi_i)$ is expected to have several minima; some of these minima are the familiar α, β, and γ conformations.

In a first attempt at studying backbone conformations, the influence of the side-chain conformation might be considered to have secondary importance on the topology of the $[\phi,\psi]$ space. A previous analysis[1] on *N*-formyl-L-serinamide that explored only a portion of the conformational energy hypersurface, $E = E(\phi_1,\psi_1,\chi_1,\chi_2)$, has shown that the change in side-chain conformation as characterized by χ_1,χ_2 leaves the topology of $[\phi,\psi]$ surface intact and that only the relative position and stability of the critical points are changed to some extent.

Transformation of the $E(\phi,\psi)$ Space

The backbone of a polypeptide chain is defined by the two dihedral angles ϕ,ψ per residue corresponding to the torsion along the CONH–CHR and CHR–CONH bonds, respectively. In the conventional description of the $E(\phi,\psi)$ surface (often called the Ramachandran map), the values of the torsional angles run between −180 and +180°. Such a definition of the coordinates leads to a surface having four low-energy fields separated by mountain ridges from each other. In the upper left quadrant of the $E(\phi,\psi)$ surface, the traditional E or extended ($\phi = -180°$, $\psi = 180°$), the β-pleated-sheet ($\phi = -150°$, $\psi = 150°$), and the γ ($\phi = -78°$, $\psi = 78°$) conformations can be found, whereas in the lower left quadrant of the surface the α-helix conformation ($\phi = -60°$, $\psi = -40°$) can be located. Minima in the two right quadrants of the Ramachandran map of L-amino acids are usually considered less important due to their relatively high energy values. In contrast to the above, for D-amino acids the relative energies of the minima are reversed and the minima located in two right-side quadrants now become important. Consequently, the experimentally determined ϕ,ψ values of globular proteins are seldomely found in the two right-hand quadrants of the Ramachandran map.

The "standard unit" ($-180° < \phi < +180°$ and $-180° < \psi < +180°$) for the $[\phi,\psi]$ space conventionally accepted by the IUPAC–IUB Commission for a Ramachandran map is shown at the lower left-hand side of Figure 1. In keeping with the above convention, it is proposed that another cut, labeled as the "topological unit" ($0° < \phi < 360°$ and $0° < \psi < 360°$) be used. (Some might feel that the term "topological unit" is misleading since it can be interpreted to imply that changing the representation will lead to a different topology. It should be clearly understood that topology is representation-independent. However, the particular representation labeled as topological unit is favored by us because it allows easy recognition of certain topological features, such as equivalences of minima on a periodic conformational surface. Consequently, the label topological unit is used, as a nickname, for our favored representation.) A map with ϕ and ψ ranging from −180° to +360° has also been found useful by Vasilescu et al.[2] An illustrative pseudo-three-dimensional PES above the two representations of Figure 1 is shown schematically in Figure 2. One of the most important consequences of using the topological unit instead of the standard unit is that all minima, that is all nine catchment regions,[3] are located in a common area on the $E(\phi,\psi)$ surface (Figure 3B). The gathering of all minima in the same flat potential energy valley (that is reminiscent to the bottom of a potential box) demonstrates more clearly the possibilities for the interconversion between different minima (Figure

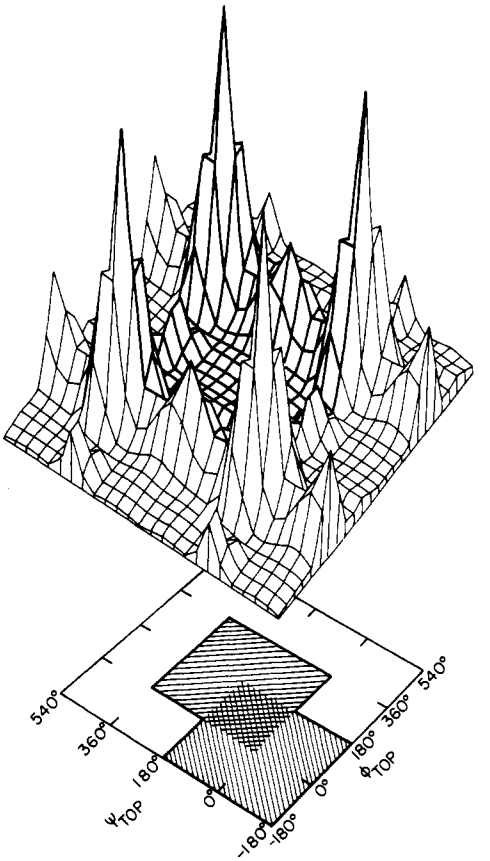


Figure 2. Pseudo-three-dimensional representation for Ac-L-Ala-$NHCH_3$ of four original (standard unit) Ramachandran maps, $E(\phi,\psi)$, and their projected domains. The topological unit is shown at the center.

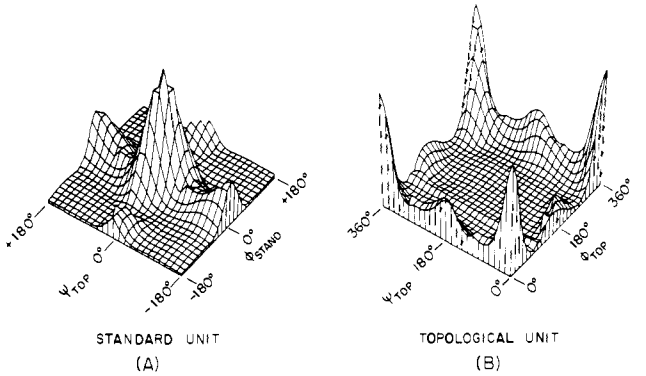


Figure 3. Pseudo-three-dimensional representations of a Ramachandran map for Ac-L-Ala-$NHCH_3$: (A) standard unit, (B) topological unit.

(1) Perczel, A.; Daudel, R.; Angyán, J. G.; Csizmadia, I. G. *Can. J. Chem.* **1989**, *68*, 1182.

(2) D. Vasilescu, D. Cabrol and A. M. Tamburro *J. Mol. Struc.* THEOCHEM **1988**, *179*, 185.

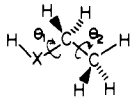
(3) Mezey, P. G. *Potential Energy Hypersurfaces*; Elsevier Science Publishers: New York, 1987; p 227.

3B). These minima are entrenched by the mountain ridges within the topological unit and are not separated by high mountain ridges as suggested by the standard unit (Figure 3A). Thus, one can see how all the stable conformations of peptides can be readily interconverted within a common region as illustrated by the topological unit.

Qualitative Multidimensional Conformational Analysis

D. H. R. Barton in his pioneering work[4a] on one-dimensional conformational analysis, carried out in the 1950s, clearly stated that staggered conformations represent minima and eclipsed conformations represent maxima on the potential energy curve. As a consequence of this, a methyl rotation has a potential energy curve of 3-fold periodicity, meaning that there are three energetically degenerate minima, due to the three equivalent staggered conformations.

However, in the same paper,[4a] Barton said the following: "In aliphatic compounds the most stable conformation is usually that in which the substituents on adjacent tetrahedral carbon atoms adopt the fully staggered conformation, the two largest groups (or, in qualification the two most strongly repelling dipoles) taking up the 180° arrangement." Subsequently, Barton applied this principle to analyze the conformation of cyclohexane with six sp^3–sp^3 C–C bonds, and with this he opened the door to multidimensional conformational analysis. In the 1960s empirical potentials were constructed to study conformations of alicyclic compounds.[4b,c] From the early 1970s, ab initio potential energy surfaces[4d,e] and ab initio potential energy hypersurfaces[4f] were generated, and sometimes analytic functions[4e,f] were fitted to the generated mesh of data points. From both the empirical potential functions[4b,c] and from the fitted analytic function,[4e,f] it is *implicitly* obvious that a potential surface may be decomposed to component potential energy curves or may be constructed from appropriate component potential energy curves. However, in 1974 the process of constructing an approximate potential energy surface, $E(\theta_1,\theta_2)$, from appropriate component potential energy curves, $E(\theta_1)$ and $E(\theta_2)$, was *explicitly* demonstrated[4g] for compounds of the following type, where X = O and S.



In the same paper[4g] it was also clearly demonstrated that the three minima associated with each of the torsional potential curves actually had lead to nine minima in a 3 × 3 array, on the potential energy surface. Thus, the topology of the idealized surfaces had been predicted from the topology of their component curves. This principle has been subsequently[4h] further elaborated. The problem with the general notion of qualitative multidimensional conformational analysis that has been used in a variety of apparently unrelated forms throughout the decades is that it does not always work and organic chemists have no prior knowledge when it might work and when it might not. Fortunately, due to the pioneering mathematical work of P. G. Mezey, in applying topology to potential energy surfaces, one is now more knowledgeable than ever.

(4) (a) Barton., D. H. R. *Q. Rev. Chem. Soc.* **1956**, *10*, 44. (b) Hendrickson, J. B. *J. Am. Chem. Soc.* **1967**, *89*, 7047. (c) Pickett, H. M.; Strauss, H. L. *J. Am. Chem. Soc.* **1970**, *92*, 7281. (d) Wolfe, S.; Tel, L. M.; Csizmadia, I. G. *Can. J. Chem.* **1973**, *51*, 2423. (e) Wolfe, S.; Schlegel, H. B.; Csizmadia, I. G.; Bernardi, F. *Can. J. Chem.* **1973**, *51*, 2423. (f) Peterson, M. R.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1978**, *100*, 6911. (g) Csizmadia, I. G. General and Theoretical aspects of the thiol group. In The chemistry of the thiol group. *The chemistry of functional groups*; Patai, S., Ed.; Wiley: New York, 1974; pp 1–109. (Cf. particularly pp 36–41 including Figures 23 and 24 as well as Table 20.) (h) Csizmadia, I. G. Multidimensional Theoretical Stereochemistry and Conformational Potential Energy Surface Topology. In *New Theoretical Concepts for Understanding Organic Reactions*; Bertrán, J., Ed.; D. Reidel Publishing: Dordrecht, The Netherlands, 1989; pp 1–31.

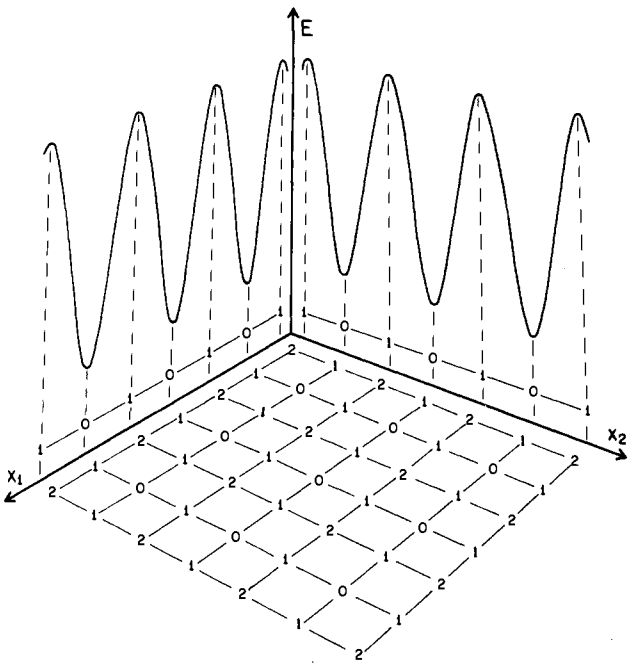


Figure 4. Schematic representation of how the ideal potential energy surface topology may be predicted from potential curves according to the sum of the λ rule: $\lambda(x_1,x_2) = \lambda(x_1) + \lambda(x_2)$ or $\lambda(\phi,\psi) = \lambda(\phi) + \lambda(\psi)$.

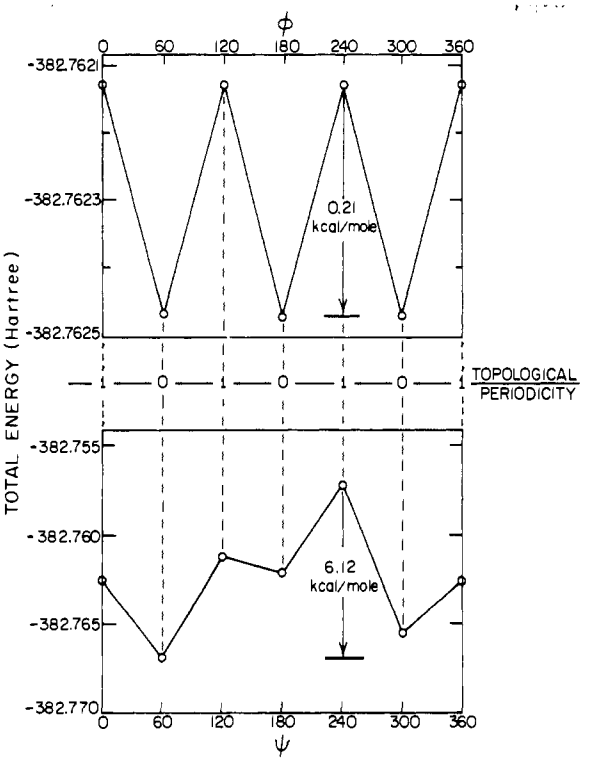
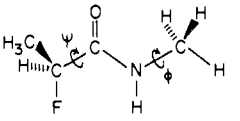


Figure 5. Model torsional potentials of $E(\phi)$ and $E(\psi)$ types in CH_3C-HFCONHCH$_3$, indicating that the rotation of a disubstituted chiral carbon, such as RHXC, has 3-fold topological periodicities (denoted 1-0-1-0-1-0-1) but nondegenerate minima while in the CH_3 rotation the 3-fold topological periodicity is coupled with energetic degeneracy.

Thus, the 1980s became different from earlier periods for conformational analysis, as in fact it became the decade of topology. One might say in connection with Mezey's 1981 paper[15] that, besides rigorous mathematical derivations, "precise sufficient conditions are given for the validity of these predictions in terms of curvatures, that are generalized force constants of interactions (eqs. 6a and 6b)[15] and also in terms of first derivatives describing the approximate alignment of mountain ridges and valley floors with the coordinate directions (eq. 31)[15], that are the constraint

on the extent of interactions between motions along different internal coordinates." It is unfortunate that the chemical importance of this mathematical paper[15] has not been emphasized earlier even though it has been referred to. In spite of the limited reliability of qualitative multidimensional conformational analysis, it has one feature that makes it still useful; it is its simplicity. Figure 4 illustrates the situation for a pair of rotors with 3-fold periodicities. Of course the pattern shown in Figure 4 is true even if the component potential curves do not have 3-fold periodicities (i.e., triply degenerate minima); it is enough if they have 3-fold topological periodicities (i.e., three nonequivalent minima). Potential curves of both types are shown in Figure 5 for the following peptide model:



$(\Psi = \phi = 0°)$

As the coupling of the two rotors in Figure 4 is ideal, the number of minima (N_0), the number of saddle points (N_1), and the number of maxima (N_2) as well as their locations are predictable from the location of the critical points associated with the component curves. This leads to the following well-known[5] alternating sum rule:

$$N_0 - N_1 + N_2 = 0$$

$$9 - 18 + 9 = 0$$

In the present paper, the nine minima ($N_0 = 9$) are labeled by the greek letters α, β, γ, δ, and ϵ subscripted by D or L. One domain, containing four minima, involves the most favored critical points for the L enantiomer (α_L, γ_L, δ_L, ϵ_L), while the other domain, containing another set of four minima, is associated with the D enantiomer (α_D, γ_D, δ_D, ϵ_D) leaving the remaining ninth critical point (β_{DL}) in a variable position as it may become β_L or β_D. The topology of the nine critical points (i.e., nine minima) is shown schematically:

γ_D	δ_D	α_L
ϵ_D	β_{DL}	ϵ_L
α_D	δ_L	γ_L

It should perhaps be reemphasized that, in the present paper, the use of the subscript L (such as in α_L) used is to denote the most favored minimum energy conformations of the L enantiomer of the amino acid and D is used for the most favorable minima of the D enantiomer. Thus, in the present notation, from the α_L conformation one can build right-handed α-helical conformation and from the α_D conformation left-handed α-helical conformation may be constructed. In general, protein chemists denote the right-handed α-helical conformation by α_R and the left-handed helical conformation by α_L. In order to avoid confusion, it is advisable to denote these enantiomeric helices as α_{RIGHT} and α_{LEFT}:

$$(\alpha_L)_n = \ldots{-}\alpha_L{-}\alpha_L{-}\alpha_L{-}\alpha_L{-}\alpha_L{-}\ldots \rightarrow \alpha_{RIGHT}$$

$$(\alpha_D)_n = \ldots{-}\alpha_D{-}\alpha_D{-}\alpha_D{-}\alpha_D{-}\alpha_D{-}\ldots \rightarrow \alpha_{LEFT}$$

Computational Methods

Peptide and protein conformations are usually studied by empirical methods due to the large dimensionality of the problem. In this work, ECEPP/2 molecular mechanics calculations[6] were performed, according to a 37 × 37, grid to obtain $E(\phi,\psi)$ surfaces each consisting of 37 × 37 = 1369 points. As expected on the basis of qualitative multidimensional conformational analysis, nine minima, associated with the typical backbone conformations (labeled α_L, α_D, β_{DL}, γ_L, γ_D, δ_L, δ_D, ϵ_L, ϵ_D), were found. These nine minima were also subjected to an ab initio MO study. Geometry optimizations by the gradient OC method using the MONSTERGAUSS program[7] at the 3-21G basis set[8] level were performed on the nine minima obtained by MM for each of *N*-formylglycinamide, *N*-formyl-L-alaninamide, and *N*-formyl-D-alaninamide. Ab initio energies, reported in this paper, are therefore for the fully optimized geometries. All optimizations were made until the largest internal forces were reduced to less than 5×10^{-4} mdyn/A.

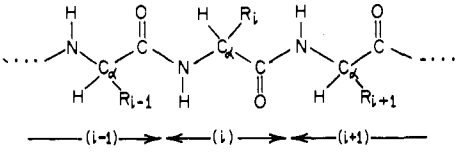
Table I. Molecular Mechanics (ECEPP/2) Results[a] for Ac-Gly-NHCH$_3$, Ac-L-Ala-NHCH$_3$, and Ac-D-Ala-NHCH$_3$ Conformations

BB[b]	ϕ_{STAND}	ψ_{STAND}	χ_1	ϕ_{TOP}	ψ_{TOP}	E_{abs}	ΔE_{rel}
			Ac-Gly-NHCH$_3$				
γ_L	−79.3	73.6		280.7	73.6	−6.33	0.00
γ_D	79.3	−73.6		79.3	286.4	−6.33	0.00
α_L	−72.7	−33.8		287.3	326.2	−5.10	1.23
α_D	72.7	33.8		72.7	33.8	−5.10	1.23
β_L	180.0	180.0		180.0	180.0	−5.08	1.25
δ_L	−168.3	52.4		191.7	52.4	−4.88	1.45
δ_D	168.3	−52.4		168.3	307.6	−4.88	1.45
ϵ_L	−90.0	150.0		270.0	150.0	−4.50	1.83
ϵ_D	90.3	−150.0		90.0	210.0	−4.50	1.83
			Ac-L-Ala-NHCH$_3$				
γ_L	−80.4	75.8	60.7	279.6	75.8	−5.18	0.00
β_L	−154.7	157.2	59.2	205.3	157.2	−4.47	0.71
α_L	−73.7	−34.9	61.6	286.3	325.1	−4.37	0.81
δ_L	−150.7	45.6	61.0	209.3	45.6	−4.08	1.10
ϵ_L	−75.5	139.0	61.9	284.5	139.0	−4.06	1.12
δ_D	−158.3	−57.5	54.4	201.7	302.5	−3.46	1.72
α_D	54.7	46.0	66.9	54.7	46.0	−2.82	2.36
ϵ_D	63.7	−174.8	80.9	63.7	185.2	−1.12	4.06
γ_D	77.7	−64.3	87.5	77.7	295.7	2.08	7.26
			Ac-D-Ala-NHCH$_3$				
γ_D	80.4	−75.8	59.3	80.4	284.2	−5.18	0.00
β_D	154.7	−157.2	60.8	154.7	202.8	−4.47	0.71
α_D	73.7	34.9	58.4	73.7	34.9	−4.37	0.81
δ_D	150.6	−45.6	59.0	150.6	314.4	−4.08	1.10
ϵ_D	75.5	−139.1	58.1	75.5	220.9	−4.06	1.12
δ_L	158.3	57.5	65.6	158.3	57.5	−3.46	1.72
α_L	−54.7	−46.0	53.1	305.3	314.0	−2.82	2.36
ϵ_L	−63.7	174.8	39.1	296.3	174.8	−1.12	4.06
γ_L	−77.7	64.3	32.5	282.3	64.3	2.08	7.26

[a] Torsion angles (ϕ, ψ, χ_1) in degrees and energy differences (ΔE_{rel}, E_{abs}) in kilocalories per mole. [b] BB = backbone conformation.

Model Compounds for Polypeptides

Most frequently the *N*-acetyl amino acid *N'*-methylamides are used as diamide systems to model peptides. The methyl group of the acetyl moiety simulates the C_α atom of $(i-1)$th amino acid residue in a peptide chain. On the other hand, the methyl bonded to the C-terminal amide nitrogen stands for the C_α of the $(i+1)$th residue.



Nevertheless, in an ab initio calculation it is practical to use the smaller formyl and primary amide groups (i.e., replacing CH_3 by H) for economic reasons.

When the positions and the relative energy order of the minima of For-Gly-NH$_2$, For-L-Ala-NH$_2$, and For-D-Ala-NH$_2$ are compared with those of Ac-Gly-NHCH$_3$, Ac-L-Ala-NHCH$_3$, and Ac-D-Ala-NHCH$_3$, respectively (see Tables I and II), it becomes evident, after the contents of Tables I and II are compared and

(5) (a) Peterson, M. R. Determination of Critical Point Geometries of Conformational Energy Hypersurfaces. Ph.D. Thesis, University of Toronto, 1980. (b) Peterson, M. R.; Csizmadia I. G.; Sharpe, R. W. *J. Mol. Stuct.*, THEOCHEM **1983**, *94*, 363.

(6) Vásquez, M.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043.

(7) Peterson, M. R., Poirier, R. A., Department of Chemistry, University of Toronto, Toronto, Canada M5S 1A1.

(8) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.*, **1980**, *102*, 939.

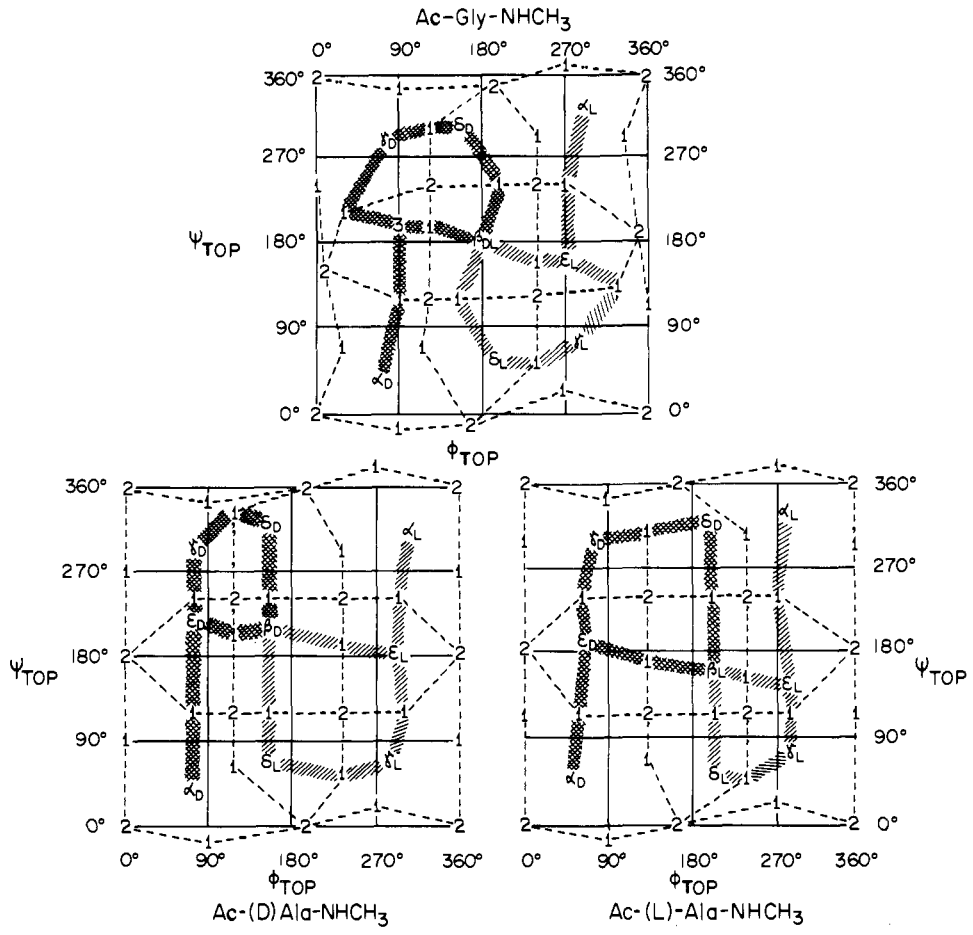


Figure 6. Ramachandran maps, $E(\phi,\psi)$, for Ac-Gly-NH_2, Ac-D-Ala-NH_2, and Ac-L-Ala-NH_2 obtained by molecular mechanics. Low-energy pathways are specified by lightly and heavily shaded lines for characterizing D and L minima, respectively.

Table II. Molecular Mechanics (ECEPP/2) Results[a] for For-Gly-NH_2, For-L-Ala-NH_2, and For-D-Ala-NH_2 Conformations

BB[b]	ϕ_{STAND}	ψ_{STAND}	χ_1	ϕ_{TOP}	ψ_{TOP}	E_{abs}	Δ_{Erel}
			For-Gly-NH_2				
γ_L	-79.5	74.4		280.5	74.4	-9.09	0.00
γ_D	79.5	-74.4		79.5	285.6	-9.09	0.00
α_L	-72.5	-35.1		287.5	324.9	-7.57	1.52
α_D	72.5	35.1		72.5	35.1	-7.57	1.52
β_L	180.0	180.0		180.0	180.0	-7.61	1.48
δ_L	-165.9	53.3		194.1	53.3	-7.43	1.66
δ_D	165.9	-53.3		165.9	306.7	-7.43	1.66
ϵ_L	-77.1	153.9		282.9	153.9	-6.89	2.20
ϵ_D	77.1	-153.9		77.1	206.1	-6.89	2.20
			For-L-Ala-NH_2				
γ_L	-80.5	76.3	60.7	279.5	76.3	-7.47	0.00
β_L	-154.8	157.7	59.4	205.2	157.7	-6.47	1.00
α_L	-73.3	-35.6	61.6	286.7	324.4	-6.33	1.14
δ_L	-150.5	47.8	60.9	209.5	312.2	-6.18	1.29
ϵ_L	-77.1	146.4	62.1	282.9	146.4	-5.96	1.51
δ_D	-159.5	-57.9	53.2	200.5	302.1	-5.18	2.29
α_D	54.8	46.3	66.9	54.8	46.3	-4.74	2.73
ϵ_D	64.4	-178.0	81.7	64.4	182.0	-1.83	5.64
γ_D	79.9	-63.3	88.8	79.9	296.7	0.81	8.28
			For-D-Ala-NH_2				
γ_D	80.5	-76.3	59.3	80.5	283.7	-7.47	0.00
β_D	154.8	-157.7	60.6	154.8	202.3	-6.47	1.00
α_D	73.3	35.6	58.4	73.3	35.6	-6.33	1.14
δ_D	150.5	-47.8	59.1	150.5	312.2	-6.18	1.29
ϵ_D	77.1	-146.4	57.9	77.1	213.6	-5.96	1.51
δ_L	159.5	57.9	66.8	159.5	57.9	-5.18	2.29
α_L	-54.8	-46.3	53.1	305.2	313.7	-4.74	2.73
ϵ_L	-64.4	178.0	38.3	295.6	178.0	-1.83	5.64
γ_L	-79.9	63.3	31.2	280.1	63.3	0.81	8.28

[a] Torsion angles (ϕ, ψ, χ_1) in degrees and energy differences (ΔE_{rel}, E_{abs}) in kilocalories per mole. [b] BB = backbone conformations.

Table III. Ab Initio SCF (3-21G) Results[a] for For-Gly-NH_2 Conformations

BB[b]	ϕ_{STAND}	ψ_{STAND}	ϕ_{TOP}	ψ_{TOP}	E	ΔE_{rel}
α_L	-83.2	-14.2	276.8	345.8	-373.641 601	4.46
α_D	83.2	14.2	83.2	14.2	-373.641 601	4.46
β_L	-180.0	-180.0	180.0	180.0	-373.647 718	0.62
γ_L	-83.9	67.8	276.1	67.8	-373.648 707	0.00
γ_D	83.9	-67.8	83.9	292.2	-373.648 707	0.00
δ_L	-126.0	25.5	234.0	25.5	-373.643 495	3.27
δ_D	126.0	-25.5	126.0	334.5	-373.643 495	3.27
ϵ_L	not found					
ϵ_D	not found					

[a] Torsion angles (ϕ, ψ) in degrees, energy (E) in hartrees, and the energy differences (ΔE_{rel}) in kilocalories per mole. [b] BB = backbone conformation.

contrasted, that the neglect of the two methyl groups does not alter the situation extensively.

Tables II–IV demonstrate the similarities as well as the differences that may exist between the MM and ab initio results obtained for For-Gly-NH_2, For-L-Ala-NH_2, and For-D-Ala-NH_2.

Results and Discussion

Topological Description of the MM Conformational Potential Energy Surface of Ac-Gly-$NHCH_3$, Ac-L-Ala-$NHCH_3$, and Ac-D-Ala-$NHCH_3$. Since the pioneering work of Schäfer et al.[9a,b] on the ab initio geometry of glycine, the diamide model system has been the subject of continuous interest. Both theoretical[10a,b,11,12]

(9) (a) Sellers, H. L.; Schäfer, L. *J. Am. Chem. Soc.*, **1978**, *100*, 7728. (b) Schafer, L.; Sellers, H. L.; Lovas, F. J.; Suenram, R. D. *J. Am. Chem. Soc.*, **1980**, *102*, 6566.

(10) (a) Schäfer, L.; Van Alsenoy, C.; Scarsdale, J. N. *J. Chem. Phys.* **1982**, *76*, 1439. (b) Klimkowski, V. J.; Schafer, L.; Momanay, F. A.; Van Alsenoy, C. *J. Mol. Struct.* **1985**, *124*, 143.

Table IV. Ab Initio SCF (3-21G) Results[a] for For-L-Ala-NH_2 and For-D-Ala-NH_2 Conformations

BB[b]	ϕ_{STAND}	ψ_{STAND}	χ_1	ϕ_{TOP}	ψ_{TOP}	E	ΔE_{rel}
			For-L-Ala-NH_2				
α_D	63.8	32.7	60.6	63.8	32.7	-412.465293	5.93
α_L			not found				
β_L	-168.4	170.9	60.0	191.6	170.9	-412.472746	1.25
γ_D	73.9	-56.7	60.5	73.9	303.3	-412.470720	2.52
γ_L	-84.4	67.7	64.7	275.6	67.7	-412.474738	0.00
δ_D	-178.6	-44.0	58.4	181.4	316.0	-412.463100	7.30
δ_L	-127.8	30.0	58.7	232.2	30.0	-412.468678	3.80
ϵ_D	67.6	-178.1	64.9	67.6	181.9	-412.461724	8.16
ϵ_L			not found				
			For-D-Ala-NH_2				
α_D			not found				
α_L	-63.9	-33.0	59.8	296.1	327.0	-412.465254	5.93
β_D	168.7	-170.4	60.5	168.7	189.6	-412.472746	1.25
γ_D	87.3	-66.7	57.5	87.3	293.3	-412.474738	0.00
γ_L	-74.0	57.9	59.8	286.0	57.9	-412.470721	2.52
δ_D	129.2	-30.0	61.7	129.2	330.0	-412.468641	3.80
δ_L	177.2	44.8	61.5	177.2	44.8	-412.463098	7.30
ϵ_D			not found				
ϵ_L	-67.1	176.6	53.2	292.9	176.6	-412.461728	8.16

[a] Torsion angles (ϕ, ψ, χ_1) in degrees, energy (E) in hartrees, and energy difference (ΔE_{rel}) in kilocalories per mole. [b] BB = backbone conformation.

and experimental[13,14] studies have been published on the conformation of achiral and chiral diamide systems derived from amino acids. However, no study to date has been reported on the topological description of the potential energy surfaces of these diamides. The whereabouts of the critical points as well as their index λ (λ = 0, 1, and 2 for minima, saddle points, and maxima, respectively) for Ac-Gly-$NHCH_3$, Ac-L-Ala-$NHCH_3$, and Ac-D-Ala-$NHCH_3$ are shown in Figure 6. The topologies of these three surfaces (Figure 6) are somewhat different from that of the ideal surface (Figure 4) as illustrated by the two types of the alternating sums:

$$N_0 - N_1 + N_2 = 0$$

$$9 - 18 + 9 = 0 \text{ (ideal surface, Figure 4)}$$

$$9 - 16 + 7 = 0 \text{ (MM surfaces, Figure 6)}$$

Mezey has shown[15,16] that the actual number of critical points (N_λ) of index λ can be related to the Betti numbers (β_λ) of index λ:

$$N_\lambda = (\prod_{i=1}^{n} m_i)\beta_\lambda = (\prod_{i=1}^{n} m_i)\binom{n}{\lambda} = m^n\binom{n}{\lambda} \qquad (3)$$

In the present case, eq 3 is valid for the idealized surface with $m^n = 3^2 = 9$. For the MM surface it is valid for N_0. However, it breaks down for λ = 1 and λ = 2 because of the annihilation of certain transition states and maxima. The number of minima and their locations are more or less where one expects them to be from a classical multidimensional conformational analysis.[4g,h] The nine minima can be related to the stable structures recognized by Scheraga et al.[17] By redrawing Scheraga's diagram from the ($\phi_{STANDARD}$,$\psi_{STANDARD}$) to the (ϕ_{TOP},ψ_{TOP}) representation, as shown in Figure 7, one can readily see that some of the 12 minima

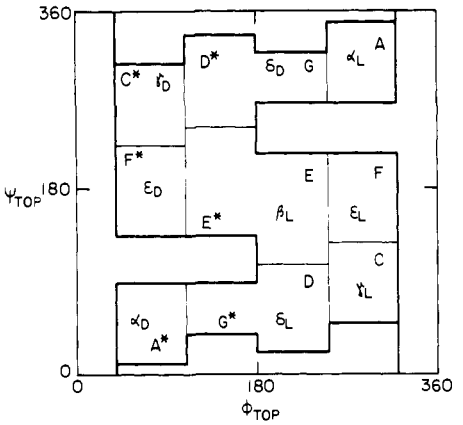


Figure 7. Correlation of Scheraga's 12 categories (A,A*, C,C*, D,D*, E,E*, F,F*, G,G*) with the currently recognized nine minima (α_L,α_D, β_L, γ_L,γ_D, δ_L,δ_D, ϵ_L,ϵ_D) on the (ϕ_{TOP},ψ_{TOP}) coordinate system. [Note that D* = δ_D = G, E* = β_L = E, and G* = δ_L = D.]

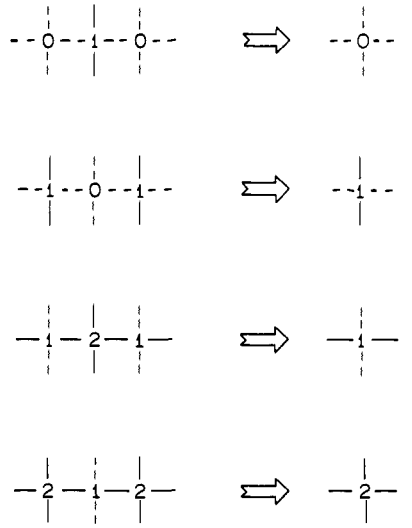


Figure 8. Selection rules for the collapse of three critical points into a single critical point.

Scheraga identified (A, C, D, E, F, G, A*, C*, D*, E*, F*, G*) are in fact identical conformations (cf. E = E*, D = G*, G = D*). The equivalence between the two sets of notation are shown below:

A	α_L	
C	γ_L	
D	δ_L	G*
	ϵ_D	F*
E	β_{DL}	E*
F	ϵ_L	
G	δ_D	D*
	γ_D	C*
	α_D	A*

Some might be convinced that Scheraga intended nothing more than to create a merely practical subdivision of the (ϕ,ψ) map into 12 regions. The present authors, however tend to believe that, in fact Scheraga's work shows incredible insight to the problem, since after taking the appropriate (ϕ_{STAND},ψ_{STAND}) → (ϕ_{TOP},ψ_{TOP}) transformation one obtains a 1 to 1 correspondence between the nine catchment regions of the ideal conformational potential energy surface and those 12 regions defined by Scheraga.

Topological Description of the ab Initio Conformational Potential Energy Surface of For-Gly-NH_2, For-L-Ala-NH_2, and For-D-Ala-NH_2. Gradient optimization for the minima on the For-Gly-NH_2 PES at the 3-21G SCF level reveals some interesting features. Table III summarized the (ϕ,ψ) values for the optimized conformers of For-Gly-NH_2. The first noticeable feature for this PES is that there are three doubly degenerate pairs of minima:

(11) Scarsdale, J. N.; Van Alsenoy, C.; Klimkowski, V. J.; Schafer, L.; Momany, F. A. *J. Am. Chem. Soc.* **1983**, *105*, 3438.
(12) Schäfer, L.; Klimkowski, V. J.; Momany, F. A.; Chuman, H.; Van Alsenoy, C. *Biopolymers* **1984**, *23*, 2335.
(13) Koyma, Y.; Shimonouchi, T. *Biopolymers* **1968**, *6*, 1037.
(14) Koyma, Y.; Shimonouchi, T.; Sato, M.; Tatsuno, T. *Biopolymers* **1971**, *10*, 1059.
(15) Mezey, P. G. *Chem. Phys. Lett.* **1981**, *82*, 100.
(16) Mezey, P. G. *Chem. Phys. Lett.* **1982**, *86*, 562.
(17) Zimmerman, S. S.; Pottle, M. S.; Nemethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.

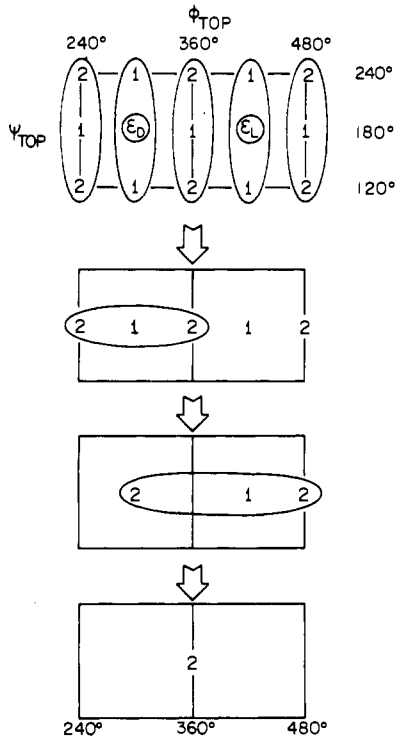


Figure 9. Schematic representation of the process associated with the annihilation of 15 critical points, including two minima (ϵ_D, ϵ_L) and the simultaneous creation of a single maximum for For-Gly-NH_2. [The process applies the selection rules presented in Figure 8.]

(α_L,α_D), (γ_L,γ_D), (δ_L,δ_D) plus a single minimum β_{LD} as illustrated below and listed in Table III.

γ_D	δ_D	α_L
.	β_{DL}	.
α_D	δ_L	γ_L

Clearly, two of the minima, previously labeled ϵ_L and ϵ_D, have disappeared from the PES.

Selection rules for the annihilation of three critical points and the simultaneous creation of a new critical point has been published earlier.[18] A graphical illustration of the selection rules operative in the modification of a PES, via the creation and annihilation of critical points, is shown in Figure 8. It perhaps should be noted that critical points are located at the various cross-sections of mountain ridges (solid lines in Figure 8) and valley floors (broken lines in Figure 8). Thus, the disappearance of a critical point implies the disappearance of either a mountain ridge or a valley floor. However, conformational changes are in fact periodic; therefore, there must be the same number of mountain ridges than valley floors in a full rotation of 0° → 360°. Consequently, if a mountain ridge disappears, this must be coupled with the synchronized disappearance of a valley floor or vice versa. Thus, when three critical points are annihilated and one critical point is created, it means effectively a reduction of the number of critical points by two as illustrated in Figure 8. None of these patterns in Figure 8 alone account for the annihilation of two minima and the simultaneous creation of a single maximum. However, it appears that with the combination of some of these patterns 15 critical points could collapse to form a single maximum, in accordance with the selection rules, as illustrated in Figure 9. The change of critical points can be enumerated in the following fashion:

order of crit pts (λ)	0	1	2
no. of crit pts annihilated	−2	−7	−6
no. of crit pts created			+1
net change of crit pts (ΔN_λ)	−2	−7	−5

(18) Ángyán, J. G.; Daudel, R.; Kucsman, Á.; Csizmadia, I. G. *Chem. Phys. Lett.* **1987**, *136*, 1.

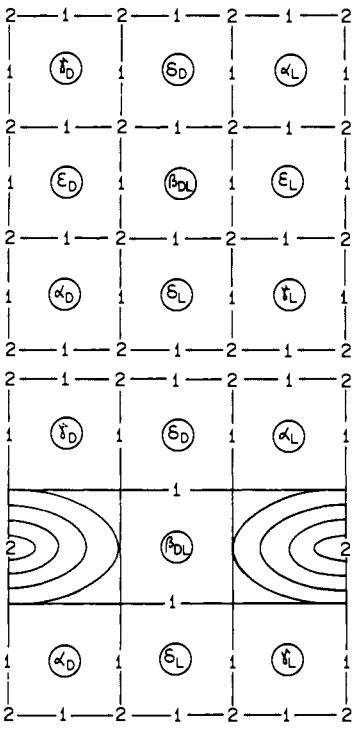


Figure 10. Comparison of a schematic representation of the ideal conformational potential energy surface topology (top) with the actual conformational potential energy surface topology (bottom) for For-Gly-NH_2.

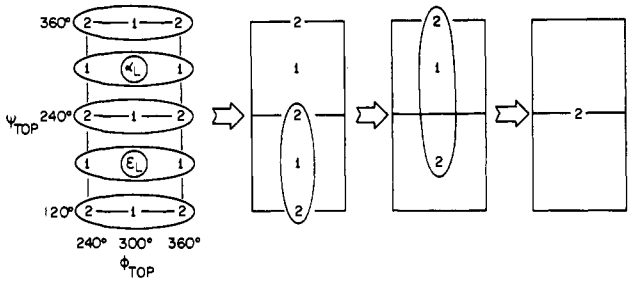


Figure 11. Schematic representation of the process associated with the annihilation of 15 critical points, including two minima (α_L, ϵ_L) and the simultaneous creation of a single maximum for For-L-Ala-NH_2. [The process applies the selection rules presented in Figure 8.]

With these net changes of critical point (ΔN_λ) values, the alternating sum rule will change in the following fashion:

$$N_0 - N_1 + N_2 = 0$$

$$9 - 18 + 9 = 0 \text{ (ideal surface, Figures 4 and 10)}$$

$$7 - 11 + 4 = 0 \text{ (SCF surface, Tables III and Figure 10)}$$

The final topology of PES is shown in the lower part of Figure 10.

For SCF gradient geometry optimization of the critical points for For-L-Ala-NH_2 and For-D-Ala-NH_2, one expects nine minima to be present as in the ideal or MM surface. However, from the ab initio method, only seven minima were found as summarized in Table IV. Again, two critical points were missing as in the case of For-Gly-NH_2. One important distinction was that a different pair of conformations was annihilated as illustrated below, which shows the pattern of the minima found in the L and D enantiomers of For-Ala-NH_2, respectively.

γ_D	δ_D	.	γ_D	δ_D	α_L
ϵ_D	β_{DL}	.	.	β_{DL}	ϵ_L
α_D	δ_L	γ_L	.	δ_L	γ_L
For-L-Ala-NH_2			For-D-Ala-NH_2		

The annihilation of 15 critical points and creation of a single maximum for For-Ala-NH_2 is illustrated in Figure 11 while the final topology is shown in Figure 12. For the D enantiomer the

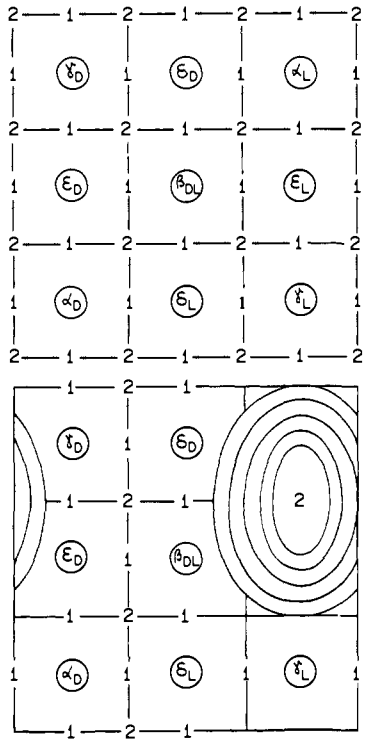


Figure 12. Comparison of a schematic representation of the ideal conformational potential energy surface topology (top) with the actual conformational potential energy surface topology (bottom) for For-L-Ala-NH_2.

situation is analogous in that it is a kind of "mirror image" of the L enantiomer.

Scheraga's work[17] clearly indicated that there are amino acid residue conformations in proteins that fall into all of the nine possible domains (catchment regions) shown by Figure 7. Thus, Scheraga's findings are in perfect agreement with the topology of conformational PES presented, which in turn is the result of multidimensional conformational analysis[4g,h] (cf. Figure 4). Consequently, the disappearance of certain minima of the ab initio PES, such as ϵ_L and ϵ_D, associated with For-Gly-NH_2 and α_L as well as ϵ_L associated with For-L-Ala-NH_2 may appear as a serious discrepancy between experiment and theory. It is therefore advisable to reiterate that the present calculations are strictly valid for For-Gly-NH_2 and For-L-Ala-NH_2 and the conclusions may not be generalized to proteins without reservations. The reason why protein chemists might find this apparent discrepancy disturbing is because they have truly believed throughout the years that eqs 1 and 2 are rigorously correct for the whole problem of protein secondary structure, rather than to a portion of the whole problem, namely backbone conformation only. No earlier report is found in the literature that specifies stable, ab initio, α_L conformation for any amino acid derivative. Our recent results[1] obtained on For-L-Ser-NH_2 reported the existence of an α_L conformation stabilized by some sort of H bonding between the backbone and the side chain. Thus, it appears that α_L conformations need to be stabilized by specific nonbonded interactions.

In fact, the instability of the α_L conformation is not a new question. Arridge and Cannon calculated[19] the interaction energy between the dipoles of peptide bonds in a helical structure. Their results showed that for the α-helix the dominant energy term was for the interaction between adjacent peptide dipoles, which is repulsive. Subsequently, Brant et al.[20] showed that a β-strand was more favorable than an α-helix. Several mechanisms have been suggested in order to explain how an inherently unstable conformation, such as an α_L, may be stabilized. Intrahelical salt

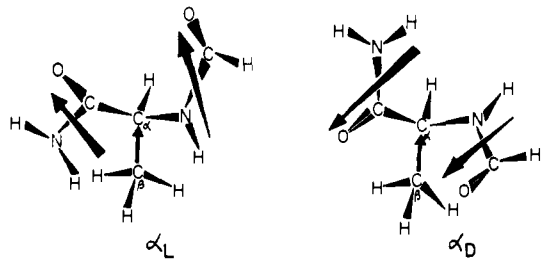


Figure 13. Orientation of group dipoles in the α_L and α_D conformations of the L-alanine derivative $HCONHCH(CH_3)CONH_2$. The figure is drawn with ECEPP/2 geometries.

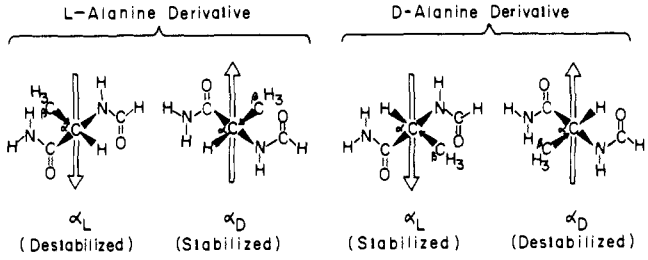


Figure 14. Schematic vectorial illustration of dipole–dipole interactions in the α_L and α_D conformers, of L-alanine and D-alanine derivative $HCONHCH(CH_3)CONH_2$, respectively. Empty arrows represent the vectorial sum of amide dipoles; the black arrows represent the C^{α}–$C^{\beta}H_3$ dipole.

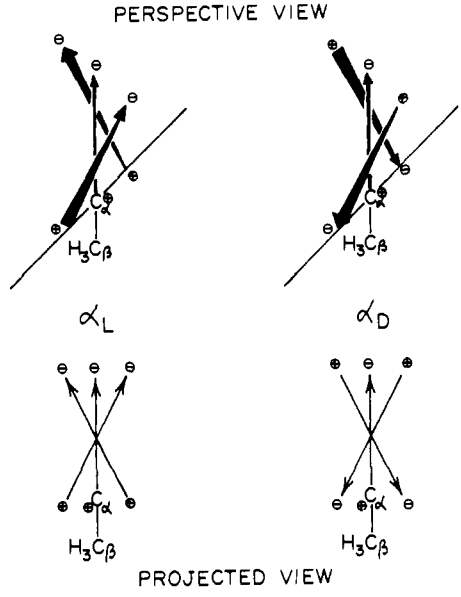


Figure 15. Schematic vectorial illustration of dipole–dipole interactions in the α_L and α_D conformers of the L-alanine derivative $HCONHCH(CH_3)CONH_2$. The unit vectors point from the positive to the negative charge, according to chemical convention.

bridges,[21,22] side-chain/backbone interactions,[23,24] and charge-charge interactions[24] have been advocated.

One may attempt to rationalize the absence of the α_L conformation on the SCF potential energy surface in the case of L-alaninediamide derivative on the basis of some of the above arguments, which have been previously published. It must be noted therefore that in the glycine derivative, where R = H, both α_L and α_D represent stable conformations. However, in the case of the alanine derivative, where a hydrogen is replaced by a methyl (R = CH_3) group, the α_L conformation disappears, while the α_D is maintained. This would suggest a side-chain/backbone in-

(19) Arridge, R. G. C.; Cannon, C. G. *Proc. R. Soc. London* **1964**, *A278*, 91-109.
(20) Brant, D. A.; Miller, W. G.; Florey, P. J. *J. Mol. Biol.* **1967**, *23*, 47–65.
(21) Bierzynski, A.; Kim, P. S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2470–2474.
(22) Sundaralingam, M.; Drendel, W.; Greaser, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7944–7947.
(23) Gray, T. M.; Matthews, B. W. *J. Mol. Biol.* **1984**, *175*, 75-81.
(24) Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5585–5587.

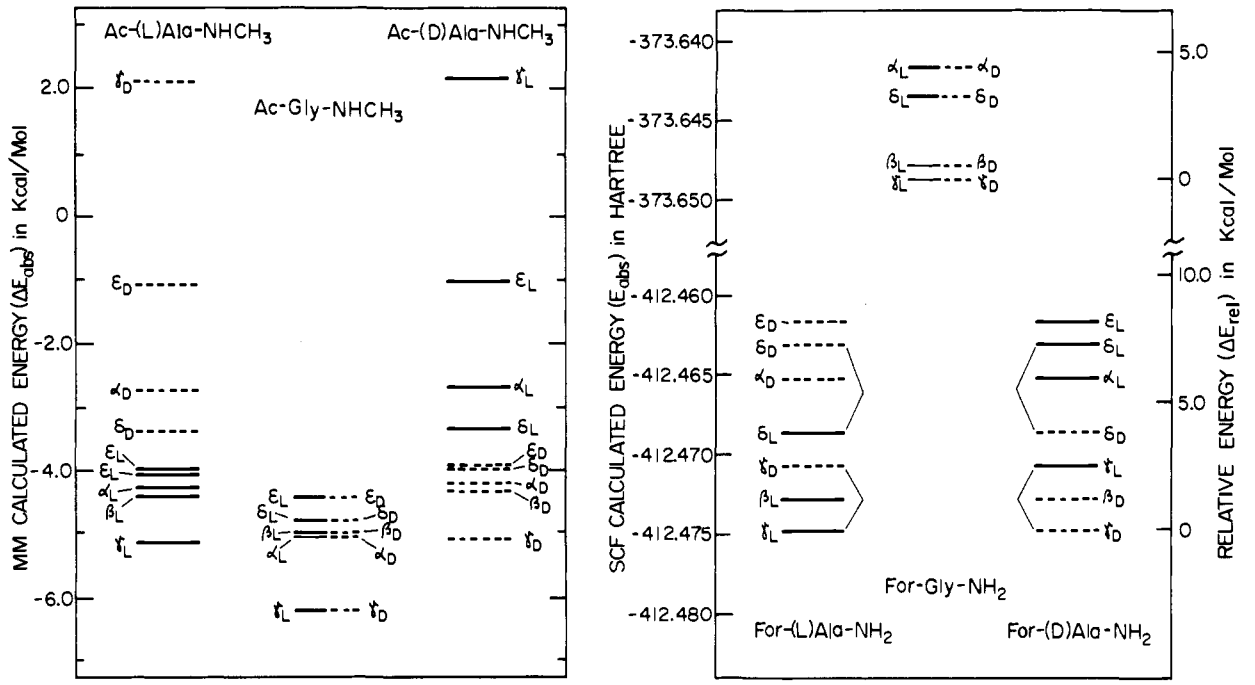


Figure 16. Relative conformational energies of glycine, L-alanine, and D-alanine diamides computed by molecular mechanics (left) and ab initio SCF methods (right).

teraction. In other words the $^{\delta+}H_3C^{\beta}-C^{\alpha\delta-}$ dipole interact with the dipoles of the two peptide bonds unfavorably in the case of α_L enforcing the destabilizing effect. In contrast, the same side-chain dipole interacts favorably with the two peptide bond dipoles, leading to stabilization in the case of the α_D conformation. These are clearly indicated, in relation to the molecular conformations, in Figures 13 and 14. A schematic view of the dipole vector model, as applied for the α_L and α_D conformations, is shown in Figure 15. This figure indicates two points: First, when all three vectors point more or less in the same directions, the like charges on each side repel each other and the system is destabilized (α_L). In contrast, when the $H_3C^{\beta}-C^{\alpha}$ dipole (which is sandwiched in between the two amide dipoles) is opposing the dipoles of the two peptide bonds, the system is stabilized (α_D). Second, due to vectorial summation of the dipole moment of the hypothetical α_L conformation, the L-alanine derivative is expected to be larger than that of the glycine derivative, which in turn is expected to be larger than that of the α_D conformation of the L-alanine derivative. The computed molecular dipoles are in agreement with such expectations:

$$\mu^{\text{L-Ala}}_{\alpha_L} > \mu^{\text{Gly}}_{\alpha_{DL}} > \mu^{\text{L-Ala}}_{\alpha_D}$$

6.98 D 6.96 D 6.56 D

Comparison of MM and ab Initio SCF Results. The topology of the MM and ab initio surfaces can be compared to each other and as well as to the ideal surface:

$$N_0 - N_1 + N_2 = 0$$

$$9 - 18 + 9 = 0 \text{ (ideal surface)}$$

$$9 - 16 + 7 = 0 \text{ (MM surface)}$$

$$7 - 11 + 4 = 0 \text{ (SCF surface)}$$

The MM results are in agreement with the ideal surface as far as the number of minima are concerned, which may be taken as an indication that MM is a faithful mathematical model of our classical additivity idea of conformational analysis. However, the discrepancy between the MM surface and the SCF surface clearly indicates that this classical model cannot describe excessive attractive or excessive repulsive interactions, which are fairly well represented in the SCF surface generated at the 3-21G basis set level. It seems reasonable to assume that the topology will remain the same when the calculations are carried out at a larger basis set level or when electron correlation is accounted for. If anything will be influenced at some higher level of calculation, it may be the ϵ_D minimum since it was not present in the corresponding glycine surface. It may be seen from Figure 16, as it is based on the data given in Tables I and II as well as Tables III and IV, that there are noticeable differences between the MM and the SCF surface energetics in addition to their similarities. The spectrum of the energy levels of the two types of surfaces do not differ widely. The maximum change in relative energy, i.e., the $\Delta(\Delta E_{rel})$ value, is 2.2 kcal/mol (at the MM level of theory) and 7.5 kcal/mol (at the SCF level of theory) for For-Gly-NH_2. The $\Delta(\Delta E_{rel})$ values for both types of surfaces (i.e., MM and SCF) fall between 8 and 9 kcal/mol for For-L-Ala-NH_2 as well as for For-D-Ala-NH_2. However, the pattern obtained by MM and the pattern obtained by the SCF calculations are markedly different, even though the corresponding γ_L conformations were found to be the most stable, by either of the two methods used for all three compounds studied. If this conclusion is generally true, then it could mean that MM may always be useful in determining the global minimum.

Although the present level of computations may easily be superseded by the use of enlarged basis sets or by the inclusion of electron correlation, nevertheless the split valence basis set (3-21G), used in the present work, is sufficiently flexible so that higher level calculations are not expected to alter markedly the relative energetic order of the minima.

Conclusion

The results of the present work clearly indicate that due to the creation and annihilation of critical points, caused by excessive repulsive or excessive attractive interactions, it may become necessary to recalibrate empirical force fields. It appears that the results of good-quality ab initio calculations might be suitable as primary standards for such a calibration. At least in the present study, MM was able to predict, for all the systems studied, the global minimum of the conformational PES. It is also clear that, in order to see at least a good portion of the total picture, the *N*-formylamides of all naturally occurring L-amino acids must be studied by ab initio methods as well as the amides of dipeptides, tripeptides, and tetrapeptides.

Acknowledgment. The ab initio computations were carried out at CIRCE (Orsey, France). We thank IBM-France and the CNRS for generous allocation of computer time within the

framework of the GS (Groupment Scientifique) "Modelisation Moleculaire". The continued financial support of the NSERC of Canada as well as the Canadian NRC and the CNRS (France) are gratefully acknowledged. This research was also supported in part by a grant from the Institute of Science Management and Informatics, Hungary and grants from the Hungarian Scientific Research Foundation (OTKA No. 1296 and 2245). Thanks are also due to the "IBM Academic Initiative in Hungary".

Note Added in Proof. After the completion of this research and after the preparation of the final draft of this manuscript, a paper was published by Professor J. A. Pople and co-workers (*Int. J. Quantum Chem. Quantum Biol. Symp.* **1989**, *16*, 311-322) on L-alaninediamides. The numerical results of the above paper and our present paper at the 3-21G basis set level are remarkably similar. However, the purposes of the two papers are obviously different and therefore they are complementary to one another.

Registry No. OHC-Gly-NH_2, 4238-57-7; OHC-L-Ala-NH_2, 134453-10-4; OHC-D-Ala-NH_2, 54046-46-7; Ac-Gly-$NHCH_3$, 7606-79-3; Ac-L-Ala-$NHCH_3$, 19701-83-8; Ac-D-Ala-$NHCH_3$, 128657-13-6; Ac-Gly-NH_2, 2620-63-5; Ac-D-Ala-NH_2, 71806-49-0; Ac-L-Ala-NH_2, 15962-47-7.

Communications to the Editor

Molecular Recognition by Circular Oligonucleotides: Increasing the Selectivity of DNA Binding

Eric T. Kool

Department of Chemistry, University of Rochester
Rochester, New York 14627

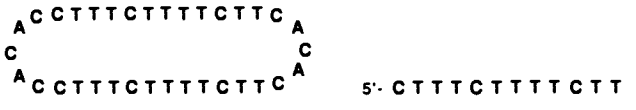
Received April 2, 1991

Sequence selectivity in the molecular recognition of DNA and RNA is an essential factor in the formation of polynucleotide secondary structure and in accurate copying of genetic information. In addition, recent interest in the possible use of oligonucleotides and analogues as therapeutic agents[1] has underscored the importance of the specificity of polynucleotide recognition.

We recently found that certain circular oligonucleotides can display strong binding affinities for single-stranded DNA and RNA by complexing the strand on two sides.[2] In this report it is shown that such a circular oligonucleotide can display higher sequence selectivity for its complement than does a standard DNA oligomer.

The circular compounds in this study were designed to bind strongly to complementary single-stranded purine sequences by forming hydrogen bonds from two sides of a circle to the central DNA target. Thus, a triple helical complex is formed,[3a-f] bounded by the two unpaired loop ends of the circle (see Figure 1). One side of a circle is complementary in the Watson-Crick sense (antiparallel), while the other side is complementary in the Hoogsteen sense (parallel).[4] This results in the formation of T-A-T and C+G-C base triads.[5] These pyrimidine-rich circles can thus be used to recognize purine sequences in single-stranded polynucleotides.

To measure the sequence selectivity of the circular ligand **1**, a set of complementary purine substrate oligomers with one variable base (X or Y) was constructed. Binding energies for the circle complexed with these oligomers were measured; the selectivity is defined as the free energy difference between binding of the correct sequence and the mismatched sequences. The selectivity obtained with the circular structure was then directly compared to the selectivity of standard linear oligomer **2**.



1 **2**

DNA oligomers were machine synthesized with standard β-cyanoethyl phosphoramidite chemistry. The circular ligand **1** was prepared from the linear precursor 5'-pTCTTTCCACACCTTTCTTTTCTTCACACTTCTTT and was cyclized by assembly around the template 5'-AAGAAAA-GAAAG, BrCN/imidazole being used to close the final bond.[2,6,7] The circular structure was confirmed by its resistance both to the 3'-exonuclease activity of T4 DNA polymerase and to calf alkaline phosphatase.

Thermal denaturation of the complexes was carried out in the presence of 10 mM $MgCl_2$, 100 mM NaCl, and 10 mM Tris·HCl (pH 7.0), with strand concentrations of 3 μM each. Free energies of association were obtained by fitting the data with a two-state model.[8,9]

Figure 1 shows the probe oligomers **1** and **2** hybridized with the variable-base (X or Y) oligomers. In the first set of four, T in the duplex is matched with X. These duplexes are then compared to the second set, in which opposing T's in the circle are paired with X. Similarly, in the third set, C in the duplex is matched with Y, while in the fourth, opposing C's in the circle are paired with Y.

Table I displays the results of the mismatch experiments. First, experiments 1-4 show the effects of a T-X mismatch on a DNA duplex. As expected, the true match (X = A) gives the most favorable complex ($-\Delta G°_{37}$ = 10.3 kcal/mol); the mismatches (X = G, C, T) result in a loss of 3.2-4.4 kcal/mol in binding energy, in good agreement with published mismatch studies.[10] Experiments 5-8, by comparison, show the effects of a T-X-T mismatch on circle complex strength. Once again, the true match

(1) (a) Toulmé, J.-J.; Hélene, C. *Gene* **1988**, *72*, 51-58. (b) Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 543-584.
(2) Prakash, G.; Kool, E. T. *J. Chem. Soc., Chem. Commun.*, in press.
(3) (a) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024. (b) Lipsett, M. N. *Biochem. Biophys. Res. Commun.* **1963**, *11*, 224-228. (c) Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, *37*, 63-80. (d) Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, *88*, 509-521. (e) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637-640. (f) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1942-1946.
(4) (a) Hoogsteen, K. *Acta Crystallogr.* **1959**, *12*, 822. (b) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645-650.
(5) Lipsett, M. N. *J. Biol. Chem.* **1964**, *239*, 1256-1260.
(6) Kanaya, E.; Yanagawa, H. *Biochemistry* **1986**, *25*, 7423-7430.
(7) The cyclization reaction contained 50 mM oligomer and template, with 200 mM imidazole hydrochloride (pH 7.0), 100 mM $NiCl_2$, and 50 mM BrCN, and was allowed to proceed for 36 h at 25 °C. The cyclic product was separated from starting material by preparative gel electrophoresis and was isolated in 58% yield.
(8) Thermal denaturation experiments were carried out in duplicate and the results averaged. Melting was monitored at 260 nm, and the resulting temperature vs absorbance curves showed a single transition from bound to free strands. Free energies of association were obtained by fitting the data with a two-state model.[9] In two cases the association energies were also determined from plots of log C_T vs $1/T_m$; good agreement (±7%) was seen between the two methods.
(9) Petersheim, M.; Turner, D. H. *Biochemistry* **1983**, *22*, 256.
(10) (a) Aboul-ela, F.; Koh, D.; Tinoco, I. *Nucleic Acids Res.* **1985**, *13*, 4811-4825. (b) Werntges, H.; Steger, G.; Riesner, D.; Fritz, H.-J. *Nucleic Acids Res.* **1986**, *14*, 3773-3791.

0002-7863/91/1513-6265$02.50/0