PERGAMON

Progress in Biophysics & Molecular Biology 71 (1999) 243–309

Progress in Biophysics & Molecular Biology

Ab initio characterization of building units in peptides and proteins

Attila G. Császár[a], András Perczel[b]

[a]*Department of Theoretical Chemistry, Eötvös University, P.O. Box 32, H-1518 Budapest 112, Hungary*
[b]*Department of Organic Chemistry, Eötvös University, P.O. Box 32, H-1518 Budapest 112, Hungary*

1. Introduction

It is hard to overestimate the revolutionary effects genetic engineering, gene technology, protein expression and other indispensable tools of protein architecture have on modern research in biochemistry and related fields. One of the basic techniques is the point mutation of a protein, which involves the exchange of one amino acid residue with another. Such a residue-based replacement is routinely used in most molecular biology laboratories. The aim of these studies is to provide a new protein with an improved 'quality'. However, not only the deletion or insertion of a new residue (and/or fragment), but also the modification of a single side chain functional group at the target position will affect the conformational properties of the new compounds. To determine the extent of these modifications and to augment the relevant experiments computational studies have been performed, for example, by Howard et al. (1975); Némethy et al. (1980); Paine and Scheraga (1986); Bruccoleri and Karplus (1987), Ripoll and Scheraga (1988); Schultz (1988); Wolfe et al. (1988) and Lambert and Scheraga (1989).

Conformational consequences of genetic engineering are seldom predictable. The relative orientation of the backbone structure (also influenced by side chains) determines the global fold of a protein. To aid our understanding of protein folding, a growing number of structures have been determined at atomic resolution by X-ray diffraction and NMR techniques (Bernstein et al., 1977; Abola et al., 1987). These studies of increasing sophistication allowed the subdivision of the structure of proteins into domains, modules, motives and other structural subunits. The main chain fold can also be categorized and divided into secondary structural elements. However, little is known about the a folding process of these structural subsets, even though continuous synthetic, modelling and computational efforts have been made for a better understanding (Mirskz and Pauling, 1936; Liquori, 1969; Aubry et al., 1974; Aubry et al., 1984; Raj et al., 1990; Stroup et al., 1990; Gething and Sambrook, 1992; Perczel et al., 1992; Lattman and Rose, 1993; Mezey, 1995).

0079-6107/99/$ - see front matter
PII: S0079-6107(98)00031-5

Dozens of observations, like those concerning the — VNTFV — sequence adopting two very different backbone conformations in ribonuclease and in erytrocruorin (Kabsch and Sander, 1984), suggest that the same atomic constitution may adopt dissimilar structures depending on the molecular environment. In a folded protein a given sequence often has a single backbone conformation. On the other hand, when the same sequence is modelled by a peptide in solution, the inherently flexible peptides will exhibit several low-energy conformers, most of them forbidden in proteins. Conformational analysis of these systems is almost impossible experimentally, even with high-field NMR spectrometers (Dyson et al., 1988; Wuthrich, 1986, 1990; Williamson, 1992; Williamson and Waltho, 1992). The key to NMR-based structure analyses is the assignment of a large number of interproton distances (Noggle and Schirmer, 1971; Neuhaus and Williamson, 1989). Since these peptides can exist in multiple conformations (Dyson et al., 1988; Wishart et al., 1991; Williamson, 1992; Williamson and Waltho, 1992), the usual structure determination, based on nuclear Overhauser effect (NOE) constraints, is not feasible: the observed interproton distances (calculated from NOE values) reflect time averaged structures (Waltho et al., 1990, 1992; Dyson et al., 1992a,b). Interpretation of molecular (e.g. vibrational) spectra of peptides and proteins (Drakenberg and Forsen, 1971; Ataka et al., 1984) also requires the knowledge of structures of most conformers present in the sample.

In response to challenges provided by experimental observations, computations have been carried out on single amino acids (NH_2–CHR–COOH), on their diamide derivatives (PCO–NHCHR–CONHQ), on more complex peptides and even on proteins at different levels of theory. For inherently flexible molecules, such as amino acids and peptides built from them, the more conformers described the better overall picture about their potential energy surface (PES) can be outlined. A large variety of molecular mechanics methods (force fields) have been introduced for the conformational analysis of peptides and proteins. All molecular mechanics methods rely on the success of their parametrization. From time to time comparative studies have been published (Roterman et al., 1989; McAllister et al., 1993). Some of the force fields HYPERCHEM (HYPERCHEM 4.5, 1994), MM+ (HYPERCHEM 4.5, 1994), AMBER (Wiener et al., 1984a,b), CHARMM (Brooks et al., 1983) and OPLS (Jorgensen and Tirado-Rives, 1988; Pranate et al., 1991), as well as the common semiempirical approaches (MINDO/3 (Bingham et al., 1975), MNDO (Dewar and Thiel, 1977), AM1 (Dewar et al., 1986) and PM3 (Stewart, 1989)) have been widely used to investigate molecules too large for nonempirical (ab initio) studies. However, the results obtained in these calculations have too often been too different from each other and from available ab initio data. Disturbingly, the number of allowed conformers, their locations and relative energies vary from one method to the other considerably. Therefore, we did not make an attempt to incorporate force-field or semiempirical results into this review. On the other hand, hundreds of inherently more accurate ab initio computations carried out on small amino acids and peptides are covered. At these higher levels of computational chemistry the first model compounds investigated included only the simplest residues; glycine (R = –H), alanine (R = $–CH_3$) and valine (R = $–CH(CH_3)_2$). Most theoretical studies have been published on glycine (for a list of studies published prior to 1992 see Császár (1992), some are available for α-alanine (e.g. Gronert and O'Hair, 1995; Császár, 1996) and valine (Shirazian and Gronert, 1997), while Schäfer and coworkers have reported an impressive series of small basis ab initio restricted Hartree–Fock (RHF)

calculations on larger amino acids (vide infra). Selected conformers of peptide models have also been the subject of ab initio studies, exploration of certain conformers of tri-, tetra-, penta-, hexa- and even hepta-amide systems (PCO–[NHCHR–CO]$_n$NHQ where $2 \leq n \leq 6$) have been reported (e.g. Böhm and Brode, 1995). Most of these studies provided detailed energetic and structural information to support spectroscopic (e.g. vibrational, CD and NMR), as well as diffraction studies, folding investigations and even protein structure predictions. Ab initio results on P–CONH–CHR–CONH–Q models, where P = Q = –H or P = Q = –CH_3, including glycine (R = –H) (Scarsdale et al., 1983; Schäfer et al., 1984; Wiener et al., 1984a,b; Head-Gordon et al., 1989; Böhm and Brode, 1991; Head-Gordon et al., 1991; Perczel et al., 1991; McAllister et al., 1993; and others), alanine (R = –CH_3) (Scarsdale et al., 1983; Schäfer et al., 1984; Wiener et al., 1984a,b; Head-Gordon et al., 1989; Böhm and Brode, 1991; Head-Gordon et al., 1991; Perczel et al., 1991; McAllister et al., 1993; and others), valine (R = –$CH(CH_3)_2$) (Viviani et al., 1993), serine (R = –CH_2OH) (Scarsdale et al., 1983; Farkas et al., 1995; Perczel et al., 1996a,b, 1997a) and phenylalanine (R = –CH_2Ph) (Farkas et al., 1996; Perczel et al., 1997b) are only a few examples where systematic topological searches have been carried out.

Finally, we give several arguments to explain why amino acids and simple model peptides form an attractive target for studies by methods of molecular electronic structure theory: (a) there is a finite number of amino acids and they have tremendous biological and biochemical significance; (b) amino acids and peptides contain a variety of intramolecular (and intermolecular) interactions most easily accessible by calculations; (c) these systems are very flexible conformationally and thus not easily amenable to experiments and (d) some of them are of tractable size even for high-level ab initio computations. To our best knowledge, the largest ab initio geometry optimization for a protein has been carried out for a local minimum of crambin (Van Alsenoy et al., 1998), a small hydrophobic protein with 46 residues and 642 atoms. A single-point ab initio energy calculation was performed for P53 with 698 atoms (Challacombe and Schwegler, 1997).

2. Theoretical armamentarium: a hierarchy of methods

The last three decades have witnessed remarkable and unforeseen advances in the theory and application of molecular quantum mechanics. Several excellent introductory and advanced textbooks have been written on the subject (e.g. Schaefer, 1977; Hehre et al., 1986; Náray-Szabó et al., 1987; Szabo and Ostlund, 1989; McWeeny, 1992; Yarkony, 1995). Therefore, a detailed treatment of modern methods of molecular electronic structure theory is not pursued here, the interested reader is referred to these volumes. Still, a short summary of the theoretical methods applied for studies on building units of peptides and proteins is given to aid readers less familiar with these techniques.

Basic to the understanding of the hierarchy of methods of electronic structure theory is the computational matrix of Fig. 1. It clearly shows that there are two fundamental approximations in methods of electronic structure theory: truncation of the one- and n-particle spaces. Extension both in the one-particle space (atomic basis sets) and n-particle space (many-electron wave function) are needed to achieve results close to the nonrelativistic limit ('exact

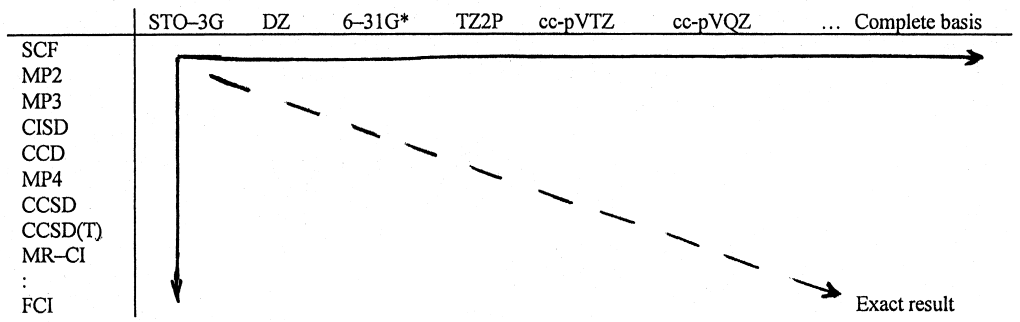


Fig. 1. Computational matrix of ab initio electronic structure theory indicating quality of one-particle space (basis set) vs. quality of the n-particle space (computational method) (for the abbreviations employed see the discussions Section 2).

result') represented, for a given basis set, by the full configuration interaction (FCI) method (vide infra).

It is noted in this respect that the electron correlation energy ($\varepsilon_{\mathrm{corr}}$) is defined as the energy difference

$$\varepsilon_{\mathrm{corr}} = E_{\mathrm{exact}} - E_{\mathrm{HF}},$$

where E_{HF} is obtained from a Hartree–Fock calculation. *Dynamic* correlation helps to keep electrons apart, it is a cumulative effect built up from individually small contributions, and usually forms the largest part of $\varepsilon_{\mathrm{corr}}$. *Nondynamic* correlation emphasizes that certain excitations are needed in the wave function for a proper zeroth-order description. Two computational strategies are based on these categories of electron correlation. One takes the single-reference (dynamic-correlation) route and introduces higher excitations until satisfactory results are obtained. The other approach introduces the relevant nondynamic correlation first through multiconfiguration (MC) approaches and augment the references by dynamic correlation via configuration interaction (CI) (or coupled cluster (CC)) schemes.

The exact basis-set correlation energy is obtained by full configuration interaction (FCI), by coupled cluster (CC) calculations with the full cluster operator $\hat{T}$, or by perturbation theory carried out to infinite order. Almost all methods of computational quantum chemistry are approximations to these techniques.

2.1. Many-electron wave functions

In computations employing methods of molecular electronic structure theory it is of great importance to determine whether single-configuration and single-reference-based methods, which are applicable with relative ease even for large molecular systems and have a 'black-box' nature, are of practical use or the at least conceptually more complex multiconfiguration (MC) and multireference (MR) methods need to be applied.

Different tests have been developed which allow estimation of the multireference character of a given electronic state of the molecule under consideration. One of the simplest and most dependable ones is the T_1 diagnostic of coupled-cluster theory (Lee and Taylor, 1989).

Calculation of T_1 values for a large number of molecules suggested that closed-shell electronic states with $T_1 < 0.02$ can adequately be described by single-reference-based electron correlation methods. For closed-shell neutral amino acids glycine, alanine and proline the T_1 values have been calculated (Császár, 1992, 1996) to be around 0.015; thus, closed-shell neutral amino acids can adequately be described by single-reference-based electron correlation techniques.

2.1.1. Hartree–Fock methods

The simplest treatment ab initio electronic structure theory offers is the Hartree–Fock, independent-particle theory (Hartree, 1928; Fock, 1930), often called self-consistent-field (SCF) theory (Hehre et al., 1986; Szabo and Ostlund, 1989). An appealing feature of HF theory is that this simple model retains the notion of molecular orbitals (MO) as delocalized one-electron functions describing movement of an electron in an average (effective) field of all the other electrons. In the SCF methods the MO's are variationally optimized in order to obtain an energetically 'best' many-electron function of a single configuration form.

There are several slightly different Hartree–Fock techniques. For closed shell species the restricted Hartree–Fock (RHF) level is usually a surprisingly adequate approximation for molecules containing atoms H–Kr, e.g. for amino acids and peptides. It accounts for a large percentage of the electronic energy of the molecule and it gives adequate wave functions. In the RHF theory it is assumed that all spin-orbitals χ_K are 'pure' space-spin products of the form $\phi_K\alpha$ or $\phi_K\beta$, where ϕ_K denotes the Kth spatial orbital, α and β are the usual spin functions ($\alpha(\omega_j)$ and $\beta(\omega_j)$, where ω_j indicates a coordinate in spin space), pairs of electrons occupy the same spatial orbital. At the spin-unrestricted Hartree–Fock (UHF) level spin-orbitals are still restricted to be of product form but α and β electrons are allowed to occupy different spatial functions. A considerable weakness of UHF theory is that the UHF wave function is contaminated by eigenfunctions of other spin multiplicity and thus it is not a pure spin eigenfunction. Similarly to the UHF method, the restricted open-shell Hartree–Fock (ROHF) method is applicable for open-shell molecules; here, however, electrons that are paired with each other are restricted, similarly to the RHF method, to occupy the same spatial orbital. Other variants of the HF method include the quasirestricted HF (QRHF) method (Rittby and Bartlett, 1988) and the generalized HF (GHF) method. The result of an SCF calculation is the total energy, the wave function consisting of MO's (canonical, localized or other) and the density, from which various properties can be calculated.

2.1.2. Multiconfiguration SCF methods

When chemical intuition or simple tests indicate a serious breakdown of the HF (single determinant) approximation, it is necessary to turn to multiconfiguration methods (e.g. multiconfiguration SCF, MCSCF, complete-active-space SCF, CASSCF, or restricted-active-space SCF, RASSCF). Multireference SCF methods are to be used when a single determinant description of the electronic state is even qualitatively incorrect. Such situations occur frequently for open-shell electronic states. To our best knowledge, such techniques have not been employed for the characterization of peptide building units.

2.1.3. Many-body perturbation theory (MBPT)

As in many branches of physics and chemistry, in computational quantum chemistry it is often expedient to use some form of perturbation theory (PT). In PT the Hamiltonian $\hat{H}$ is assumed to differ only slightly from an operator $\hat{H}_0$ with known eigenvalues and eigenfunctions: $\hat{H} = \hat{H}_0 + \hat{H}'$, where $\hat{H}'$ is a small 'perturbation' of the 'unperturbed Hamiltonian' $\hat{H}_0$. The perturbation treatment may converge poorly (or does not converge at all), and a complete set of eigenfunctions, even of the unperturbed Hamiltonian, may not be generally available.

Møller and Plesset (1934) proposed a perturbation treatment of atoms and molecules in which the unperturbed wave function is the Hartree–Fock function; this form of MBPT is called Møller–Plesset (MP) perturbation theory. When the Møller–Plesset perturbation theory is carried out to second order (MP2), it defines the simplest method besides density functional theory (DFT; vide infra) which incorporates electron correlation. MP2 calculations usually provide about 90% of the electron correlation energy.

2.1.4. Coupled-cluster (CC) methods

The coupled-cluster (CC) method was introduced into molecular electronic structure theory in the 1960's (Čížek, 1969). Efficient single-reference coupled-cluster procedures have been developed which are based on several types of reference wave functions, including UHF, ROHF, QRHF (Bartlett and Stanton, 1994) or Brueckner determinants (Handy et al., 1989). In the vicinity of equilibrium structures these methods have now reached a high degree of sophistication (Bartlett and Stanton, 1994; Yarkony, 1995).

The fundamental equation of CC theory is

$$\psi = e^{\hat{T}} \quad \phi_0,$$

where ψ is the exact nonrelativistic ground-state molecular electronic wave function; in most applications ϕ_0 is a variant of the normalized ground-state Hartree–Fock wave function, the $\exp(\hat{T})$ operator is defined by its usual Taylor-series expansion

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \cdots = \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!}$$

and the cluster operator $\hat{T}$ is defined as the sum of n-tuple excitation operators $\hat{T}_n$; for example, $\hat{T}_1$ converts the reference Slater-determinant into a linear combination of all singly excited Slater determinants. The effect of the $\exp(\hat{T})$ operator is to express ψ as a linear combination of Slater determinants that include ϕ_0 and all possible excitations of electrons from occupied to virtual spin-orbitals, exactly the same way as in a full CI calculation. The aim of a CC calculation is to find the mixing coefficients (called amplitudes); once these are known, the wave function ψ is known. In the usual applications of CC theory, only certain excitation operators are included in the cluster operator. Inclusion of $\hat{T}_1$ and $\hat{T}_2$ gives the CC singles and doubles (CCSD) method. The CCSD(T) approach (Raghavachari et al., 1989) — which includes a perturbative treatment for the effect of triple excitations — has provided accurate relative electronic energies (e.g. Yarkony, 1995; Császár et al., 1998a), as well as excellent results for a wide range of molecular properties (Bartlett and Stanton, 1994; Yarkony,

1995). It remains the most popular method for high-accuracy calculations. For most properties of general interest, the SCF, MP2 and CCSD(T) methods provide the most useful hierarchy of approximations of higher and higher accuracy.

In the equation-of-motion (EOM) formalism excited state energies $\{\varepsilon_k\}$ and wave functions $\{\psi_k\}$ are obtained by diagonalizing a similarity-transformed Hamiltonian

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}$$

where $\hat{T}$ is the cluster operator of the reference state and

$$\bar{H}\psi_k = \varepsilon_k \psi_k.$$

Several implementations of the method exist at the EOM-CCSD level (e.g. Stanton and Bartlett, 1993).

2.1.5. Configuration interaction (CI) methods

The following equation defines the single-reference CI wave function in general:

$$\psi_{\mathrm{CI}} = \phi_0 + \sum_{i,a} C_i^a \phi_i^a + \sum_{\substack{i<j \\ a<b}} C_{ij}^{ab} \phi_{ij}^{ab} + \ldots,$$

that is in addition to the reference (HF) wave function ϕ_0 the total wave function includes single, double and higher excitations, where, e.g.

$$\phi_i^a = \mathscr{A}(n)(\varphi_1(1) \ldots \varphi_a(i) \ldots \varphi_n(n))$$

corresponds to replacing the occupied SCF orbitals φ_i by an unoccupied orbital φ_a, ϕ_i^a, ϕ_{ij}^{ab}, $\phi_{ijk}^{abc} \ldots$ represent single, double, triple, etc. excitations and $\mathscr{A}(n)$ is the antisymmetrization operator. The CI coefficients C_i^a, C_{ij}^{ab}, etc. are optimized variationally to give the lowest electronic energy; evidently, $E_{\mathrm{CI}} < E_{\mathrm{HF}}$.

For an n-electron system the full CI (FCI) method is defined as the wave function that includes all possible excitations through order n. With a complete basis set the FCI method would become the so-called 'complete' CI technique, which is the exact solution to the nonrelativistic Schrödinger equation.

Although the conceptually relatively simple CI methods, which are variational in nature in contrast to nonvariational CC methods, have been in use from the early days of computational electronic structure theory, in the last few years they find less favor among users. This is due to the lack of size-extensivity of truncated CI treatments which makes studies of many-electron systems problematic.

2.1.6. Density functional methods

Density-functional (Kohn–Sham) theory (DFT, Parr and Yang, 1989) can be viewed as if the HF exchange term is replaced by an exchange-correlation functional. The functionals normally employed are integrals of some function of the density and the density gradient.

The most widely employed functional, the so-called B3LYP functional is a modification of the hybrid functional proposed by Becke (1993) and incorporates the exact exchange energy with the Lee et al. (1988) correlation potential. Although even simple DFT techniques have been employed successfully for the calculation of a number of properties, correct treatment of weak inter- and intramolecular interactions, abundant in amino acids and peptides, provides great challenge for the functionals to be employed. Nevertheless, use of DFT techniques in studies on amino acids have become widespread in recent years (e.g. Barone et al., 1995; Császár, 1996; Sirois et al., 1997).

2.2. The one-particle basis

All traditional quantum chemical calculations, whether HF, CI, MP or CC, start with the selection of a one-particle basis set. A large number of optimal basis sets have been developed for the efficient calculation of (relative) energies and different molecular properties (Dunning, 1989; Almlöf and Taylor, 1991).

The usual choice for the form of basis functions is the Cartesian Gaussian-type function (GTF), whose definition is

$$g_l(\mathbf{r}) = N x_a^i y_a^j z_a^k \exp[-\alpha(\mathbf{r} - \mathbf{r}_a)^2],$$

where N is a normalization constant, i, j and k are nonnegative integers, the orbital exponents α are taken to be positive and the basis function is centered on atom a. When $i + j + k$ is equal to 0, 1 and 2, the GTF is called s-type, p-type and d-type, respectively. Larger values of $i + j + k$ define basis functions of higher angular momenta.

It is more efficient to take each molecular basis function as a linear combination of a small number of Gaussians according to

$$\chi_r = \sum_l c_{lr} g_l,$$

where the g_l's are Cartesian Gaussians centered on the same atom and have the same i, j, k values to one another but different α's and χ_r is the rth contracted Gaussian-type function (CGTF). The contraction coefficients c_{lr} and the orbital exponents α have been preoptimized for a given basis set and they are held fixed during the actual quantum chemical calculation.

In the following description is given of basis sets most often employed for species of biological interest.

2.2.1. Minimal basis sets

When CGTFs are used as basis functions, a minimal basis set (MBS) consists of one contracted function for each core and valence atomic orbital (AO) (Hehre et al., 1986).

2.2.2. Split-valence basis sets

The 3-21G, 6-31G and 6-311G basis sets (Hehre et al., 1986) are the most commonly employed split-valence basis sets, which can generally be denoted as *K-LMN*G. They provide considerably more accurate results than MBS's both for energies and molecular properties. In

split-valence basis sets each inner-shell (core) AO (e.g. 1*s* for atoms Li–Ne) is represented by a single CGTF that is a linear combination of K primitive Gaussians; each valence-shell AO (e.g. 2*s* and 2*p* for Li–Ne) is represented either by two sets of basis functions with L and M primitive Gaussians, respectively (resulting in a valence double zeta (DZ) basis, where $N = 0$), or by three sets of basis functions comprised of L, M and N primitive Gaussians, respectively (resulting in a valence triple zeta (TZ) basis).

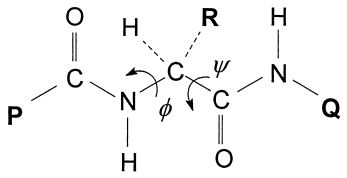
The split-valence basis sets are often augmented with *polarization* functions (usually denoted by *; e.g. 6-31G*) on one or more atoms, for polarization functions the $l = i + j + k$ angular momentum quantum number is greater than the maximum l in the corresponding free atom. These functions are not normally needed for the conceptual bonding description of the atom but give extra radial and angular flexibility to the wave function. Such higher angular momentum functions are especially critical when electron correlation is treated. Split-valence basis sets are also often extended with *diffuse* functions (usually denoted by $+$; e.g. 6-311++G**), which are low angular momenta functions (most simply *s* and *p*) and have small orbitals exponents and so fall off most slowly as r increases. Diffuse functions are needed to describe lone electron pairs and longer-range interactions, e.g. H-bonds.

2.2.3. Correlation-consistent basis sets

Among the basis sets developed for *correlated-level* electronic structure calculations one of the most successful sets are the correlation-consistent (cc) families of basis sets developed by Dunning and coworkers (Dunning, 1989; Kendall et al., 1992; Woon and Dunning, 1993). These basis sets, usually denoted as (aug)-cc-p(C)V*X*Z (X = D, T, Q, 5, 6), where (aug) means the addition of diffuse functions, (C) the addition of functions developed to describe core-valence correlation important for high-accuracy calculations and X is the cardinal number of the basis, approach completeness in a systematic way.

2.3. Empirically parametrized thermochemical schemes

Since extrapolation schemes to the basis-set and FCI limits considered involve very expensive computations, design of schemes which allow extrapolation to the limits based on relatively inexpensive calculations are of considerable practical value. The two most widely used methods are the Gaussian-2 (G2) approach and the complete-basis-set (CBS) model (Petersson et al., 1988; Ochterski et al., 1995). The widespread use of these schemes has also been aided by their efficient inclusion in the Gaussian program packages (Frisch et al., 1995) of electronic structure theory. Among the best contemporary schemes for general thermochemical predictions is the Gaussian-2 (G2) approach and its modifications (Curtiss et al., 1991, 1993; Raghavachari and Curtiss, 1995). In these techniques MP2, MP4 and QCISD(T) energies from basis sets up to *spdf* quality are employed with certain additivity approximations, and an empirical correction depending on the number of paired/unpaired electrons is utilized to account for deviations from the ab initio limit. The G2(MP2) approach (Curtiss et al., 1993) gives predictions of similar accuracy to G2 at a much reduced cost.



Scheme 1.

2.4. Multidimensional conformational analysis (MDCA)

To describe the main chain conformational characteristics of model peptides, the systematic scanning of the $E = E(\phi, \psi)$ surface (see Scheme 1) is needed (Head-Gordon et al., 1991). The computation of the entire surface ($0^\circ \leq \phi \leq 360^\circ$ and $0^\circ \leq \psi \leq 360^\circ$) requires 144 points, if the grid search is executed at 30° increments. Nevertheless, the resulting 2-D Ramachandran surface provides considerable rewards (see Fig. 2). This approach, however, can hardly be used at higher dimensions. For example, the grid search for HCO–L-Val–NH_2 [$E = E(\phi, \psi, \chi_1)$] would require calculations at $12 \times 12 \times 12 = 1728$ points. Furthermore, the complete structure analysis of HCO–L-Ser–NH_2 or HCO–L-Ala–L-Ala–NH_2 requires as many as $(12^4) = 20{,}736$ grid points. To fulfil the dream of ab initio conformational research in peptides and proteins, one must find a radically less laborious strategy.

Based on previously described theories (Peterson and Csizmadia, 1982; Mezey, 1987; Csizmadia, 1989), the use of multidimensional conformational analysis (MDCA) has been

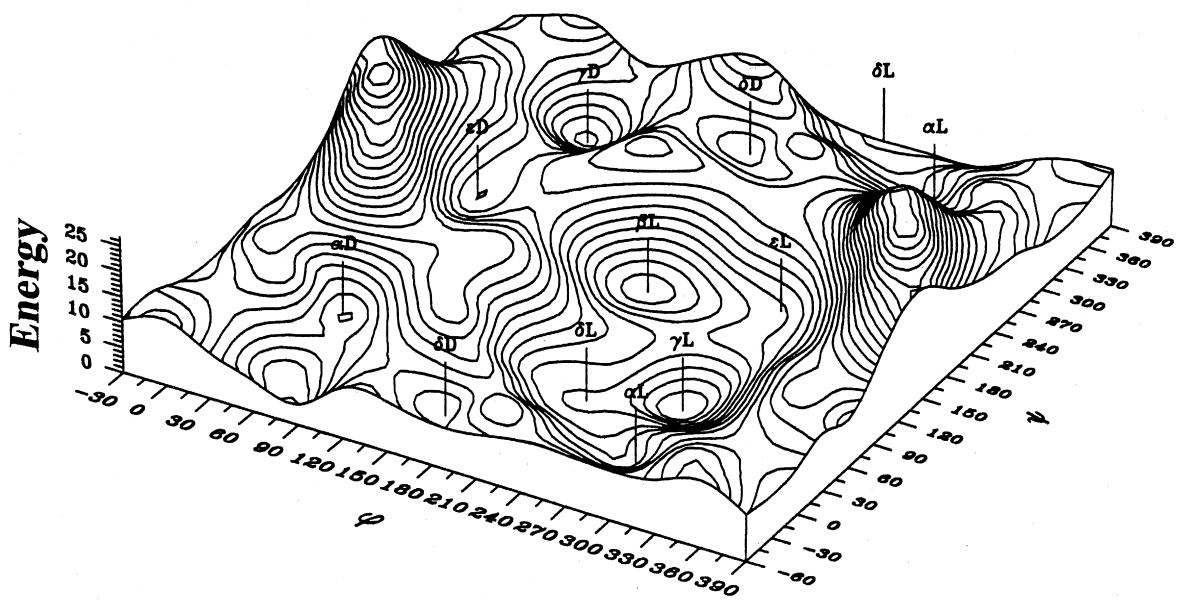


Fig. 2. The 3-21G RHF Ramachandran plot of HCO–L-Ala–NH_2. The seven minima are labelled according to Perczel et al. (1991) using subscripted Greek letters: α_D, β_L, γ_L, γ_D, δ_L, δ_D and ε_D. The expected locations of conformers α_L and ε_L are also marked.

suggested for the location of minimum-energy structures of peptides (Perczel et al., 1991). As a test of MDCA, minimization of the nine input geometries located by MDCA on the Ramachandran surface of HCO–L-Ala–NH_2 resulted in the same stationary points (Perczel et al., 1991) as the full grid search carried out by Head-Gordon et al. (1991). Prompted by the success of MDCA, a systematic grid search was then carried out for the serine-containing diamide molecule HCO–L-Ser–NH_2 (Farkas et al., 1995; Perczel et al., 1996a,b, 1997a). Even though the hydroxymethyl side chain of this molecule can interact strongly with the backbone atoms, the systematic search over more than 1000 3-21G RHF grid points (see Fig. 3) resulted in the same number of minima as the simple minimization of the 81 initial geometries predicted by MDCA. Nevertheless, MDCA is not a foolproof strategy (Ramek and Schäfer, 1998). Note that one can also fine-tune the approximate locations of initial points within a catchment region by slightly shifting the appropriate input ϕ, ψ or χ values based on previous ab initio results. For example, in the case of the 'fragile' δ_L backbone conformer (vide infra), instead of the use of the MDCA-recommended $\phi = -180^\circ$, $\psi = +60^\circ$ values, the $\phi = -150^\circ$, $\psi = +40^\circ$ could be the preferred choice. We have to note that the success of any minimization depends not only the correct location of the input geometry, but also on the applied minimization algorithm.

3. Structures of neutral amino acids

The interest in the shapes and spectra of the conformationally flexible free amino acids, the building blocks of peptides, the backbones of proteins, stems from at least three reasons: (a) the search for the origin and signs of life in cool interstellar space, which can be aided by careful laboratory investigations of the structures and the related signals of these biomolecules; (b) the desire to establish the intrinsic tautomeric and conformational energetics and the underlying potential energy hypersurfaces of these species which should help to determine the same characteristics in polypeptides and proteins and (c) to stimulate and to provide vital data for the development of better, more efficient and/or reliable computational methods, whether they are nonempirical (like correlated-level ab initio and DFT techniques) or empirical (like molecular mechanics) in nature.

Detection of amino acids (like the simplest one, glycine) on other planets or in interstellar clouds would have an enormous impact not only on interstellar chemistry and physics but also on relevant fields of biophysics and molecular biology including theories about the origin of life on Earth. There are more than 100 species identified in interstellar space and their number is growing steadily. Among these species are the molecules CH_3NH_2 and HCOOH, which could form glycine, as well as complex chain molecules. Searches for glycine in interstellar space were made in both the cm and mm ranges (Brown et al., 1979; Hollis et al., 1980; Snyder, 1983; Combes et al., 1996), but the detection of glycine cannot be confirmed.

Characterization of the structural properties of conformationally flexible, free molecules, such as amino acids, is only achieved when the full ensemble of possible conformations resulting from rotations about relevant bonds is considered. Because amino acids (for their names and abbreviations employed in this article see Table 1) are capable of a variety of intramolecular H-bonding and other secondary bonding interactions, they exhibit an unusually

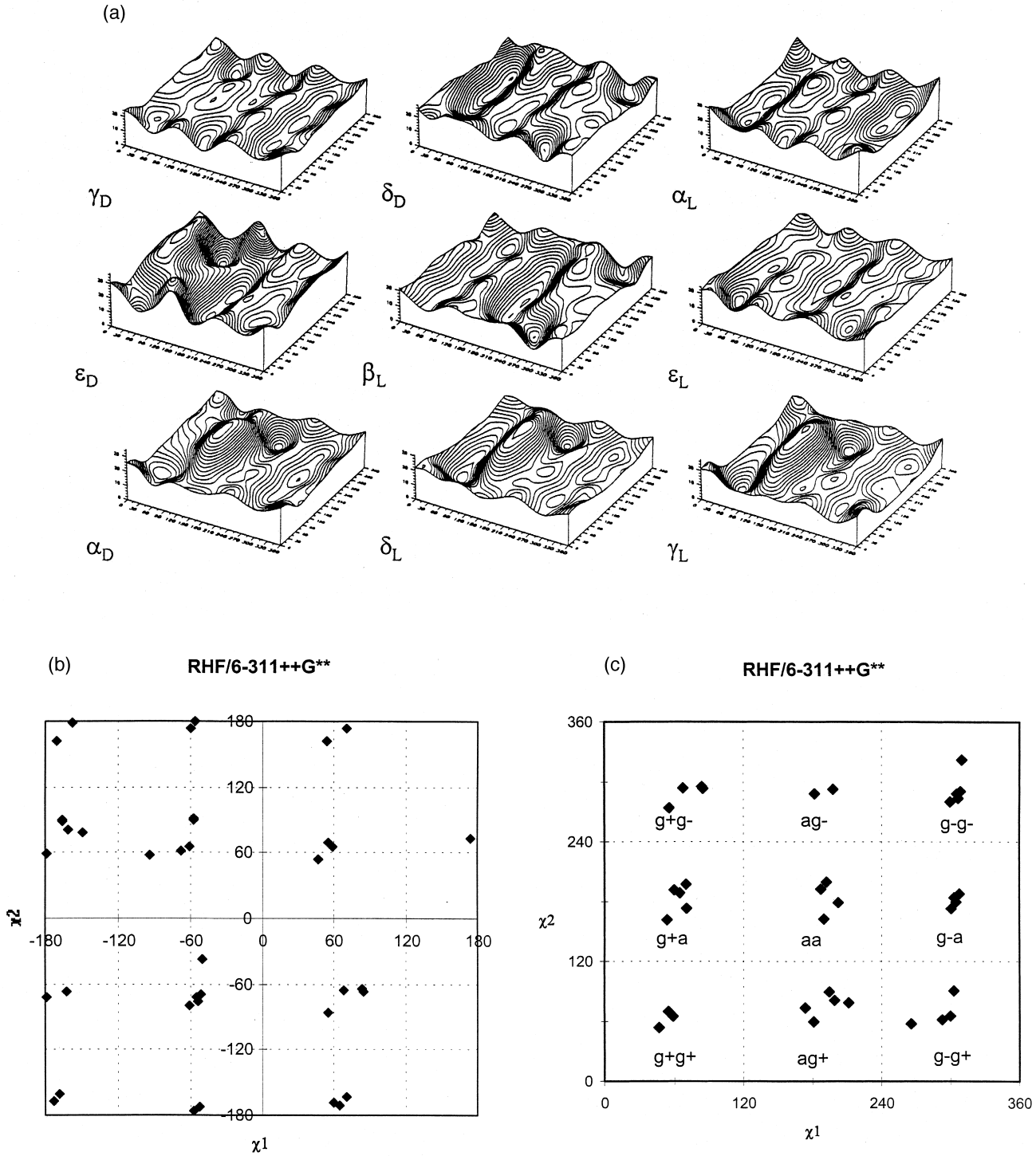


Fig. 3. (a) The nine $E = E(\chi_1, \chi_2)$ potential energy surfaces of HCO–L-Ser–NH_2 determined at the 3-21G RHF level. The appropriate backbone conformers associated with each side-chain maps are denoted by their conformational code: α_L, α_D, β_L, γ_L, γ_D, δ_L, δ_D, ε_L and ε_D. (b) Location of the χ_1, χ_2 values of HCO–L-Ser–NH_2 determined at the 6-311 + + G^{**} RHF level using the $-180^\circ \leq \phi \leq 180^\circ$ and $-180^\circ \leq \psi \leq 180^\circ$ representation. (c) Location of the χ_1, χ_2 values of HCO–L-Ser–NH_2 determined at 6-311 + + G^{**} RHF level using the $0^\circ \leq \phi \leq 360^\circ$ and $0^\circ \leq \psi \leq 360^\circ$ representation.

Table 1
Names and abbreviations of the 20 neutral α amino acids that occur naturally in proteins

Name	R	Abbreviation	Name	*R*	Abbreviation
Glycine	–H	Gly	aspartic acid	$-CH_2COO^-H^+$	Asp
Alanine	$-CH_3$	Ala	asparagine	$-CH_2CO(NH_2)H$	Asn
Valine	$-CH(CH_3)_2$	Val	glutamic acid	$-CH_2CH_2COO^-H^+$	Glu
Leucine	$-CH_2CH(CH_3)_2$	Leu	glutamine	$-CH_2CH_2CO(NH_2)H^+$	Gln
Isoleucine	$-CH(CH_3)CH_2CH_3$	Ile	lysine	$-(CH_2)_4NH_3^+$	Lys
Serine	$-CH_2OH$	Ser	arginine	$-(CH_2)_3NHC(NH_2)_2^+$	Arg
Threonine	$-CH(OH)CH_3$	Thr	histidine	$-(CH_2)(C_3H_3N_2)$	His
Cysteine	$-CH_2SH$	Cys	phenylalanine	$-CH_2(C_6H_5)$	Phe
Methionine	$-CH_2CH_2SCH_3$	Met	tyrosine	$-CH_2(C_6H_4)OH$	Tyr
Proline[a]		Pro	tryptophane	$-CH_2(C_8H_6N)$	Trp

[a]The zwitterionic focus of the amino acids will be denoted by the letters ZW attached to the abbreviated name of the amino acid. Therefore, usual notations, such as ^+N_3N–Ala–COO^- or $^+H_2$–Ala–O^- will not be employed here.
[b] No $-NH_2$ group.

large number of stable conformers of similar energy. Experimental characterization of all these conformers is impossible. Therefore, theoretical investigation of the structures of neutral amino acids is of special importance. When developments in the appropriate hardware and software allowed, almost the whole armamentarium of molecular electronic structure theory, as detailed in Section 2, has been employed for the investigation of the structure and related spectra of amino acids. Nevertheless, calculations have long been hindered by the size of these systems and by the demonstrated problems with entry-level theoretical calculations on these species, including (a) the inability of the restricted Hartree–Fock (RHF) level of theory to predict the relative energies of certain H-bonded conformers in these systems (Császár, 1992) and (b) conflicting results of geometry optimizations concerning minima vs. transition states in the different regions of the PES (Frey et al., 1992; Ramek et al., 1997).

3.1. Glycine

Following the labeling scheme introduced in (Császár, 1992), the conformers of the simplest amino acid, glycine, are numbered according to the increasing relative energy of the eight possible C_s symmetry structures. All the structures identified on the PES of the lowest singlet electronic state of Gly are indicated in Fig. 4; p denotes conformers of C_s and n those of C_1 symmetry. Eight conformers have been identified on the PES, among them there are five distinct conformers whose relative energy is less than 3 kcal mol^{-1} (see Table 2). Five of the eight minima have a nonplanar heavy-atom structure (Császár, 1992), which results from a balance between stabilizing intramolecular H-bonds and destabilizing steric strain and lone-pair electron–repulsion interaction.

No dependable experimental structural data exist for but the two lowest-energy conformers of Gly. The results of microwave (MW) studies (Suenram and Lovas, 1978; Brown et al., 1978; Schäfer et al., 1980; Suenram and Lovas, 1980; Godfrey and Brown, 1995) on Gly are in agreement with a gas electron diffraction (GED) study (Iijima et al., 1991) and the various

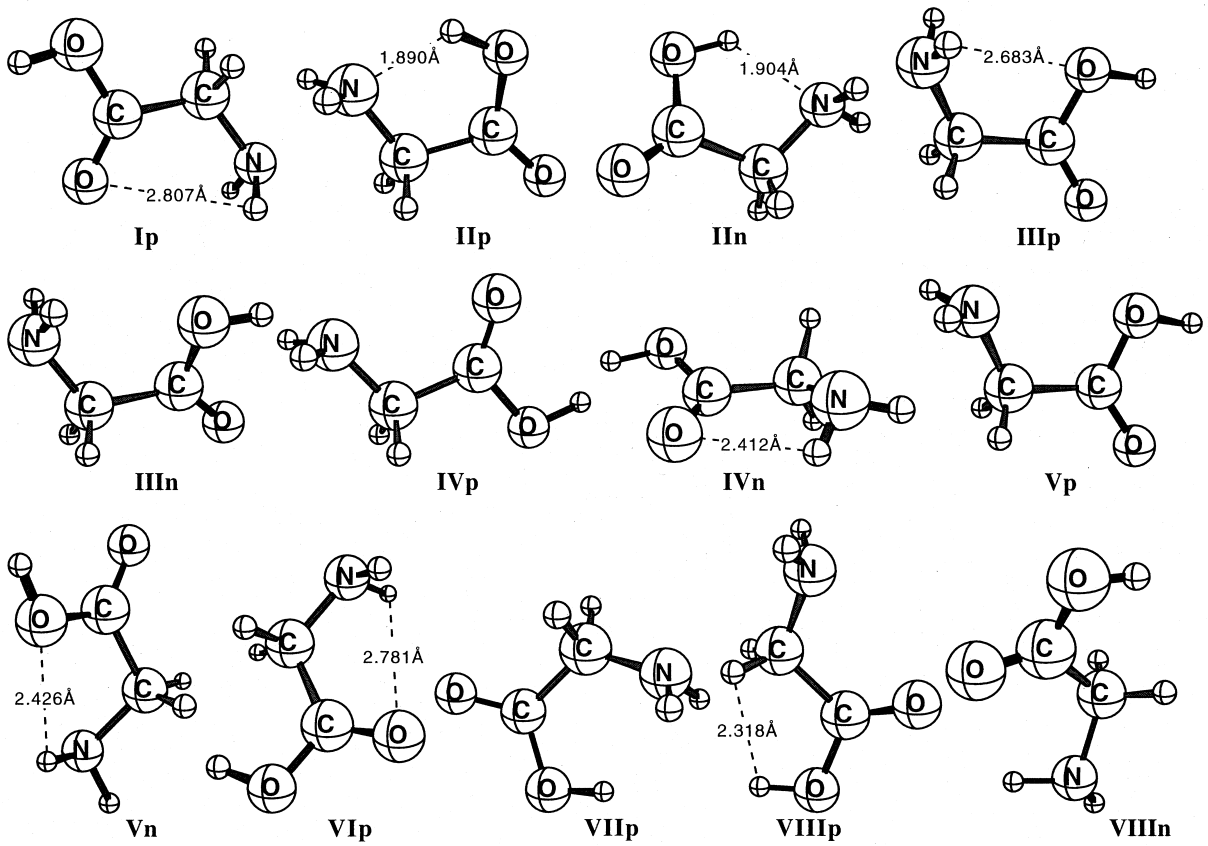


Fig. 4. Stationary points investigated on the PES of neutral glycine.

Table 2
Relative energies, in kcal mol^{-1}, of the conformers of Gly[a]

Method	Ip	IIn	IIIn	IVn	Vn	VIp	VIIp	VIIIn
{6s5p3d2f/4s2p1d} SCF	0.00	3.26	2.15	1.59	2.88	5.69	7.61	7.38
6-311 + + G^{**} MP∞	0.00	0.64	1.37	1.31	2.17	5.48	6.90	6.92
6-311 + + G^{**} CCSD	0.00	1.24	1.55	1.33	2.23	5.68	7.35	
6-311 + + G^{**} CCSD(T)	0.00	0.79	1.46	1.30	2.18	5.55		
Final prediction	0.00	0.49	1.60	1.23	2.51	4.72	5.76	6.04
Expt.[b]	0.00	1.40 ± 0.43						
Expt.[c]	0.00		1.70					
Expt.[d]	0.00	2.0 ± 0.2						
Expt.[e]	0.00	1.3–1.6	0.9–1.5					

[a] Most of the relative energies, in particular, the final predictions, are taken from Császár (1992). The reference geometries employed for the electronic energy calculations have been optimized at the 6-311 + + G^{**} MP2(full) level. No ZPVE corrections have been added. [b]Suenram and Lovas (1978). [c]Iijima et al. (1991). [d]Lovas et al. (1995). [e]Reva et al. (1995).

Table 3
Rotational, quartic centrifugal distortion, nuclear electric quadrupole constants and dipole moments for conformers Ip and IIp of free glycine[a]

Parameter	**Ip**		**IIp**	
	Experimental[b]	Theoretical[c]	Experimental[b]	Theoretical[c]
A	10,341.5279(49) [10,298]	10 279.0	10 130.2	10 175.1
B	3876.1806(23) [3868]	3877.0	4 071.5	4 076.3
C	2912.3518(16) [2911]	2 908.1	3 007.5	3 010.9
Δ_J	0.7703(55)		0.718(19)	
Δ_{JK}	3.834(29)		4.54(13)	
Δ_K	3.24(88)		5.0(12)	
δ_J	0.1874(28)		0.1759(80)	
δ_K	0.80(13)		1.98(42)	
eQq_{aa} (^{14}N)	−1.208(9)		1.773(2)	
eQq_{bb} (^{14}N)	−0.343(8)		−3.194(4)	
eQq_{cc} (^{14}N)	1.552(10)		1.421(4)	
μ_a	0.911(3)		5.372(34)	
μ_b	0.697(5)		0.93(10)	

[a] The rotational constants (A, B and C) are given in MHz, the A-reduced quartic centrifugal distortion (QCD) constants (Δ_J, Δ_{JK}, Δ_K, δ_J and δ_K) in kHz, the components of the dipole moment (μ_a, μ_b) in debye. The rotational constants given in brackets correspond to B_z values, i.e. rotational constants corrected for vibrational effects (Császár, 1992). [b]The experimental constants are taken from Lovas et al. (1995). [c]Most of the theoretical values are taken from Császár (1992), the rotational constants were obtained at the 6-311++G** DFT(B3LYP) level.

electronic structure calculations (Császár, 1992, 1995; Hu et al., 1993) which predict that the most stable conformer is Gly-**Ip**, which has a planar heavy-atom structure and two equal N–H···O H-bonds. As far as conformer **II** is concerned, the MW data indicate that this conformer has C_s symmetry with a planar heavy-atom framework (i.e. it is **IIp**). At the highest levels of ab initio theory employed thus far, the optimized equilibrium structure is of C_1 symmetry (**IIn**) resulting from a slight out-of-plane twist of the planar heavy-atom skeleton and structure **IIp** is a saddle point between the two equivalent **IIn** structures. However, zero-point vibrational energy (ZPVE) data (vide infra) suggest that the effective ground-state structure of **II** is probably planar which explains the extremely good agreement between the measured and calculated rotational constants of **IIp** (see Table 3). For conformer **III** the PES is extremely flat around **IIIp,** thus even calculations at the highest levels leave the question open whether the C_s or the C_1 form corresponds to a minimum. The experimental and the best related theoretical structural data are collected in Tables 2 and 3.

Based on intensity measurements on the $13_{5,8}$–$12_{5,7}$ rotational transitions of both conformers, the energy difference between conformers **Ip** and **IIp** was estimated to be 1.40 ± 0.43 kcal mol^{-1} by Suenram and Lovas (1980). In their GED study Iijima et al. (1991) arrived at the value of 1.70 kcal mol^{-1} for the energy difference between conformers **Ip** and **III.** When the several pitfalls associated with the use of the GED technique to obtain energy differences between rotational conformers (Hedberg, 1988) and some anomalies in that study (e.g. the complete neglect of the second most stable conformer of glycine, **II,** in the GED

analysis) are considered, the agreement between the theoretical values of ~1.6 kcal mol^{-1} and the experimental value is fortuitous. While high-level theoretical results suggest a small energy difference between conformers **Ip** and **IIp**, only about 0.5 kcal mol^{-1} (see Table 2), Lovas et al. (1995) revised the previous best experimental estimate of the energy difference upwards to about 2 kcal mol^{-1}. Further studies are needed to resolve this apparent discrepancy. Elaborate theoretical studies (Császár, 1992; Hu et al., 1993) predict a low relative energy for Gly-**IIIp.** In this regard Godfrey and Brown (1995) point out that their limited observation (only **Ip** and **IIp**) could be due to relaxation of the other conformers to **Ip** in the expanding gas jet suggesting that interconversion between the conformers is hindered only by low energy barriers. A recent theoretical study (Godfrey et al., 1996) of the barriers to interconversion between glycine conformers indicates that this explanation is likely correct.

3.2. Alanine

There are three possible methyl substitutions in Gly, on C_α–H, on N–H and on O–H; the resulting molecules are called α-alanine, N-methyl-glycine (sarcosine) and glycine methyl ester, respectively. The GED studies of the structures of glycine methyl ester and alanine methyl ester (Ewbank et al., 1987) indicate that the lowest-energy form of these compounds have a bifurcated H-bond between NH_2 and the carbonyl group. The calculated relative energies of the lowest energy conformers of these molecules are given in Table 4. As expected, the most stable of the three molecules is α-alanine (Ala).

Ala is a particularly important amino acid because it is the smallest one with a chiral α-carbon atom and, in peptides and proteins, the conformational properties of many other residues are well modeled by this simple system. The stable conformers of Ala can be built up from the eight stable conformers, three of C_s and five of C_1 symmetry, of Gly. The C_s symmetry conformers of Gly (**Ip, VIp** and **VIIp**) have equivalent C_α hydrogens. Upon methyl substitution this equivalence is maintained, resulting in only one stationary point on the PES of Ala. On the other hand, the C_1 symmetry conformers of Gly have nonequivalent C_α hydrogens and upon methyl substitution this nonequivalence seems to be maintained resulting in two different Ala conformers. Thus, the close correspondence between the conformational behavior of Ala and Gly results in 13 conformers of Ala (Cao et al., 1995; Császár, 1995, 1996). Following the notation employed for Gly the conformers of Ala are numbered according to their relative energies (see Fig. 5).

In entry-level (4-21G RHF) ab initio calculations Sellers and Schäfer (1978, 1979) and Siam et al. (1984) examined two Ala conformers of low relative energy and the N–C–C=O torsional

Table 4
Relative energies, in kcal mol^{-1}, of the lowest energy conformers of α-alanine, sarcosine and glycine methyl ester[a]

Method	α-Alanine	Sarcosine	Glycine methyl ester
3-21G* SCF	0.00	43.8	41.90
6-311++G** DFT (B3LYP)	0.00	8.91	10.25

[a] No ZPVE corrections have been appended.

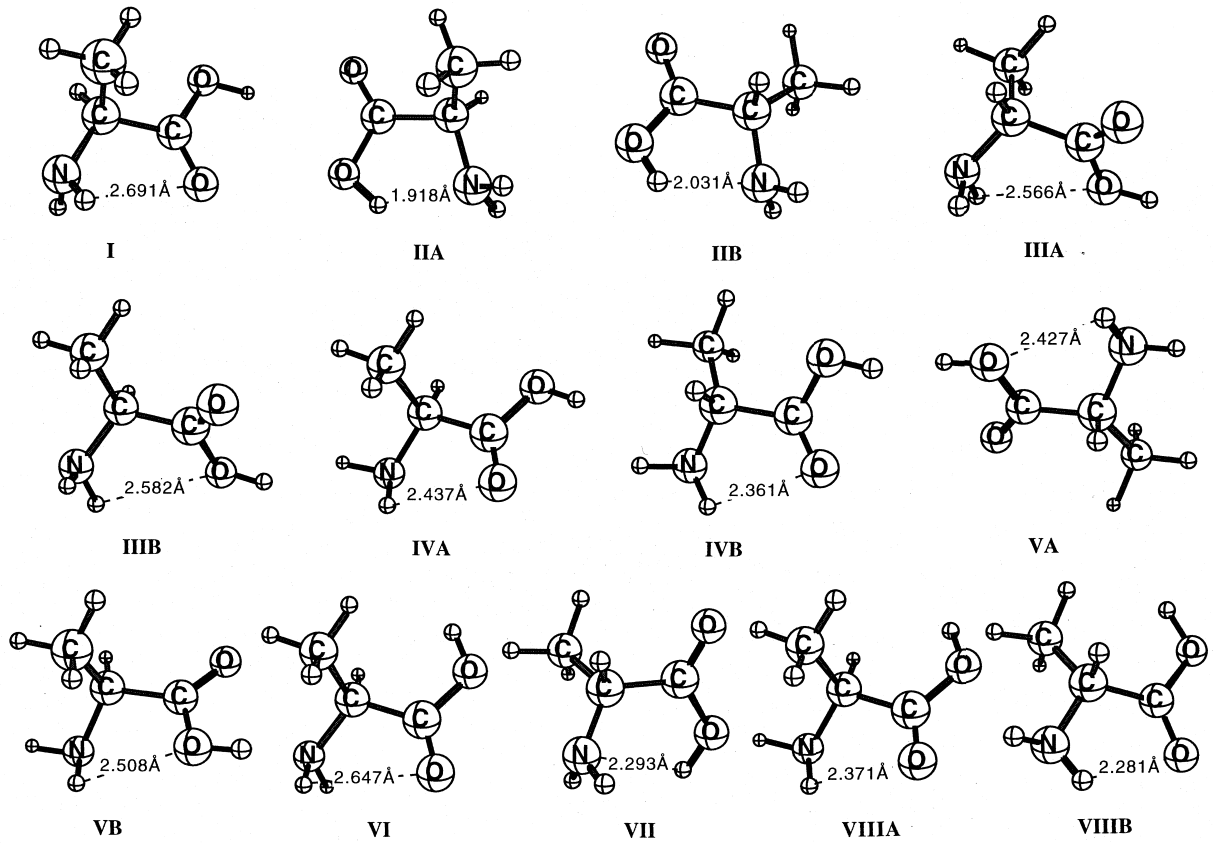


Fig. 5. Conformers of neutral α-alanine.

potential energy curve. In an extensive study at the RHF level Godfrey et al. (1993) reported 6-31G** RHF relative energies and several molecular parameters of six conformers of Ala (conformers **I, II, III, IV, V** and **VI** of Godfrey et al. (1993) correspond to conformers **I, IIA, IIB, IIIA, IIIB** and **VI,** respectively, of Fig. 5). In their detailed study Gronert and O'Hair (1995) located 10 conformers of Ala and characterized them by 6-31G* RHF and MP2 calculations. Their conformers 1–10 correspond to **I, IIA, IIIA, IIB, IVA, IVB, IIIB, VA, VB** and **VI,** respectively, of this review. Cao et al. (1995) determined all 13 conformers, their conformers 1–13 correspond to **I, IIB, IIA, IIIA, IIIB, IVA, IVB, VA, VB, VI, VII, VIIIA** and **VIIIB,** respectively. In summary, all theoretical studies indicate that introduction of the methyl group for Ala to replace one of the hydrogens of Gly has a rather small effect on either the geometry or the conformational preferences of Ala.

The millimeterwave (MMW) spectrum of Ala has been identified and analysed by Godfrey et al. (1993), who determined rotational and quartic centrifugal distortion constants and dipole moments for two conformers corresponding to those identified previously for Gly (see Table 5).

The only experimental attempt so far to determine geometries for Ala conformers is the GED study of Iijima and Beagley (1991), who determined structural parameters for Ala-I. They concluded, based on model refinements assuming internal rotation only around the C–C

Table 5
Experimental and theoretical rotational and quartic centrifugal distortion constants and dipole moments for conformers I and IIA of free α-alanine[a]

Parameter	**I**		**IIA**	
	Experimental[b]	Theoretical[c]	Experimental[b]	Theoretical[c]
A	5 066.1	5055.4	4973.1	4951.6
B	3 100.9	3037.8	3228.3	3218.0
C	2 264.0	2262.7	2307.8	2284.5
D_J	1.567	1.552	1.411	1.253
D_{JK}	−1.064	−0.931	−0.335	−0.157
D_K	0.87	1.038	0.84	0.561
d_1	−0.571	−0.584	−0.408	−0.335
d_2	−0.446	−0.446	−0.395	−0.335
μ_a	0.622	0.6	4.924	5.4
μ_b	1.60	1.2	1.4	1.4
μ_c	0.339	0.5	0.279	0.5

[a] The rotational constants (A, B and C) are given in MHz, the S-reduced quartic centrifugal distortion (QCD) constants (D_J, D_{JK}, D_K, d_1 and d_2) in kHz, the components of the dipole moment (μ_a, μ_b and μ_c) in debye. [b]The experimental constants are taken from Godfrey et al. (1993); for experimental uncertainties see the original publication. [c]The theoretical values are taken from Császár (1996), the rotational constants were obtained at the 6-311++G** DFT(B3LYP) level, the QCD constants at the 6-311++G** RHF level after a crude scaling employing the scaled quantum mechanical (SQM) method (Pulay et al., 1983; Allen et al., 1992) and scale factors of 0.9 for all stretches and 0.8 for all bends.

bond, that "the vapor of α-alanine consists of one conformer with a high potential barrier around the C–C bond", and consequently carried out their structural refinement for only one conformer. This finding is in clear contrast to indications of the MMW study of Godfrey et al. (1993) and detailed ab initio studies (Gronert and O'Hair, 1995; Császár, 1995, 1996), which suggest the coexistence of several conformers at the temperature of the GED experiment.

Studies by low-resolution photoelectron spectroscopy (Debies and Rabalais, 1974; Klasnic, 1976) did not yield any conformational information about Ala. On the other hand, Godfrey et al. (1993) estimated relative abundances of certain Ala conformers at the temperature of their MMW experiment, about 530 K, and determined approximate lower limits for conformational energies relative to that of Ala-**I** as ∼2.3 kcal mol^{-1} for Ala-**IIA,** ∼2.0 kcal mol^{-1} for Ala-**IIIA** and ∼3.4 kcal mol^{-1} for Ala-**VI.** The relative energies of the conformers of Ala obtained from ab initio calculations are collected in Table 6.

3.3. Serine

Serine is a good representative of those amino acids which contain the added complexity of a side chain capable of strong hydrogen-bonding interactions. Substitution of an OH group for one of the H atoms of the methyl side chain of Ala introduces new rotators, those around the C_α–C_β and C_β–O bonds, resulting in a dramatic increase in the number of possible conformers. The presence of three H-bond donors and four H-bond acceptors in Ser allows for the

Table 6
Theoretical relative energies, in kcal mol^{-1}, of the nine lowest energy conformers of Ala[a]

method	I	IIA	IIB	IIIA	IIIB	IVA	IVB	VA	VB
aug–cc–pVTZ SCF	0.00	2.64	2.82	1.38		1.40	1.52	1.85	
cc–pVTZ MP2	0.00	−0.27	0.02	1.08	1.25	1.15	1.44	1.84	2.24
6-31G** MP4	0.00	0.58	0.54	1.34	1.24	1.41	1.57	1.84	2.34
6-31G** CCSD(T)	0.00	0.64	0.60	1.34	1.25	1.41	1.56	1.77	2.32

[a] Most of the relative energies are taken from Császár (1996), the reference geometries have been optimized at the 6-311++G** MP2(full) level. No ZPVE corrections have been added.

existence of the following types of H-bonds: (a) bifurcated H-bonds between NH_2 and C=O, similar to that found in the most stable conformers of Gly and Ala; (b) simple H-bonds, like N–H to sidechain O–H and to carboxylic acid O–H and C=O; (c) carboxylic acid O–H to C=O, to sidechain O–H and NH_2; and (d) sidechain O–H to carbonyl oxygen, to nitrogen and carbonyl OH.

Quantum chemical calculations involving Ser include studies of protonation (Mezey et al., 1979; Wright and Borkman, 1980) and lattice energies (Voogd et al., 1981), parity violating energy differences between its enantiomers (Tranter, 1985, 1986), and the structures of many of its conformations (Van Alsenoy et al., 1981, 1988; Gronert and O'Hair, 1995).

In the most detailed structural investigation so far, Gronert and O'Hair (1995) located 51 conformers of Ser at the 6-31G* RHF level. They also determined vibrational frequencies, IR intensities, rotational constants, and dipole moments for all of the conformers. Van Alsenoy et al. (1988) located 14 conformers of Ser at the 4-21G RHF level, including eight of the 10 most stable structures found by Gronert and O'Hair (1995). If the conformers of Van Alsenoy et al. (1988) are numbered by Roman numerals in the order of relative stability, conformations I–X correspond to conformations 1, 2, 4, 3, 5, 6, 17, 9, 7 and 24 of Gronert and O'Hair (1995).

As repeatedly observed for amino acids, the conformers of Ser vary in energy by less than ~12 kcal mol^{-1} (see Table 7 for relative energies of five selected conformers). The three most stable conformers (see Fig. 6 for their bonding arrangements) lie within 0.25 kcal mol^{-1}, while altogether 33 conformers have been identified to lie within a 4 kcal mol^{-1} range.

Simultaneous combinations of different H-bonds form bi-, tri- and tetracyclic systems. As in Gly and Ala, the two N–H bonds in the most stable conformer of Ser were found (Gronert and O'Hair, 1995) to be almost symmetrically disposed to the carbonyl oxygen atom, while the

Table 7
Theoretical relative energies, in kcal mol^{-1}, of five low-energy conformers of serine[a]

Method	1	2	3	4	7
6-31 + G* RHF	0.00	1.5	0.7	1.0	1.7
6-31 + G* MP2	0.00	0.1	0.2	1.2	1.5

[a] Numbering of the conformers and their relative energies are taken from Gronert and O'Hair (1995), the reference geometries have been optimized at the 6-31G* RHF level. No ZPVE corrections have been appended.

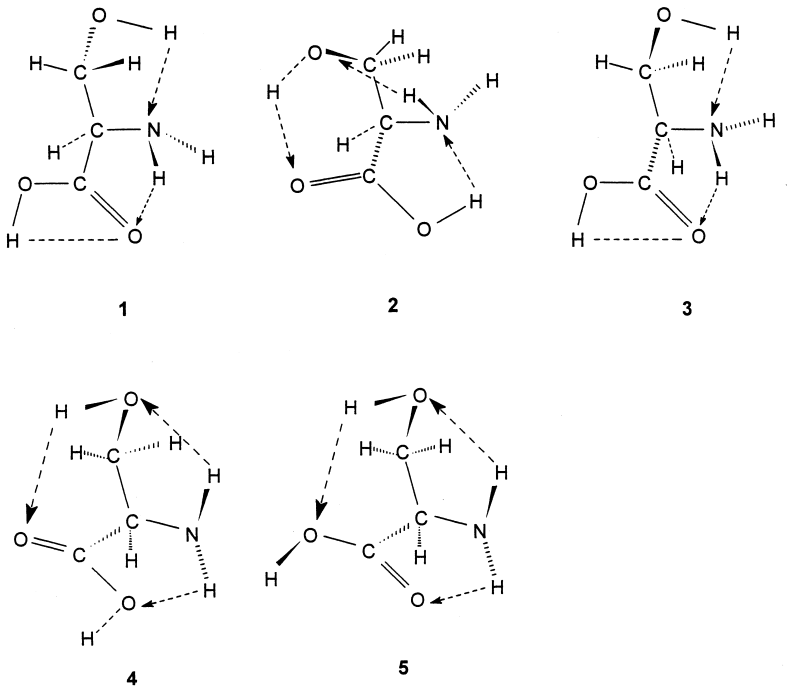


Fig. 6. Conformers of neutral serine.

side chain –OH formed a H-bond with the $-NH_2$ lone pair. Therefore, the most stable conformation of Ser can be considered to be a tetracyclic 11-membered ring. As in Gly and Ala, there is a strong preference for H-bonding to the carboxyl proton. Interactions of the side-chain hydroxyl with the nitrogen and the carbonyl are preferred over interaction with carbonyl OH.

An important side-result of the ab initio study of Gronert and O'Hair (1995) is that the semiempirical AM1 method is incapable of characterizing the relative energies of the conformers of Ser. This adds another example to the list of failures of empirical and semiempirical methods in predicting the subtle changes in the conformational preferences of neutral amino acids.

3.4. Cysteine

In principle, Cys could contain the same conformers as Ser. However, since the interactions in Cys are weaker than in Ser (the thiol group of the side-chain has relatively poor H-bonding characteristics), the barriers separating the conformers become smaller and in some cases may disappear. As a result, Cys exhibits fewer unique conformations than Ser.

To our best knowledge no experimental studies are available on the structures of the conformers of Cys. Ab initio calculations include studies on conformational behavior (Schäfer et al., 1990a,b; Gronert and O'Hair, 1995) and on various physical properties (e.g. proton affinities and ionization potentials (Wright and Borkman, 1980); and comparison of some

PCILO and RHF results (Laurence and Thomson, 1981)). Schäfer et al. (1990a,b) investigated 10 conformers at the 4-21G RHF level and established conformational trends. Gronert and O'Hair (1995) located 42 conformers at several ab initio levels of theory, including 6-31G* RHF and 6-31+G* MP2. The conformers vary in energy by less than ~10 kcal mol^{-1}. The five most stable conformers of Cys lie within 1 kcal mol^{-1}, while altogether 33 conformers were identified to lie within a 4 kcal mol^{-1} range. The best available theoretical results on the relative energies of some conformers of Cys are collected in Table 8.

Possibly due to the much weaker interactions present in Cys, the RHF and MP2 single-point energy results (Gronert and O'Hair, 1995; see also Table 8) are drastically different for a number of conformers. While in Gly and Ala inclusion of electron correlation tends to decrease the energy differences between the conformers, in the case of Cys it increases the energy differences in almost all cases. This observation also serves as a warning that the theoretical results obtained with small basis sets and simple ab initio methods might change considerably once more extended techniques are employed.

It is also noteworthy that, in contrast to Ser, the global minimum of Cys seems to have a H-bonding interaction between the carboxyl proton and the NH_2 group; however, conformers with syn carboxyls are nearly as stable.

3.5. Proline

The important role played by Pro, an unique amino acid in which the side chain is linked to the amino nitrogen atom, and its derivatives in enzymes and in proteins and peptide hormones is well known. Some of the studies on proline-containing peptides focused on tight turns, specifically on the β turn with a 1:4 hydrogen-bond and the γ turn with a 1:3 H-bond. While β turns are a common occurrence in proteins and peptides, γ turns are less frequent. Némethy and Printz (1972) first proposed the γ turn as a possible feature of polypeptide and protein conformation.

The structural features of Pro have been studied by experimental methods (X-ray crystallography, NMR, IR and CD spectroscopies) and by theoretical calculations (molecular mechanics (DeTar and Luthra, 1977; Ramek et al., 1997), PCILO (Cabrol et al., 1979), and ab initio (Sapse et al., 1987; Ramek et al., 1997; Császár et al., 1998b) methods).

Ramek et al. (1997) recently investigated 10 distinct conformers of Pro at the RHF level. The results of that study were extended by Császár et al. (1998b).

Table 8
Theoretical relative energies, in kcal mol^{-1}, of five low-energy conformers of cysteine[a]

Method	1	2	3	4	5
6-31+G* RHF	0.00	−1.4	−1.0	−0.4	−0.7
6-31+G* MP2	0.00	0.3	0.4	0.5	1.0

[a] Numbering of the conformers and their relative energies are taken from Gronert and O'Hair (1995), the reference geometries have been optimized at the 6-31G* RHF level. No ZPVE corrections have been appended.

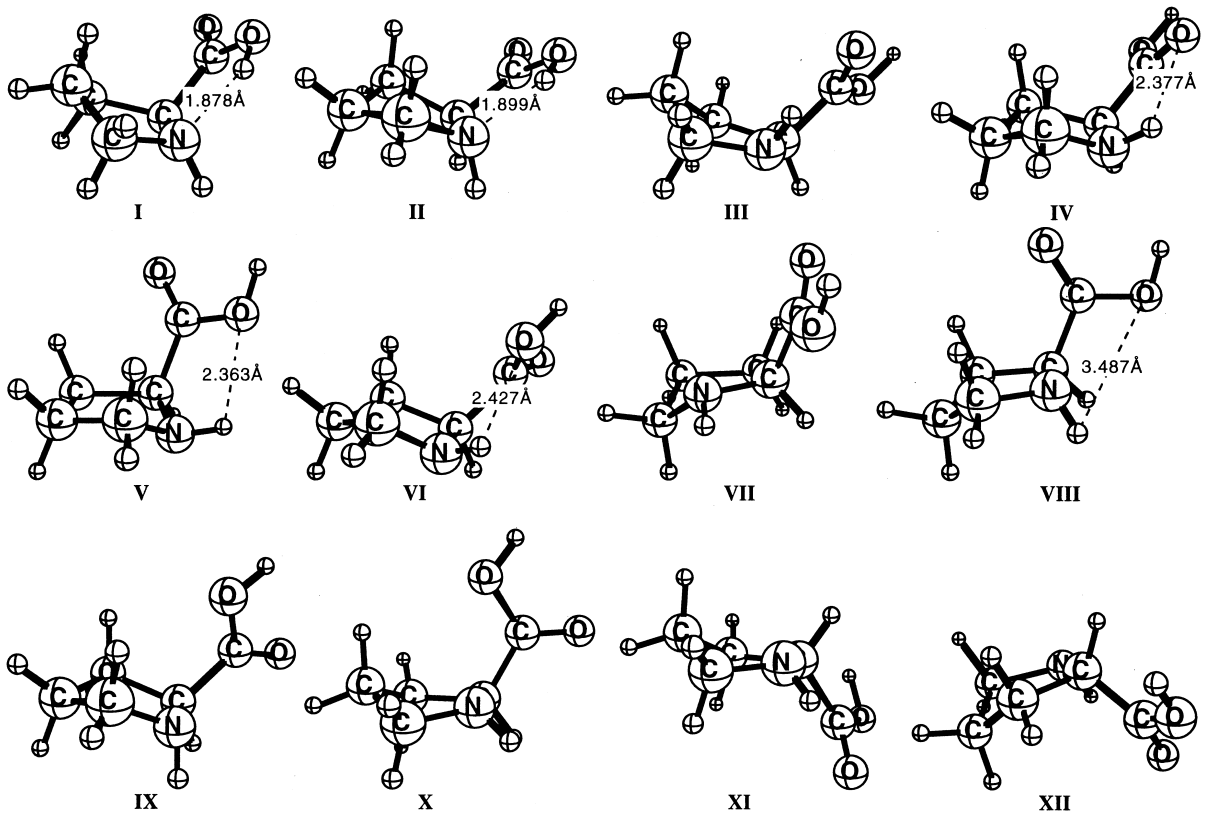


Fig. 7. Conformers of neutral proline.

By assuming a planar ring, the conformers of Pro can be characterized by the following torsion angles: HOCO, describing the *cis*/*trans* arrangement of the –COOH functional group, NCCO, making C=O or C–O closer to the N atom and H–N–C–C(OOH), describing the position of NH relative to C–C(OOH). The flexible 5-membered ring can deviate several ways

Table 9
Theoretical relative energies, in kcal mol^{-1}, of conformers of proline[a]

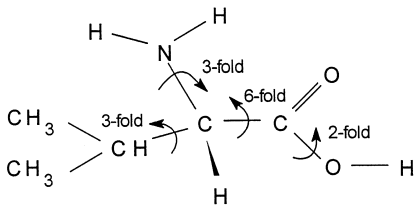
Method	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
6-311++G^{**} RHF	0.00	0.32	−0.20	−0.74	1.25	1.31	0.77	1.40	1.81	1.46		
cc-pVTZ RHF	0.00	0.37	−0.87	−1.50	0.70	0.71	0.22	1.18	1.40	0.95	4.13	4.68
6-311++G^{**} B3LYP	0.00	0.47	2.04	1.83	3.43	3.57	3.54	3.91	4.31	4.36	6.90	7.12
cc-pVTZ MP2	0.00	0.76	2.04	2.12	3.61	3.84	4.15	4.45	4.74	4.83	6.73	6.64
6-311+G^{*} MP4	0.00	0.46	1.73	1.50	2.83	3.05	3.11	3.41	3.73		7.09	

[a] Data are taken from Császár et al. (1998b) and Ramek et al. (1997); the reference geometries were optimized at the 6-311++G^{**} B3LYP level, except for the 6-311++G^{**} RHF energies, which were obtained at 6-311++G^{**} optimized geometries. No ZPVE corrections have been appended. For numbering of the conformers see Fig. 7.

from planarity, adding to the conformational freedom of Pro and increasing the number of minima. Theoretical relative energy estimates for the minima, depicted in Fig. 7, are given in Table 9. The ab initio studies (Ramek et al., 1997; Császár et al., 1998b) found that (a) not all of the conformers investigated correspond to minima on the PES of proline studied at different levels of RHF theory; (b) the lowest-energy conformer pairs **I/II** and **III/IV** have the same H-bonding characteristics, they differ only in the ring pucker angle; (c) there are three different ways the ring might distort from planarity and (d) at correlated levels of theory only the pair **I/II** has low relative energy, relative energies for all the other conformers increases markedly when compared to the often inferior RHF results. Ramek et al. (1997) also investigated reaction paths and potential barriers connecting stationary points on the PES of proline.

3.6. Valine

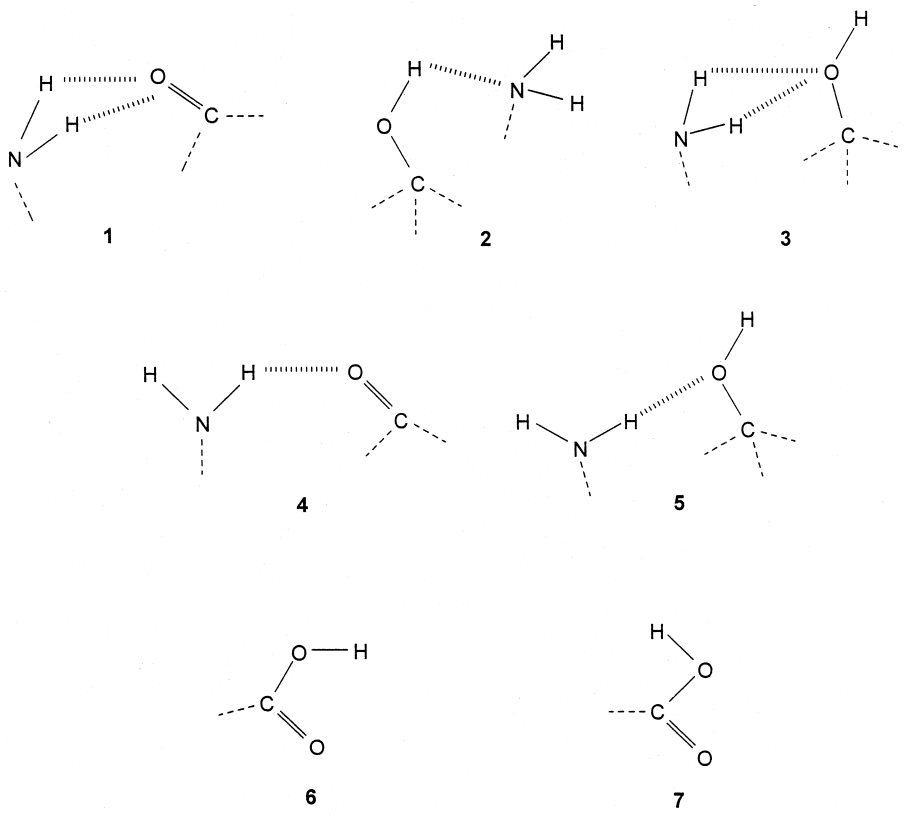
There are three H-bond acceptors (C=O, C(O)–OH and $-NH_2$) and two H-bond donors (–COOH and $-NH_2$) in Val (see Scheme 2), leading again to several H-bonding interactions. Seven geometries of Val were optimized at the RHF/4-21G level by Schäfer et al. (1990a,b). Conformational effects associated with the χ_1 (C^{α}–C^{β}) torsion were also determined. A survey of 108 structures of Val (Shirazian and Gronert, 1997) led to 26 unique conformers at the 6-31G* RHF level. Somewhat surprisingly, it was established (Shirazian and Gronert, 1997) that the isopropyl side chain has a significant effect on the stability of the conformers, even when the H-bonding interactions are essentially identical. Therefore, it appears that, unlike in the case of Ala, in Val there is a complex interplay between the orientation of the side chain and the H-bonding properties of the functional groups. Conformers 1, 3 and 4 of Val correspond to conformers **I, IIA** and **IIIA** of Ala. Energy results obtained at the RHF and MP2 levels also support earlier studies (Császár, 1992, 1995, 1996) that (a) MP2 energy results are substantially different from their RHF counterparts providing much improved predictions and (b) it does not appear that geometry optimizations at the MP2 level are required to obtain correct structures.



Scheme 2.

3.7. Other amino acids

Ten geometries of threonine were optimized at the RHF/4-21G level by Schäfer et al. (1990a,b) in their series of ab initio calculations on amino acids. The global minimum among the conformers considered has very similar characteristics to the global minimum of Ser, the –



Scheme 3.

OH group forms a relatively strong H-bond with the lone pair of the $-NH_2$ group. Structural differences obtained for the diastereomers of Thr are also discussed at some length.

Preliminary structural and energy results for arginine have been reported at the 6-31G* DFT(B3LYP) and MP2 levels of theory (Price et al., 1997). As admitted, these authors may not have located the lowest-energy neutral form of Arg. The high basicity of the guanidine side-chain functionality seems to stabilize the zwitterion, suggesting that Arg might exist as a zwitterion in the gas phase. Nevertheless, further experimental and theoretical work is warranted before this unusual result can be accepted.

We are not aware of further theoretical studies concerning the structures of any other nonionized amino acid.

From the short summary of theoretical results given in this chapter it is obvious that there is still lot of room for ab initio structural investigations of amino acids. These studies are, however, hindered considerably by the proven conformational flexibility of these systems and the need to include electron correlation to obtain meaningful results. Nevertheless, it is no doubt many more calculations will follow those correlated-level calculations published during the last few years.

Table 10
Computational matrix for the calculation of the E/Z rotamer separation energy of HCOOH[a]

	RHF	MP2	MP4	CCSDT
cc-pVDZ	5.70	5.36	5.15	5.19
cc-pVTZ	4.97	4.54	4.38	4.44
cc-pVQZ	4.77	4.34	4.20	4.28
cc-pV5Z	4.72	4.27	4.15	4.22
cc-pV6Z	4.70	4.27	4.13	4.21
Extrapolation limit	4.69	4.26	4.12	4.20

[a] Results are taken from Császár et al. (1998a).

3.8. Qualitative considerations

Energy estimates (Gronert and O'Hair, 1995; Császár, 1996), based on an approximate, qualitative picture obtained by completely neglecting the effects of steric strain and lone electron pair repulsion and assigning independent energy contributions to the different possible H-bonds and the carboxylic functional group, clearly assess the importance of the related energy effects as follows (see Scheme 3): (a) the O–H···N H-bond, 2, has the largest stabilization effect; (b) the stabilization contribution of the N–H···O H-bond, 1, is only about half as large as that of the O–H···N H-bond; (c) the N–H···OH bifurcated H-bond, 3, seems to be even weaker; (d) there is a sizeable energy difference between the E and Z –COOH arrangements (6 and 7) favoring the Z form.

It is thus clear that factors influencing the conformational energetics of simple neutral amino acids include not only various H-bonding interactions but also the isomerization energy of the –COOH group. The prototype of this moiety is formic acid (HCOOH). Microwave relative intensity measurements by Hocking (1976) place the E form 3.90 ± 0.09 kcal mol^{-1} above its predominant, naturally occurring (Z) counterpart. This value remains uncontested and has been accepted in calibrations of molecular mechanics force fields (Allinger et al., 1992). As detailed in Table 10, focal-point analysis (Allen and Császár, 1993) results (Császár, 1996; Császár et al., 1998a) put the vibrationless E/Z isomerization energy of formic acid to 4.19 kcal mol^{-1}. The effect of zero-point vibrations (ZPVE) on the rotamer separation is –0.22 kcal mol^{-1}, thus high-level ab initio results give direct support for the measured value.

4. Zwitterions of amino acids

All empirical, semiempirical and nonempirical theoretical calculations (e.g. Vishveshwara and Pople, 1977; Palla et al., 1980; Wright and Borkman, 1980; Bonaccorsi et al., 1984; Jensen and Gordon, 1995) agree that the neutral forms of simple amino acids are more stable than their zwitterion (ZW) analogs in the gas phase. A counterexample is provided by the extended Hückel (EH) method, which produced the opposite result in the case of glycine (Imamura et al., 1969); evidently, neglect of the electron–electron repulsion in the EH method leads to a gross exaggeration of the polar nature of zwitterions. On the other hand, it must be mentioned

that a recent experimental and theoretical study on Arg (Price et al., 1997) claims that the high basicity of the guanidine side chain stabilizes the zwitterion as compared to the neutral form suggesting that the most stable form of Arg is zwitterionic even in the gas phase.

4.1. Structures

The rotational potential of the backbone of the zwitterions of amino acids can be defined, following the convention adopted by Scheraga (1968), by the dihedral angle pair (ϕ, ψ), where $\phi = \angle H–N–C^{\alpha}–C'$ and $\psi = \angle N–C^{\alpha}–C'–O$. The angle ϕ indicates the rotation of the NH_3 fragment, while ψ the rotation of the CO_2 fragment. In the case of glycine, the (ϕ, ψ) = (0°, 0°) conformation has a H pointing toward the CO_2 fragment and all the heavy atoms are in the same plane.

Experimental geometries for the zwitterionic forms of amino acids adopted in solids have been obtained by the techniques of X-ray and neutron diffraction. Table 11 summarizes the references where the Cartesian coordinates of the atoms can be found.

Note that X-ray diffraction does not resolve the coordinates of the H atoms properly. On the other hand, neutron diffraction provides not only definitive stereochemical information on H-atom positions but accurate heavy-atom positions.

Table 11
Papers containing geometries for the zwitterions of natural amino acids in the crystalline state[a]

Amino acid	Method	Reference
Glycine	XD	Albrecht and Corey (1939); Jonsson and Kvick (1972)
Alanine	XD	Levy and Corey (1941); Simpson and Marsh (1966)
	ND	Lehmann et al. (1972)
Valine	XD	Torii and Iitaka (1970)
Leucine	XD	Golic and Hamilton (1972)
Isoleucine	XD	Torii and Iitaka (1971)
Arginine	ND	Lehmann et al. (1973)
Asparagine	XD	Ramanadham et al. (1972)
Lysine	XD	Koetzle et al. (1972)
Glutamine	XD	Koetzle et al. (1973)
Glutamic acid	XD	Sequeira et al. (1972)
Aspartic acid	XD	Rao (1973)
Proline	XD	Mitsui et al. (1969)
Histidine	XD	Eddington and Harding (1974)
Tryptophan	XD	Pasternak (1950)
Tyrosine	XD	Mostad et al. (1971)
	ND	Frey et al. (1973a)
Threonine	XD	Shoemaker et al. (1950)
Serine	XD	Frey et al. (1973b)
Cysteine	XD	Jones et al. (1974)
Methionine	XD	Mathieson (1952)

[a] XD means X-ray diffraction and ND neutron diffraction.

4.2. Energy considerations

In the simplest possible picture two factors govern the energy preferences of the zwitterions of amino acids in the gas phase: electrostatic attraction between the charged ends and repulsion of the hydrogens attached to the N and C atoms. In polar solvents interaction with the solvent molecules has a considerable impact on the structural properties of the zwitterions.

Classical calculations on the zwitterion of glycine (GlyZW) gave (Shipman and Christoffersen, 1973), the most stable structure around $(\phi, \psi) = (60^\circ, 0^\circ)$. In this case, the $(0^\circ, 0^\circ)$ structure may be viewed as a transition state connecting two $(60^\circ, 0^\circ)$ conformations. The $\psi = 60^\circ$ conformations are considerable less stable than the $\psi = 0^\circ$ conformations. Note that certain RHF calculations, using small basis sets (Palla et al., 1980; Bonaccorsi et al., 1984) produced a global minimum around $(\phi, \psi) = (0^\circ, 0^\circ)$. Thus, it is not completely clear which form of GlyZW is more stable, much higher level calculations than available were needed to make a firm conclusion. Nevertheless, it has been shown by detailed ab initio investigations (Jensen and Gordon, 1991; Yu et al., 1992; Ding and Krogh-Jespersen, 1992) that there is no minimum on the gas-phase PES corresponding to GlyZW. With basis sets which contain polarization functions on H atoms, the intramolecularly bonded zwitterion species collapse to Gly at all levels of theory studied. Therefore, the fully relaxed molecular structure of GlyZW in the gas phase cannot be determined by geometry optimizations.

At the highest levels of ab initio theory, including a G2(MP2) study (Yu et al., 1995), the lowest energy form of neutral glycine is computed to be more stable than the zwitterion by $\sim$20 kcal mol^{-1} (Jensen and Gordon, 1991; Yu et al., 1995). We are not aware of any other detailed ab initio studies on zwitterionic forms of amino acids in the gas phase but the work of Price et al. (1997) on Arg.

4.3. Solvation effects

A problem of considerable interest for the understanding of the chemistry of peptides and proteins is the determination of the most stable configurations of the macromolecules in solution (most importantly, in water). There is, of course, a complementary but related problem: structural organization of the solvent around these macromolecules. Ab initio theory offers two basic techniques to study solvation (environmental) effects. One can either put the system in a cavity surrounded by a continuum characterized by some macroscopic property (e.g. the dielectric constant), or can place discrete solvent molecules around the species studied (Tomasi and Persico, 1994). While the first approach allows to account for important nonspecific and long-range electrostatic interactions, the second makes possible the study of specific solvent effects.

In water, at room temperature, GlyZW is favored over Gly by a free energy and enthalpy of 7 and 10 kcal mol^{-1}, respectively (Wada et al., 1992; Watanabe et al., 1997).

Several ab initio studies have been published on the conformational stability of GlyZW in water employing continuum models. Bonaccorsi et al. (1984) calculated a (ϕ, ψ) map of GlyZW in water by partial geometry optimization and concluded that GlyZW is more stable than Gly. Full geometry optimizations, using the continuum model of Onsager, have been carried out on GlyZW by Ding and Krogh-Jespersen (1996). Kikuchi et al. (1990, 1994)

Table 12
Calculated relative energies (in kcal mol^{-1}) of representative (ϕ, ψ) conformations of amino acid zwitterions in water[a]

Amino acid	(0°, 0°)	(60°, 0°)	(60°, 90°)	(0°, 90°)
GlyZW	0.7	0.0	5.7	9.0
AlaZw	0.7	0.0	5.9	9.1
SerZW	0.6	0.0	7.1	10.3

[a] Taken from Kikuchi et al. (1994); calculated at the HF level by the generalized Born-formula with the 6-31G* basis set and $\varepsilon = 79$, where ε is the dielectric constant of the continuum (water).

employed continuum models at the MNDO and HF levels. The particular question why glycine is zwitterionic in aqueous solution has been answered in a detailed study by Tortonda et al. (1996).

Tortonda et al. (1996, 1998) investigated Gly, GlyZW, and the transition state connecting them by the ellipsoidal cavity model. While Gly-**Ip** is more stable than Gly-**IIp** in the gas, in solution Gly-**IIp** is favored by some 3 kcal mol^{-1} at the 6-31 + G** MP2 level due to the much larger dipole moment (see Table 3) of Gly-**IIp** (a very substantial dipole contribution to the solvation energy has been observed). Although not at the RHF level but at the MP2 level zwitterion formation is favored both from thermodynamic and kinetic points of view.

Other amino acid zwitterions have also been studied by continuum models. The results (see Table 12) suggest that the ϕ rotation requires a small activation energy, while the ψ rotation has an appreciable activation energy; the $\psi = 0$ conformation is stable.

As far as the absolute hydration energies presented in Table 13 are concerned, the negative of the hydration energy of GlyZW has been estimated to be 50–99 kcal mol^{-1} (Bonaccorsi et al., 1984; Kikuchi et al., 1994) by different theoretical methods. Comparison of data presented in Tables 12 and 13 suggests the following characteristics for the relative solvation energies of the different conformations: (a) the twisted forms ($\psi = 90°$) are stabilized more by the solvent than the planar ($\psi = 0°$) forms and (b) the solvation energy of the (60°, 0°) conformation is larger than the (0°, 0°) conformation.

We are not aware of any quantitative experimental information regarding the properties of zwitterion–water complexes. Clementi et al. (1977) considered the interaction energy of one molecule of water with the naturally occurring amino acids. Computational results on the interaction between the zwitterions of glycine, alanine, and serine with water have also been

Table 13
Calculated hydration energies (in kcal mol^{-1}) of representative structures of amino acid zwitterions[a]

Amino acid	(0°, 0°)	(60°, 0°)	(60°, 90°)	(0°, 90°)
GlyZW	−86.4	−89.3	−93.1	−94.4
AlaZw	−85.1	−87.7	−91.0	−92.5
SerZW	−93.8	−96.3	−102.3	−103.8

[a] Taken from Kikuchi et al. (1994). The energies have been calculated at standard geometries both in the gas phase ($\varepsilon = 1$) and in solution ($\varepsilon = 79$); the solvation energy is taken as the energy difference between the two.

reported by others (Förner et al., 1981; Deprick-Cote et al., 1992; Tortonda et al., 1996; Kikuchi et al., 1997; Natanabe et al., 1997; Okuyama-Yoshida et al., 1998).

The zwitterionic amino acid–water complexes provide information on pairwise solute–solvent interactions. Data obtained on these species could be useful in calibrating potential energy functions employed in future molecular mechanics/dynamics simulations on the properties of amino acids in water.

The most detailed theoretical study performed on the 1:1 GlyZW–water complex is due to Ding and Krogh-Jespersen (1992). They investigated 13 possible structures of various GlyZW–water arrangements. The conformers examined featured all possible H-bonding interactions between GlyZW and the water molecule: a single H-bond with the $CO_2^{-\delta}$ group as the acceptor and the H_2O molecule as the donor, a single H-bond with the $NH_3^{+\delta}$ group as the donor and H_2O as the acceptor, two unsymmetrical H-bonds with the CO_2^- group as the acceptor and H_2O acting as a double proton donor and multiple H-bonds involving both charged glycine terminal groups and the water molecule acting as both a proton donor and acceptor. The total H-bond energy appears to be substantial (15–16 kcal mol^{-1}) but, in agreement with all available theoretical results (Jensen and Gordon, 1995; Tortonda et al., 1996), it is not sufficient to render the 1:1 GlyZW–water complex lower in energy than a 1:1 nonionized glycine–water complex. It must also be noted that an analysis of the 6-31+G** MP2 PES of Gly–ZW·H_2O suggested (Tortonda et al., 1996) that there is no stationary point on the surface.

It is impossible to carry out a theoretical vibrational analysis on GlyZW since this species is nonexistent in the gas phase. The probably nonexistent GlyZW·H_2O complex of Ding and Krogh-Jespersen (1996) represented an equilibrium structure at the RHF surface, thus its vibrational spectrum could be calculated and compared with the experimental aqueous solution spectra of GlyZW (Furic et al., 1992). The calculated and experimental vibration frequency values compared favorably. An even better agreement can be found if the continuum model is employed to mimic hydration effects (Tortonda et al., 1996). According to the calculations the primary spectroscopic effects of GlyZW–water interactions should be found in the fundamental modes below 500 cm^{-1}, but they are, unfortunately, all computed to possess low Raman scattering intensities.

Jensen and Gordon (1995) studied the effect of solvation by one or two water molecules on mechanisms for proton transfer in GlyZW. Some of their important conclusions are as follows: (a) one or two water molecules are not sufficient to stabilize zwitterionic structures for which direct intramolecular proton transfer is possible; (b) two water molecules stabilize GlyZW; and (c) the barrier to proton transfer in dihydrated glycine appears to be at most a few kcal mol^{-1}. Tortonda et al. (1996) also pointed out that the energy barrier for the intramolecular formation of GlyZW is only about 2 kcal mol^{-1}, while the barrier to intermolecular proton transfer is about 15 kcal mol^{-1}. Therefore, it seems that even if a large number of water molecules surrounded GlyZW, the intramolecular mechanism would be the preferred one.

Carozzo et al. (1978) investigated the interaction of three distinct conformations of SerZW with one water molecule at the RHF level employing certain simplifications. While the results might change significantly if electron correlation was included in the calculation, in an interesting modeling approach the water molecule was placed at a number of distances and orientations relative to SerZW, so as to scan in reasonable number of energy points on the

interaction energy surface. The interacting atoms have been divided into 'classes' representing different electronic environments. Simple Lennard–Jones-type pair potentials have been constructed between pairs of atoms belonging to specified classes. Their conclusions are very similar to those obtained for GlyZW.

Naturally, the difference in energy between neutral and zwitterionic amino acids in the gas phase has not been determined experimentally. However, the calculated energy differences can be used together with the experimental heats of sublimation and heats of solution to provide estimates of the energies associated with the transition of the solid amino acid to the gaseous zwitterion ($^{+}A^{-}_{(s)} \rightarrow {}^{+}A^{-}_{(g)}$) and the solvation energy of the gaseous zwitterion ($^{+}A^{-}_{(g)} \rightarrow {}^{+}A^{-}_{(aq)}$).

The enthalpy change for the reaction $Gly_{gas} \rightarrow GlyZW_{sol}$ has been determined by Gaffney et al. (1977) on the basis of experimental data by exploiting the relation

$$\{\Delta H(Gly_{gas} \longrightarrow GlyZW_{sol}) = \Delta H_{diss} - \Delta H_{sub},$$

where ΔH_{diss} and ΔH_{sub} are the solution and sublimation heats of solid glycine. The resulting value is -19.2 ± 1 kcal mol^{-1} at 298 K (Gaffney et al., 1977), where the uncertainty regards the sublimation heat only, a relatively higher, but not quantified error, being associated with the dissolution heat.

Detailed knowledge of the conformational behavior and the interactions of amino acids (whether neutral or zwitterionic) in solution, especially in water, is of extreme importance to understanding hydration of proteins and the role of water in biochemical and biological systems. Therefore, despite the apparent successes ab initio calculations had in the modeling of the structures of amino acids in water, many more calculations at considerably higher levels of theory including more sophisticated continuum models will undoubtedly appear before this far-reaching goal is achieved.

5. Vibrational spectroscopy of amino acids

Vibrational spectroscopy is a nonintrusive experimental technique which provides important information about the structures of the ionic forms of amino acids in solution and in the crystal and about the structure and presence of different conformers of amino acids in the gas phase, especially at higher resolution.

5.1. Solid state and solution

Most of the research on the vibrational spectra of amino acids was performed in solution and in the solid state. Some of the relevant papers published on this topic are collected in Table 14.

Results of inelastic incoherent neutron scattering (IINS) supplemented by IR and Raman scattering studies have also been reported on valine (Pawlukojc et al., 1995), leucine (Pawlukojc et al., 1996) and isoleucine (Pawlukojc et al., 1997) in the solid state. These investigations resulted in vibrational bands below 800 cm^{-1}, which allow studies of vibrations involved in H-bonding interactions.

Table 14
Raman vibrational studies of different amino acids in aqueous solution

Amino acid	Reference
Glycine	Edsall (1936); Takeda et al. (1958)
Alanine	Edsall (1936); Takeda et al. (1958)
Valine	Garfinkel (1958)
Leucine	Garfinkel (1958)
Isoleucine	Garfinkel (1958)
Arginine	Garfinkel (1958)
Asparagine	Navarrete et al. (1997)
Lysine	Garfinkel (1958)
Proline	Garfinkel (1958)
Histidine	Garfinkel and Edsall (1958)
Threonine	Garfinkel (1958)
Serine	Garfinkel (1958)
Cysteine	Garfinkel and Edsall (1958)
Methionine	Garfinkel (1958)

In solution (for references of experimental work see Table 14), each amino acid can be studied in three forms: the protonated cation NH_3^+–CHR–COOH, the deprotonated anion NH_2–CHR–COO^- and the zwitterion $NH_3^{+\delta}$–CHR–$COO^{-\delta}$. The different forms of the amino acids can be studied in neutral, acidic and alkaline solutions. Infrared studies in ordinary aqueous solution are rendered extremely difficult by the strong absorption of the water used as solvent. Raman spectra, on the other hand, are readily obtained using either H_2O or D_2O as solvent and a wider range of conditions can be observed. The vibrational spectrum of GlyZW in aqueous solution, approximated by a continuum model, was calculated by Tortonda et al. (1996). The calculated and experimental bands compare favorably, suggesting that similar theoretical techniques might be of considerable use in the modeling of solvated amino acids and peptides.

5.2. Gas phase

Ab initio calculation of vibrational spectra can be accomplished most easily in the gas phase. Furthermore, the high apparent accuracy of the results obtained even at the simplest HF level is well established (Pulay et al., 1979, 1983; Allen et al., 1992). The calculated spectra are also characteristic for vibrational spectra measured in low-temperature inert gas matrices. Consequently, the vibrational spectra of neutral amino acids have been calculated in a number of recent papers.

5.2.1. Glycine

Reva et al. (1995) and Stepanian et al. (1998) investigated the vibrational spectra of Gly in low-temperature inert-gas (Ne, Ar and Kr) matrices in the infrared (IR) in order to determine the presence and provide relative energy estimates of (some) glycine conformers. Based on relevant assignments of the IR spectra they deduced relative energy estimates for Gly-**I, -IIp**

Table 15
Relevant theoretical (6-311++G** MP2(full)) harmonic vibrational frequencies (cm^{-1}), infrared intensities (km mol^{-1}) and zero-point vibrational energies (ZPVE, cm^{-1}) of conformers of glycine[a]

ν_i	**Ip**		**IIn**		**IVn**		**Vn**	
	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity
3	3562	3	3549	302	3559	3	3558	2
6	1825	265	1848	323	1831	258	1832	289
9	1422	23	1437	374	1472	32	1451	46
13	1191	56	1171	2	1159	210	1154	137
	1147	264						
17	845	77	867	101	869	72	868	79
					858	95	846	70
ΔZPVE	0		+ 138		+ 17		+ 42	

[a] Numbering of the normal modes (ν_i) of **Ip** was changed to follow those of the C_1 symmetry conformers. The relative zero-point vibrational energies (ΔZPVE) are measured with respect to the ZPVE of **Ip,** 17,723 cm^{-1}. A scale factor of 0.97 is perhaps a reasonable estimate for theoretical frequency and ZPVE corrections.

and **-IIIp** (see Fig. 4), as detailed in Table 2. Correlated-level ab initio results (Császár, 1992) suggest (relevant vibrational results are summarized in Table 15) that the rough relative energy predictions reported by Reva et al. (1995) for three conformers of neutral glycine are considerably less accurate than claimed: for Gly-**II** and **-III** the estimates seem to be too high and too low, respectively. Two points of general interest about the study of Reva et al. (1995): (i) neglect of the presumably third most stable conformer, Gly-**IVn,** could have a significant effect on the relative energy predictions that are simply based on measurements in the C=O stretch (ν_6) region and (ii) the convenient assumption based on RHF results "that in the C=O str region the relative intensities of the multiplet components are proportional to the fractions of the corresponding conformers in the matrix" might not hold resulting in severe distortion of the relative energy estimates.

5.2.2. Alanine

The plethora of information contained in the 6-311++G** RHF theoretical vibrational frequencies and infrared intensities (Császár, 1996) presented in Table 16 for Ala would allow interpretation of a carefully executed experimental investigation of the gas-phase (low-temperature inert-gas matrix) vibrational spectrum. To our best knowledge, no such study has been performed.

The results summarized in Table 16 reveal that there are some normal modes which do not change significantly from one conformer to another (including ν_8(C=O str), which changes between 1993 (**I**) and 2031 (**VIIIA**) cm^{-1}, and the usually rather intense band ν_{22}), while some normal modes having high IR intensity shift considerably. Most notably, conformer **IIA** has characteristic, high-intensity bands: ν_{13} at 1527 cm^{-1} and ν_{25} at 782 cm^{-1}. High-intensity normal modes whose frequency does not change from one conformer to another should serve as indicators of the presence of alanine, while characteristic shifts should help identification of alanine conformers in its gas.

Table 16
6-311 + + G** RHF vibrational frequencies, infrared intensities and zero-point vibrational energies (ZPVEs) of α-alanine conformers[a]

ν_i	I		IIA		IIB		IIIA		IIIB		IVA		IVB		VA		VB		VI		VII		VIIIA		VIIIB	
	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity	frequency	intensity
1	4115	124	4034	208	4033	221	4117	130	4117	127	4114	128	4118	11	4119	130	4119	132	4175	87	4170	107	4180	94	4193	92
2	3810	8	3825	13	3822	14	3821	9	3822	44	3827	12	3839	14	3829	10	3822	8	3809	10	3871	15	3827	15	3842	20
3	3736	4	3745	7	3735	6	3738	4	3734	3	3741	5	3752	6	3747	3	3738	3	3735	6	3764	7	3741	8	3754	9
8	1993	431	2019	440	2016	422	1994	440	1997	403	2000	412	1996	436	2005	415	2005	459	2026	377	2018	386	2031	358	2026	394
12	1552	13	1536	25	1541	154	1540	10	1554	2	1572	29	1553	19	1571	15	1546	34	1554	9	1535	12	1565	30	1546	10
13	1525	4	1527	396	1536	151	1520	13	1522	51	1528	4	1540	7	1537	39	1529	5	1531	2	1502	24	1532	5	1541	41
17	1290	198	1315	3	1318	7	1324	178	1293	261	1269	286	1311	148	1294	298	1307	111	1267	6	1289	3	1267	37	1282	10
18	1249	151	1233	17	1249	16	1237	18	1242	73	1257	13	1248	162	1248	60	1204	76	1235	44	1245	26	1239	18	1241	40
22	976	125	947	93	949	112	943	150	947	139	917	173	913	107	942	65	918	153	981	83	874	17	922	179	911	117
25	691	16	782	141	790	142	665	44	657	75	693	36	658	74	646	63	659	56	701	10	596	30	708	15	637	14
ΔZPVE	0		+ 122		+ 108		+ 10		−4		−25		−32		−18		−24		−72		−124		−97		−98	

[a] Frequencies (ν_i) in cm^{-1}, intensities in km/mol, relative zero-point vibrational energies (ΔZPVE), measured with respect to the ZPVE of I, 25,407 cm^{-1}, in cm^{-1}. All theoretical values were obtained at the 6-311++G** RHF level at the respective fully optimized reference geometries. The scale factor of 0.89 might be a reasonable estimate for frequency (and consequently ZPVE) corrections.

Table 17
Theoretical (6-311++G** B3LYP) harmonic vibrational frequencies (cm^{-1}) and infrared intensities (km mol^{-1}) of some bands of high IR intensity of three conformers of proline[a]

ν_i	**I**		**II**		**III**		Expt.[b]
	Frequency	Intensity	Frequency	Intensity	Frequency	Intensity	Frequency
1	3563	7.2	3559	6.6	3756	74.1	3559
2	3393	353.8	3423	326.5	3530	25.2	3393
10	1834	382.1	1833	390.9	1800	295.4	1789, 1766
14	1444	118.8	1443	28.9	1489	10.6	
15	1426	249.7	1423	355.3	1381	54.0	1412
24	1164	37.5	1193	30.8	1166	127.9	1142
25	1116	28.8	1102	48.4	1122	204.8	1109, 1105
30	926	84.8	928	17.5	935	7.5	
32			900	55.1	889	3.3	
34	854	7.3	832	6.2	780	91.7	
35	801	4.9	786	22.4	733	39.2	
36	735	16.5	731	11.3	689	58.5	
37	630	22.7	667	8.7	619	20.0	

[a] A scale factor of 0.98 is perhaps a reasonable estimate for theoretical frequency corrections. [b]Reva et al. (1994).

The frequency data of Table 16 also contain important information about the internal bondings of the conformers and about the relative strength of the different H-bonds. For example, the O–H stretch (ν_1) frequency shifts considerably from its free value of about 4115 cm^{-1} (I) to 4034 cm^{-1} (**IIA**) when involved in H-bonding. Some of the shift, however, is due to the change from a *Z* to an *E* COOH arrangement, which itself (Hu et al., 1993) produces an about 60 cm^{-1} shift. (Note that the upward shift in the predominantly C=O stretch mode, ν_8, from α-alanine I to II can also be observed in formic acid). This change is also in line with the almost 0.007 Å shorter C=O bond length in *E*-formic acid as compared to *Z*-formic acid. A less pronounced shift in the N–H stretch frequencies (ν_2 and ν_3) can be observed upon H-bonding, also indicating that the N–H···O H-bond is considerably weaker than the O–H···N H-bond.

5.2.3. Proline

Reva et al. (1994) investigated the IR spectrum of Pro trapped in an Ar matrix (some of their experimental frequencies together with ab initio data are given in Table 17). Their most important conclusions are as follows: (a) the spectra recorded do not exhibit the bands corresponding to the zwitterionic form of proline; (b) as evidenced by a high-intensity OH stretch band at 3559 cm^{-1} and by the C=O stretch bands in the region 1760–1790 cm^{-1}, proline exists in its neutral form in the matrix. There are two intensive bands in the C=O stretch region, which have been assigned to two conformers based on AM1 semiempirical calculations. The three conformers considered by Reva et al. (1994) correspond to Pro-**III, -V** and -**XII** of Fig. 7. Thus, they did not include in their study Pro-**I** and Pro-**II**, the most stable conformers. Therefore, reinvestigation of the relevant findings of this study is warranted. To

help, several harmonic frequencies and IR intensities of Pro-**I,** -**II** and -**III** are given in Table 17.

As it is most apparent from the studies summarized briefly in this chapter, we are still a long way off from understanding the vibrational spectra of free amino acids. This line of research is hindered by the low volatility and the conformational freedom of neutral amino acids. While the existing experimental studies helped to establish that the amino acids investigated are not zwitterionic in an 'interaction-free' environment, they suffered from poor choices made for theoretical modeling. Nevertheless, interplay of theory and experiment is expected to play a fundamental role in unraveling the complex vibrational spectra of at least the lowest-energy conformers of most amino acids.

6. Selected thermochemical data for amino acids

In order to understand the energetics (stability, reaction energies, etc.) of amino acids in the solid state, in solution and in the gas phase, it is advantageous to know different types of so-called thermochemistry data.

As detailed below, several experimental techniques (e.g. static and rotating bomb calorimetry) as well as different methods of computational quantum chemistry have been employed to obtain relevant thermochemical data for amino acids. Occasional comparison of experimental and calculated thermochemistry data sheds light on the precision of the two approaches.

6.1. Enthalpies of formation

The standard enthalpy of formation of a substance is the standard reaction enthalpy for the formation of the compound from its elements in their reference state. These data are especially useful since one can regard a reaction as proceeding by decomposing the reactants into elements and then forming the same elements into products. Therefore, the reaction enthalpy for the overall reaction can be considered as the sum of the two enthalpies.

Experimental enthalpies of formation, $\Delta H^0_{f,\mathrm{solid}}$ and entropies, S^0_{solid}, of amino acids in the solid state have been collected in Table 18. The few available gas-phase data are given in Table 19. We are not aware of any theoretical calculation of these quantities.

6.2. Gas-phase acidities

The gas-phase acidity (proton-donor ability) of a species AH can be measured according to the deprotonation reaction

$$\mathrm{AH} \xrightarrow{\Delta G^0_{\mathrm{acid}}} \mathrm{A}^- + \mathrm{H}^+.$$

The acidity values are of fundamental importance in understanding the reactivity of AH. Chemical intuition suggests that deprotonation of α-amino acids occurs at the carboxyl group.

Table 18
Experimental enthalpies of formation ($\Delta H^0_{f,\mathrm{solid}}$/kcal mol^{-1}) and entropies (S^0_{solid}/cal mol^{-1} K^{-1}) of amino acids in the solid state

Amino acid	$\Delta H^0_{f,\mathrm{solid}}$	S^0_{solid}	Reference
Glycine	-126.3	24.7	Ngauv et al. (1977)
Alanine	-134.0 ± 0.4	28.4	Contineanu and Marchidau (1984)
Valine	-147.7	42.8	Hutchens et al. (1963)
Leucine	-154.6	50.6	Hutchens et al. (1963)
Isoleucine	-153.1 ± 0.5	49.7	Hutchens et al. (1963); Wu et al. (1993)
Threonine	-185.5 ± 0.3	–[a]	Wu et al. (1993)
Proline	-121.3 ± 0.6	39.2	Sabbah and Laffitte (1978)

[a] Not available.

Locke (1981) and Locke and McIver (1983), using the pulsed equilibrium ion cyclotron resonance (ICR) technique, as well as O'Hair et al. (1992), using the kinetic method of Cooks and Kruger (1977), determined the gas-phase acidities of almost all of the α-amino acids. The experimentally determined gas-phase acidities as well as some computational results are presented in Table 20.

It is important to note that experimental methods measure $\Delta G^0_{\mathrm{acid}}$ directly, while methods of electronic structure theory reflect $\Delta H^0_{\mathrm{acid}}$. Furthermore, the calculations assume that the amino acids and carboxylates adopt their lowest energy conformations; however, this may not be the case during the unimolecular decomposition. Nevertheless, one can estimate the experimentally determined quantities from the theoretical $\Delta H^0_{\mathrm{acid}}$ estimates by assuming that entropy effects are constant throughout the series. For example, O'Hair et al. (1992) estimate the difference $\Delta H^0_{\mathrm{acid}} - \Delta G^0_{\mathrm{acid}}$ to be 6.9 kcal mol^{-1}.

The data of Table 20 indicate that a methyl group does not have a substantial effect on the acidity, but that hydroxymethyl and thiomethyl groups greatly increase the acidity of an amino acid. The effects that alkyl substituents have on the acidity of amino acids is qualitatively parallel to those of related carboxylic acids and alcohols and they can be discussed in terms of dipolar, polarizability and hyperconjugation effects.

6.3. Gas-phase proton affinities

Smaller oligopeptides can be ionized by chemical ionization by fast atom bombardment (FAB) ionization, by electrospray ionization or by matrix-assisted laser desorption ionization

Table 19
Experimental enthalpies of formation ($\Delta H^0_{f,\mathrm{gas}}$/kcal mol^{-1}) of amino acids in the gas phase

Amino acid	$\Delta \mathrm{H}^0_{f,\mathrm{gas}}$	Reference
Glycine	-93.3 ± 1.1	Ngauv et al. (1977)
Alanine	-99.1 ± 1.0	Ngauv et al. (1977)
Proline	-87.52 ± 0.96	Sabbah and Laffitte (1978)

Table 20
Experimental gas-phase acidities (ΔG^0_{acid}/kJ mol^{-1}) of amino acids relative to that of glycine

Amino acid	O'Hair et al. (1992)	Locke and McIver (1983)	Locke (1981)
Glycine	0.0	0.0	0.0
Alanine	−3.5	−1.7	−7.5
Proline	−7.0		−10.5
Valine	−11.0		−11.7
Leucine	−12.0		−10.9
Isoleucine	−13.5		
Lysine	−19.0		
Tryptophan	−21.5		
Phenylalanine	−23.0		
Tyrosine	−23.5		
Methionine	−26.0		−24.3
Cysteine	−38.0		
Serine	−39.0		
Threonine	−41.5		
Arginine	−42.5		
Asparagine	−43.0		
Glutamine	−43.0		
Histidine	−46.0		

and the resulting species analyzed by mass spectrometry (MS) or by tandem MS. Variants of the FAB-MS technique proved to be useful for sequence analysis of oligopeptides, potentially helping the structure elucidation of proteins (Biemann and Martin, 1987).

Addition of a proton to a peptide or protein may occur at several sites. To better understand the complex processes due to the large number of functional groups which can simultaneously interact with a proton, protonation energetics of the amino acids is mandatory.

The techniques of high-pressure mass spectrometry (HPMS; Meot-Ner et al., 1979), ion cyclotron resonance spectrometry (Locke and McIver, 1983), and reaction bracketing (Gorman et al., 1992), as well as kinetic approaches (Bojesen, 1987; Li and Harrison, 1993) have been employed in order to determine the proton affinities (PA) of the 20 α-amino acids. Despite all these efforts, there is still considerable uncertainty in the PA values of most of the amino acids. The different experimental approaches have different weaknesses (Harrison, 1997). For example, the kinetic method provides reasonable PAs for simple α-amino acids but it underestimates the PAs of amino acids containing a second functional group. This is partly caused by intramolecular H-bonding, which stabilizes the protonated species and might even render the kinetic method invalid due to entropic effects. Correspondingly, while Bojesen (1987) reports the relative order of PAs for all 20 amino acids, his data appear to be problematic for aspartic acid, asparagine, glutamic acid and glutamine. It is also well known that the reaction bracketing techniques have uncertainties on the order of 2.5–3 kcal mol^{-1}.

In the gas phase, protonation of nonionic amino acids, except lysine, histidine and arginine, is expected to occur on nitrogen rather than on oxygen:

$$NH_2 - CHR - COOH + H^+ \longrightarrow {}^+NH_3 - CHR - COOH.$$

Table 21
Gas-phase proton affinities (in kcal mol^{-1}) of the α-amino acids[a]

Amino acid	Experimental				Theoretical		
	Hunter and Lias (1997)	Kin[b]	ICR[c]	HPMS[d]	Wright and Borkman (1980)	G2	G2(MP2)
Glycine	211.8	211.0	212.0	209.9	222.30	210.1	210.2
Alanine	215.5	213.6	214.8	213.9	225.78		
Cysteine	215.9	213.8	214.7		219.76		
Serine		215.0	217.2		219.99		
Valine	217.6	215.7	217.4	215.6			
Aspartic acid			217.1				
Leucin	218.6	216.1	218.5	216.2			
Threonine	220.6	216.6	218.8		221.04		
Isoleucine	219.4	216.8	219.3				
Phenylalanine	220.6	217.4	220.5	216.8			
Methionine	223.2	218.2	221.8				
Tyrosine	221.3	217.8	222.7				
Asparagine		218.7	220.2				
Proline	220.0	219.4	223.1	220.1			
Glutamic acid	224.1		216.9				
Tryptophane		220.8	225.8				
Glutamine			218.8				
Lysine			230.7				
Histidine			232.3				
Arginine			> 232				

[a] The order of the proton affinities of α-amino acids corresponds to that reported by Campbell et al. (1994). [b]Kinetic approach (Li and Harrison, 1993). [c]ICR = ion cyclotron resonance (Locke and McIver, 1983). Entropy values estimated at 0.6 cal mol^{-1} K^{-1}. [d]HPMS = high-pressure mass spectrometry (Meot-Ner et al., 1979). Relative to $PA(NH_3)$ = 204.0 kcal mol^{-1}.

The experimental and theoretical proton affinities (PA) of the α-amino acids are collected in Table 21. The order of the amino acids corresponds to the order reported by Campbell et al. (1994). This order is very similar to that deduced by Isa et al. (1990) based on collisionally activated decomposition mass spectra. Data about protonation on other functional groups of amino acids have not been obtained experimentally. However, scattered data from theoretical calculations do exist for glycine and alanine (Zhang et al., 1993).

Simple chemical intuition suggests that there are two possible low-energy minima on the PES of $GlyH^+$, both can be obtained from Gly-Ip while keeping its C_s symmetry: one form has the bifurcated H-bond of Gly-Ip, i.e., the added H atom points away from C=O, while the other is obtained by a 60° rotation of the NH_3^+ unit. According to 6-311 + + G^{**} B3LYP geometry optimizations, the first form is more stable by some 3.6 kcal mol^{-1}. 6-311 + + G^{**} B3LYP calculations on $AlaH^+$ gave basically the same results: the same structure containing a bifurcated H-bond is the most stable form on the PES of $AlaH^+$. Given the reference geometries, different levels of ab initio theory can be employed for the calculation of the gas-phase protonation energies of amino acids. Representative results obtained for the $GlyH^+$–Gly system are collected in Table 22.

Table 22
Computational matrix for the determination of theoretical gas-phase protonation energies, in kcal mol^{-1} of glycine[a]

System	B3LYP	RHF	MP2	MP3	MP4
$GlyH^+$–Gly					
3-21+G^*		220.85	219.77		
6-311++G^{**}	218.40	222.10	219.40	221.39	220.44
cc-pVTZ		223.99	220.69		

[a] The reference geometries have been optimized at the 6-311++G^{**} B3LYP level. All 1*s* core and the corresponding $1s^*$ virtual orbitals have been kept frozen during all correlated-level ab initio calculations.

Proximity of the electron-donating methyl group of α-alanine to the positively charged nitrogen atom, as well as the stabilizing charge-induced dipole polarization interaction, make $AlaH^+$ a relatively more stable protonated species than $GlyH^+$. The difference between the experimental proton affinities of glycine and α-alanine in the gas phase has been determined to be 4.0 kcal mol^{-1} (Meot-Ner et al., 1979) and 3.7 kcal mol^{-1} (Hunter and Lias, 1997), while Locke and McIver (1983) predict a difference of only 2.8 kcal mol^{-1}.

As indicated by the computational results of Wright and Borkman (1980) in Table 21, when the side chain contains electron-withdrawing groups, e.g. in serine, cysteine and threonine, their effect is just the opposite to that of the electron-donating methyl group. The decrease in the proton affinity of the amino acids serine, cysteine, and threonine relative to that of glycine is 1–2 kcal mol^{-1}. The available experimental results do not support this observation, further theoretical and experimental studies are needed to resolve this apparent discrepancy.

Finally we note that the gas-phase basicity of most amino acids is about 8 kcal mol^{-1} lower than their proton affinities (PA) at room temperature, since the eutropy change is likely to arise from the loss of the translational entropy of H^+.

6.4. C^{α}–H and C^{α}–C^{β} bond dissociation energies

Oxidative damage to proteins, initiated in many cases by OH radicals, has been implicated in several pathological disorders (Stadman, 1993). It is of considerable interest to identify which amino acid residues are most susceptible to oxidative damages. The side groups and the C^{α} center of the amino acid residues are the two major points of attack.

Although free neutral amino acids do not exist in solvents and thus have no direct biological significance, they are important as models and for the purposes of reference. Therefore, Rauk et al. (1997) employed isodesmic reactions and ab initio energies obtained at the 6-31G^* DFT(B3LYP) level to obtain bond dissociation energies (BDE) $D_{C^{\alpha}C}$, $D_{C^{\beta}H}$ and $D_{C^{\alpha}C}$ for the neutral amino acids glycine, alanine, serine and threonine, as well as on some model peptide systems. The energies obtained in that study are collected in Table 23.

There does not appear to be any mechanism, biochemical or otherwise, that would lead to C^{α}–C bond dissociation, although these bonds are expected to be weak for the same reason as the C^{α}–H bonds.

Table 23
Bond dissociation energies (in kcal mol^{-1}, at 298 K) of α-amino acids[a]

Compound	$D_{C^{\alpha}H}$	$D_{C^{\beta}H}$	$D_{C^{\alpha}C}$
Glycine	331.0		
Alanine	317.4	427.3	269.0 (292)
Serine	327.2	384.6	261.4
Threonine	327.8	376.8	250.3

[a] From 6-31G* DFT(B3LYP) + 0.95 H ZPVE level calculations and isodesmic reactions (Rauk et al., 1997).

7. Photoelectron spectroscopy of amino acids

The low-resolution outer-valence photoelectron (PES) spectrum of Gly has been observed by Debies and Rabalais (1974) and by Cannington and Ham (1983) using He(I) and He(II) resonance radiation, respectively. The PES spectra of alkyl amino acids has been recorded by Klasnic (1976). HeI and HeII spectra has been recorded for many more amino acids (Cannington and Ham, 1983; Campbell et al., 1994). The valence orbital electron densities of glycine have been measured by electron momentum spectroscopy (EMS) (Neville et al., 1996).

Photoelectron spectroscopy studies found three peaks in the low binding energy region of Gly, at 10.0, 11.1 and 12.1 eV. They are attributed to the nitrogen lone pair orbital, n_N, (the highest occupied molecular orbital, HOMO), to the oxygen lone pair orbital, n_O and to π_{CO}, respectively. As the theoretical data presented in Table 24 show, these peaks belong to valence vertical ionizations $\tilde{X}$, $\tilde{A}$ and $\tilde{B}$, respectively, of Gly-**Ip.** The two experimental studies (Debies and Rabalais, 1974; Cannington and Ham, 1983) made the same assignments for the symmetry of the second and third orbitals (a′ and a″, respectively) but differed on the symmetry of the HOMO, with Debies and Rabalais predicting an out-of-plane (a″) orbitals and Cannington and Ham predicting an in-plane (a′) orbital. Evidently, theoretical calculations support the assignment of Cannington and Ham (1983). For Gly-**IIp** the HOMO electron density is predominantly located on the carbonyl oxygen. This difference in the nature of the HOMO of **Ip** and **IIp** is most likely a result of conformer **IIp** being the only one of the five most stable glycine conformers containing an intramolecular H-bond to the N atom. This stabilizes the nitrogen lone-pair electron density and consequently increases the binding energy of the

Table 24
The first four vertical ionization potentials (in eV) of glycine obtained from EOMIP-CCSD calculations[a]

Symm.	Order	**Ip**			**IIp**	**IIIp**	**IVn**
		PBS	TZ2P/TZP	TZ2P	TZ2P	TZ2P	TZ2P/TZP
A'	$\tilde{X}$	9.92	9.88	9.85	9.90	9.58	9.58
	$\tilde{A}$	10.87	10.91	10.89	11.36	11.41	10.80
A''	$\tilde{B}$	12.04	12.06	12.06	11.31	11.92	12.02
	$\tilde{C}$	13.54	13.55	13.56	13.83	13.71	13.91

[a] The reference geometries employed have been optimized at the 6-311++G** B3LYP level.

electrons in this orbital. Concurrently, the carbonyl oxygen nonbonding orbital of **IIp** is destabilized by the unfavorable anti orientation of the –COOH group. It must be stressed that data in Table 24 suggest that the low-energy conformers of glycine have vertical ionization potentials sufficiently different for allowing their detection at medium or high resolution.

In their pioneering study, Debies and Rabalais (1974), using photoelectron spectroscopy, have determined the first vertical ionization potential of Ala to be 9.8 eV. As Campbell et al. (1994) have also pointed out, the photoelectron spectra of amino acids with complex side chains can be difficult to interpret since the peaks for ionization from the amine nitrogen lone pair orbital and the side-chain orbitals tend to overlap. For example, the spectrum of Phe (Campbell et al., 1994) clearly displays this problem as the nitrogen lone pair and aromatic phenyl π orbitals have virtually the same ionization energies. Nevertheless, the first three vertical(adiabatic) ionization energies of Phe could be determined as 8.9(8.5), 9.3(8.8) and 9.7(9.2) eV. Accurate theoretical calculations, similar to the ones reported in Table 24, should prove indispensable in unraveling the photoelectron spectra of complex amino acids.

As far as ab initio calculations are considered, the first ionization energies of the amino acids glycine, alanine, serine, cysteine and threonine have been determined by Wright and Borkman (1980) employing both Koopman's theorem and the $E(\mathrm{A}) - E(\mathrm{A}^+)$ energy difference technique. It is also noteworthy that the G2(MP2) results of Yu et al. (1995) predicted the ionization potential of Gly to be only 9.27 eV.

As pointed out by Campbell et al. (1994), an almost linear correlation exists between proton affinities and nitrogen lone electron pair ionization energies of those amino acids which protonate on the amino group. The correlation is observed for tyrosine and cysteine, where n_N is not the HOMO. Upon protonation, Lys, Met and Trp form internal H-bonds, which enhance their proton affinities (see Table 21) thus causing deviation from linearity. Nevertheless, such a linear relationship should prove useful in obtaining estimates for proton affinities of peptides, for which direct studies are not amenable due to the low volatility of the samples (Robin, 1975).

8. Structures of model peptides

Structures of peptides and proteins, their folding and motion is most often described using a simplified rigid model with fixed bond lengths and bond angles. In this model three torsion angles (ϕ, ψ and ω) per amino acid residue are employed to describe the main chain folding (see Fig. 1). Knowledge of the first two variables is often sufficient, since ω, describing the rotation of the amide bond, is typically 180° (seldomly 0°). The resulting two-dimensional potential energy surface, $E = E(\phi, \psi)$, is the so-called Ramachandran map (see Figure 2). It is used for the visualization of the energy cost when changing the relative orientation of two adjacent amide bonds.

In studies aimed at understanding the main chain fold of a protein in term of its backbone structure, certain 'patterns' have been recognized and were classified as typical or 'ordered' structures. The rest of the main chain fold is called untypical secondary structural elements. The misleading term of 'random conformation' was also introduced for these residual structural units. The term 'ordered conformers' is further divided into 'periodic' and 'aperiodic'

backbone conformers. The most common type of the former is the α-helix, the β-pleated sheet and the poly-proline II (Liquori, 1969). All periodic conformers should have rather similar consecutive (ϕ, ψ) values. In a right-handed α-helical structure the values of ϕ and ψ are around -54° and -45°, respectively. On the other hand, in a left-handed α-helix (seldom observed in proteins) the same variables should have angles $+54$ and $+45^\circ$, respectively. Additional helix-like backbone structures have also been found in proteins, such as the 3_{10}-helix ($\phi \approx -60^\circ$ and $\psi \approx -30^\circ$) and the poly-proline II (or collagen helix) ($\phi \approx -60^\circ$ and $\psi \approx +120^\circ$). Beside these helices, two forms of extended backbone conformers make up a significant amount of the skeleton: the parallel and the antiparallel β-pleated sheets. These conformers are again made from periodic subunits, such as $\phi \approx -150^\circ$ and $\psi \approx +150^\circ$ or $\phi \approx -130^\circ$ and $\psi \approx +130^\circ$ (Karle, 1979, 1981; Karle et al., 1983; Kaiser and Kézdy, 1984; Fesik et al., 1991; Weber et al., 1991). All these secondary structures are composed from the repetition of the same diamide subunit and therefore are called 'homo-conformers'. On the other hand, most of the hairpin conformers (also known as β-turns, β-loops, etc.) are composed of two different diamide conformers. Therefore, these secondary structural elements are 'ordered' but 'aperiodic' conformers. Consequently, an unambiguous conformational description of β-turns requires the definition of four torsional angle values; e.g. for type I β-turn: $\phi_i \approx -60^\circ$, $\psi_i \approx -30^\circ$, $\phi_{i+1} \approx -90^\circ$ and $\psi_{i+1} \approx 0^\circ$. These quartets describe the relative orientation of three consecutive amide groups (Venkatachalam, 1968) (see Table 25).

Table 25
Approximate backbone torsional values of homo- and hetero-conformers in proteins

Homo-conformers	ϕ_i [a]	ψ_i		
right-handed 3_{10}-helix	−60	−30		
right-handed α-helix	−54	−45		
right-handed π-helix[b]	−45	−54		
left-handed 3_{10}-helix[b]	60	30		
left-handed α-helix	54	45		
left-handed π-helix[b]	45	54		
poly-proline II	−60	120		
parallel β-sheets	−130	130		
antiparallel β-sheets	−150	150		
Hetero-conformers	ϕ_i	ψ_i	ϕ_{i+1}	ψ_{i+1}
β-turn type I	−60	−30	−90	0
β-turn type I′	60	30	90	0
β-turn type II	−60	120	80	0
β-turn type II′	60	−120	−80	0
β-turn type III	−60	−30	−60	−30
β-turn type III′	60	30	60	30
β-turn type VIa	−60	120	−90	0
β-turn type VIb	−120	120	−60	0
β-turn type VIII	−60	−30	−120	120

[a] Numbers are approximate values found in protein structures determined by X-ray diffraction. [b]Expected but not yet observed.

Among the $3n-6$ internal coordinates needed to describe the structure and motion of an n-atomic molecule, there are typically $n-3$ dihedral angles. Since macromolecules, like peptides, contain a large number of atoms, $n-3$ is too large for systematic conformational searches. To understand at least the main chain conformational properties of a peptide or protein fragment, the identification of stationary points on its energy surface, $E = E(x)$ ($x = (\phi_1, \psi_1, \phi_2, \psi_2, \ldots, \phi_k, \psi_k)$, where k is the number of amino acid residues), is required. The more minima on this dihedral energy surface are known, the better our chance is to have a valuable conformational description of the target molecule. For example, all minima of the HCO–L–Ala–L–Ala–NH_2 triamide could be characterized on the appropriate $E = E(\phi_1, \psi_1, \phi_2, \psi_2)$ 4-D hypersurface.

Characterization of 'Ramachandran-type' potential energy surfaces (PES) or $E = E(\phi, \psi)$ maps, forms the foundation of most systematic computational approaches on peptides and proteins (Sasisekharan, 1962; Ramachandran et al., 1963; Ramachandran et al., 1966; Ramachandran and Sasisekharan, 1968). According to this model, the conformational properties of a peptide, made up from k residues, are based on the knowledge of k number of 2-D surfaces. This strategy is straightforward but obviously ignores: (a) the structure-modifying

A

ψ \ φ	60°	180°	300°
300°	g^+g^-	ag^-	g^-g^-
180°	g^+a	aa	g^-a
60°	g^+g^+	ag^+	g^-g^+

The 9 legitimate conformers predicted by MDCA

B

ψ \ φ	60°	180°	300°
300°	γ_D	δ_D	α_L
180°	ε_D	β_L	ε_L
60°	α_D	δ_L	γ_L

MDCA predicted 9 legitimate conformers using the conformational symbolism suggested for peptide building units

C

ψ \ φ	60°	180°	300°
300°	γ_D	δ_D	$[\alpha_L]$
180°	ε_D	β_L	{.}
60°	α_D	δ_L	γ_L

Optimized *ab initio* minima of HCO–Xxx–NH_2 (Xxx = Gly, Ala, Val, and Phe). The α_L is put in bracket since it may vanish. The {} stands for the ε_L minima not yet found in P–CONH–CHR–NHCO–Q systems.

Scheme 4.

effects of nearest neighbors and (b) those interactions that originate from far-lying molecular fragments. Nevertheless, this model proved to be a useful starting point of all ab initio computational approaches.

8.1. Notation

Two adjacent amide bonds incorporating a chiral center could have a maximum of $6\times6 = 36$ stable orientations. However, topology analysis of the full Ramachandran map $E = E(\phi, \psi)$ resulted in a maximum number of only nine minima. This is understandable, since the torsional potential along both ϕ and ψ have three minima (g^+, a and g^-). Therefore, there must be nine *legitimate* conformers denoted by $g^+g^+;(+60°, +60°)$, $ag^+;(+180°, +60°)$, $g^-g^+;(+300°, +60°),\ldots, g^-g^-;(+300°, +300°)$ according to the IUPAC-IUB convention (see Scheme 4).

Several research groups attempted to divide the Ramachandran surface into subconformational regions and denoted the resulting catchment regions with different symbols. The works of Richardson and Richardson (1989) and Rooman et al. (1992) suggested a useful strategy to cluster different backbone conformers. Zimmerman et al. (1977), as well as Thornton et al. (1995), performed several studies aimed at assigning and labeling peptide conformations. In particular, Zimmerman et al. (1977) suggested 16 catchment regions (A, B, C, D, E, F, G, H and A^*, B^*, C^*, D^*, E^*, F^*, G^*, H^*). In the approach used by Karplus (1996), the Ramachandran surface has 12 distinct regions, labelled as α_L, α_R, β_S, β_P, γ, γ', δ_L, δ_R, ε, ε', ε'' and ζ. The suggested subdivision of the $E = E(\phi, \psi)$ surface by Karplus is similar to the one proposed by Efimov (1993). All these approaches depend on X-ray data. Perczel et al. (1991) introduced the following set of abbreviations for labelling the backbone minima of peptides and proteins: α_L, α_D, β_L, γ_L, γ_D, δ_L, δ_D, ε_L and ε_D (see Fig. 8). Although this labelling of the backbone conformers is optional, it is consistent, unambiguous and is based entirely on peptide topology. Furthermore, this notation incorporates as much as possible from the 'jargon' of peptide and protein chemistry. Ideally, all nomenclature pinpoints to centers of conformational clusters and therefore able to characterize different backbone orientations. An important difference between a peptide-topology-based clustering and one that derives locations from a frequency analysis of observed peptide conformers in proteins is that the former will not change with the improvement of the database. Although all notations have their advantages, in this review, where peptide topology is in focus, the topology-based notation (Perczel et al., 1991) will be employed.

It must be stressed that the maximum number of backbone conformers expected in peptides for each residue is the important feature, which is nine (Perczel et al., 1991) and not the way that they are labeled. Preliminary ab initio calculations carried out on selected monopeptide models resulted in seven and not nine minima, the missing ones being α_L and ε_L. On the other hand, both α_L and ε_L minima have been found in HCO–L-Ala–L-Ala–NH_2 (Perczel et al., 1993; Perczel and Csizmadia, 1995). Furthermore, all nine minima have been found in proteins with known X-ray structures (Perczel and Csizmadia, 1995).

Rotation about two bonds connected to a prochiral center produces nine conformers from which eight occurs as four doubly degenerate pairs (α, γ, δ and ε). This is the case for the achiral *N*-formyl-glycine-amide (cf. Scheme 5).

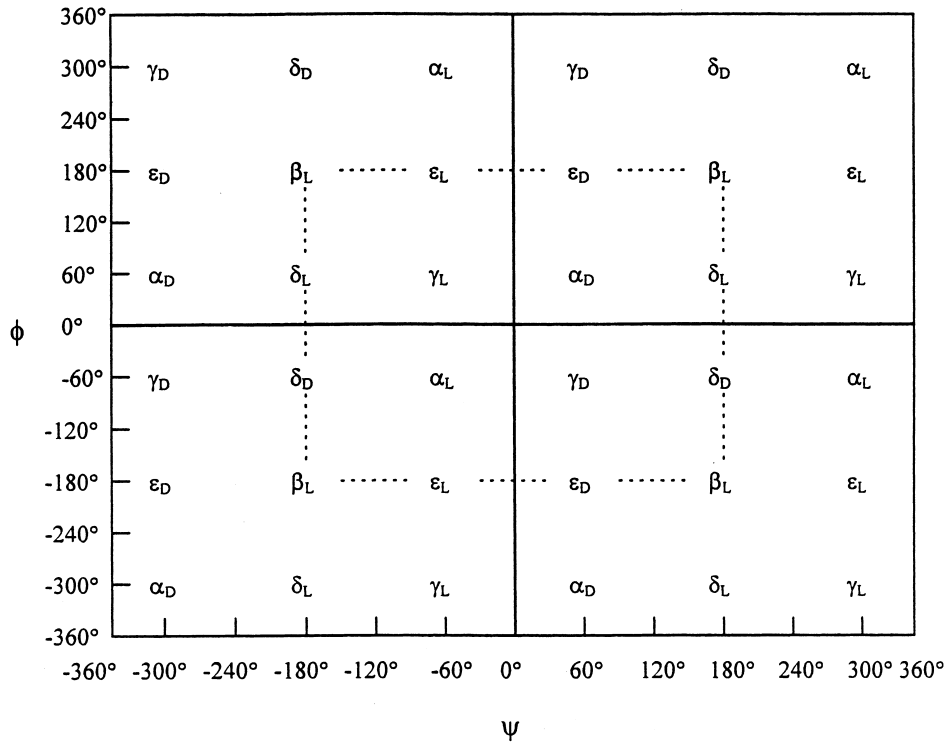
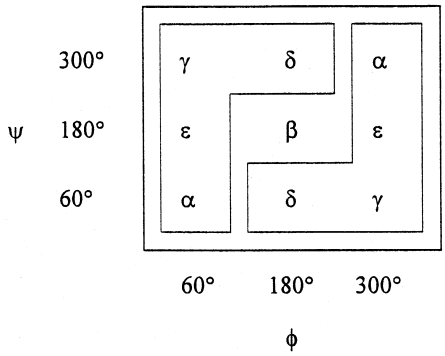


Fig. 8. The idealized potential energy surface of a single amino acid diamide. Location of the MDCA-predicted minima are specified by their terms using subscripted Greek letters.

For chiral amino acids (19 out of the 20 natural amino acids, see Table 1) this double degeneracy disappears. In the case of L-amino acids the four conformers at the right-hand side will be favored (i.e. lowered in energy) with respect to their counterparts at the left side. Therefore, minima with subscript L have lower energy than their D counterparts and appear with a higher probability. When protein X-ray data are analyzed, all the frequently assigned protein secondary structural elements are to be denoted with symbols having subscript L. For example, the common right-handed α-helix is $(\alpha_L)_n$, while the β-sheet is $(\beta_L)_n$. This right-hand side of the Ramachandran map may be referred to as the L-valley. Accordingly, the left-hand side section may be referred to as the D-valley (cf. Scheme 6). (Of course, the opposite is true for a D-amino acid diamide, where the D-valley is the favored one).

The notation just introduced is useful in several respect: it reflects the rotational topology of the peptide units, adheres to traditional convention (α-structures: $(\alpha_L)_n$, β-structures: $(\beta_L)_n$, inverse γ-turns = $:\gamma_L$) and reflects the 'symmetry' of the Ramachandran surface and the relative energetics of the appropriate pairs of conformers (i.e. L amino acid residues favor the L-valley). Furthermore, this convention provides additional features for fine tuning of a structure: e.g. $((\alpha_L)_n)3_{10}$ or $[(\alpha_L)_n]^{\pi}$ could stand for 3_{10} or π-helices, obviously members of the α-helix 'family'. The only slight problem with this notation is when we specify right-handed and left-handed helical conformations: $(\alpha\text{-helix})_{right} = (\alpha_L)_n$ and $(\alpha\text{-helix})_{left} = (\alpha_D)_n$. This could lead to some misunderstanding, but it should be mentioned that the official notation of IUPAC-IUB for the

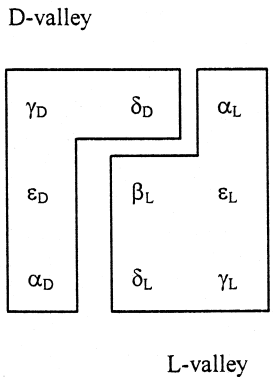


Scheme 5.

two forms of helical structures is *P*- or *M*- and not *R*- or L-, like sometimes used to describe right-handed and left-handed helices. It is hoped, however, that the practical utility and uniformity of this notation overrides this marginal problem.

When specifying numerical values for ϕ and ψ, we use the IUPAC-IUB convention (IUPAC-IUB, 1970), i.e. $-180^\circ \leq \phi \leq 180^\circ$ and $-180^\circ \leq \psi \leq 180^\circ$. Occasionally, however, we also employ the traditional cut ($0^\circ \leq \phi \leq 360^\circ$ and $0^\circ \leq \psi \leq 360^\circ$) suggested by Ramachandran and Sasisekharan (1968). These two cuts are clearly shown in Fig. 8, where the central broken square is the IUPAC-IUB convention and the four equivalent quadrants are that of the traditional cut.

Gly has only main-chain conformational properties. In Ala all three H^β protons are equivalent; therefore, all three side-chain orientations about χ_2($g-$, *a* and $g+$) are expected to be degenerate. The constitution of all the other naturally occurring amino acid residues, except proline, could be derived from Ala with the replacement of one or two H^β(s) by one or two functional group(s). In 14 residues only a single H^β is replaced. Only in three out of the 20 natural amino acids are two H^β(s) substituted: Val, Ile and Thr. In all the other 17 amino acids distinct χ_1 rotamers are expected. In what follows, all side-chain conformers are denoted using the combination of $g+ :=$ *gauche* +, $a :=$ *anti* and $g- :=$ *gauche* −. The conformational



Scheme 6.

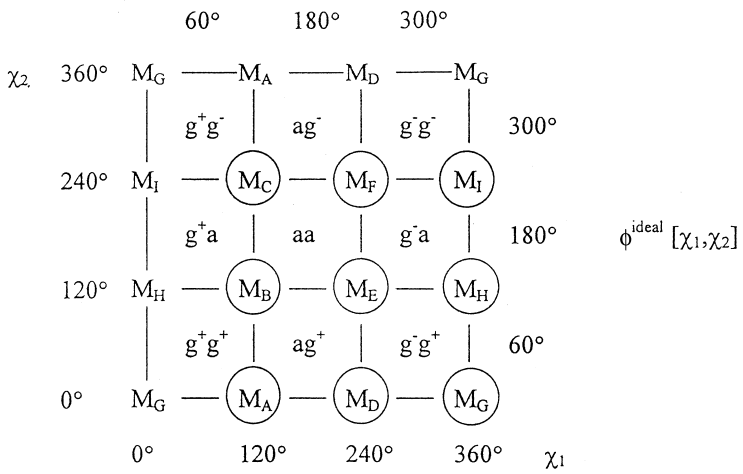
consequences of χ_2 rotation (with the additional side-chain rotations χ_3, χ_4, etc.) are complex. The side chains (R'–$C^{\beta}H_2$–) of 17 amino acids are expected to have a maximum number of 9 different conformers when analyzing the $E(\chi_1, \chi_2)$ surfaces. Three stable orientations are expected about the rotation of the C^{α}–C^{β} covalent bond (χ_1 torsional angle) where each and every minimum is coupled with an additional three minima arising from the three possible stable orientations associated with the rotation about the C^{β}–R'^{γ} (χ_2 torsional angle) bonds. In an idealized case, the locations of the above nine minima on a side-chain surface are presumed to be: $g+\ g+$;(60°, 60°), $g+\ a$;(60°, 180°),... , $g-\ g-$(−60°, −60°). On the same surface, where an idealized topological behavior is expected, nine unique maxima (M_A;(120°, 0°), M_B;(120°, 120°), M_C;(120°, 240°),... , M_H;(360°, 120°), M_I;(360°, 240°)) should be present. The topology of such an (ϕ^{ideal} (χ_1, χ_2)) is shown in Scheme 7. The nine unique maxima are circled, the uncircled maxima are due to topological repetitions.

Depending on the conformational type of the backbone, annihilation of some side-chain conformers may well be expected due to specific interactions between side-chain/side-chain or side-chain/backbone atoms.

8.2. Peptide conformers (P–CO–Xxx–NH–Q)

The 'diamide approximation' has been widely accepted for modeling selected features of peptides and proteins (Sasisekharan, 1962; Ramachandran et al., 1963; Leach et al., 1966; Gibson and Scheraga, 1966; Ramachandran and Sasisekharan, 1968; Ponnuswamy and Sasisekharan, 1971; Lewis et al., 1973; Pullman and Pullman, 1974).

Structural results from diffraction experiments indicate that amino acid diamides often adopt backbone values close to the β_L, γ_L and δ_L regions of the Ramachandran map. These three minima are close to each other on the (ϕ, ψ) surface and have similar relative energies. Based on ab initio calculations of alanine diamide, the β_L and δ_L minima are shifted on the PES away from their ideal locations toward the γ_L location (Head-Gordon et al., 1991; Böhm and Brode, 1991; McAllister et al., 1993; Perczel and Csizmadia, 1995 (see Table 27 vide infra)). Due to



Scheme 7.

this shift all three minima are located close to each other in a common catchment region, also referred to in protein chemistry as the β-region. This area is quoted during topology analyses as the 'grand canyon' of the Ramachandran surface.

Neither NMR nor CD (or IR) spectroscopy resulted in structural data suggesting a unique backbone conformation for peptides. On the contrary, a conformational mixture is typically present in any solvent. If limited for the β-region, this spectroscopic observation agrees with the above mentioned topology and energetic considerations: low energy barriers calculated between selected conformers could explain why conformational mixtures are always present in solution. However, this can only be generalized if all minima could be incorporated in the description. In what follows, we will trace and compare the conformational ensemble of selected amino acid diamides.

8.2.1. HCO–Gly–NH_2

As many as eight minima exist on the PES of neutral Gly (Császár, 1992). Achiral HCO–Gly–NH_2 has a symmetric Ramachandran surface, so that $\alpha_L = \alpha_D$, $\delta_L = \delta_D$ and $\varepsilon_L = \varepsilon_D$, resulting in five distinct structures (Head-Gordon et al., 1991; Perczel et al., 1991). Only three of the five conformers could be located by ab initio calculations (see Table 26).

Computations for HCO–Gly–NH_2 and CH_3CO–Gly–$NHCH_3$ were carried out by several research groups (Scarsdale et al., 1983; Schäfer et al., 1984; Wiener et al., 1984a,b; Head-Gordon et al., 1989; Böhm and Brode, 1991; Frey et al., 1992; Ramek et al., 1991; McAllister et al., 1993; Endrédi et al., 1997). Saddle points were also determined at the 3-21G and 6-31+G* RHF levels by Head-Gordon et al. (1991) for the simplest glycine diamide model compound.

8.2.2. HCO–L-Ala–NH_2

Ala, the smallest chiral amino acid, has 13 conformers in the gas (Cao et al., 1995; Császár, 1996). Its diamide, when studied at the 3-21G RHF level, has only little more than half of the backbone conformers than the amino acid. The alanine diamide model is probably the most intensively investigated system (Scarsdale et al., 1983; Schäfer et al., 1984; Weiner et al., 1984a,b; Head-Gordon et al., 1989, 1991; Böhm and Brode, 1991; Ramek et al., 1991; Perczel et al., 1991; Frey et al., 1992; McAllister et al., 1993; Endrédi et al., 1997). Selected conformational parameters of HCO–L-Ala–NH_2 with relative energies, calculated at four different levels of theory, are given in Table 27.

Table 26
Selected conformational parameters of HCO–Gly–NH_2 with relative energies obtained at the 3-21G(6-31+G*) RHF levels of theory

conformer	ϕ	ψ	ΔE
β_L	−180(−180)	180(180)	0.65(−0.58)
$\gamma_L = \gamma_D$	−83.3(−85.2)	64.7(67.4)	0.00(0.00)
$\delta_L = \delta_D$	−121.9(not found)	25.2	3.27

[a] The dihedral angles ϕ and ψ are in degrees, ΔE is in kcal mol^{-1}.

Table 27
Selected conformational parameters of HCO–L-Ala–NH_2 with relative energies, calculated at four different levels of theory

bb[a] conf RHF/3-21G[b], RHF/6-31 + G*[c], RHF/6-311 + + G**[d], MP2/6-311 + + G**[e]	ϕ	ψ	ΔE
γ_L	−84.5	+ 67.3	0.00
	−85.8	+ 78.1	0.00
	−86.2	+ 78.8	0.00
	−82.8	+ 80.6	0.00
β_L	−168.3	+ 170.6	+ 1.25
	−155.6	+ 160.2	+ 0.19
	−155.1	+ 161.0	+ 0.11
	−157.1	+ 163.2	+ 1.21
δ_L	−128.1	+ 29.8	+ 3.83
	−110.4	+ 12.0	+ 2.24
	−112.8	+ 13.2	+ 2.22
	not found		
α_L	not found at all		
ε_L	not found at all		
α_D	+ 63.8	+ 32.7	+ 5.95
	+ 69.5	+ 24.9	+ 4.73
	+ 69.0	+ 26.9	+ 4.56
	+ 63.1	+ 35.5	+ 3.88
γ_D	+ 74.0	−57.4	+ 2.53
	+ 75.1	−54.1	+ 2.56
	+ 75.3	−55.4	+ 2.54
	+ 74.4	−49.1	+ 2.19
δ_D	−178.6	−44.1	+ 7.31
	−165.6	−40.7	+ 5.52
	−165.2	−42.1	+ 5.39
	−166.0	−39.9	+ 5.45
ε_D	+ 67.2	−171.9	+ 8.16
	not found		
	not found		
	not found		

[a] bb: backbone conformer type (ϕ and ψ are in degree, ΔE in kcal mol^{-1}). [b]From Endrédi et al. (1997). [c]From Head-Gordon et al. (1991). [d]From Endrédi et al. (1997). [e]From Endrédi et al. (1997).

Upon comparing the location and the relative (ϕ, ψ) shifts as a function of the applied level of theory, it becomes evident that each minimum remains in its own catchment region. At the 6-311++G** MP2 level two minima vanish: δ_L and ε_D. The latter (ε_D) is predicted (MDCA) to be located close to ($\phi = 60°$, $\psi = 180°$). Indeed, using the 3-21G RHF approach Head-Gordon et al. (1991), as well as Perczel et al. (1991), determined the actual location at ($\phi = 67.5°$, $\psi = -177.3°$) and ($\phi = 67.6°$, $\psi = -178.1°$), respectively. Upon investigation of this region of the Ramachandran surface (see Fig. 2), ε_D is found in a basin-like, extremely shallow valley. This property of the surface helps to explain the minor conformational differences between the calculated structures for ε_D using the same 3-21G RHF approach. The local environment of δ_L

has a rather similar shape to ε_D. Due to low-energy barriers, both the δ_L and ε_D minima are 'fragile'. In fact, neither of these two conformers could be located as stationary points on the 6-311++G** MP2 surface (see Table 27). The following conclusions can be drawn upon relative energies (3-21G RHF and 6-311++G** MP2) about the 5 minima, α_D, β_L, γ_L, γ_D and δ_D, relative to γ_L: (a) the relative order of the minima is conserved ($E_{\gamma L} < E_{\beta L} < E_{\gamma D} < E_{\alpha D} < E_{\delta D}$), (b) the relative energy differences do change significantly and (c) the larger the relative difference at the 3-21G RHF level, the more the appropriate 6-311++G** MP2 difference decreases.

On the basis of these observations the 6-311++G** MP2 Ramachandran surface would perhaps look like the one in Fig. 2, with the following modifications: (a) the energy axis is reduced by some 30% and (b) the regions of δ_L and ε_D could be a mountain-side on the 6-311++G** MP2 surface, like α_L and ε_L, which are located on slopes of the 3-21G RHF Ramachandran map (Fig. 2).

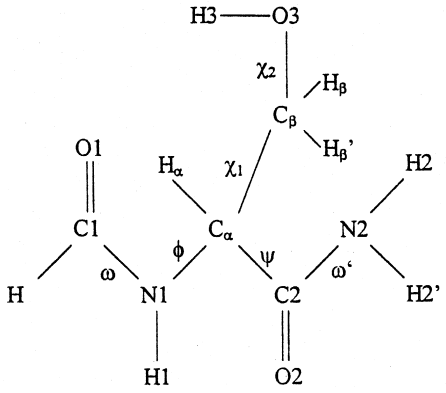
Saddle points on the Ramachandran surface of alanine diamide have been determined by Head-Gordon et al. (1991) at the 3-21G and 6-31+G* RHF levels and could be used to estimate the energy requirements of conformational interconversion.

8.2.3. HCO–L-Val–NH$_2$

Val is a typical representative of those three amino acids (Val, Ile and Thr) which contain only a single H^β. If Val is derived from Ala, two H^βs have to be replaced by two methyl groups. Both methyl groups of the isopropyl side chain will adopt a $g+$ orientation. Therefore, investigation of the minima on the $E(\phi, \psi, \chi_1)$ PES could provide a rather complete conformational picture of this model compound. Out of the 27 expected conformers (three χ_1 orientations for each backbone conformer), 20 were determined at the 3-21G RHF level (Viviani et al., 1993). With the side chain having a $g+$ or $g-$ orientation, all seven backbone conformers (α_D, β_L, γ_L, γ_D, δ_L, δ_D and ε_D) found for HCO–L-Ala–NH$_2$ could be located for HCO–L-Val–NH$_2$. However, if χ_1 has an *anti* orientation, then not only the α_L and ε_L conformers but also the δ_L backbone orientation vanishes. This suggests that an apolar side chain can interact with the peptide backbone to such an extent that an otherwise stable stationary point, δ_L, can disappear.

8.2.4. HCO–L-Ser–NH$_2$

The catalytic mechanism of aspartylglucosaminidase, which was found to be similar to the well-known mechanism of serine proteases, has been the subject of ab initio analysis (Perakyla and Rouvinen, 1996; Perakyla and Kollman, 1997). Quantum chemical studies have also been performed for the catalytic triad in the heart of serine proteases. Such investigations relied on conformational properties of simpler models, like certain conformers of HCO–L-Ser–NH$_2$ (Scarsdale et al., 1983; Siam et al., 1987; Perczel et al., 1990). In these studies the side chain was mostly kept in its $+$*syn-clinal* ($g+$, $g+$) conformation. The restriction of side-chain motion was based on the hypothesis that perturbation originating from the side-chain will not affect the topology of the (ϕ, ψ) surface. This concept was not confirmed by NMR, X-ray and computational studies (Karle, 1979; Aubry et al., 1984; Perczel et al., 1990). The strong coupling between side-chain and backbone conformation initiated additional studies (Farkas et al., 1995; Perczel et al., 1996a,b, 1997a; Jákli et al., 1998a,b).



Scheme 8.

In the case of HCO–L-Ser–NH_2 (Scheme 8), the $E = E(\phi, \psi, \chi_1, \chi_2)$ potential energy hypersurface is expected to have $3^4 = 81$ stationary points (see Fig. 3). Only 44 structures have been located at the 3-21G RHF level (Perczel et al., 1996a,b). Upon reoptimization (see Fig. 9) using a larger 6-311++G^{**} basis set (Jákli et al., 1998a,b), a total of eight conformational migrations are observed: $\beta_L(g-, a) \Rightarrow \gamma_L(g-, a)$, $\delta_L(g-, a) \Rightarrow \alpha_L(g-, a)$, $\delta_L(a, g-) \Rightarrow \gamma_L(a, g-)$, $\gamma_L(g, +) \Rightarrow \beta_L(g, +)$, $\beta_D(g-g, -) \Rightarrow \gamma_D(g-g, -)$, $\gamma_D(g, -) \Rightarrow \alpha_D(g, -)$, $\delta_D(g-g, +) \Rightarrow \gamma_D(g-g, +) \Rightarrow \gamma_D(g-, g+) \Rightarrow \gamma_D(g-, g+)$ and $\delta_D(a, a) \Rightarrow \beta_L(a, a)$ (Fig. 10). The observed migrations occurred on the (ϕ, ψ) and not on the (χ_1, χ_2) potential energy surface.

It is noteworthy to compare the dihedral angles of the optimized structures, as presented in Table 28.

The trend between the relative energy differences calculated at different levels of theory plotted as a function of conformational type is shown in Fig. 11.The relative energy differences between calculations at the RHF/6-311++G^{**}//RHF/3-21G and RHF/6-311++G^{**} levels are small, indicative of only minor conformational changes.

8.2.5. HCO–L-Phe–NH_2

Investigation of the electronic spectra of poly(L-phenylalanine) (Matsubara et al., 1997), together with vibrational, CD and VCD spectral analysis of some phenylalanine-based neurotransmitters (Sun and Bernstein, 1996) are interesting examples of the research involving Phe analogs. The role played by Phe43 in the binding of CD4 and gp120 point mutants (Székely et al., 1996), physicochemical properties of Phe-containing coenzymes (Hayashi et al., 1993), stereospecificity in the binding of Phe in the active site of thermolysin (Ghosh and Edholm, 1994), aromatic ring stacking problems (Singh and Thornton, 1992; Ghalem et al., 1994; Yamauchi, 1995) and other equally important biochemical problems (Moews and Kretsinger, 1975; Blow, 1976; Warne and Morgan, 1978; Hamgauer et al., 1984; Singh and Thornton, 1985; Finer-Moore et al., 1989; Sapse et al., 1992; Moebius et al., 1992) provide several reasons to investigate the structural properties of the Phe residue. Clearly, deeper understanding of these topics can be gained by a detailed knowledge of the conformational properties of Phe.

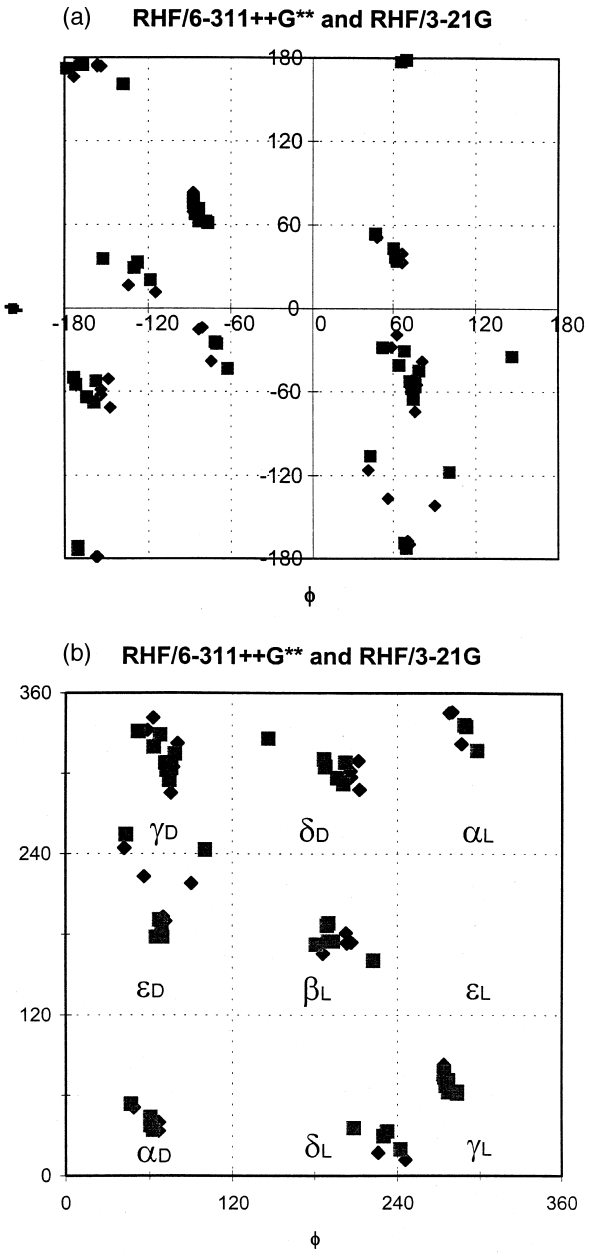


Fig. 9. Location of the (ϕ, ψ) values of HCO–L-Ser–NH_2 determined at two different level of theory (3-21G and 6-311++G^{**} RHF) A. The $-180° \leq \phi \leq 180°$ and $-180° \leq \psi \leq 180°$ representation. B. The traditional $0° \leq \phi \leq 360°$ and $0° \leq \psi \leq 360°$ representation reflecting more to peptide backbone topology.

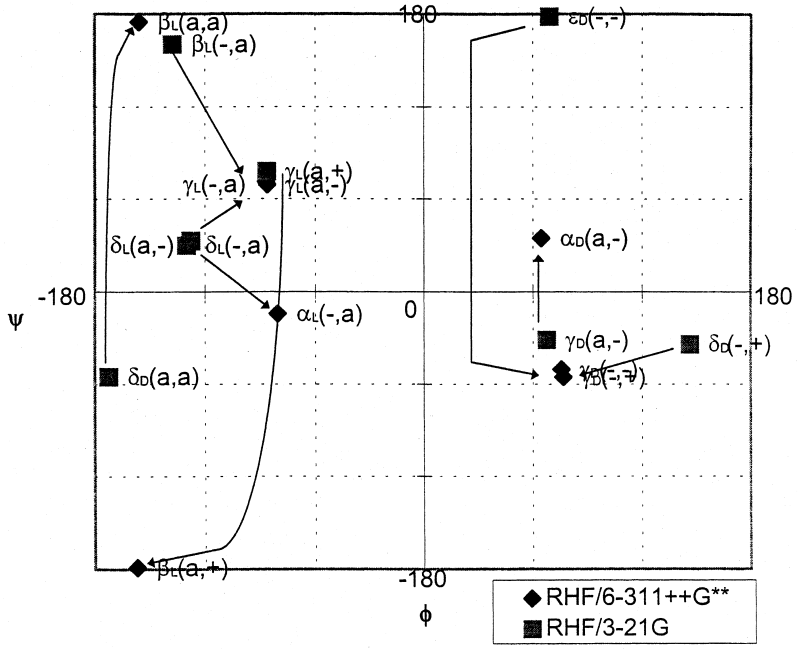


Fig. 10. Location of the (ϕ, ψ) values of the 8 migrated HCO–L-Ser–NH_2 conformers. The 3-21G RHF minima (input structures '■') migrated to another backbone type '◆' when reoptimized at the 6-311++G^{**} RHF level.

Phe is expected to have similar conformational features as the other three naturally occurring aromatic amino acid residues, Tyr, Trp and His. As the simplest model, several ab initio investigations have been completed in recent years on the structure of HCO–L-Phe–NH_2 (Farkas et al., 1996; Perczel et al., 1997b; Jákli et al., 1998a,b). MDCA predicts 27 structures for HCO–L-Phe–NH_2 (Scheme 9): nine backbone conformers each adopting three different χ_1 orientations with a single χ_2 ($\chi_2 \approx g+$). Ab initio structure determinations at the 3-21G RHF level resulted in 19 different conformers for HCO–L-Phe–NH_2 (Perczel et al., 1997b). In Fig. 12 the 3-21G RHF results are compared with those obtained at 6-31+G^* RHF, RHF/3-21G//RHF/6-31+G^* and B3LYP/6-311++G^{**}//RHF/3-21G levels of theory (Jákli et al., 1998a,b).

Three conformers out of the 19 migrated $\{(\beta_L(g-) \Rightarrow \gamma_L(g-),\ \varepsilon_D(g+) \Rightarrow \gamma_D(g+)$ and $\varepsilon_D(g-) \Rightarrow \gamma_D(g-)\}$ when structures were reoptimized at the 6-31+G^* RHF level (see Table 29 and Fig. 13). The conformational building unit of the right-handed helix-like structure (α_L) and that of the poly-proline II (ε_L) are missing on all of the $E = E(\phi, \psi)$ surfaces.

8.2.6. Problematic secondary structure building units

When comparing the topology of the $E = E(\phi, \psi)$ surface of HCO–L-Ala–NH_2 calculated at the 3-21G RHF level with the idealized surface, the most obvious difference is that the expected α_L and ε_L minima are missing. Furthermore, in peptide models (P–CONH–CHR–CONH–Q), such as HCO–Gly–NH_2, HCO–L-Val–NH_2 and HCO–L-Phe–NH_2, the same anomaly has been found. These structures have been optimized at higher levels of theory by using a rigorous grid search, but the results were the same: two out of the nine expected

Table 28
Selected ab initio conformational parameters of For–L-Ser–NH_2 obtained at the 6-311++G^{**} RHF level[a]

conf.[b]	ϕ	ψ	χ_1	χ_2
$\beta_L(-,+)$	−174.0	165.5	−94.4	57.6
$\beta_L(+,-)$	−153.3	174.0	67.3	−66.3
$\beta_L(+,a)$	−156.7	173.3	70.1	−162.8
$\beta_L(a,a)$	−155.9	174.9	−170.4	162.6
$\beta_L(a,+)$	−157.3	−179.1	−165.7	89.5
$\delta_L(-,-)$	−134.1	17.1	−51.7	−69.5
$\delta_L(+,a)$	−114.4	11.8	53.4	161.5
$\gamma_L(-,-)$	−86.9	73.1	−54.9	−72.0
$\gamma_L(-,a)$	−87.0	71.1	−57.3	−175.9
$\gamma_L(-,+)$	−86.7	82.8	−67.5	61.6
$\gamma_L(a,-)$	−86.5	69.2	−178.3	−71.9
$\gamma_L(+,+)$	−85.4	72.2	54.4	69.8
$\alpha_L(a,a)$	−73.9	−38.0	−172.9	−167.4
$\alpha_L(-,a)$	−80.8	−14.2	−52.6	−172.3
$\alpha_L(-,-)$	−82.6	−14.9	−53.6	−76.4
$\varepsilon_D(-,+)$	56.0	−137.0	−60.3	65.4
$\varepsilon_D(a,a)$	71.3	−170.2	−157.8	179.0
$\varepsilon_D(a,+)$	69.7	−167.1	−161.1	81.0
$\varepsilon_D(+,-)$	89.9	−142.0	84.6	−66.6
$\varepsilon_D(+,a)$	41.8	−115.6	64.5	−171.3
$\gamma_D(-,-)$	75.3	−50.5	−60.6	−80.0
$\gamma_D(-,a)$	77.1	−55.5	−59.8	173.0
$\gamma_D(-,+)$	76.5	−55.4	−57.4	90.6
$\gamma_D(a,a)$	75.3	−74.6	−168.1	−160.5
$\gamma_D(a,+)$	74.2	−61.4	−179.2	59.6
$\gamma_D(+,-)$	80.2	−37.7	83.5	−64.9
$\gamma_D(+,a)$	58.5	−27.7	70.5	173.1
$\gamma_D(+,+)$	62.7	−18.6	46.5	53.8
$\alpha_D(-,a)$	66.4	33.2	−55.6	179.4
$\alpha_D(+,+)$	48.3	50.8	58.4	65.1
$\alpha_D(a,-)$	64.3	34.7	−162.3	−6.,4
$\alpha_D(a,+)$	66.5	39.8	−148.9	78.6
$\delta_D(-,-)$	−148.4	−50.7	−50.6	−37.9
$\delta_D(a,+)$	−147.5	−72.0	173.6	73.4
$\delta_D(+,-)$	−153.9	−62.8	55.3	−86.1
$\delta_D(+,a)$	−154.0	−58.3	59.7	−168.4

[a] Values are from Jákli et al. (1998a,b). [b]The backbone conformers are labelled according to the set of abbreviation introduced in Perczel et al. (1991): α_L, α_D, β_L, γ_L, γ_D, δ_L, δ_D, ε_L and ε_D.

minima vanished. From X-ray crystallography we know that a large amount of proteins have α-helix ($(\alpha_L)_n$) and/or poly-proline II ($(\varepsilon_L)_n$) secondary structural elements. The question arises whether the lack of these two minima, α_L and ε_L, on the ab initio Ramachandran surfaces is a result of the limitations of the 'dipeptide approximation'. If so, this observation suggests that the backbone conformations of polypeptides may not be represented properly by the conformers of diamide models.

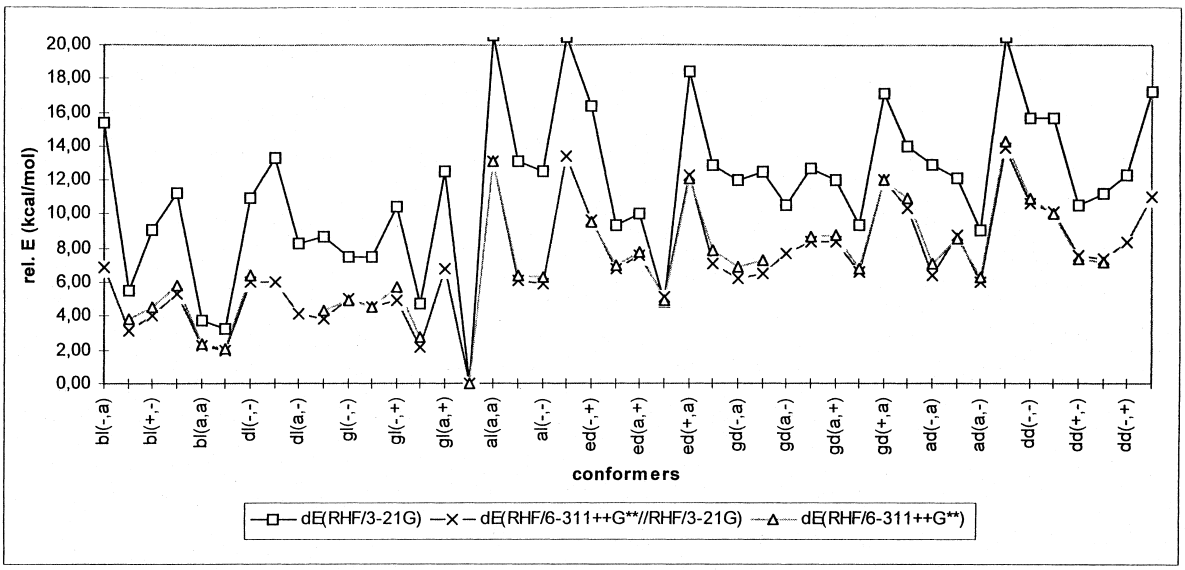


Fig. 11. Comparison of relative energy differences of HCO–L-Ser–NH_2 obtained at the 3-21G RHF, 6-311++G^{**} RHF and RHF/6-311++G^{**}//RHF/3-21G levels of theory.

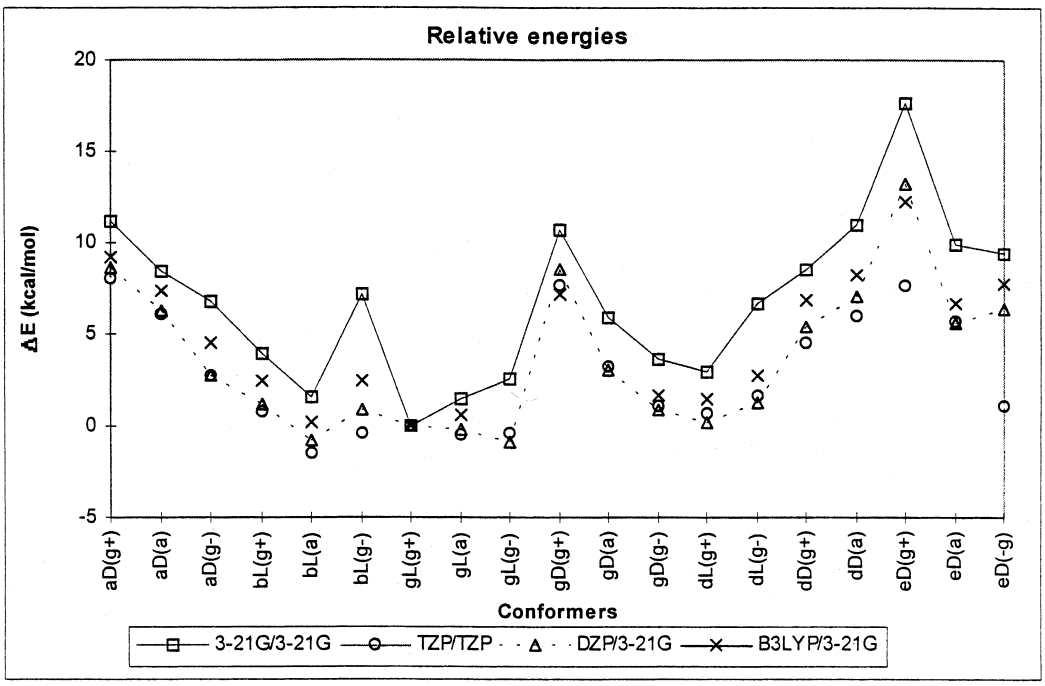
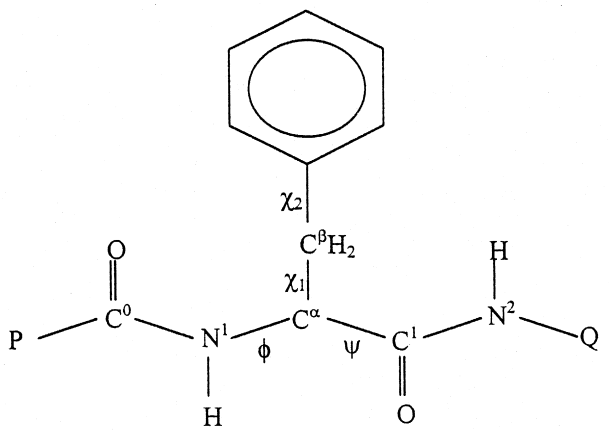


Fig. 12. Comparison of relative energy differences obtained at the 3-21G RHF, 6-31 + G^{*} RHF, RHF/6-31+G^{*}// RHF/3-21G and B3LYP/6-311++G^{**}//RHF/3-21G levels of theory.



Scheme 9.

Table 29
Selected ab initio conformational parameters of HCO–L-Phe–NH_2 obtained at the 3-21G (6-31+G*) RHF levels of theory[a]

Conformer	ϕ	ψ	χ_1	χ_2	ΔE
α_D(g+)	50.1 (43.1)	41.3 (50.3)	52.1 (44.6)	82.1 (78.1)	8.02 (11.16)
α_D(a)	66.6 (60.6)	37.9 (41.1)	−139.6 (−144.6)	93.5 (101.9)	6.07 (8.39)
α_D(g−)	69.6 (68.6)	27.0 (28.2)	−57.0 (−56.7)	105.1 (98.9)	2.76 (6.75)
β_L(g+)	−156.1 (−175.9)	168.1 (174.7)	57.3 (53.4)	86.6 (84.4)	0.74 (3.90)
β_L(a)	−156.5 (−166.6)	152.2 (173.9)	−167.9 (−146.6)	71.2 (68.5)	−1.44 (1.54)
β_L(g−) ⇒ γ_L(g−)[c]	−86.9 (−130.3)	79.4 (160.9)	−59.1 (−61.8)	114.3 (91.5)	−0.43 (7.11)
γ_L(g+)	−85.4 (−83.0)	52.5 (59.8)	44.9 (40.7)	80.1 (76.3)	0.00 (0.00)
γ_L(a)	−85.1 (−85.6)	89.2 (72.5)	−168.4 (−162.0)	84.7 (94.5)	−0.49 (1.51)
γ_L(g−)	−86.9 (−84.4)	79.4 (66.8)	−59.1 (−55.1)	114.3 (111.5)	−0.43 (2.59)
γ_D(g+)	55.2(34.9)	−18.1(−8.6)	68.7(68.7)	81.3(84.8)	7.68(10.72)
γ_D(a)	74.7(74.5)	−72.2(−65.2)	−167.7(−172.6)	80.7(88.9)	3.22(5.87)
γ_D(g−)	75.9(76.3)	−54.5(−57.6)	−59.9(−57.2)	105.4(100.3)	1.04(3.66)
δ_L(g+)	−123.3(−123.1)	15.7(25.6)	54.4(52.4)	82.1(83.1)	0.68 (2.98)
δ_L(g−)	−83.3 (−119.6)	−15.7 (23.2)	−59.2 (−60.0)	113.7 (102.3)	1.63 (6.66)
δ_D(g+)	−164.3(−173.9)	−32.0(−31.1)	62.2(58.0)	95.2(89.7)	4.56(8.54)
δ_D(a)	−154.8(−169.9)	−57.3(−49.6)	−174.5(−149.4)	74.8(76.2)	5.94(10.96)
ε_D(g+) ⇒ γ_D(g+)	55.2 (34.0)	−18.1 (−134.3)	68.7 (71.6)	81.3 (175.7)	7.68 (17.61)
ε_D(a)	68.5 (67.6)	−169.2 (−177.6)	−153.8 (−158.3)	59.1 (64.6)	5.64 (9.89)
ε_D(g−) ⇒ γ_D(g−)	75.9 (71.6)	−54.5 (174.4)	−59.9 (−58.4)	105.4 (97.7)	1.04 (9.46)

[a] Values are taken from Jákli et al. (1998a,b). [b]The backbone conformers are labelled according to the set of abbreviation introduced in Perczel et al. (1991): α_L, α_D, β_L, γ_L, γ_D, δ_L, δ_D, ε_L and ε_D. [c]The observed conformational migrations at RHF/6-31+G* level of theory are noted such as: input 3-21G RHF structure ⇒ output 6-31+G* RHF conformer.

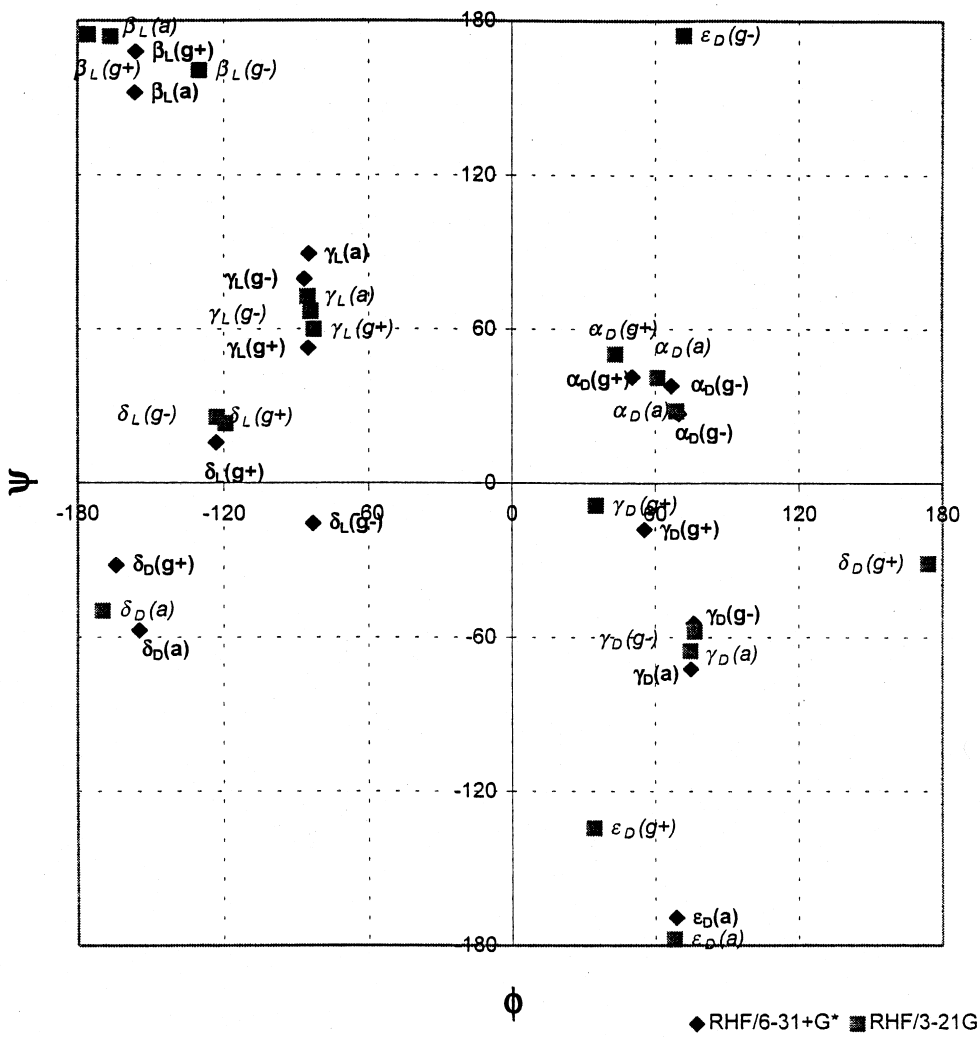


Fig. 13. Location of the (ϕ, ψ) values of HCO–L-Phe–NH_2 backbone conformers calculated at two levels of theory: '■', 3-21G RHF; and '◆', 6-31+G* RHF.

To be useful, the diamide approximation should provide a suitable model for the building units of the well-known secondary structural elements of proteins. After performing a grid search in the neighborhood of $\phi = -54°$, $\psi = -45°$, Schäfer et al. (1984) found that no minimum exists in that region of the Ramachandran surface of HCO–L-Ala–NH_2. This observation, while first handled with some criticism, initiated several studies (Klimkowski et al., 1985; Siam et al., 1987; Head-Gordon et al., 1991; Perczel et al., 1991, 1992, 1993, 1994a,b; Van Alsenoy et al., 1993). In the early 90s Head-Gordon et al. (1991) obtained the first complete ab initio Ramachandran surface ($-180° \leq \phi \leq +180°$, $-180° \leq \psi \leq +180°$ with a step size of 30°) and found again only seven out of the expected nine minima. Therefore, the conformational building unit of the right-handed helical structure turned out not to be a

stationary point on the PES of HCO–L-Ala–NH_2. A grid search with significantly smaller step size, mapping the ($\phi \approx -60^\circ + 30^\circ$, $\psi \approx -30^\circ + 30^\circ$) region concluded that in the region, where the minimum is expected, a mountain side can be found. Similar topology was found for the appropriate area of HCO–L-Val–NH_2. Furthermore, in phenylalanine diamide no right-handed helical building unit could be found. It appears that in the case of the smaller methyl (α-alanine), the larger isopropyl (valine) and the aromatic side-chain containing diamide models these two stationary points are not present. It seems that unfavorable interactions between backbone/side-chain atoms destabilize the main-chain orientation of α-helix (α_L) and poly-proline II (ε_L). To solve this anomaly within the limits of the diamide approximation is a challenge.

Favorable side-chain/backbone interaction(s), present, for example, in Ser and Thr, are expected to stabilize conformations annihilated in other model compounds. Note that neither Ser nor Thr adopt the helix-like backbone structure in proteins more often than Val or Phe (Karplus, 1996). The $\alpha_L(a, a)$, $\alpha_L(g-, a)$ and $\alpha_L(g-, g-)$ structures were the first examples showing that even a simple amino acid diamide (HCO–L-Ser–NH_2) may adopt the 'conformational monomer' of the right-handed helical structure. A study with systematically increased basis sets (3-21G, 4-21G, 6-31G*, 6-311 + + G**) has shown that although these structures could be slightly shifted, they are always present (Perczel and Csizmadia, 1995). Obviously, the hydroxymethyl side chain has an impact on the backbone that even a helix-like backbone structure is stabile. Although vanishes in the case of HCO–Xxx–NH_2 (where Xxx = Gly, Ala, Val, Phe), three α_L main-chain folds were assigned in the case of Xxx = Ser (and Thr).

Reoptimization employing a larger basis set resulted in more realistic relative energies (Jákli et al., 1998a,b) but the conformational parameters of the helix-like building subunits were a bit distorted toward the δ_L conformational region, also called the bridge region. This observation supports the assumption that the conformational monomer of a 'helix-like' structure can be stable for amino acids with polar side chains even before the repetitive (i, $i + 3$)- or (i, $i + 4$)-type backbone/backbone H-bond network system could be built up.

During investigation of the conformational properties of Thr and Cys, two different results were obtained. The 'extra' methyl group located at the β carbon atom of the side-chain of Thr does not have an influence on its main chain; the α_L conformers could be located. Almost identical structural parameters were determined for HCO–L-Thr–NH_2 than for HCO–L-Ser–NH_2. The same type of minimization performed in the case of HCO–L-Cys–NH_2 provided nonhelical conformers (Perczel and Csizmadia, 1995), possibly owing to the decreased H-bond ability of the –SH group. We also mention that the conformational monomer of the right-handed helix-like structure was found as a minimum in the case of the monohydrated HCO–L-Ala–NH_2 complex (Perczel et al., 1995).

As pointed out repeatedly, the conformational monomer of the poly-proline II ($(\varepsilon_L)_n$) secondary structural element could be obtained for none of the simple amino acid diamides, all efforts failed to locate this main-chain orientation. Therefore, the question remains whether the diamide approximation holds only for eight out of the nine backbone conformers?

8.3. Amino acid triamides (P–CO–Xxx–Yyy–NH–Q)

Using the grid-search technique, Antohi et al. (1996) performed a comprehensive analysis of L-Pro–L-Ala, L-Pro–D-Ala and L-Pro–Gly. The calculations have been carried out using the 6-31G basis set, the results obtained support the current view that the presence of a D-Ala residue in the $(i + 2)$th position favors a type II β-turn over a type I β-turn structure. These calculations are examples for the transition between diamide and triamide models, since they incorporate two amino acid residues but have only a single amide bond. Undoubtedly, the most simple triamide peptide model made from chiral amino acid residues is HCO–L–Ala–L–Ala–NH_2.

There has been constant interest in the conformational properties of *N*-formyl–L-alanyl–L-alaninamide (Van Alsenoy et al., 1993; Schäfer et al., 1993; Perczel et al., 1993, 1994a,b). The first studies incorporated only the inverse γ-turn backbone conformer of this molecule. This was followed by the calculation of some bent forms. The complete set of stationary points of HCO–L-Ala–L-Ala–NH_2 (49 backbone conformers) was reported by Perczel et al. (1994a,b). The previously computed 49 structures were minimized from 81 MDCA-predicted input locations on $E = E(\phi_1, \psi_1, \phi_2, \psi_2)$. Further investigations provided two additional stable conformers at the 3-21G RHF level. A comprehensive analysis of the 51 geometries and their relative energies provides the following conclusions: (a) 30 different β-turns have been assigned based on previously applied selection criteria (τ- and *d*-values) (Perczel et al., 1993) and (b) all stable structures are made of the combinations of previously described minima found for amino acid diamides: α_L, α_D, β_L, γ_L, γ_D, δ_L, δ_D, ε_L and ε_D. Although the α_L and ε_L conformers are not stationary points on the Ramachandran surface of HCO–L-Ala–NH_2, in the case of HCO–L-Ala–L-Ala–NH_2 both backbone orientations appear as subconformers of stationary points (e.g. $\alpha_L\delta_L$ and $\gamma_L\varepsilon_L$). This observation confirms that amino acid diamides (P–CONH–CHR–CONH–Q) are suitable models to estimate conformational building units of proteins. Results of ab initio calculations performed on alanine-containing triamides suggest that all MDCA predicted minima (all 9) on the Ramachandran surface will indeed appear as stable backbone substructures of the molecule. These results have relevance to the description of peptide folding, peptide and protein spectroscopy (e.g. CD, i.r. and NMR) and protein structure prediction.

8.4. Oligo- and polypeptides (P–CO–Xxx_n–NH–Q, where $n \geq 3$)

It is of interest to note that besides linear peptide models, cyclic forms, such as dioxopiperazine, *cyclo*-tetra-, *cyclo*-penta-, *cyclo*-hexa-peptides, have also been investigated by means of ab initio techniques. For example, 3-21G and DZP RHF calculations (Böhm and Brode, 1995) were used for the determination of four low-energy conformers of cyclohexaglycine. Böhm and Brode (1995) demonstrated that the double-type-I β-turn and the double-type-II β-turn conformers are stable. Although this study resulted in fascinating observations, it concentrated only on selected conformers.

Based on 3-21G RHF calculations on peptide models with systematically increased residue length, (HCO–L-$(Ala)_n$–NH_2), the selection of some homo-conformers was observed. With the increase of n the two helical forms ($(\alpha_L)_n$ and $(\alpha_D)_n$) become more and more favored. On the

other hand, the stability of less familiar homo-conformers (e.g. $(\delta_D)_n$ and $(\varepsilon_D)_n$) continuously decreases (Fig. 14). Besides the helix-like secondary structural elements, the conformational building unit of β-sheets $(\beta_L)_n$ is also favored, even though the H-bond network, operative between two antiparallel β-sheets, cannot be incorporated into this model calculation. An important structural feature of helical units is their periodic H-bond network. The H-bond pattern of a 3_{10} helix incorporates 10 atoms ($CO_i \ldots H_{Ni+3}$), while in the conventional α-helix 13 atoms ($CO_i \ldots H_{Ni+4}$) are involved. According to X-ray data analyses of globular proteins the 3–10-helix can not only be found as an individual secondary structural element but also as subconformers located at the C- or N-terminus of normal α-helices. In the latter situation these sharper 3_{10} helices form the first or the last turn(s) of α-helix. In the steeper 3_{10} helix the appropriate ϕ torsional angle value is around $-60°$, while ψ is around $-30°$. In a 'standard' α-helix $\phi \approx -54°$ and $\psi \approx -45°$. The reason why an α-helix sequence is adopted by a protein and

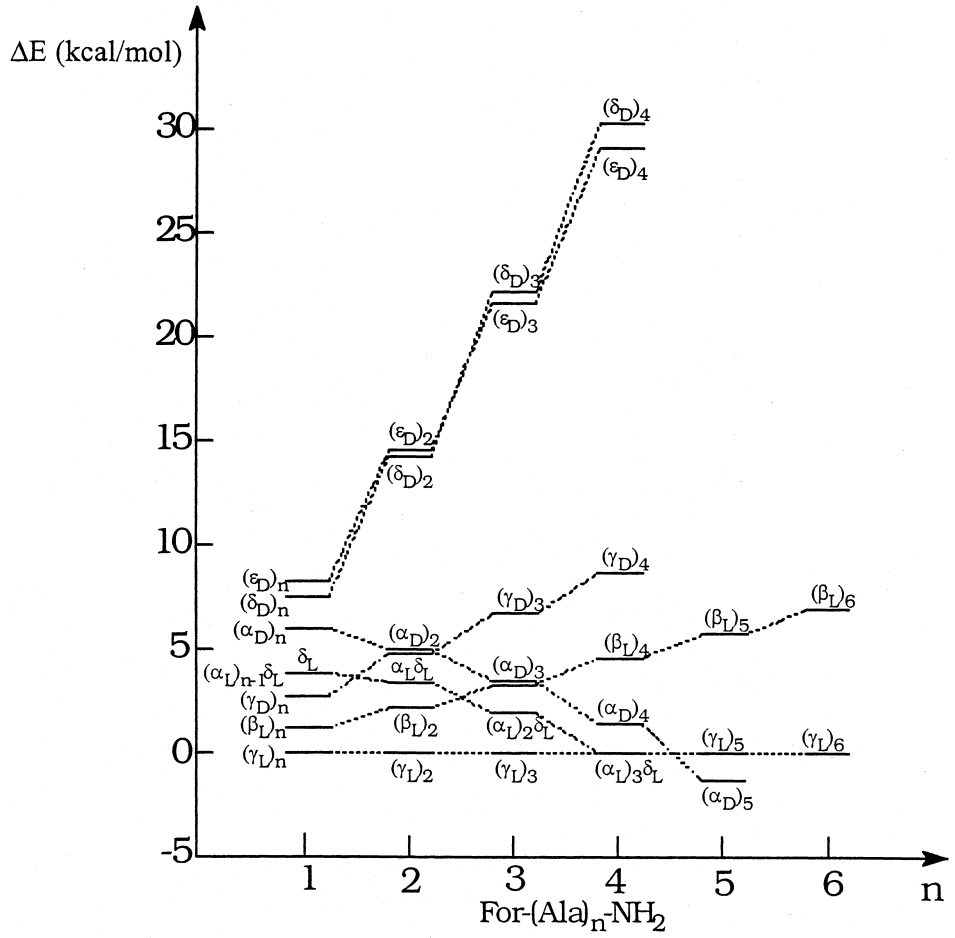


Fig. 14. Comparison of 3-21G RHF relative energy differences of HCO–L-$(Ala)_n$–NH_2 homo-conformers. Helices become more favored conformers with the increase of the length of peptides. Familiar secondary structural elements are among the most stable conformers after a long enough sequence.

the question of 'helix signal(s)' in proteins have also been investigated. A systematic study on helical conformers (Perczel et al., 1994a,b) of HCO–L-$(Ala)_n$–NH_2 showed that optimization started from the $(\alpha_L)_n$ conformation converged to the $(\alpha_L)_{n-1}\delta_L$ backbone structure at the 3-21G RHF level. All these backbone structures incorporate the δ_L subconformation at the carboxylic end of the oligopeptide chain. Furthermore, the last two residues have $\alpha_L\delta_L$ main-chain fold, which is the well-known and stable type I β-turn. Analyzing the H-bond network system of the optimized $(\alpha_L)_{n-1}\delta_L$ conformation of the HCO–(L-Ala)$_n$–NH_2 molecule a 3_{10} helix was assigned. These ab initio calculations confirmed the expectation that the formation of a helical segment is strongly coupled with the build-up of a systematic H-bond network system.

9. Prospects

It is sincerely hoped that the present review has demonstrated convincingly even to nonspecialists that ab initio techniques of computational quantum chemistry matured to the level whereby their intrinsic accuracy make them useful tools in the hands of biochemists who previously had to rely on more approximate techniques such as molecular mechanics. The continued exponential growth in computing power coupled with more and more sophisticated quantum chemistry techniques and programs will certainly result in a growing number of applications aimed at describing the building units of peptides and proteins. Therefore, the future seems just as bright as the beginning was: one of the earliest calculations (Sellers and Schäfer, 1978) employing methods of molecular electronic structure theory on the modeling problem of the structures of amino acids and model peptides provided results which contradicted those available from relevant experiments (Brown et al., 1978) of the time: an ab initio geometry optimization predicted the existence of a hidden conformation state and at the same time was decisive in guiding new experiments which led to the discovery of that state.

Acknowledgements

The original research of the authors incorporated in this review has been supported by grants from the Hungarian Scientific Research Foundation (OTKA F013799, T017604, T017192, and T024044), by the Hungarian Ministry of Culture and Education (MKM/517 and FKFP 017/1997) and by the Hungarian Academy of Sciences (AKP96/2-427 2.4).

References

Abola, E.E., Bernstein, F.G., Bryant, S.H., Koetzle, I.F., Weng, J., 1987. In: Allen, F.H., Bergerhoff, G., Sievers, R. (Eds.), Crystallographic Database–Information Content, Software System, Scientific Applications, in Data Commission of the International Union of Crystallography. Bonn/Cambridge/Chester, p. 107.

Albrecht, G., Corey, R.B., 1939 J. Am. Chem. Soc. 61, 1087–1103.

Allen, W.D., Császár, A.G., Horner, D.A., 1992 J. Am. Chem. Soc. 114, 6834–6849.

Allen, W.D., Császár, A.G., 1993 J. Chem. Phys. 98, 2983–3015.

Allinger, N.J., 1992 J. Am. Chem. Soc. 114, 1.
Almlöf, J., Taylor, P.R., 1991 Adv. Quantum Chem. 22, 301.
Antohi, O., Naider, F., Sapse, A.M., 1996 J. Mol. Struct. (THEOCHEM) 360, 99–108.
Ataka, S., Takeuchi, H., Tasumi, M.J., 1984 J. Mol. Struct. 113, 147.
Aubry, A., Marraud, M., Protas, J., Neel, J., 1974 Compt. Rend. Acad. Sci. Paris 287c, 163.
Aubry, A., Ghermani, N., Marraud, M., 1984 Int. J. Peptide Protein Res. 23, 113.
Barone, V., Adamo, C., Lelj, F., 1995 J. Chem. Phys. 102, 364–370.
Bartlett, R.J., Stanton, J.F., 1994. In: Lipkowitz, K.B., Boyd, D.B. (Eds.), Reviews of Computational Chemistry. VCH, New York.
Becke, A.D., 1993 J. Chem. Phys. 98, 5648–5652.
Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Jr., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M., 1977 J. Mol. Biol. 112, 535.
Biemann, K., Martin, S.A., 1987 Mass Spectrum. Rev. 6, 1.
Bingham, R.C., Dewar, M.J.S., Lo, D.H., 1975 J. Am. Chem. Soc. 97, 1302–1307.
Blow, D.M., 1976 Acc. Chem. Res. 9, 145.
Böhm, H.J., Brode, S., 1991 J. Am. Chem. Soc. 113, 7129–7135.
Böhm, H.J., Brode, S., 1995 J. Comp. Chem. 16, 146–153.
Bojesen, G., 1987 J. Am. Chem. Soc. 109, 5557–5558.
Bonaccorsi, R., Palla, P., Tomasi, J., 1984 J. Am. Chem. Soc. 106, 1945–1950.
Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983 J. Comput. Chem. 4, 187–217.
Bruccoleri, R.E., Karplus, M., 1987 Biopolymers 26, 137.
Brown, R. D., Godfrey, P. D., Storey, J. W. V., Bassez, M. P., 1978. J. Chem. Soc. Chem. Commun. 547–548.
Brown, R.D., Godfrey, P.D., Storey, J.W.V., Bassez, M.P., Robinson, B.J., Batchelor, R.A., McCulloch, M.G., Rydbeck, O.E.H., Hjalmarson, A.G., 1979 Mon. Not. R. Astron. Soc. 186, 5.
Cabrol, D., Brock, H., Vasilescu, D., 1979 Int. J. Quantum Chem. 6, 365.
Campbell, S., Marzluff, E.M., Rodgers, M.T., Beauchamp, J.L., Rempe, M.E., Schwinck, K.F., Lichtenberger, D.L., 1994 J. Am. Chem. Soc. 116, 5257–5264.
Cannington, P.H., Ham, N.S., 1983 J. Electron Spectrosc. Relat. Phenom. 32, 139–151.
Cao, M., Newton, S.Q., Pranata, J., Schäfer, L., 1995 J. Mol. Struct. (THEOCHEM) 332, 251–267.
Carozzo, L., Corongiu, G., Petrongolo, C., Clementi, E., 1978 J. Chem. Phys. 68, 787.
Challacombe, M., Schwegler, E., 1997 J. Chem. Phys. 106, 5526–5536.
Čížek, J., 1969 Adv. Chem. Phys. 14, 35.
Clementi, E., Cavallone, F., Scordamaglia, R., 1977 J. Am. Chem. Soc. 99, 5531–5545.
Contineanu, I., Marchidau, D.I., 1984 Rev. Rom. Chim. 29, 43–48.
Cooks, R.G., Kruger, T.L., 1977 J. Am. Chem. Soc. 99, 1279–1281.
Combes, F., Nguyen-Q-Rieu, Wlodarczak, G., 1996 Astron. Astrophys.
Curtiss, L.A., Raghavachari, K., Trucks, G.W., Pople, J.A., 1991 J. Chem. Phys. 94, 7221–7230.
Curtiss, L.A., Raghavachari, K., Pople, J.A., 1993 J. Chem. Phys. 98, 1293–1298.
Császár, A.G., 1992 J. Am. Chem. Soc. 114, 9569–9575.
Császár, A.G., 1995 J. Mol. Struct. 346, 141–152.
Császár, A.G., 1996 J. Phys. Chem. 100, 3541–3551.
Császár, A. G., Allen, W. D. and Schaefer III, H. F., 1998. J. Chem. Phys., 108, 9751.
Császár, A.G., Kelterer, A.-M., Ramek, M., 1998. Unpublished results.
Csizmadia, I.G., 1989. In: Csizmadia, I.G., Bertran, J.D. (Eds.), New Theoretical Concept for Understanding Organic Reactions. Reidel.
Debies, T.P., Rabalais, J.W., 1974 J. Electron Spectrosc. 3, 315–322.
Deprick-Cote, B., Lauglet, J., Caillet, J., Berges, J., Kassab, E., Constanciel, R., 1992 Theor. Chim. Acta 82, 435.
DeTar, D.F., Luthra, N.P., 1977 J. Am. Chem. Soc. 99, 1232–1244.
Dewar, M.J.S., Thiel, W., 1977 J. Am. Chem. Soc. 99, 4899–4907.
Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., Stewart, J.J.P., 1986 J. Am. Chem. Soc. 107, 3902–3909.
Ding, Y., Krogh-Jespersen, K., 1992 Chem. Phys. Lett. 199, 261.
Ding, Y., Krogh-Jespersen, K., 1996 J. Comp. Chem. 17, 338–349.
Drakenberg, T. and Forsen, S., 1971. J. Chem. Soc. Chem. Commun. 1404.
Dunning, T.H., 1989 J. Chem. Phys. 90, 1007–1023.
Dyson, H.J., Rance, M., Houghten, R.A., Lerner, A., Wright, E., 1988 J. Mol. Biol. 201, 161.
Dyson, H.J., Merutka, G., Waltho, J.P., Lerner, R.A., Wright, P.E., 1992a J. Mol. Biol. 226, 795.
Dyson, H.J., Sayre, J.R., Merutka, G., Shin, H.C., Lerner, R.A., Wright, P.E., 1992b J. Mol. Biol. 226, 819.
Efimov, A.V., 1993 Prog. Biophys. Mol. Biol. 60, 201–239.

Eddington, P., Harding, M.H., 1974 Acta Crystallogr. Sect. B. 30, 204.
Edsall, J.T., 1936 J. Chem. Phys. 4, 1.
Endrédi, G., Perczel, A., Farkas, O., McAllister, M.A., Csonka, G.I., Ladik, J., Csizmadia, I.G., 1997 J. Mol. Struct. (THEOCHEM) 391, 15–26.
Ewbank, J.D., Klimkowski, V.J., Siam, K., Schäfer, L., 1987 J. Mol. Struct. 160, 275–285.
Farkas, O., Perczel, A., Marcoccia, J.F., Hollósi, M., Csizmadia, I.G., 1995 J. Mol. Struct. (THEOCHEM) 331, 27–36.
Farkas, O., McAllister, M.A., Ma, J.H., Perczel, A., Hollósi, M., Csizmadia, I.G., 1996 J. Mol. Struct. (THEOCHEM) 396, 105–114.
Fesik, S.W., Gampe, R.T., Eaton, H.L., Gemmecker, G., Olejniczak, E.T., Neri, P., Holyman, T.F., Egan, D.A., Edalji, R., Simmer, R., Helfrich, R., Hochlowski, J., Jackson, M., 1991 Biochemistry 30, 6574.
Finer-Moore, J., Bazan, J.F., Rubin, J. and Stroud, R.M., 1989. In: Fasman, G.D. (Ed.), Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York.
Fock, V.A., 1930 Z. Phys. 15, 126.
Förner, W., Otto, P., Bernhardt, J., Ladik, J.J., 1981 Theor. Chim. Acta 60, 269.
Frey, M.N., Koetzle, T.F., Lehmann, M.S., Hamilton, W.C., 1973a J. Chem. Phys. 58, 2547–2556.
Frey, M.N., Lehmann, M.S., Koetzle, T.F., Hamilton, W.C., 1973b Acta Crystallogr. Sect. B 29, 876.
Frey, R.F., Coffin, J., Newton, S.Q., Ramek, M., Cheng, V.K.W., Momany, F.A., Schäfer, L., 1992 J. Am. Chem. Soc. 114, 5369.
Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T.A., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Cioslowski, J., Stefanov, B.B., Nanayakkara, A., Challacombe, M., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Replogle, E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., DeFrees, D.J., Baker, J., Stewart, J.J.P., Head-Gordon, M., Gonzalez, C., Pople, J.A. Gaussian 94. Gaussian, Inc., Pittsburgh, PA, 1995.
Gaffney, J.S., Pierce, R.C., Friedman, L., 1977 J. Am. Chem. Soc. 99, 4293–4298.
Garfinkel, D., 1958 J. Am. Chem. Soc. 80, 3827–3831.
Garfinkel, D., Edsall, J.T., 1958 J. Am. Chem. Soc. 80, 3823–3826.
Gething, M.-J., Sambrook, J., 1992 Nature 335, 33.
Ghalem, S., Fan, B.T., Xlao, L., Lapluye, G., 1994 J. Chim. Phys. Phys.-Chem. Biol. 91, 113–125.
Ghosh, I., Edholm, O., 1994 Biophys. Chem. 50, 237–248.
Gibson, K.D., Scheraga, H.A., 1966 Biopolymers 4, 709.
Godfrey, P.D., Firth, S., Hatherley, L.D., Brown, R.D., Pierlot, A.P., 1993 J. Am. Chem. Soc. 115, 9687–9691.
Godfrey, P.D., Brown, R.D., 1995 J. Am. Chem. Soc. 117, 2019–2023.
Godfrey, P.D., Brown, R.D., Rodgers, F.M., 1996 J. Mol. Struct. 376, 65–81.
Golic, L., Hamilton, W.C., 1972 Acta Crystallogr. Sect. B. 26, 1265.
Gorman, G.S., Speir, J.P., Turner, C.A., Amster, I.J., 1992 J. Am. Chem. Soc. 114, 3986.
Gronert, S., O'Hair, R.A.J., 1995 J. Am. Chem. Soc. 117, 2071–2081.
Hamgauer, D.G., Monzigo, A.F., Mathews, B.W., 1984 Biochemistry 23, 5730.
Handy, N.C., Pople, J.A., Head-Gordon, M., Raghavachari, K., Trucks, G.W., 1989 Chem. Phys. Lett. 164, 185.
Harrison, A.G., 1997 Mass Spectrum. Rev. 16, 201–217.
Hartree, D.R., 1928 Proc. Cambridge Phil. Soc. 24, 328.
Hayashi, H., Mizuguchi, H., Kagamiyama, H., 1993 Biochemistry 32, 812–818.
Head-Gordon, T., Head-Gordon, M., Frisch, M.J., Brooks, C.L., III, Pople, J.A., 1989 Int. J. Quantum Chem. Quantum Biol. 16, 311–322.
Head-Gordon, T., Head-Gordon, M., Frisch, M.J., Brooks, C.L., III, Pople, J.A., 1991 J. Am. Chem. Soc. 113, 5989–5997.
Hedberg, K., 1988. In: Hargittai, I., Hargittai, M. (Eds.), Stereochemical Applications of Gas-phase Electron Diffraction. VCH, New York, pp. 347–366.
Hehre, W.J., Radom, L., Schleyer, P.v.R., Pople, J.A., 1986. Ab Initio Molecular Orbital Theory. Wiley, New York.
Hocking, W.H., 1976 Z. Naturforsch. A 31, 1113.
Hollis, J.M., Snyder, L.E., Suenram, R.D., Lovas, F.J., 1980 Astrophys. J. 241, 1001.
Howard, J.C., Ali, A., Scheraga, H.A., Momany, F.A., 1975 Macromolecules 8, 607.
Hutchens, J.O., Cole, A.G., Stout, J.W., 1963 J. Phys. Chem. 67, 1128–1130.
Hu, C.-H., Shen, M., Schaefer, H.F., III, 1993 J. Am. Chem. Soc. 115, 2923–2929.
HYPERCHEM 4.5, 1994. Hypercube, Inc. 419 Phillip St., Waterloo, Ont. Canada N2L 3X2. The HyperMM+ force field is derived from the public domain code developed by Allinger, N.L. (1977) J. Am. Chem. Soc. 99, 8127–8134.
Hunter, E.P., Lias, S.G., 1997 Proton Affinity Evaluation, in NIST Standard Reference Database, Number 69, NIST, Gaithersburg.
Iijima, K., Tanaka, K., Onuma, S., 1991 J. Mol. Struct. 246, 257–266.
Iijima, K., Beagley, B., 1991 J. Mol. Struct. 248, 133.
Imamura, A., Fujita, H., Nogata, C., 1969, Bull. Chem. Soc. Jpn. 3118–3123.
Isa, K., Omote, T., Amaya, M., 1990 Org. Mass. Spetrom. 25, 620.

IUPAC-IUB Commission on Biochemical Nomenclature, 1970 Biochemistry 9, 3471–3479.
Jákli, I., Perczel, A., Farkas, O., Hollósi, M., Csizmadia, I.G., 1998a. J. Mol. Struct. (THEOCHEM), in press.
Jákli, I., Perczel, A., Farkas, O., Csizmadia, I.G., 1998b. J. Comp. Chem., in preparation.
Jensen, J.H., Gordon, M.S., 1991 J. Am. Chem. Soc. 113, 7917.
Jensen, J.H., Gordon, M.S., 1995 J. Am. Chem. Soc. 117, 8159.
Jones, D.D., Bernal, L., Frey, M.N., Koetzle, T.F., 1974 Acta Crystallogr. Sect. B 30, 1220.
Jonsson, P.-G., Kvick, A., 1972 Acta Crystallogr. Sect. B. 28, 1827.
Jorgensen, W.L., Tirado-Rives, J., 1988 J. Am. Chem. Soc. 110, 1657–1666.
Karle, I.L., 1977. In: Gross, E., Meienhofer, J. (Eds.), Peptide Structure and Biological Function. Wiley, New York.
Kabsch, W., Sander, C., 1984 J. Mol. Biol. 81, 1075.
Kaiser, E.T., Kézdy, F.J., 1984 Science 223, 249.
Karle, I.L., 1979 J. Am. Chem. Soc. 101, 181.
Karle, I.L., 1981. In: Gross E., Meienhofer, J. (Eds.), Peptides. Acad. Press, New York.
Karle, I.L., Karle, J., Mastropaolo, D., Camerman, A., Camerman, N., 1983. In: Hruby, V.J., Rich, D.H. (Eds.), Peptides: Structure and Function. Pierce Chem. Co., Rockford, IL.
Karplus, P.A., 1996 Protein Sci. 5, 1406.
Kendall, R.A., Dunning, T.H., Jr., Harrison, R.J., 1992 J. Chem. Phys. 96, 6796.
Klasnic, L.J., 1976 J. Electron Spectrosc. Relat. Phenom. 8, 161.
Kikuchi, O., Natsui, T., Kozaki, T., 1990 J. Mol. Struct. (THEOCHEM) 207, 103.
Kikuchi, O., Matsuoka, T., Sawahata, H., Takahashi, O., 1994 J. Mol. Struct. (THEOCHEM) 305, 79–87.
Kikuchi, O., Watanabe, T., Ogawa, Y., Takase, H., Takahashi, O., 1997 J. Phys. Org. Chem. 10, 145–151.
Klimkowski, V.J., Schäfer, L., Momany, F.A., van Alsenoy, C., 1985 J. Mol. Struct. (THEOCHEM) 124, 143.
Koetzle, T.F., Lehman, M.S., Verbist, J.J., Hamilton, W.C., 1972 Acta Crystallogr. Sect. B. 28, 3207.
Koetzle, T.F., Frey, M.N., Lehman, M.S., Hamilton, W.C., 1973 Acta Crystallogr. Sect. B. 29, 2571.
Lambert, M.H., Scheraga, H.A., 1989 J. Comp. Chem. 10, 770–817.
Lattman, E.E., Rose, G.D., 1993 Proc. Natl. Acad. Sci. 90, 439.
Laurence, P.R., Thomson, C., 1981 Theor. Chim. Acta 58, 121.
Leach, S.J., Némethy, G., Scheraga, H.A., 1966 Biopolymers 4, 369.
Lee, C., Yang, W., Parr, R.G., 1988 Phys. Rev. B 37, 785–789.
Lee, T.J., Taylor, P.R., 1989 Int. J. Quantum. Chem. Symp. 23, 199.
Lehmann, M.S., Koetzle, T.F., Hamilton, W.C., 1972 J. Am. Chem. Soc. 94, 2657–2660.
Lehmann, M.S., Verbist, J.J., Hamilton, W.C., Koetzle, T.F., 1973 J. Chem. Soc. Perkins Trans. 2, 133.
Levy, H.A., Corey, R.B., 1941 J. Am. Chem. Soc. 63, 2095–2108.
Lewis, P.N., Momany, F.A., Scheraga, H.A., 1973 Isr. J. Chem. 11, 121.
Li, X., Harrison, A.G., 1993 Org. Mass. Spectrum. 28, 366.
Liquori, A.M., 1969 Quant. Rev. Biophys. 2, 65.
Locke, M.J., 1981. Ph.D. thesis, University of California, Irvine, CA.
Locke, M.J., McIver, R.T., 1983 J. Am. Chem. Soc. 105, 4226.
Lovas, F.J., Kawashima, Y., Grabow, J.-U., Suenram, R.D., Fraser, G.T., Hirota, E., 1995 Astrophys. J. 455, L201–L204.
McWeeny, R., 1992. Methods of Molecular Quantum Mechanics, 2nd ed. Academic Press, London.
Mathieson, A.L., 1952 Acta Crystallogr. 5, 332.
Matsubara, T., Shinohara, H., Sisido, M., 1997 Macromolecules 30, 2651–2656.
McAllister, M.A., Perczel, A., Császár, P., Viviani, W., Rivail, J.L., Csizmadia, I.G., 1993 J. Mol. Struct. (THEOCHEM) 290, 161–179.
Meot-Ner, M., Hunter, E.P., Field, F.H., 1979 J. Am. Chem. Soc. 101, 686.
Mezey, P.G., Ladik, J.J., Suhai, S., 1979 Theor. Chim. Acta 51, 323.
Mezey, P.G., 1987. Potential Energy Hypersurfaces. Elsevier.
Mezey, P.G., 1995 Struct. Chem. 6, 261–270.
Mirskz, A.E., Pauling, L., 1936 Proc. Natl. Acad. Sci. 22, 439.
Mitsui, Y., Tsuboi, M., Iitaka, Y., 1969 Acta Crystallogr. Sect. B 25, 2182.
Moebius, U., Clayton, L.K., Abraham, S., Harrison, S.C., Reinherz, E.L., 1992 J. Exp. Med. 176, 507.
Moews, P.C., Kretsinger, R.H., 1975 J. Mol. Biol. 91, 201.
Møller, C., Plessett, M.S., 1934 Phys. Rev. 46, 618.
Mostad, A., Nissen, H. M., Rømming, C., 1971. Tetrahedron Lett. 2131.
Náray-Szabó, G., Surján, P.R., Ángyán, J.G., 1987. Applied Quantum Chemistry. Reidel, Dordrecht.
Natanabe, T., Hashimoto, K., Takase, H., Kikuchi, O., 1997 J. Mol. Struct. (THEOCHEM) 397, 113–119.
Navarrete, J.T.L. et al, 1997 J. Raman Spectry. 28, 501.

Némethy, G., Printz, M.P., 1972 Macromolecules 6, 755.
Némethy, G., Miller, M.H., Scheraga, H.A., 1980 Macromolecules 13, 914.
Neuhaus, D., Williamson, M., 1989. The Nuclear Overhauser Effect in Structural and Conformational Analysis. VCH Publications.
Neville, J.J., Zheng, Y., Brion, C.E., 1996 J. Am. Chem. Soc. 118, 10533–10544.
Ngauv, S.N., Sabbah, R., Laffitte, M., 1977 Thermochim. Acta 20, 371–380.
Noggle, J.H., Schirmer, R.E., 1971. The Nuclear Overhauser Effect: Chemical Applications. Acad. Press, New York.
O'Hair, R.A.J., Bowie, J.H., Gronert, S., 1992 Int. J. Mass Spectrom. Ion Proc. 117, 23–36.
Ochterski, J.W., Petersson, G.A., Montgomery, J.A., 1995 J. Am. Chem. Soc. 117, 11299.
Okuyama,-Yoshida, N., Nagaoka, M., Yamabe, T., 1998 J. Phys. Chem. A 102, 285–292.
Paine, G.H., Scheraga, H.A., 1986 Biopolymers 25, 1547.
Palla, P., Petrongolo, C., Tomasi, J., 1980 J. Phys. Chem. 84, 435.
Pasternak, R.A., 1950 Acta Crystallogr 9, 341.
Parr, R. G., Yang, W., 1989. Density-functional Theory of Atoms and Molecules. Oxford University Press, Oxford.
Pawlukojc, A., Bobrowicz, L., Natkaniec, I., Leciejewicz, J., 1995 Spectrochim. Acta 51A, 303.
Pawlukojc, A., Leciejewicz, J., Natkaniec, I., 1996 Spectrochim. Acta 52A, 29.
Pawlukojc, A., Bajdor, K., Dobrowolski, J.Cz., Leciejewicz, J., Natkaniec, I., 1997 Spectrochim. Acta 53A, 927.
Perakyla, M., Kollman, P.A., 1997 J. Am. Chem. Soc. 119, 1189–1196.
Perakyla, M., Rouvinen, J., 1996 Eur. J. Chem. 2, 1548–1551.
Perczel, A., Daudel, R., Ángyán, J.G., Csizmadia, I.G., 1990 Can. J. Chem. 68, 1882–1888.
Perczel, A., Ángyán, J.G., Kajtár, M., Viviani, W., Rivail, J.L., Marcoccia, J.F., Csizmadia, I.G., 1991 J. Am. Chem. Soc. 113, 6256–6265.
Perczel, A., Foxman, B.M., Fasman, G.D., 1992 Proc. Nat. Acad. Sci. 89, 8210.
Perczel, A., McAllister, M.A., Császár, P., Csizmadia, I.G., 1993 J. Am. Chem. Soc. 115, 4849–4858.
Perczel, A., McAllister, M.A., Császár, P., Csizmadia, I.G., 1994a Can. J. Chem. 72, 2050.
Perczel, A., Endrédi, G., McAllister, M.A., Farkas, O., Császár, P., Ladik, J., Csizmadia, I.G., 1994b J. Mol. Structure (THEOCHEM) 331, 5–10.
Perczel, A., Csizmadia, I.G., 1995 Int. Rev. Phys. Chem. 14, 127–168.
Perczel, A., Farkas, O., Csizmadia, I.G., 1995 J. Am. Chem. Soc. 117, 1653–1655.
Perczel, A., Farkas, O., Csizmadia, I.G., 1996a J. Comp. Chem. 17, 821–834.
Perczel, A., Farkas, O., Csizmadia, I.G., 1996b J. Am. Chem. Soc. 118, 7809–7817.
Perczel, A., Farkas, O., Marcoccia, J.F., Csizmadia, I.G., 1997a Int. J. Quant. Chem. 61, 797–814.
Perczel, A., Farkas, O., Császár, A.G., Csizmadia, I.G., 1997b Can. J. Chem. 75, 1120–1130.
Petersson, G.A., Bennett, A., Tensfeldt, T.G., Al-Laham, M.A., Shirley, W.A., Mantzaris, J., 1988 J. Chem. Phys. 89, 2193.
Peterson, M.R., Csizmadia, I.G., 1982 Progr. Theor. Org. Chem. 3, 190.
Ponnuswamy, P.K., Sasisekharan, V., 1971 Biopolymers 10, 565.
Pranate, J., Wierschke, S., Jorgensen, W.L., 1991 J. Am. Chem. Soc. 113, 2810.
Price, W.D., Jockusch, R.A., Williams, E.R., 1997 J. Am. Chem. Soc. 119, 11988–11989.
Pulay, P., Fogarasi, G., Pang, F., Boggs, J.E., 1979 J. Am. Chem. Soc. 101, 2550–2560.
Pulay, P., Fogarasi, G., Pongor, G., Boggs, J.E., Vargha, A., 1983 J. Am. Chem. Soc. 105, 7037–7047.
Pullman, B., Pullman, A., 1974 Adv. Prot. Chem. 28, 347.
Raghavachari, K., Trucks, G.W., Pople, J.A., Head-Gordon, M., 1989 Chem. Phys. Lett. 157, 479.
Raghavachari, K., Curtiss, L.A., 1995. In: Yarkony, D.R. (Ed.), Modern Electronic Structure Theory. II. World Scientific, Singapore.
Raj, P.A., Soni, S.D., Ramasubbu, N., Bhandary, K.K., Levine, M.J., 1990 Biopolymers 30, 73.
Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963 J. Mol. Biol. 7, 95.
Ramachandran, G.N., Venkatachalam, C.M., Krimm, S., 1966 Biophys. J. 6, 849.
Ramachandran, G.N., Sasisekharan, V., 1968 Adv. Protein Chem. 23, 283–438.
Ramanadham, M., Kikka, S.K., Chidambaram, R., 1972 Acta Crystallogr. Sect. B. 28, 3000.
Ramek, M., Cheng, V.K.W., Frey, R.F., Newton, S.Q., Schäfer, L., 1991 J. Mol. Struct. (THEOCHEM) 235, 1.
Ramek, M., Kelterer, A.M., Nikolic, S., 1997 Int. J. Quant. Chem. 65, 1033–1045.
Ramek, M., Schäfer, L., 1998 Can. J. Chem. in print.
Rao, S.T., 1973 Acta Crystallogr. Sect. B. 29, 1718.
Rauk, A., Yu, D., Armstrong, D.A., 1997 J. Am. Chem. Soc. 119, 208–217.
Reva, I.D., Stepanian, S.G., Plokhotnichenko, A.M., Radchenko, E.D., Sheina, G.G., Blagoi, Y.P., 1994 J. Mol. Struct. 318, 1.
Reva, I.D., Plokhotnichenko, A.M., Stepanian, S.G., Ivanov, A.Y., Radchenko, E.D., Sheina, G.G., Blagoi, Y.P., 1995 Chem. Phys. Lett. 232, 141–148.
Richardson, J.S., Richardson, D.C., 1989. In: Fasman, G.D. (Ed.), Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York.

Ripoll, D.R., Scheraga, H.A., 1988 Biopolymers 27, 1283.
Rittby, M., Bartlett, R.J., 1988 J. Phys. Chem. 92, 3033.
Robin, M.B., 1975. Higher Excited States of Polyatomic Molecules, vol. II. Academic Press, New York.
Rooman, M.J., Kocher, J.P.A., Widak, S.J., 1992 Biochemistry 31, 10226–10238.
Roterman, I.K., Lambert, M.H., Gibson, K.D., Scheraga, H.A., 1989 J. Biomol. Struct. Dyn. 7, 421–453.
Sabbah, R., Laffitte, M., 1978 Bull. Soc. Chim. Fr. 1, 50–52.
Sapse, A.-M., Mallah-Levy, L., Daniels, S.B., Erickson, B.W., 1987 J. Am. Chem. Soc. 109, 3526–3529.
Sapse, A.M., Schweitzer, B.S., Dicker, A.P., Bertino, J.R., Frecer, V., 1992 Int. J. Pept. Protein Res. 39, 18.
Sasisekharan, V., 1962. Stereochemical criteria for polypeptide and protein structures. In: Ramanaathan, N. (Ed.), Collagen, Madras, India. Wiley, pp. 39–78.
Scarsdale, J.N., Van Alsenoy, C., Klimkowski, V.J., Schäfer, L., Momany, F.A., 1983 J. Am. Chem. Soc. 105, 3438–3445.
Schaefer, H.F. (Ed.), 1977. Methods of Electronic Structure Theory. Plenum, New York.
Schäfer, L., Sellers, H.L., Lovas, F.J., Suenram, R.D., 1980 J. Am. Chem. Soc. 102, 6566.
Schäfer, L., Klimkowski, V.J., Momany, F.A., Chuman, H., van Alsenoy, C., 1984 Biopolymers 23, 2335.
Schäfer, L., Siam, K., Klimkowski, V.J., Ewbank, J.D., van Alsenoy, C., 1990a J. Mol. Struct. (THEOCHEM) 204, 361–372.
Schäfer, L., Kulp-Newton, S.Q., Siam, K., Klimkowski, V.J., van Alsenoy, C., 1990b J. Mol. Struct. (THEOCHEM) 209, 373–385.
Schäfer, L., Newton, S.Q., Cao, M., Peeters, A., Van Alsenoy, C., Wolinski, K., Momany, F.A., 1993 J. Am. Chem. Soc. 115, 272.
Scheraga, H.A., 1968 Adv. Phys. Org. Chem. 6, 103.
Shipman, L.L., Christoffersen, R.E., 1973 Theor. Chim. Acta 31, 75–82.
Shoemaker, D.P., Donohue, J., Shomaker, V., Coray, R.B., 1950 J. Am. Chem. Soc. 72, 2328–2349.
Schultz, G.E., 1988 Annu. Rev. Biophys. Chem. 17, 1–21.
Sellers, H.L., Schäfer, L., 1978 J. Am. Chem. Soc. 100, 7728–7729.
Sellers, H.L., Schäfer, L., 1979 Chem. Phys. Lett. 63, 609.
Sequeira, A., Rajagopal, H., Chidrambaram, R., 1972 Acta Crystallogr. Sect. B. 28, 2514.
Shirazian, S., Gronert, S., 1997 J. Mol. Struct. (TAEOCHEM) 397, 107–112.
Siam, K., Klimkowski, V.J., Ewbank, J.D., Van Alsenoy, C., Schäfer, L., 1984 J. Mol. Struct. (THEOCHEM) 110, 171.
Siam, K., Klimkowski, V.J., van Alsenoy, C., Ewbank, J.D., Schäfer, L., 1987 J. Mol. Struct. (THEOCHEM) 152, 261.
Simpson, H.J., Jr., Marsh, R.E., 1966 Acta Crystallogr. 20, 550.
Singh, J., Thornton, J.M., 1985 FEBS 191, 1.
Singh, J., Thornton, J.M., 1992. In: Atlas of Protein Side-chain Interactions. IRL Press, Oxford.
Sirois, S., Proynov, E.I., Nguyen, D.T., Salahub, D.R., 1997 J. Chem. Phys. 107, 6770–6781.
Snyder, L.E., Hollis, J.M., Suenram, R.D., Lovas, F.J., Brown, L.W., Buhl, D., 1983 Astrophys. J. 268, 123.
Stadman, E.R., 1993 Annu. Rev. Biochem. 62, 797.
Stanton, J.F., Bartlett, R.J., 1993 J. Chem. Phys. 98, 7029–7039.
Stepanian, S.G., Reva, I.D., Radchenko, E.D., Rosado, M.T.S., Duarte, M.L.T.S., Fausto, R., Adamowicz, L., 1998 J. Phys. Chem. A 102, 1041–1054.
Stewart, J.J.P., 1989 J. Comp. Chem. 10, 221.
Stroup, A.N., Cole, L.B., Dhingra, M.M., Gierach, L.M., 1990 Int. J. Peptide Protein Res. 36, 531.
Suenram, R.D., Lovas, F.J., 1978 J. Mol. Spectrosc. 72, 372–382.
Suenram, R.D., Lovas, F.J., 1980 J. Am. Chem. Soc. 102, 7180–7184.
Sun, S., Bernstein, E.R., 1996 J. Am. Chem. Soc. 118, 5086–5095.
Szabo, A., Ostlund, N.S., 1989. Modern Quantum Chemistry, rev. ed. McGraw-Hill, New York.
Székely, Z., Kónya, Z., Becskei, A., Goldering, W.P.D., Perczel, A., Penke, B., Molnár, J., Michejda, C.F., Aszalos, A., Csizmadia, I.G., 1996 J. Mol. Struct. (THEOCHEM) 367, 159–186.
Takeda, M. et al, 1958 J. Am. Chem. Soc. 80, 3813.
Thornton, J., Jones, D.T., MacArthur, M.W., Orengo, C.M., Swindells, M.B., 1995 Phil. Trans. R. Soc. Lond. B 348, 71–79.
Tomasi, J., Persico, M., 1994 Chem. Rev. 94, 2027.
Torii, K., Iitaka, Y., 1970 Acta Crystallogr. Sect. B. 26, 1317.
Torii, K., Iitaka, Y., 1970 Acta Crystallogr. Sect. B. 27, 2237.
Tortonda, F.R., Pascual-Almir, J.L., Silla, E., Tuñón, I., 1996 Chem. Phys. Lett. 260, 21–26.
Tortonda, F.R., Pascual-Aliuir, J.-L., Silla, E., Tuñón, I., Ramírez, F.J., 1998 J. Chem. Phys. 109, 592–602.
Tranter, G.E., 1985 Mol. Phys. 4, 825.
Tranter, G.E., 1986 J. Theor. Biol. 119, 467.
Van Alsenoy, C., Scarsdale, J.N., Sellers, H.L., Schäfer, L., 1981 Chem. Phys. Lett. 181, 124.
Van Alsenoy, C., Kulp, S., Siam, K., Klimkowski, V.J., Ewbank, J.D., Schäfer, L., 1988 J. Mol. Struct. 181, 169–178.
Van Alsenoy, C., Cao, M., Newton, S.Q., Teppen, B., Perczel, A., Csizmadia, I.G., Momany, F.A., Schäfer, L., 1993 J. Mol. Struct. 286, 149.

Van Alsenoy, C., Yu, C.-H., Peeters, A., Martin, J.M.L., Schäfer, L., 1998 J. Phys. Chem. 102, 2246.
Venkatachalam, C.M., 1968 Biopolymers 6, 1425.
Vishveshwara, S., Pople, J.A., 1977 J. Am. Chem. Soc. 99, 2422–2426.
Viviani, W., Rivail, J.L., Perczel, A., Csizmadia, I.G., 1993 J. Am. Chem. Soc. 115, 8321–8329.
Voogd, J., Derissen, J.L., van Duijneveldt, F.B., 1981 J. Am. Chem. Soc. 103, 7701–7706.
Wada, G., Tamura, E., Okina, M., Nakamura, M., 1992 Bull. Chem. Soc. Jpn. 55, 3064.
Waltho, J.P., Fehér, V.A., Wright, P.E., 1990. In: Villafranca, J.J. (Ed.), Current Research in Protein Chemistry. Acad. Press.
Waltho, J.P., Fehér, V.A., Dyson, H.J., Wright, P.E., 1992. Biochemistry.
Warne, P.K., Morgan, R.S., 1978 J. Mol. Biol. 118, 273.
Watanabe, T., Hashimoto, K., Takase, H., Kikuchi, O., 1997 J. Mol. Struct. (THEOCHEM) 397, 113–119.
Weber, C., Wider, G., Van Freyberg, B., Traber, R., Braun, W., Widmer, H., Wutrich, K., 1991 Biochemistry 30, 3563.
Wiener, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profetajr, S., Wiener, P., 1984a J. Am. Chem. Soc. 106, 765–784.
Wiener, S.J., Singh, U.C., O'Donnel, T.J., Kollman, P., 1984b J. Am. Chem. Soc. 106, 6243–6245.
Williamson, M.P., 1992. In: Jones, C., Mulloy, B., Thomas A.H. (Eds.), Methods in Molecular Biology Humana Press, New Jersey.
Williamson, M.P., Waltho, J.P., 1992 Chem. Soc. Rev. 227, 1.
Wishart, D.S., Sykes, B.D., Richards, M., 1991 FEBS 293, 72.
Wolfe, S., Bruder, S., Weaver, D.F., Yang, K., 1988 Can. J. Chem. 66, 2703.
Woon, D.E., Dunning, T.H., Jr., 1993 J. Chem. Phys. 98, 1358–1371.
Wright, L.R., Borkman, R.F., 1980 J. Am. Chem. Soc. 102, 6207–6210.
Wuthrich, K., 1986. NMR of Proteins and Nucleic Acids. Wiley and Sons, New York.
Wu, D., Zhu, Y., Gao, Z. and Qu, S. 1993 Wuahu Daxue Xuebao Ziran Kexuabu, 78–82.
Wuthrich, K., 1990 J. Biol. Chem. 265, 22059.
Yamauchi, O., 1995 Pure Appl. Chem. 67, 297–304.
Yarkony, D.R. (Ed.), 1995. Modern Electronic Structure Theory, vol. I and II. World Scientific, Singapore.
Yu, D., Armstrong, D.A., Rauk, A., 1992 Can. J. Chem. 70, 1792.
Yu, D., Rauk, A., Armstrong, D.A., 1995 J. Am. Chem. Soc. 117, 1789–1796.
Zhang, K., Zimmerman, D.M., Chung-Phillips, A., Cassaday, C.J., 1993 J. Am. Chem. Soc. 115, 10812.
Zimmerman, S.S., Pottle, M.A., Némethy, G., Scheraga, H.A., 1977 Macromolecules 10, 1–9.