PEPTIDE POTENTIAL ENERGY SURFACES AND PROTEIN FOLDING

Torrens, F.

Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P. O. Box 22085, E-46071 València, Spain.

FAX: +34-963543274, E-mail: Francisco.Torrens@uv.es

Castellano, G.

Departamento de Ciencias Experimentales, Facultad de Ciencias Experimentales, Universidad Católica de Valencia *San Vicente Mártir*, Guillem de Castro-106, E-46003 València, Spain.

Received February 10th, 2006. In final form March 14th, 2006
Dedicated to Prof. Imre G. Csizmadia on the occasion of his 75th birthday

Abstract

This report outlines the utility of a 3D→1D transformation of peptide conformation, which leads to a linearized notation of protein secondary and tertiary structures that may be used for an objective description of protein folding. The method is intended to be descriptive and not to be predictive. It is established from first principles that the idealized 2D-ψ–ϕ map must have nine minima. It is obvious to ask whether all these nine conformations are actually occurring in proteins. The objective is to repeat a previous analysis of 258 proteins determined using program ECEPP2, with the improved ECEPP2 + polarization. An analysis is performed on 258 proteins with known X-ray structure. The proteins contain 56 495 amino-acid residues with well-defined ϕ and ψ angles. The minima are identified with the aid of the nine ECEPP2 minima of Ac–Ala–NHMe with ϕ and ψ ±40° tolerance. ECEPP2 is improved with the inclusion of the interacting induced-dipole polarization model, SIMPLEX-MS-3 geometry optimization and the calculation of the dipole moment from the point distribution of net charges. The analysis of 258 proteins determined using ECEPP2 is repeated with the improved ECEPP2 + polarization. The relative frequency of occurrence of those conformations energetically favoured for enantioners g^-g^-, etc. in the ψ–ϕ map of the backbone conformations of amino acids decreases as: $g^-a/g^+a > g^-g^+/g^+g^- > g^-g^-/g^+g^+ >> ag^+/ag^- > aa$. For the amino acids, the same preference diminishes as: Pro >> Ile > Val > Leu > Thr > Met > Ala > Glu > Phe > Trp > Tyr > Gln > Lys > Ser > Cys > Arg > Asp > His > Asn > Gly. The strong preference of Pro is in agreement with its character of α-helix and β-sheet breaker, and β-turn and random-coil former. The analysis of 258 proteins determined using ECEPP2 is repeated with the improved ECEPP2 + polarization and there is a good agreement between the two. Achiral Gly relative frequencies of occurrence are close to one. Pro is the amino acid with the greatest (g^-g^-, etc.)/(g^+g^+, etc.) preference and with the greatest influence on protein conformation. Pro is the amino acid with the largest P_{global} conformational parameter. The original software used in the investigation is available from the author.

Resumen

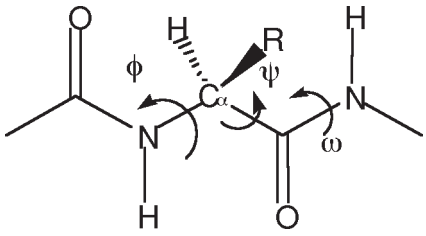
Este reporte reseña la utilidad de una transformación de conformación de péptido 3D→1D, que conduce a una notación linealizada de estructuras de proteínas secundarias y terciarias la cual puede ser usada para una descripción objetiva del plegamiento de proteínas. El método tiene la intención de ser descriptivo y no predictivo. Desde los primeros principios se ha establecido que el mapa 2D-ψ–ϕ idealizado debe tener nueve mínimos. Es obvia la pregunta, entonces, si todas las nueve conformaciones ocurren realmente en proteínas. El objetivo es repetir un análisis previo, realizado con el programa ECEPP2 en 258 proteínas, con estructura de Rayos-X conocida, utilizando el mejorado ECEPP2 + polarización. Estas proteínas contienen 56496 residuos amino ácidos con ángulos ϕ y ψ bien definidos. Los mínimos son identificados con la ayuda de los nueve mínimos obtenidos para Ac–Ala–NHMe por ECEPP2 con tolerancia ± 40° para ϕ y ψ. ECEPP2 es mejorado con la inclusión del modelo de polarización de dipolo inducido SIMPLEX-MS-3en la optimización de geometrías y el cálculo del momento dipolar a partir de la distribución puntual de cargas netas. La frecuencia relativa de ocurrencia de aquellas conformaciones energéticamente favorecidas por los enantiómeros g^-g^-, etc. en el mapa ψ–ϕ de las conformaciones del esqueleto de amino ácidos decrece como: $g^-a/g^+a > g^-g^+/g^+g^- > g^-g^-/g^+g^+ >> ag^+/ag^- > aa$. Para los amino ácidos, la misma preferencia disminuye en el sentido: Pro >> Ile > Val > Leu > Thr > Met > Ala > Glu > Phe > Trp > Tyr > Gln > Lys > Ser > Cys > Arg > Asp > His > Asn > Gly. La fuerte preferencia de Pro está de acuerdo con su carácter rompedor de hélices alfa y capas beta y formador de giros beta y ovillos aleatorios. El análisis de 258 proteínas determinadas utilizando ECPP2 se repitió utilizando el mejorado ECEPP2 + polarización y hay buen acuerdo entre los dos métodos. Las frecuencias relativas de ocurrencia de Gly aquiral son próximas a uno. Pro es el amino ácido con la mayor preferencia (g^-g^-, etc.)/(g^+g^+, etc.) y con mayor influencia en la conformación de proteínas. Pro es el amino ácido con el mayor parámetro conformacional P_{global}. El software original utilizado en la investigación está disponible por parte del autor.

Introduction and Notation

Multidimensional conformational analysis (MCA) allows predicting, from the topology of the potential energy curves (PEC), the topology of the potential energy surface (PES) if the molecular system is ideal [1–3]. In the case of three-fold periodicity the 3×3 = 9 minima are energetically degenerate. This case is operative for two $–CH_3$ rotors as may be occurring in propane, and in molecules with two equivalent $–CH_3$ groups. If the component PECs continue to have three minima, but these minima are energetically non-degenerate, the resultant PES will have nine non-equivalent minima. In the case of the ideal PES, it was possible to make a statement that all nine minima have the same energy value; in the non-ideal case, it is possible to make an analogous statement that all nine minima have different energy values. However, it is not possible to predict what the energy spectrum of these nine minima might be, and what the relative stability of these minima could be. Nevertheless, by making an intuitive guess, it is suggested an order for the relative stabilities of the diagonal elements:

$$E(O_2) > E(O_1) > E(O_0) \quad (1)$$

where E is the energy. What is important to note is that PES for a single peptide unit (*cf.* Scheme 1)



Scheme 1

may be represented as:

$$E = E(\phi, \psi) \tag{2}$$

if ω is constant (usually $\omega = 180^{o}$). Nevertheless, taken into account that, from the viewpoint of the torsional potential, the ϕ and ψ rotations are demonstrated to be practically free, the corresponding Ramachandran (ψ–ϕ) maps are determined by the non-bonding and hydrogen bonding (H-bonding) interactions, for each amino acid in a specific way.

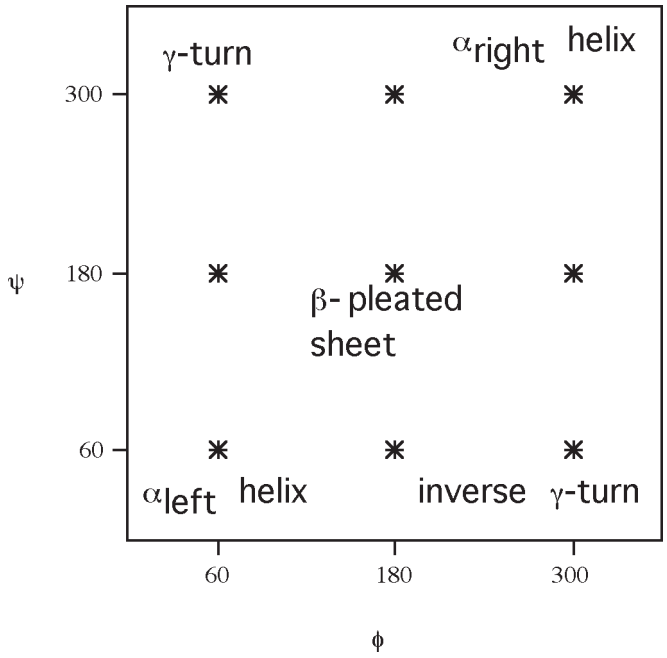


Figure 1. *Idealized PES topology for a single amino-acid residue indicating the five minima already identified in the protein literature.*
(The idealized location of the minima is specified by stars.)

Nine minima are expected to be present on the surface (*cf*. Figure 1). However, only five out of the nine minima have been recognized earlier in the literature, which are labelled as left-handed helix, right-handed helix, extended-like conformation, γ-turn and inverse γ-turn.

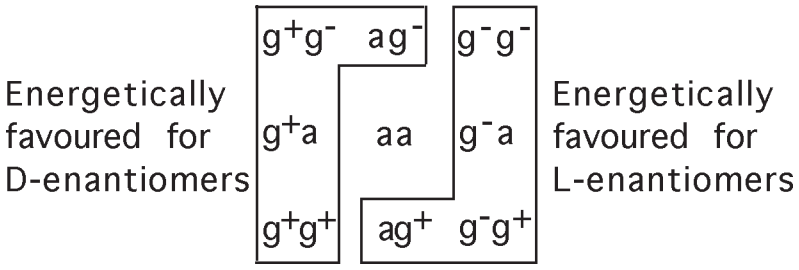
In Figure 1 both ϕ and ψ vary between zero and 360°. However, protein chemists adopted a range for both ϕ and ψ that runs between –180° and 180°, covering both clockwise and counter-clockwise rotations, which may be labelled as standard (STD):

$$-180^{o} \leq \phi_{STD} \leq 180^{o} \qquad (3)$$
$$-180^{o} \leq \psi_{STD} \leq 180^{o}$$

The representation is more useful, as topological (TOP) relationships can be recognized with a greater ease:

$$0^{o} \leq \phi_{TOP} \leq 360^{o} \qquad (4)$$
$$0^{o} \leq \psi_{TOP} \leq 360^{o}$$

More important is the fact that, apart from the central minimum (β-pleated sheet), the minima occur in pairs. Thus, the remaining unassigned four minima (Figure 1) could be regarded as two pairs of minima. Apart from the *aa* conformation, the most important, that is the energetically most favoured, conformations for the L-enantiomer are at the *extreme* and *lower right* of Scheme 2 (*g*, gauche, *a*, anti), and for the D-enantiomer the most favoured conformations are at the *upper* and *extreme left*. The topological relationship of the two families of conformations ($\{g^{-}g^{-},g^{-}g^{+},ag^{+},g^{-}a\}$ and $\{g^{+}g^{+},g^{+}g^{-},ag^{-},g^{+}a\}$) is illustrated (Scheme 2).



Scheme 2

In order to refer to the as of yet unassigned conformations, the midpoint at the *top* is labelled as ag^{-} and the midpoint at the *bottom* is labelled as ag^{+}. The midpoint at the *left* is labelled as $g^{+}a$ and the midpoint at the *right* is labelled as $g^{-}a$. Utilizing the labels used previously to denote the location of the minima, the following arrangement is obtained. For glycine (Gly) where no chiral centre exists, the *aa* conformation is to be located at the geometric centre [4]. For L-amino acids, the position of the *aa* conformation is shifted towards the *lower-right* hand corner. Similarly, for D-amino acids the position of the *aa* conformation is shifted towards the *upper-left* hand corner of the idealized topological scheme (Scheme 2) which represents only a different cut of the PES as illustrated by the broken lines in Figure 2.

For certain molecular residues, molecular computations established the actual location of the nine minima (Scheme 2). The values of ϕ and ψ deviate somewhat from the ideal values. Table 1 lists these numerical values for *N*-formylalaninamide (For–Ala–NH_2) [3]. Typical absolute errors for *folded* gauche–gauche $g^{-}g^{-}$–$g^{+}g^{+}$–$g^{-}g^{+}$–$g^{+}g^{-}$, completely-extended *fully*-planar anti–anti *aa* and *semifolded* gauche–anti ag^{+}–ag^{-}–$g^{-}a$–$g^{+}a$ are 15.9, 11.3 and 17.1°, respectively (14.8° on average). In particular, the extended *aa* conformation shows a smaller error, the *semifolded ag* conformations, a greater error, and the *folded gg* conformations, an intermediate

error. Therefore, the error is smaller for extended and *folded* structures and greater for *semifolded* structures.

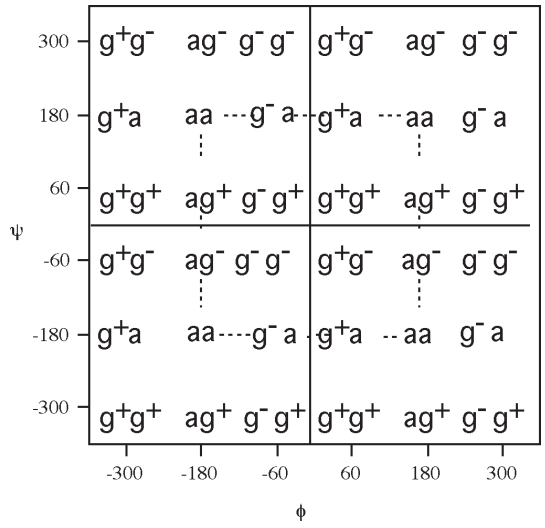


Figure 2. *Idealized PES topology for a single amino-acid residue involving two complete cycles of rotation in both* ϕ *and* ψ *(g, gauche, a, anti).*

In a PES associated with an ideal molecular system, minima, saddle points and maxima occur in a predictable regular pattern. It is customary to denote these critical points with the number of negative eigenvalues of the Hessian matrix, with elements:

$$H_{ij} = \frac{\partial^2 E}{\partial x_i \, \partial x_j} \tag{5}$$

where $[x_i, x_j]$ are any pair of the total of n variables including $[\phi, \psi]$. The number of negative eigenvalues of the Hessian is usually referred to by the index λ of the critical point. For ordinary surfaces n varies between zero and two ($0 \leq \lambda \leq 2$):

$$\lambda = 0 \text{ for minima}$$

$$\lambda = 1 \text{ for saddle points} \tag{6}$$

$$\lambda = 2 \text{ for maxima}$$

For potential energy hypersurfaces (PEHS):

$$0 \leq \lambda \leq n \tag{7}$$

for minima $\lambda = 0$, for maxima $\lambda = n$, and in between are located the transition-state points with a variety of indices ranging from one to $n - 1$. Figure 3 again shows an ideal surface as applied to a single peptide residue.

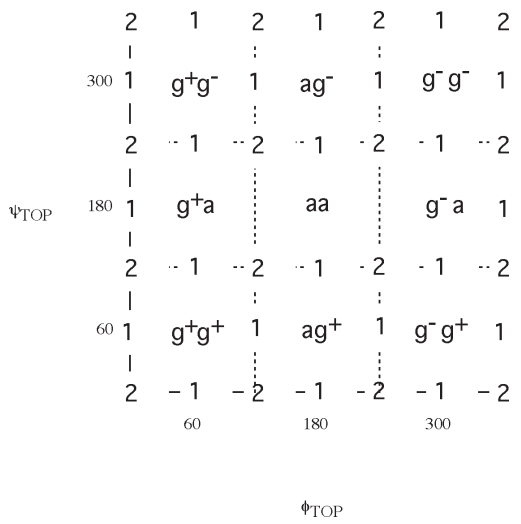


Figure 3. *The topology of an idealized two-dimensional (2D)-ψ–ϕ map containing the a priori predicted nine minima for a single amino-acid residue (…–CONH–CHR–CONH–…). The horizontal and vertical dashed lines represent low lying mountain ridges that separate the nine distinctly different catchment regions (g, gauche, a, anti). [Notice that the topologically (TOP) useful regions of ϕ and ψ are given in a 0–360° range.] Numerals indicate the expected location of saddle points (λ = 1) and maxima (λ = 2).*

In Figure 3 the minima are not labelled by 0 but by the letters introduced earlier g^-g^-, *aa*, g^-g^+, *etc*., but critical points of higher indices are denoted by their λ values, *viz*. 1, and 2. There are two points to note about Figure 3. (1) The minima are separated from each other by mountain ridges containing maxima and saddle points. Each valley contains a single minimum and these valleys are normally referred to, after Mezey [5], as catchment regions. (2) In Figure 3, the indices of the PES may be calculated from the indices of the appropriate PEC if Mezey's criteria are fulfilled:

$$\lambda\,(\chi_1, \chi_2) = \lambda\,(\chi_1) + \lambda\,(\chi_2) \tag{8}$$

It was established from first principles that the idealized 2D-ψ–ϕ PES (Table 1 and Figure 3) must have nine minima. It is, therefore, an obvious question to ask whether all these conformations are actually occurring in proteins. Perczel *et al*. [6] analyzed 258 proteins with known X-ray structure [7,8], which contained 56 495 amino-acid residues with well-defined ϕ and ψ angles. They identified the minima with the aid of those of *N*-acetyl-*N'*-methylalaninamide (Ac–Ala–NHMe) determined with the ECEPP2 method [9,10], allowing a ±40° tolerance in the ϕ and ψ values. Perczel *et al*. [11] concluded the following. (1) The *non-assigned* conformations are quite large, indicating that Ac–Ala–NHMe may not be as good a model to mimic a single amino-acid residue in a protein than hitherto might have been believed. (2) Gly has the greatest number of non-assigned cases implying that the alanine (Ala) derivative, which has a side chain, may be a much better model to all amino-acid residues with side chains than Gly, which has no

side chain. (3) Since Gly is achiral, instead of nine only five unique conformations occur ($g^-g^- = g^+g^+, g^-g^+ = g^+g^-, ag^+ = ag^-, g^-a = g^+a$). The actual finding is not all that far from expectation: $g^-g^- = 850$, $g^+g^+ = 631$, $g^-g^+ = 79$, $g^+g^- = 160$, $ag^+ = 62$, $ag^- = 45$, $g^-a = 388$ and $g^+a = 324$. The actual degeneracy is lost in the 1799 non-assigned conformations. (4) Phenylalanine (Phe) has no g^+g^- conformation, and proline (Pro) has no g^+a and g^+g^- conformations. All other amino-acid residues do occur in all the possible nine conformations. One of the authors, F.T., met Prof. Csizmadia during his postdoctoral stage on protein modelling, working for the Centre National de la Recherche Scientifique (CNRS) Molecular Modelling Scientific Group, IBM–CNRS–Université de Nancy I (1991-1992). In our joint collaboration with Prof. Rivail to molecular modelling, he always advised with good courage and mood, constantly trying to extend the *ab initio* quantum chemical picture of the subject. In earlier publications, the dipeptide model *N*-formylglycinamide (For–Gly–NH_2) was studied with molecular mechanic polarizing force fields implemented in MM2 [12,13] and ECEPP2 [14]. The aim of the present study is to repeat a previous analysis of 258 proteins determined using ECEPP2, with the improved ECEPP2 + polarization. Section 2 describes the computational method. Section 3 present and discusses the calculation results. Section 4 summarizes the conclusions.

Computational Method

A frequently used molecular mechanics method for peptides is the empirical conformational energy program for peptides version 2 (ECEPP2) [9,10]. The force field describes the molecular *steric* energy as a sum of the electrostatic, non-bonded, torsional, cystine torsional and loop-closing energy components. ECEPP2 provides the following functionalities: (1) study of linear polypeptides and those polypeptides that include intramolecular disulphide bonds, (2) calculation of the conformational energy for any sequence of residues and any set of dihedral angles, (3) comparison of the relative energies of the different conformations of a given polypeptide; (4) a standard file of residues is provided, which includes 26 amino acid residues and 20 terminal groups; (5) the user can eventually provide complementary residues or replace the standard residues by its own. The auxiliary program chemical modelling application platform (CMAP, B. T. Luke, IBM) can serve as an access platform to ECEPP2, for which it offers the following functionalities: (1) aid in the preparation of the data and job-command-language needed for the submission of an ECEPP2 work, (2) gateway with all the other programs to which CMAP gives access and (3) visualization of the studied polypeptide. CMAP integrates ECEPP2 as calculation program: ECEPP2 calculations of *reasonable* size can then be interactively executed under CMAP. The following improvements have been implemented in ECEPP2 [14]: (1) inclusion of the interacting induced-dipole polarization model by the method of Applequist [15], (2) geometry optimization by SIMPLEX-MS-3 algorithm [16] and (3) calculation of the dipole moment from the point distribution of atomic net charges. The modifications have been also implemented in programs molecular mechanics (MM2) [17] and molecular mechanics extended for coordination complexes of transition metals (MMX) [18–21].

Two methods for the calculation of the effect of the induced dipole moments on the polarization energy term have been proposed, *viz.* the polarization procedure by non-interacting induced dipoles (NID), and the polarization scheme by interacting induced dipoles (ID) [12–14]. NID assumes scalar isotropic atomic polarizabilities. ID allows the interaction of the induced

dipole moments by means of tensor effective anisotropic atomic polarizabilities. The atomic polarizabilities used (NID) and obtained (ID) for For–Gly–NH_2 (*cf*. Table 2) show that for ECEPP2, the total molecular polarizabilities are greater with ID than with NID. The atomic polarizabilities of the H, C and N atoms are greater with ID; however, the atomic contributions from the O atom are greater with NID. For the five ID minima, similar atomic and total molecular polarizabilities are obtained. For MM2, the total molecular polarizabilities are greater for NID than for ID. The atomic polarizabilities of the N and O atoms are greater with NID; however, the atomic contributions from the H and C atoms are greater with ID. For *aa* and g^-g^+, similar ID total molecular polarizabilities are obtained. Effective atomic and total molecular polarizabilities increase in the order *gg* < *ag* < *aa*, *i.e*. *folded* < *semifolded* < extended conformation.

A previous analysis of 258 proteins determined using ECEPP2 has been repeated with ECEPP2 + polarization. The set of Protein Data Bank (PDB) structures is the same used by Perczel *et al*. [6,11]. The use of ECEPP2 + polarization followed two strategies: (1) double scan of the idealized 2D-ψ–ϕ maps (just as Perczel *et al*. used ECEPP2) and (2) geometry optimization of the ϕ–ψ angles with SIMPLEX–MS–3. Both plans reached the same set of minima.

Calculation Results and Discussion

ECEPP2 + polarization has been applied to the calculation of the five minima of the conformational PES of For–Gly–NH_2. The minima were described by Perczel *et al*. [3] with ECEPP2 (grid geometry optimization) and *ab initio* (second-derivatives optimization). The g^-g^-, g^-g^+, ag^+ and g^-a minima are *folded* conformations while the fully-planar *aa* minimum is all-*trans* extended. The ECEPP2 + polarization calculations have been optimized with SIMPLEX MS-3. The total energy differences (*cf*. Table 3) are compared with MM2 and *ab initio* SCF 3-21G references [3]. Five structures are found with the ECEPP2 methods, two with the MM2 methods and four with *ab initio*. The three types of methods show only *aa* and g^-g^+ structures at the same time. These are the only minima with MM2, as well as the two main minima with ECEPP2 and *ab initio*. The ECEPP2 + polarization relative energies of the local *aa* minimum are in agreement with the reference calculations, lying between the MM2+ID and *ab initio*. Intramolecular H-bonds contribute to the stabilization of the g^-g^+ conformers. The local g^-g^- and *aa* minima are stabilized by one H-bond forming a five-membered ring N–H…N (g^-g^-) or N–H…O (*aa*); the global g^-g^+ and local ag^+ minima show two shared H-bonds forming a five-membered ring N–H…N and closing a seven-membered ring N-H…O (g^-g^+), or forming two shared five-membered rings N–H…N (ag^+); the local g^-a minimum shows no H-bond.

Table 1. *Optimized ϕ, ψ Torsional Angle Pairs for For–Ala–NH_2 and the Idealized Torsional Angle Pairs*

	Optimized values		Idealized values	
Conformational classification[a]	ϕ	ψ	ϕ	ψ
g^-g^-	-66.6	-17.5	-60	-60
g^+g^+	61.8	31.9	60	60
aa	-167.6	169.9	-180	180
g^-g^+	-84.5	68.7	-60	60
g^+g^-	74.3	-59.5	60	-60
ag^+	-126.6	26.5	-180	60
ag^-	-179.6	-43.7	-180	-60
g^-a	-74.7	167.8	-60	180
g^+a	64.7	-178.6	60	-180

[a] *g*, gauche; *a*, anti.

There are 20 naturally occurring amino acids. A total of 18 of them have the same type of backbone folding, *i.e*. nine discrete conformations (Table 1 and Figure 3). The two other amino acids are exceptions. One exception is Pro, which is built into proteins like any other amino acid, but its N atom is locked in a five-membered ring. For Pro, ϕ can only be in the vicinity of –60º and, therefore, only three backbone conformations are possible, *viz*. g^-g^-, g^-a, and g^-g^+. The other unique amino acid is Gly, which is achiral. In the case of Gly, double degeneracy occurs in its conformational PES ($g^-g^- = g^+g^+$, $g^-g^+ = g^+g^-$, $ag^+ = ag^-$, $g^-a = g^+a$). Pro is fundamentally different from all the other 18 chiral amino acids in more than one respect: (1) the R group forms a five-membered ring with the backbone; (2) there is no peptidic N–H group in the residue to be involved in H-bonding; (3) since there are two C atoms connected to the N atom, there is a greater chance of *cis*/*trans* isomerization in the peptide bond.

Table 2. *Atomic Polarizabilities (in Å^3) Used (ECEPP2+NID)[a] and Obtained (ECEPP2+ID)[b] in the Calculation of the Polarization Energy for For–Gly–NH_2 conformations*

		ECEPP2+NID[a]	ECEPP2+ID[b]					MM2+NID	MM2+ID	
Residue	At.	g^-g^-,aa,g^-g^+, ag^+,g^-a	g^-g^-,[c] g^+	aa[c]	g^-g^+	ag^+	g^-a	aa, g^-	aa	g^-g^+
Formyl	H	0.407	1.666	1.676	1.665	1.669	1.669	0.407	1.671	1.665
	O	1.395	0.361	0.369	0.362	0.366	0.361	1.395	0.364	0.362
	C	0.075	0.635	0.642	0.635	0.636	0.634	0.075	0.636	0.634
Glycine	N	0.628	1.056	1.050	1.052	1.054	1.050	2.255	0.212	0.211

(to be continued)

	HN	0.092	0.100	0.100	0.099	0.101	0.100	0.010	0.162	0.162
	CA	1.027	1.349	1.353	1.349	1.351	1.351	1.027	1.349	1.345
	HA	0.407	1.698	1.716	1.713	1.710	1.720	0.407	1.684	1.685
	HA	0.407	1.688	1.716	1.686	1.699	1.697	0.407	1.684	1.676
	C	0.075	0.628	0.628	0.628	0.628	0.628	0.075	0.628	0.627
	O	1.395	0.357	0.358	0.357	0.360	0.358	1.395	0.357	0.356
Carboxyl	N	0.628	1.057	1.073	1.059	1.055	1.060	2.255	0.213	0.211
$-NH_2$										
	H2	0.092	0.098	0.102	0.099	0.098	0.101	0.010	0.163	0.159
	H2	0.092	0.097	0.099	0.097	0.098	0.098	0.010	0.159	0.157
Total	–	6.721	10.792	10.883	10.802	10.824	10.827	9.729	9.281	9.250

[a] NID: polarization by non-interacting induced dipoles.
[b] ID: polarization by interacting induced dipoles.
[c] *g*, gauche, *a*, anti, $g^-g^- = g^+g^+$, $g^-g^+ = g^+g^-$, $ag^+ = ag^-$, $g^-a = g^+a$.

Table 3. *Molecular mechanics (ECEPP2) results for For–Gly–NH_2 conformations. Number of H-bonds and total energy differences in kJ·mol^{-1}*

Backbone conformation[a]	No. of H-bonds	ECEPP2	ECEPP2+NID[b]	ECEPP2+ID[c]	MM2	MM2+NID	MM2+ID	Ref.[d]
g^-g^-	1	6.3	6.1	0.2	–[e]	–[e]	–[e]	18.6
aa	1	6.1	7.0	9.2	21.5	22.0	16.2	2.6
g^-g^+	2 (shared)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ag^+	2 (shared)	6.9	7.9	5.9	–[e]	–[e]	–[e]	13.7
g^-a	0	9.2	6.3	9.0	–[e]	–[e]	–[e]	–[e]

[a] *g*, gauche, *a*, anti, $g^-g^- = g^+g^+$, $g^-g^+ = g^+g^-$, $ag^+ = ag^-$, $g^-a = g^+a$.
[b] NID: polarization by non-interacting induced dipoles.
[c] ID: polarization by interacting induced dipoles.
[d] Reference: *ab initio* SCF 3-21G (optimized geometry) taken from Reference 3.
[e] A dash (–) indicates no local minimum for this conformation.

Notice that all nine conformations do occur in proteins (*cf.* Table 4) [11]. For symmetric conformational pairs, *e.g.* g^-g^- (which is capable of producing a right-handed helix) and g^+g^+ (which is capable of generating a left-handed helix) of a given L-amino acid, *e.g.* Ala, g^-g^- is more stable than g^+g^+. In proteins, therefore, the frequency of occurrence of the Ala residue in the g^-g^- conformation (2593 right-handed helix in 4894 conformations) is greater than that of g^+g^+ (54 left-handed helix in 4894 conformations); therefore, the ratio of frequencies of occurrences g^-g^-/g^+g^+ is much greater than unity (2593/54 = 48.019 >> 1). The only exception is achiral Gly, where g^-g^- is the specular image of g^+g^+ with the same energy. In proteins, therefore, the frequencies of occurrences of Gly in g^-g^- (850 right-handed helix in 4798 conformations) and g^+g^+ (631 left-handed helix in 4798 conformations) have practically identical relative abundance (*i.e.* g^-g^- H" g^+g^+, $g^-g^-/g^+g^+ = 850/631 = 1.347 \approx 1$). The ratio is closer to unity in the total

$(g^-g^- + g^-g^+ + ag^+ + g^-a)/(g^+g^+ + g^+g^- + ag^- + g^+a) = 1.189 \approx 1$. In general, the relative frequency of occurrence of these energetically favoured conformations of the 20 residues in proteins decreases as: $g^-a/g^+a > g^-g^+/g^+g^- > g^-g^-/g^+g^+ >> ag^+/ag^- > aa/aa = 1$. There is good agreement between ECEPP2 and ECEPP2 + polarization results. The results for all the amino acids relative to Gly are also calculated. For the 20 amino acids, there are g^-g^-/g^+g^+, *etc*. preferences, which diminish as: Pro >> Ile > Val > Leu > Thr > Met > Ala > Glu > Phe > Trp > Tyr > Gln > Lys > Ser > Cys > Arg > Asp > His > Asn > Gly. In particular, Pro is largely the amino acid with the gratest value of total $(g^-g^- + g^-g^+ + ag^+ + g^-a)/(g^+g^+ + g^+g^- + ag^- + g^+a)$ relative to Gly (934.561), while the 19 other amino acids show this ratio in the range 1–42. Again, Pro is the amino acid with the greatest g^-g^-/g^+g^+, *etc*. ratios, because the Pro ring serves to intrinsically restrict its ϕ dihedral angle *ca*. –60º. This is consistent with the fact that Pro strongly favours ϕ dihedral angles *ca*. –60º [Pro conformations are fairly tightly clustered in the range $\phi = (-63 \pm 15)^\circ$] [22]. Therefore, Pro greatly influences protein conformation.

Table 4. *Relative Frequency of Occurrence of the Backbone Conformations*[a] *of Amino-Acid (AA) Residues in Proteins*

E.	AA	g^-g^-/g^+g^+	g^-g^+/g^+g^-	ag^+/ag^-	g^-a/g^+a	Tot.num./ tot. den.	g^-g^-/g^+g^+ rel. Gly	g^-g^+/g^+g^- rel. Gly	ag^+/ag^- rel. Gly	g^-a/g^+a rel. Gly	T.n./t.d. rel. Gly
1	Ala	48.019	4.083	1.814	66.083	24.903	35.647	8.270	1.317	55.183	20.949
2	Arg	12.911	21.800	5.700	101.000	16.108	9.584	44.152	4.137	84.340	13.550
3	Asn	3.028	13.750	4.833	41.667	4.814	2.248	27.848	3.508	34.794	4.050
4	Asp	18.309	14.333	3.106	37.167	14.446	13.592	29.030	2.255	31.036	12.152
5	Cys	25.421	25.333	0.800	64.600	17.346	18.871	51.308	0.581	53.944	14.591
6	Gln	21.357	14.333	2.000	62.167	19.333	15.855	29.030	1.452	51.912	16.263
7	Glu	53.538	7.929	1.481	33.417	24.608	39.744	16.058	1.075	27.905	20.700
8	Gly	1.347	0.494	1.378	1.198	1.189	1.000	1.000	1.000	1.000	1.000
9	His	11.962	15.200	4.167	48.500	12.241	8.880	30.785	3.024	40.500	10.297
10	Ile	69.714	55.500	2.333	183.000	49.500	51.753	112.405	1.694	152.814	41.639
11	Leu	42.333	26.727	2.649	68.071	31.009	31.426	54.131	1.922	56.843	26.084
12	Lys	24.915	9.095	2.200	51.583	18.521	18.496	18.421	1.597	43.075	15.580
13	Met	28.077	58.000	1.857	69.000	24.957	20.843	117.468	1.348	57.619	20.993
14	Phe	25.419	∞	2.778	196.000	23.950	18.870	∞	2.016	163.670	20.146
15	Pro	936.000	∞	3.000	∞	1111.000	694.842	∞	2.177	∞	934.561
16	Ser	24.329	6.321	3.286	39.667	17.503	18.060	12.803	2.385	33.124	14.723
17	Thr	83.467	12.500	1.327	63.667	25.067	61.962	25.316	0.963	53.165	21.086
18	Trp	130.667	4.000	2.250	158.000	23.731	97.001	8.101	1.633	131.938	19.962
19	Tyr	15.694	48.667	5.067	146.333	21.509	11.651	98.565	3.677	122.196	18.093
20	Val	115.714	46.000	1.057	113.375	35.554	85.901	93.165	0.767	94.674	29.908

[a] *g*, gauche, *a*, anti.

Table 5. *Conformational Parameters of the Backbone Conformations of Various Amino-Acid Residues in Proteins*

Entry	Amino acid	P_α [a]	P_β [b]	P_t [c]	P_c [d]	P_{global} [e]	P_α rel. Gly	P_β rel. Gly	P_t rel. Gly	P_c rel. Gly	P_{global} rel. Gly
1	Ala	1.42	0.83	0.66	0.66	-0.93	2.491	1.107	0.423	0.465	-0.560
2	Arg	0.98	0.93	0.95	1.20	0.24	1.719	1.240	0.609	0.845	0.145
3	Asn	0.67	0.89	1.56	1.33	1.33	1.175	1.187	1.000	0.937	0.801
4	Asp	1.01	0.54	1.46	1.09	1.00	1.772	0.720	0.936	0.768	0.602
5	Cys	0.70	1.19	1.19	1.07	0.37	1.228	1.587	0.763	0.754	0.223
6	Gln	1.11	1.10	0.98	0.79	-0.44	1.947	1.467	0.628	0.556	-0.265
7	Glu	1.51	0.37	0.74	0.87	-0.27	2.649	0.493	0.474	0.613	-0.163
8	Gly	0.57	0.75	1.56	1.42	1.66	1.000	1.000	1.000	1.000	1.000
9	His	1.00	0.87	0.95	0.92	0.00	1.754	1.160	0.609	0.648	0.000
10	Ile	1.08	1.60	0.47	0.78	-1.43	1.895	2.133	0.301	0.549	-0.861
11	Leu	1.21	1.30	0.59	0.66	-1.26	2.123	1.733	0.378	0.465	-0.759
12	Lys	1.16	0.74	1.01	1.05	0.16	2.035	0.987	0.647	0.739	0.096
13	Met	1.45	1.05	0.60	0.61	-1.29	2.544	1.400	0.385	0.430	-0.777
14	Phe	1.13	1.38	0.60	0.81	-1.10	1.982	1.840	0.385	0.570	-0.663
15	Pro	0.57	0.55	1.52	1.45	1.85	1.000	0.733	0.974	1.021	1.114
16	Ser	0.77	0.75	1.43	1.27	1.18	1.351	1.000	0.917	0.894	0.711
17	Thr	0.83	1.19	0.96	1.05	-0.01	1.456	1.587	0.615	0.739	-0.006
18	Trp	1.08	1.37	0.96	0.82	-0.67	1.895	1.827	0.615	0.577	-0.404
19	Tyr	0.69	1.47	1.14	1.19	0.17	1.211	1.960	0.731	0.838	0.102
20	Val	1.06	1.70	0.50	0.66	-1.60	1.860	2.267	0.321	0.465	-0.964
21	mean	1.00	1.03	0.99	0.99	-0.05	1.754	1.371	0.636	0.694	-0.031

[a] P_α: conformational parameter for the α-helix.
[b] P_β: conformational parameter for the β-sheet.
[c] P_t: conformational parameter for the β-turn.
[d] P_c: conformational parameter for random coil.
[e] $P_{global} = -P_\alpha - P_\beta + P_t + P_c$.

Table 6. *Conformational Assignment for the First Eleven Residues of a Brookhaven Protein Data Bank Protein*

Resd. No.	Amin. Acid	ω	ϕ	ψ	χ_1	χ_2	χ_3	χ_4	Conformational assignment
1	Met	–	-169.082	158.512	-138.927	178.567	-177.661	-179.430	*aa*
2	Val	-173.242	-107.521	139.671	156.028	170.935[a]	170.148[b]	–	g^-a
3	Leu	178.235	-157.078	85.782	-148.991	145.809	-175.891[c]	-178.671[d]	ag^+
4	Thr	161.617	-74.386	150.934	-168.274	-179.685[a]	-168.911[b]	–	g^-a
5	Val	176.353	-136.295	134.941	174.424	-173.167[a]	-161.362[b]	–	*aa*

6	Thr	-164.820	172.599	156.311	165.695	149.269[a]	-169.563[b]	–	*aa*
7	Leu	-178.544	-144.883	64.536	-131.655	69.193	159.835[c]	148.049[d]	ag^+
8	Asn	139.486	-128.851	86.671	-114.970	-23.599	-175.643[d]	–	ag^+
9	Pro	129.860	–	108.488	–	–	–	–	g^-g^+
10	Ala	-178.255	-117.695	164.137	-172.995	–	–	–	g^-a
11	Leu	172.214	-97.690	105.868	-160.058	163.966	178.083[c]	-172.238[d]	g^-g^+

[a] χ_{21}. [b] χ_{22}. [c] χ_{31}. [d] χ_{32}.

For the different conformations, the g^-g^-/g^+g^+, *etc*. comparative frequencies of occurrences relative to Gly (*cf*. Figure 4) show that g^-a/g^+a is the conformational parameter with the greatest variability.

For the different conformations (Table 4), the trend lines of the g^-g^-/g^+g^+, *etc*. comparative frequencies of occurrences relative to Gly are shown in Figure 5. Two data for Pro have been eliminated to obtain better detail. Again, g^-a/g^+a shows the greatest variability. The slope of the trend lines decrease as: $g^-g^-/g^+g^+ \approx$ total $>> g^-a/g^+a > g^-g^+/g^+g^- >> aa = 0 \approx ag^+/ag^-$.

Cluster analysis (CA) [23] was applied to the amino-acid residues in proteins. CA involved grouping the amino acids into clusters using hierarchical cluster analysis (HCA) [24]. There are many reasons why one might want to cluster a database of molecular structures [25–28]. A program has been written using the IMSL [29] subroutine CLINK to carry out HCA, based upon either a *distance* or a *similarity matrix*. Both single- and complete-linkage HCAs allow building the *dendrogram* (binary tree) for the amino acids, corresponding to frequencies of occurrence of the backbone conformations and their ratios $\{g^-g^-, g^+g^+, aa, g^-g^+, g^+g^-, ag^+, ag^-, g^-a, g^+a, g^-g^-/g^+g^+, g^-g^+/g^+g^-, ag^+/ag^-, g^-a/g^+a\}$ [30]. Both HCAs perform a binary taxonomy of the amino acids that separates first both units in class 1 (Gly and Pro, *cf*. Figure 6 *top*), then class 2 (nine units, *viz*. Ala, Arg, Asn, Asp, Gln, Glu, His, Leu, and Lys, *middle*) and, finally, class 3 (nine units, *viz*. Cys, Ile, Met, Phe, Ser, Thr, Trp, Tyr, and Val, *bottom*). In particular, Pro (class 1) is the first separated amino acid.

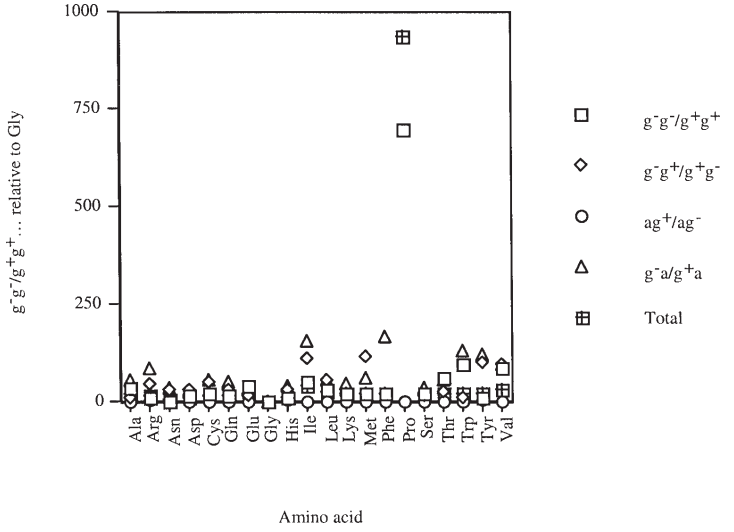


Figure 4. *Comparative frequency of occurrence of conformations of amino acids relative to Gly (g, gauche, a, anti).*

From both HCAs, the radial tree for the amino acids relating to $\{g^-g^-, g^+g^+, aa, g^-g^+, g^+g^-, ag^+, ag^-, g^-a, g^+a, g^-g^-/g^+g^+, g^-g^+/g^+g^-, ag^+/ag^-, g^-a/g^+a\}$ separates first both units in class 1 (Gly and Pro, *cf.* Figure 7 *middle*), then class 2 (nine units, *viz.* Ala, Arg, Asn, Asp, Gln, Glu, His, Leu, and Lys, *bottom*) and, finally, class 3 (nine units, *viz.* Cys, Ile, Met, Phe, Ser, Thr, Trp, Tyr, and Val, *top*). Again, Pro (class 1) is separated first. The classes correspond to the dendrogram (Figure 6).

Using the known structure of 29 proteins as determined *via* X-ray crystallography, Chou and Fasman calculated the probabilities of α-helix [31], β-sheet [32], β-turn (sharp turn connecting β-strands) [33] and random coil (*cf.* Table 5). The conformational parameters P_α, P_β, P_t and P_c were defined as the frequency with which a particular residue is found in a structure, relative to the average frequency for all amino acids being found in that structure. By definition, the means $\langle P_\alpha \rangle = \langle P_\beta \rangle = \langle P_t \rangle = \langle P_c \rangle = 1$. In this study, a new conformational parameter $P_{\text{global}} = -P_\alpha - P_\beta + P_t + P_c$ is proposed. By definition, the mean $\langle P_{\text{global}} \rangle = 0$. Notice that $\langle P_{\alpha \text{ relative to Gly}} \rangle \neq 1$, *etc.*; furthermore, by definition, $\langle P_{\text{global relative to Gly}} \rangle = 0$.

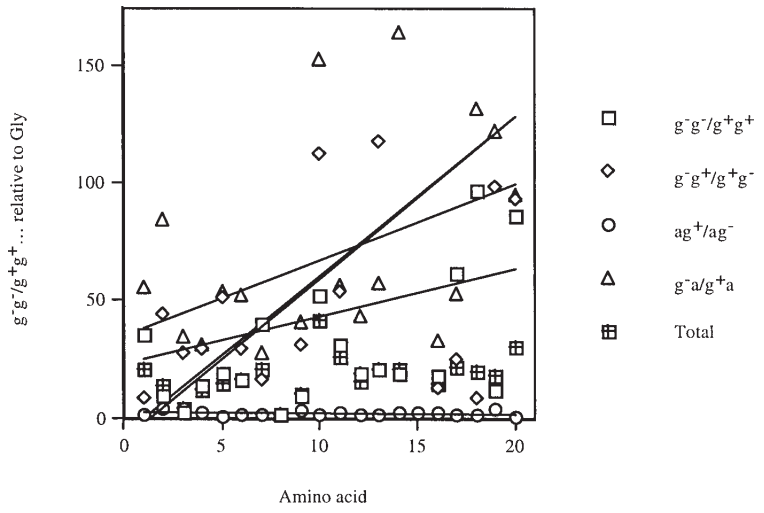


Figure 5. *Trend line of comparative frequency of occurrence of conformations relative to Gly (g, gauche, a, anti).*

In particular, it can be seen from the conformational parameters that Pro is a strong α-helix breaker, strong β-sheet breaker, strong β-turn former and strong random-coil former. This is consistent with the fact that Pro plays a particular role in peptide and protein structural biology as a β-turn-promoting unit. Pro, of course, is an imino, not amino, acid. The ring structure prevents H-bonding on the amide N atom, as well as makes its occurrence rare in β-sheet and α-helix. Instead, Pro along with Gly is more commonly found in β-turns [34–37], as well as rigid extended structural proteins, *e.g.* collagen and cuticle. Pro never participates directly in catalysis due to the chemical inertness of its methylene groups ($-CH_2-$), though it may line a substrate pocket or provide rigidity to an active site. Peptide bonds other than those with Pro have a double bond character, and two consecutive C^α are generally *trans* with respect to this plane of the adjacent amide bond. Pro still emulates this double bond angle *via* steric hindrance, with the ω dihedral

angle seldom varying by more than 15º from peptide-planar. While the *cis* amide conformation is not sterically forbidden for non-Pro amino acids in short peptides, the *trans*/*cis* ratio of the adjacent amide bond is nonetheless *ca*. 1000/1. For Pro, the *cis* imidic conformation (relative to the preceding residue) is less unfavourable, and the ratio of the adjacent imidic bond approaches 4/1. Although both conformers are in equilibrium [38,39], the activation energy is so high (*ca*. 80 kJ·mol^{-1} in model compounds [40]) that unassisted attainment of equilibrium can take minutes at physiological temperatures, much longer or never in large proteins [41].

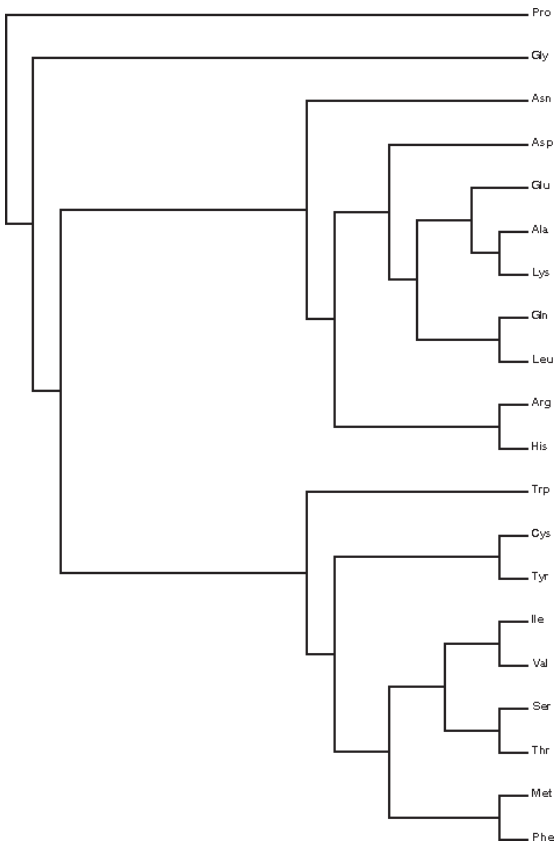


Figure 6. *Dendrogram for the amino-acid residues in proteins according to frequency of occurrence and relative to Gly.*

The strong structural character of Pro (Table 5) is in agreement with its strong relative frequency of occurrence of the conformations (Table 4). Therefore, a new conformational parameter is proposed to maximize this difference: $P_{\text{global}} = -P_{\alpha} - P_{\beta} + P_{t} + P_{c}$. The physical meaning of P_{global} is that this descriptor is high for an amino acid that is a strong α-helix breaker, strong β-sheet breaker, strong β-turn former and strong random-coil former. For the different amino acids, P_{global} decreases as: Pro > Gly > Asn > Ser > Asp > Cys > Arg > Tyr > Lys > His > Thr > Glu > Gln > Trp > Ala > Phe > Leu > Met > Ile > Val. As expected, Pro is the amino acid with the greatest value of P_{global}. The results for all the amino acids relative to achiral Gly are also calculated. The comparative conformational parameters relative to Gly (*cf*. Figure 8) shows that P_{global} is the conformational parameter with the greatest variability. The strong preference of Pro

is in agreement with its character of strong α-helix breaker, strong β-sheet breaker, strong β-turn former and strong random-coil former. A new conformational parameter P_{global} maximizes Pro distinguished character.

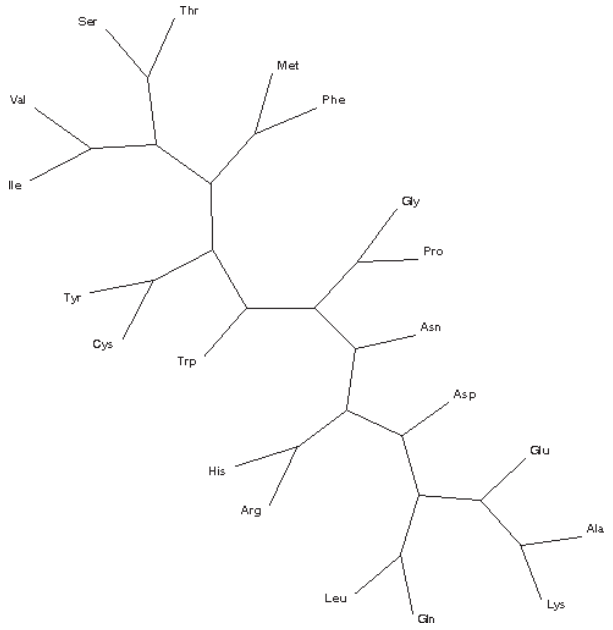


Figure 7. *Radial tree for the amino-acid residues in proteins according to frequency of occurrence and relative to Gly.*

The trend lines of comparative conformational parameters for the amino acids relative to Gly (Entries 1–20 in Table 5) are illustrated in Figure 9. The slopes of the trend lines decrease as: $P_\beta >> 0 > P_c > P_t > P_\alpha > P_{global}$; however, their absolute slopes diminish as: $P_\beta > P_{global} > P_\alpha > P_t > P_c$.

Figure 10 displays the variations of the conformational parameter P_{global} *vs*. the relative frequency of occurrence of the $(g^-g^-+g^-g^++ag^++g^-a)/(g^+g^++g^+g^-+ag^-+g^+a)$ amino-acid residues and P_{global} relative to Gly *vs*. the relative frequency of occurrence of the $(g^-g^-+g^-g^++ag^++g^-a)/(g^+g^++g^+g^-+ag^-+g^+a)$ amino acids relative to Gly. In both representations, the datum for Pro has been eliminated to obtain better deatil and fit. In particular, P_{global} drops quicker than $P_{global\ relative\ to\ Gly}$.

The regressions turn out to be, respectively:

$$P_{global} = 1.507 - 0.0776\left(g^-g^- + g^-g^+ + ag^+ + g^-a\right)\Big/\left(g^+g^+ + g^+g^- + ag^- + g^+a\right) \qquad r = 0.839 \tag{9}$$

$$P_{global\ rel.\ Gly} = 0.908 - 0.0556\left(g^-g^- + g^-g^+ + ag^+ + g^-a\right)\Big/\left(g^+g^+ + g^+g^- + ag^- + g^+a\right)_{rel.\ Gly} \tag{10}$$

$$r = 0.839$$

As expected, the correlation coefficient is equal after both ordinates and abscissas are divided by their corresponding values for Gly.

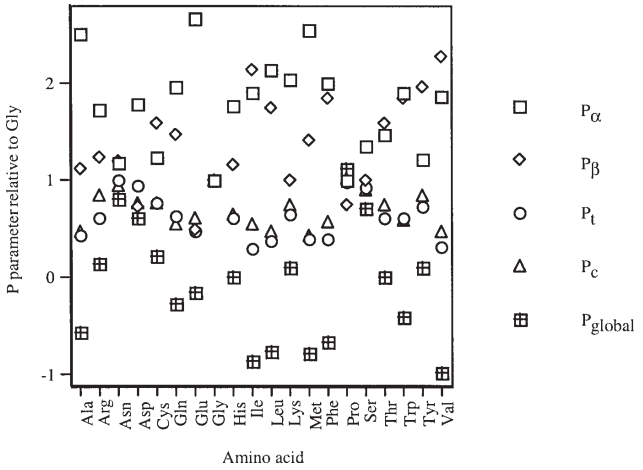


Figure 8. *Comparative conformational parameters of the backbone conformations of amino acids relative to Gly.*

From both HCAs, the radial tree for the amino acids relating to $\{P_{\alpha},P_{\beta},P_{t},P_{c},P_{global}\}$ separates the five units in class 1 (Asn, Asp, Gly, Pro and Ser, *cf*. Figure 11 *top*), class 2 (six units, *viz*. Ala, Arg, Glu, His, Lys, and Met, *bottom*) and, finally, class 3 (nine units, *viz*. Cys, Gln, Ile, Leu, Phe, Thr, Trp, Tyr, and Val, *left*), in moderate agreement with the dendrogram and radial tree obtained form the frequencies of occurrence of the backbone conformations and their ratios (Figures 6–7). In particular, the best agreement is observed for class 1, which includes Pro.

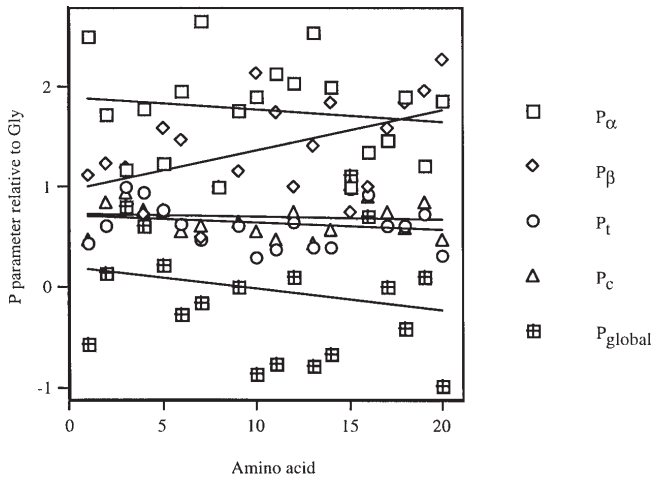


Figure 9. *Comparative conformational parameters of the backbone conformations of amino acids relative to Gly.*

The 4D-ψ–ϕ map clearly indicates the practicality of the assignment of the backbone conformation of a peptide. Scheme 3 shows the explicit form of the linearized notation of backbone conformation for cytochrome b_5 (PDB code 2B5C) [11]. The notation can be directly converted to a numerical format in a semi quantitative way, by using the idealized PES topology (Table 1 and Figure 3); *e.g*. Ala-ag^{+} corresponds to $\phi \approx 180^{o}$ and $\psi \approx 60^{o}$.

10

Ala-Val-Lys-Tyr-Tyr-Thr-Leu-Glu-Gln-Ile-Glu-Lys-His-Asn-Asn-Ser-Lys-

δ_L - ε_L - ε_L - ε_L - β - ε_L - α_L - α_L - α_L - α_L - α_L - α_L - ε_L - δ_L - β - α_L - α_L -

20 30

Ser-Thr-Trp-Leu-Ile-Leu-Hys-Tyr-Lys-Val-Tyr-Asp-Leu-Thr-Lys-Phe-Leu-

β - ε_L - β - β - ε_L - β - α_D - α_D - ε_L - ε_L - β - γ_L - γ_L - α_L - α_L - α_L - α_L -

40 50

Glu-Glu-Hys-Pro-Gly-Gly-Glu-Glu-Val-Leu-Arg-Glu-Gln-Ala-Gly-Gly-Asp-

α_L - α_L - γ_L - α_L - α_L - ε_D - α_L - α_L - α_L - α_L - α_L - α_L - δ_L - ε_L - α_D - β - ε_L -

60

Ala-Thr-Glu-Asp-Phe-Glu-Asp-Val-Gly-Hys-Ser-Thr-Asp-Ala-Arg-Glu-Leu-

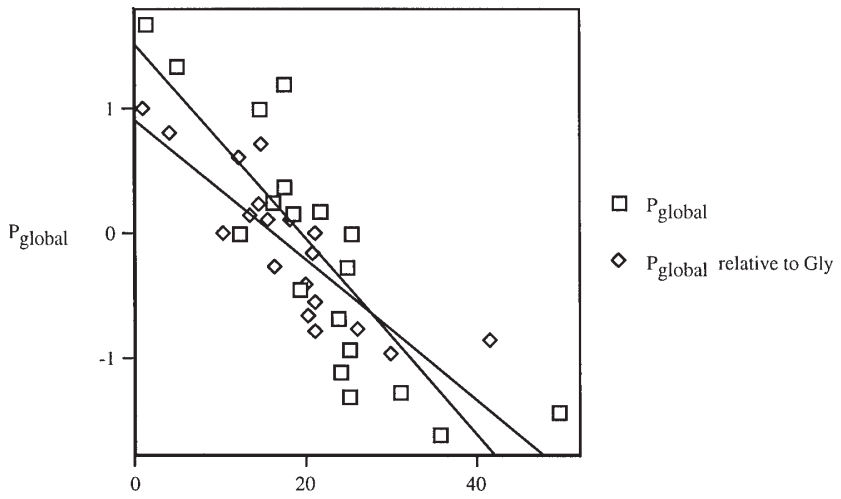
δ_D - α_L - α_L - α_L - α_L - α_L - α_L - δ_L - α_D - ε_L - β - α_L - α_L - α_L - α_L - α_L - α_L -

70 80

Ser-Lys-Thr-Phe-Ile-Ile-Gly-Glu-Leu-Hys-Pro-Asp-Asp-Arg-Ser-Lys-Ile-

α_L - α_L - α_L - δ_L - ε_L - α_L - ε_D - β - ε_L - ε_L - α_L - α_L - α_L - α_L - α_L - α_L - α_L - *Scheme 3*

The conformational assignment for the backbone conformation of the first eleven residues of a protein in PDB (*cf*. Table 6) shows all the ω angles indicating that all the residues are in the *trans* conformation. The *trans* conformation is due to the trapping of



Relative frequency of occurrence of $(g^- g^- + g^- g^+ + ag^+ + g^-a)/(g^+ g^+ + g^+ g^- + ag^- + g^+a)$ or relative to Gly

Figure 10. *Variation of the conformational parameter* P_{global} *vs. the frequency of occurrence of the amino-acid residues*

this conformation *via* hydrophobic, helix and sheet formation. The average ω angle is 165º, 15º off from planarity (180º). In particular, the ω angle is *ca.* 130º for Pro9, which is the residue with the greatest deviation from planarity.

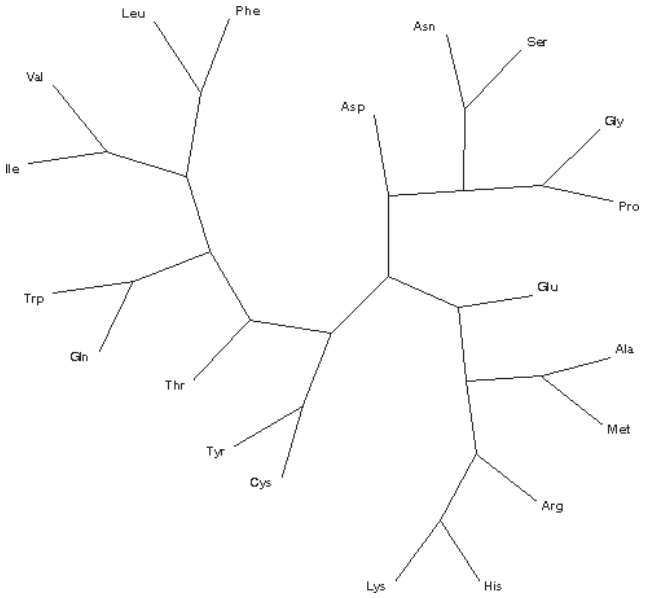


Figure 11. *Radial tree for the amino-acid residues in proteins according to conformational parameter* P_{global}.

Conclusions

From the precedent results and discussion, the following conclusions can be drawn.

1. An objective method based on quantitative geometric data has revealed to be useful to analyzing the description and classification of protein secondary and tertiary structures. The objective is to repeat a previous analysis of 258 proteins determined using ECEPP2, with the improved ECEPP2 + polarization, and there is good agreement between the two.

2. All nine conformations do occur in proteins. The relative frequency of occurrence of those conformations energetically favoured for L-enantiomers in the idealized ψ–ϕ map of the backbone conformations of the 20 amino acids in proteins decreases as: $g^-a > g^-g^+ > g^-g^- >> ag^+ > aa$. For the diverse amino acids, the same preference diminishes as: Pro >> Ile > Val > Leu > Thr > Met > Ala > Glu > Phe > Trp > Tyr > Gln > Lys > Ser > Cys > Arg > Asp > His > Asn > Gly. Achiral Gly relative frequencies of occurrence are close to one. Pro is the amino acid with the greatest g^-g^-/g^+g^+ preferences and greatly influences protein conformation. Pro and pseudo-prolines (ΨPro) are applied in peptide-based drug and pro-drug design, molecular recognition studies, as well as protein folding and self-aggregation processes [42,43].

Acknowledgment

The authors acknowledge financial support from the Spanish MEC DGI (Project No. CTQ2004-07768-C02-01/BQU) and Generalitat Valenciana (DGEUI INF01-051, INFRA03-047 and OCYT GRUPOS03-173).

References

[1] Csizmadia, I.G., *General and Theoretical Aspects of the Thiol Group*, in: Patai, S. (Ed.), *The Chemistry of the Thiol Group*, Wiley, New York, **1974**, 1.

[2] Csizmadia, I.G., *Multidimensional Theoretical Stereochemistry and Conformational Potential Energy Surface Topology*, in: Bertrán, J.; Reidel, D. (Eds.), *New Theoretical Concepts for Understanding Organic Reactions*, Dordrecht, **1989**, 1.

[3] Perczel, A.; Ángyán, J.G.; Kajtár, M.; Viviani, W.; Rivail, J.L.; Marcoccia, J.F.; Csizmadia, I.G., *J. Am. Chem. Soc*. **1991**, *113*, 6256.

[4] Mitchell, J.B.O.; Smith, J., *Proteins* **2003**, *50*, 563.

[5] Mezey, P.G.; *Potential Energy Hypersurfaces*, Elsevier, Amsterdam, **1987**, 227.

[6] Perczel, A.; Kajtár, M.; Marcoccia, J.F.; Csizmadia, I.G., *J. Mol. Struct. (Theochem)* **1991**, *232*, 291.

[7] Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Mayer, E.F.Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M., *J. Mol. Biol*. **1977**, *112*, 535.

[8] Abola, E.E.; Bernstein, F.C.; Bryant, S.H.; Koetzle, T.F., J. Weng, *Protein Data Bank*, in: Allen, F.H.; Bergerhoff, G.; Sievers, R. (Eds.), *Crystallographic Database: Information Content, Software System, Scientific Applications*, Data Commission of the International Union of Crystallography, Bonn–Cambridge–Chester, **1987**, 107.

[9] Némethy, G.; Pottle, M.S.; Scheraga, H. A., *J. Phys. Chem*. **1983**, *87*, 1883.

[10] Sippl, M.J.; Némethy, G.; Scheraga, H.A., *J. Phys. Chem*. **1984**, *88*, 6231.

[11] Perczel, A.; Viviani, W.; Csizmadia, I.G., *Peptide Conformational Potential Energy Surfaces and Their Relevance to Protein Folding*, in: Bertrán, J. (Ed.), *Molecular Aspects of Biotechnology: Computational Models and Theories*, Kluwer, Dordrecht, **1992**, 39.

[12] Torrens, F.; Voisin, C.; Rivail, J.L., *Electric Polarization in a Force Field for the Study of Dipeptide Models*, in: Glowinski, R. (Ed.), *Computing Methods in Applied Sciences and Engineering*, Nova Science, New York, **1991**, 249.

[13] Torrens, F.; Sánchez-Marín, J.; Rivail, J.L., *An. Fís. (Madrid)* **1994**, *90*, 197.

[14] Torrens, F., *Mol. Simul*. **2000**, *24*, 391.

[15] Applequist, J., *J. Phys. Chem*. **1993**, *97*, 6016.

[16] Walters, F.H.; Parker Jr., L.J.; Morgan, S.L.; Deming, S.N., *Sequential Simplex Optimization*, CRC, Boca Raton, **1991**.

[17] Torrens, F.; Ruiz-López, M.; Cativiela, C.; García, J.I.; Mayoral, J.A., *Tetrahedron* **1992**, *48*, 5209.

[18] Torrens, F., *Polyhedron* **2003**, *22*, 1091.

[19] Torrens, F., *Int. J. Quantum Chem*. **2004**, *99*, 963.

[20] Torrens, F., *J. Inclusion Phenom. Mol. Recognit. Chem*. **2004**, *49*, 37.

[21] Torrens, F., *Molecules* **2004**, *9*, 632.

[22] MacArthur, M.W.; Thornton, J.M., *J. Mol. Biol*. **1991**, *218*, 397.

[23] Tryon, R.C., *J. Chronic Dis*. **1939**, *20*, 511.

[24] Jarvis, R.A.; Patrick, E.A., *IEEE Trans. Comput*. **1973**, *C22*, 1025.

[25] McGregor, M.J.; Pallai, P.V., *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 443.

[26] Doman, T.N.; Cibulskis, J.M.; Cibulskis, M.J.; McCray, P.D.; Spangler, D.P., *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 1195.

[27] Turner, D.B.; Tyrrell, S.M.; Willett, P., *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 18.
[28] Reynolds, C.H.; Druker, R.; Pfahler, L.B., *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 305.
[29] *Integrated Mathematical Statistical Library (IMSL)*, IMSL, Houston, **1989**.
[30] Page, R.D.M., *Program TreeView*, Universiy of Glasgow, **2000**.
[31] Chou, P.Y.; Fasman, G.D., *Biochemistry* **1974**, *13*, 211.
[32] Chou, P.Y.; Fasman, G.D., *Trends Biochem. Sci*. **1977**, *2*, 128.
[33] Chou, P.Y.; Fasman, G. D., *Annu. Rev. Biochem*. **1978**, *47*, 251.
[34] Rose, G.D.; Gierasch, L.M.; Smith, J.A., *Adv. Protein Chem*. **1985**, *37*, 1.
[35] Müller, G.; Gurrath, M.; Kurz, M.; Kessler, H., *Proteins: Struct., Funct., Genet*. **1993**, *15*, 235.
[36] Richardson, J.S., *Adv. Protein Chem*. **1981**, *34*, 116.
[37] Smith, J.A.; Pease, L.G., *CRC Crit. Rev. Biochem*. **1980**, *8*, 315.
[38] Higgins, K.A.; Craik, D.J.; Hall, J.G.; Andrews, P.R., *Drug Design Deliv*. **1988**, *3*, 159.
[39] Weißhoff, H.; Wieprecht, T.; Henklein, P.; Frömmel, C.; Antz, C.; Mügge, C., *FEBS Lett*. **1996**, *387*, 201.
[40] Stein, R.L., *Adv. Quantum Chem*. **1993**, *11*, 1.
[41] Scherer, G.; Kramer, M.L. Schutkowski, M.; Reimer, U.; Fischer, G., *J. Am. Chem. Soc*. **1998**, *120*, 5568.
[42] Dumy, P.; Keller, M.; Ryan, D.E.; Rohwedder, B.; Wöhr, T.; Mutter, M., *J. Am. Chem. Soc*. **1997**, *119*, 918.
[43] Keller, M.; Sager, C.; Dumy, P.; Schutkowski, M.; Fischer, G.S.; Mutter, M., *J. Am. Chem. Soc*. **1998**, *120*, 2714.