Peptide Models XXXI. Conformational Properties of Hydrophobic Residues Shaping the Core of Proteins. An *Ab Initio* Study of N-Formyl-L-Valinamide and N-Formyl-L-Phenylalaninamide

PÉTER HUDÁKY,[1] IMRE JÁKLI,[1] ATTILA G. CSÁSZÁR,[2] ANDRÁS PERCZEL[1]

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

Received 10 July 2000; accepted 12 December 2000

ABSTRACT: Employing introductory (3-21G RHF) and medium-size (6-311++G** B3LYP) *ab initio* calculations, complete conformational libraries, containing as many as 27 conformers, have been determined for diamide model systems incorporating the amino acids valine (Val) and phenylalanine (Phe). Conformational and energetic properties of these libraries were analyzed. For example, significant correlation was found between relative energies from 6-311++G** B3LYP and single-point B3LYP/6-311++G**//RHF/3-21G calculations. Comparison of populations of molecular conformations of hydrophobic aromatic and nonaromatic residues, based on their *ab initio* relative energies, with their natural abundance indicates that, at least for the hydrophobic core of proteins, the conformations of Val (Ile, Leu) and Phe (Tyr, Trp) are controlled by the local energetic preferences of the respective amino acids.
 J Comput Chem 22: 732–751, 2001

Keywords: hydrophobic residues; peptides and proteins; *ab initio* calculations

Correspondence to: A. Perczel; e-mail: perczel@para.chem.elte.hu; WWW site: http://www.szerves.chem.elte.hu

Contract/grant sponsor: the Hungarian Scientific Research Fund; contract/grant numbers: OTKA T017604, T024044, and T032486.

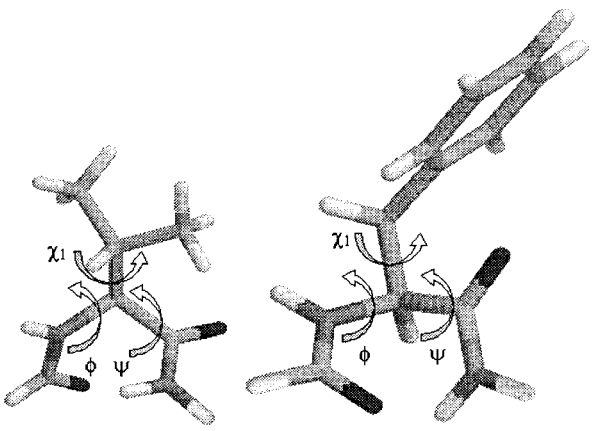
Introduction

It is customary[1] to describe the spatial arrangement of proteins, peptides, and their models incorporating the peptide unit —NH—CHR—C′O— through torsional angle pairs [ϕ,ψ], where ϕ describes torsion around the N—C and ψ describes torsion around the C—C′ bonds. (See Scheme 1 for definitions of torsions of the two model compounds investigated in this work, For-L-Val-NH_2 and For-L-Phe-NH_2.) Because the torsional potentials along the periodic variables ϕ and ψ may have three minima for an alpha amino acid residue, one expects[2] nine characteristic backbone conformers for each peptide unit (see Fig. 1 for the Ramachandran surface of For-L-Val-NH_2). The [ϕ, ψ] values of residues found in hundreds of nonhomologous proteins, whose structures have been investigated by X-ray diffraction,[3] revealed the existence of all nine backbone conformers, though with very different abundances. Conformers forming the broad β-region and structures corresponding to the helical subspace (3_{10}- and/or 4_{13}-helices) of the Ramachandran surface are observed most often, while D-type conformers are rare for amino acid residues of S-chirality.

The simplest models for investigating the conformational behavior of proteins and peptides are provided by amino acid diamides For-*Xxx*-NH_2 and Ac-*Xxx*-NHMe. *Ab initio* geometry optimizations on systems with *Xxx* = Gly, Ala, Val, Ser, Thr, Phe, and His, recently reviewed in refs. 4 and 5, found in considerably fewer minima than the maximum number allowed. For example, at the 3-21G(6-311++G**) RHF level two(three) backbone conformers of For-L-Ala-NH_2 could not be located. Nevertheless, existence of the "missing" minima has been verified by *ab initio* techniques for larger peptide models.[6] Due to its hydrophobic side chain, Val is a typical representative of those amino acid residues that have a single H^{β} (Val, Leu, Ile, and Thr). Formally, the side chain of Val, —$CH(CH_3)_2$, is derived from the methyl side chain of Ala: two H^{β}s are replaced by two methyl groups. Typically both methyl groups adopt a *gauche*$^+$ orientation ($\chi_2 = 60°$ and $\chi_2' = 60°$). Consequently, it is meaningful to restrict the systematic conformational analysis of a valine dipeptide model (e.g., For-L-Val-NH_2 of this study) to three torsional variables, ϕ, ψ, and χ_1, with χ_2 and χ_2' both constrained at about 60°. The resulting $E(\phi, \psi, \chi_1)$ potential energy hypersurface (PES) is expected to have 27 characteristic conformers (three χ_1 orientations for each of the nine backbone conformers). Nevertheless, for For-L-Val-NH_2, only 20 conformers could be located previously at the 3-21G RHF level.[7]

Similarly to valine, phenylalanine (Phe) can also be derived from alanine: one of the H^{β}s is replaced by a phenyl group. Therefore, phenylalanine has two vicinal H^{β}s (H^{β}_A and H^{β}_B), and distinct χ_1 rotamers are expected. Thus, for For-L-Phe-NH_2 the three predicted χ_1 orientations (*gauche*$^+$, *anti*, and *gauche*$^-$) will differ energetically. Conformational behavior of the χ_2 (C^{β}—C^{γ}) rotation is expected to be complex;[8] nevertheless, in practice, χ_2 typically adopts a *gauche*$^+$ orientation with $\chi_2 \approx 60°$. On the $E(\phi, \phi, \chi_1)$ PES, out of the 27 expected conformers of For-L-Phe-NH_2, a previous *ab initio* investigation[9] at the 3-21G RHF level resulted in only 19 conformers, while reoptimization at the 6-31+G* RHF level resulted[10] in the disappearance of three additional minima.

The molecular size of For-L-Val-NH_2 and For-L-Phe-NH_2 enables full geometry optimization at restricted Hartree–Fock (RHF),[11] density functional theory (DFT, e.g., B3LYP),[12] and second-order Møller–Plesset perturbation theory (MP2)[13] levels. It is common belief[14–16] that DFT(B3LYP) calculations provide structural data of similar quality to those obtained from the more time consuming MP2 approach, at least if better quality basis sets (such as 6-311++G**) are applied. Therefore, in this study only RHF and DFT methods were employed, providing geometries and energies associated with the minima of the PES of For-L-



SCHEME 1. HCO-L-Val-NH_2 and HCO-L-Phe-NH_2 model systems with the definition of torsional variables primarily determining the folding of these amino acid types in proteins.

Val-NH_2 and For-L-Phe-NH_2. Systematic geometry optimizations of all conformers by coupled cluster (CC) computations is not yet realistic and most likely not worth the effort. From full/constrained geometry optimizations, complete conformational libraries were determined for these peptide models to trace systematic structural changes induced by the use of different *ab initio* techniques. Energetic data obtained from these calculations allow both the establishment of certain computational characteristics and correlation between the relative energies and natural abundance of conformations of Val (Ile, Leu) and Phe (Tyr, Trp) residues in proteins.

Obviously, in these models several well-known phenomena are neglected, such as interresidue interactions, long-range effects, hydration, etc. These effects with additional ones influencing the conformational properties of linear polymers must be incorporated in the future. However, the present study focuses on the conformational properties of structural and conformational building units of proteins, where long-range interactions are not operative. Adhering to the present limitations, it is, however, possible to correlate the relative energy of a conformer and its relative probability observed in an ensemble of proteins.

Methodological Details

AB INITIO COMPUTATIONS

In this study all *ab initio* computations were performed with the program package Gaussian94.[17]

In preliminary studies,[7,9,18] 20 conformers of For-L-Val-NH_2 and For-L-Phe-NH_2 were found as minima at the 3-21G RHF level. In the present study, all these structures were reoptimized with tighter convergence criteria and seven additional $[\phi, \psi]$-constrained 3-21G RHF structures were optimized to complete the conformational libraries of the valine and phenylalanine diamide model systems. For the 27 structures, further full/constrained geometry optimizations were performed at the 6-311++G** B3LYP (For-L-Val-NH_2) and 6-31+G* RHF (For-L-Phe-NH_2) levels. All 27 structures were subsequently used for single-point energy calculations at the 6-31+G* and TZ2P[19] RHF and 6-311++G** B3LYP levels. Description of theoretical levels A1 through E2 for For-L-Val-NH_2 and A2 through F2 for For-L-Phe-NH_2, resulting from different combinations of full/constrained geometry optimizations and energy calculations, is given in Table I.

TABLE I.
Description of *Ab Initio* Calculations and the Resulting Notation of Computational Levels Employed for the Model Systems For-L-Val-NH_2 and For-L-Phe-NH_2.[a]

Model	Method	Energy Calculation	Optimized Geometry Used	Conformer Categories
Val	A1	3-21G RHF	3-21G RHF	20 full opt. + 7[ϕ, ψ]-constr. opt.
	A2	3-21G RHF	3-21G RHF	20 full opt.
	B1	6-31+G* RHF	3-21G RHF	20 full opt. + 7[ϕ, ψ]-constr. opt.
	B2	6-31+G* RHF	3-21G RHF	20 full opt.
	C1	6-311++G** B3LYP	3-21G RHF	20 full opt. + 7[ϕ, ψ]-constr. opt.
	C2	6-311++G** B3LYP	3-21G RHF	20 full opt.
	D1	6-311++G** B3LYP	6-311++G** B3LYP	18 full opt. + 9[ϕ, ψ]-constr. opt.
	D2	6-311++G** B3LYP	6-311++G** B3LYP	18 full opt.
	E1	TZ2P RHF	6-311++G** B3LYP	18 full opt. + 9 [ϕ, ψ]-constr. opt.
	E2	TZ2P RHF	6-311++G** B3LYP	18 full opt.
Phe	A2	3-21G RHF	3-21G RHF	19 full opt.
	B2	6-31+G* RHF	3-21G RHF	20 full opt.
	C2	6-311++G** B3LYP	3-21G RHF	17 full opt.
	D2	6-31+G* RHF	6-31+G* RHF	16 full opt.
	E2	6-311++G** RHF	6-31+G* RHF	16 full opt.
	F1	TZ2P RHF	6-31+G* RHF	16 full opt. + 11[ϕ, ψ]-constr.
	F2	TZ2P RHF	6-31+G* RHF	16 full opt.

[a] In the designation of the levels, index 1 means that all possible conformations (27 for both model compounds) are considered, while index 2 indicates that only those conformers corresponding to minima are taken into account.

Geometric and energetic results for For-L-Val-NH_2 are collected into Table II. The relative energies of Table II are referenced to two different global minima, $E[\gamma_L(g^+)]$ and $E[\delta_L(a)]$ for the 3-21G RHF and 6-311++G** B3LYP levels, respectively. For geometric and corresponding energetic results for For-L-Phe-NH_2 see refs. 9 and 10.

DATABASES

A total of 11,103 Val, 8911 Ile, 13,474 Leu, 6390 Phe, 5868 Tyr and 2374 Trp residues were collected from 650 proteins having a homology level equal to or lower than 25%.[20] All entries correspond to high-resolution X-ray structures taken from the 1998 issue of the PDB.[21]

NOMENCLATURE FOR BACKBONE AND SIDE-CHAIN CONFORMERS

The Ramachandran map[1] $E = E(\phi, \psi)$ can be divided into conformational subregions (catchment regions) in several ways. For the torsional angle pair ϕ and ψ, multidimensional conformational analysis (MDCA)[2] predicts nine catchment regions, as depicted in Scheme 2 and Figures 1 and 2 for For-L-Val-NH_2. Following IUPAC–IUB recommendations[21] the *gauche*$^+$ (g^+), *anti* (a), and *gauche*- (g^-) desgruptors could be used for notation of the conformers. Nevertheless, we introduce, following previous recommendations,[2] the following shorthand notation for the typical main-chain folds: $\alpha_L \equiv (g^-, g^-)$, $\alpha_D \equiv (g^+, g^+)$, $\beta_L \equiv (a, a)$, $\gamma_L \equiv (g^-, g^+)$, $\gamma_D \equiv (g^+, g^-)$, $\delta_L \equiv (a, g^+)$, $\delta_D \equiv (a, g^-)$, $\varepsilon_L \equiv (g^-, a)$, and $\varepsilon_D \equiv (g^+, a)$ (Scheme 2).

In For-L-Val-NH_2, which contains two geminal C^γ carbons (C^γ_A and C^γ_B) plus a proton (H^β) attached to C^β, distinct χ_1 rotamers are expected. [The β carbon atom is a prochiral center next to the chiral C^α (Scheme 1).] Thus, the three predicted orientations (g^+, a, and g^-) about χ_1 will differ energetically. The value of χ_1 can be defined by using either the N—C^α—C^β—H^β or the N—C^α—C^β—C^γ_1 torsional variable, the latter being used in PDB files. For completeness, both sets of data are reported in Table II.

TABLE II.
Structural and Energetic Results for Selected *Ab Initio* Conformers of For-L-Val-NH_2 at the 3-21G RHF (A1) and 6-311++G B3LYP (D1) Levels of Theory.**

Method	$[\phi, \psi]$	Conf.[a]	s.c. Conf. in PDB[b]	ω_0	ϕ	ψ	ω_1	χ_1	E/E_h	ΔE kcal mol^{-1}	$p(i)$ (%)[c]
A1	*const.*	$\alpha_L(g^+)$	(a)	−177.5	*−54.0*	*−45.0*	−179.6	61.5	−490.102146	8.05	0
D1	*const.*			−177.1	*−54.0*	*−45.0*	171.8	54.8	−495.998247	6.21	0
A1	*const.*	α_L(a)	(g^-)	−176.1	*−54.0*	*−45.0*	179.2	176.0	−490.100929	8.81	0
D1	*const.*			−176.4	*−54.0*	*−45.0*	171.3	176.6	−495.997491	6.69	0
A1	*const.*	$\alpha_L(g^-)$	(g^+)	−175.2	*−54.0*	*−45.0*	179.0	−54.8	−490.101750	8.30	0
D1	*const.*			−174.7	*−54.0*	*−45.0*	170.4	−54.2	−495.997523	6.67	0
A1	opt.	$\alpha_D(g^+)$	(a)	174.0	60.3	40.9	−180.0	76.4	−490.104220	6.75	0
D1	opt.			171.8	63.2	39.1	−173.1	83.2	−495.996506	7.31	0
A1	opt.	α_D(a)	(g^-)	174.9	47.3	44.6	−179.3	178.7	−490.102403	7.89	0
D1	opt.			172.7	51.7	33.7	−174.0	−179.8	−495.996649	7.22	0
A1	opt.	$\alpha_D(g^-)$	(g^+)	174.4	50.0	43.1	−179.8	−44.9	−490.101067	8.73	0
D1	opt.			172.3	51.2	39.6	−173.1	−42.8	−495.994158	8.78	0
A1	opt.	$\beta_L(g^+)$	(a)	176.4	−137.5	143.5	178.6	66.5	−490.109634	3.35	0
D1	opt.			175.0	−118.8	125.8	−177.3	59.2	−496.005447	1.70	4
A1	opt.	β_L(a)	(g^-)	174.0	−142.4	163.5	177.3	172.9	−490.112089	1.81	2
D1	opt.			170.9	−131.1	162.2	−179.3	178.8	−496.005843	1.45	5
A1	opt.	$\beta_L(g^-)$	(g^+)	179.0	−163.2	157.4	178.6	−55.2	−490.111804	1.99	2
D1	opt.			176.3	−151.9	157.6	−179.8	−49.6	−496.005546	1.63	4
A1	*const.*	$\delta_L(g^+)$	(a)	−174.6	*−130.0*	*30.0*	176.5	80.8	−490.107363	4.78	0
D1	*const.*			−172.6	*−130.0*	*30.0*	172.0	83.7	−495.999985	5.12	0
A1	opt.	δ_L(a)	(g^-)	−174.2	−125.5	28.9	175.8	178.9	−490.109724	3.29	0
D1	opt.			−172.3	−114.0	12.1	171.0	177.2	**−496.008149**	**0.00**	61
A1	opt.	$\delta_L(g^-)$	(g^+)	−173.6	−137.0	36.1	175.9	−49.2	−490.108232	4.23	0
D1	opt.			−171.1	−112.1	4.3	171.7	−43.0	−496.005114	1.90	2

TABLE II.
(Continued)

Method	$[\phi, \psi]$	Conf.[a]	s.c. Conf. in PDB[b]	ω_0	ϕ	ψ	ω_1	χ_1	E/E_h	ΔE kcal mol^{-1}	$p(i)$ (%)[c]
A1	opt.	$\delta_D(g^+)$	(a)	174.9	−136.7	−59.9	179.8	67.2	−490.100556	9.05	0
D1	opt.			174.0	−126.9	−68.0	−175.4	60.0	−495.997371	6.76	0
A1	opt.	δ_D(a)	(g^-)	172.1	−170.4	−46.6	−176.5	141.4	−490.097556	10.93	0
D1	*const.*			176.1	*−144.0*	*−54.0*	−174.6	169.0	−495.993469	9.21	0
A1	opt.	$\delta_D(g^-)$	(g^+)	173.6	−175.5	−34.6	−178.0	−55.4	−490.102318	7.94	0
D1	opt.			172.0	−159.3	−40.1	−172.9	−55.1	−495.996816	7.11	0
A1	*const.*	$\varepsilon_L(g^+)$	(a)	173.2	*−60.0*	*120.0*	−179.3	60.6	−490.104901	6.32	0
D1	*const.*			169.4	*−60.0*	*120.0*	−172.0	60.2	−496.003059	3.19	0
A1	*const.*	ε_L(a)	(g^-)	174.8	*−60.0*	*120.0*	−179.7	172.8	−490.103813	7.00	0
D1	*const.*			170.5	*−60.0*	*120.0*	−172.7	172.0	−496.001053	4.45	0
A1	*const.*	$\varepsilon_L(g^-)$	(g^+)	175.9	*−60.0*	*120.0*	−179.2	−62.6	−490.103519	7.19	0
D1	*const.*			171.3	*−60.0*	*120.0*	−171.8	−58.6	−496.000913	4.54	0
A1	opt.	$\varepsilon_D(g^+)$	(a)	−166.4	75.2	152.6	179.2	61.9	−490.097830	10.76	0
D1	opt.			−168.2	85.1	133.3	172.6	61.4	−495.992649	9.73	0
A1	opt.	ε_D(a)	(g^+)	−164.9	70.5	170.3	179.8	−152.6	−490.097383	11.04	0
D1	*const.*			−166.9	*84.1*	*152.0*	171.8	−158.1	−495.990260	11.23	0
A1	opt.	$\varepsilon_D(g^-)$	(g^+)	−168.1	75.4	162.6	178.4	−17.6	−490.095291	12.35	0
D1	opt.			−168.7	83.7	160.6	174.0	−18.7	−495.990273	11.22	0
A1	opt.	$\gamma_L(g^+)$	(a)	−175.9	−86.6	71.4	−177.4	65.9	**−490.114973**	**0.00**	50
D1	opt.			−179.8	−83.9	82.4	−170.3	65.8	−496.006780	0.86	14
A1	opt.	γ_L(a)	(g^-)	−174.5	−84.9	62.7	−179.8	173.2	−490.114404	0.36	28
D1	opt.			−176.5	−83.2	62.0	−179.9	170.4	−496.005901	1.41	6
A1	opt.	$\gamma_L(g^-)$	(g^+)	−174.2	−85.3	65.6	−179.0	−56.9	−490.113923	0.66	17
D1	opt.			−177.2	−83.8	74.1	−173.4	−57.8	−496.004839	2.08	2
A1	opt.	$\gamma_D(g^+)$	(a)	175.8	74.2	−61.5	−179.4	58.0	−490.111455	2.21	1
D1	opt.			177.4	72.8	−63.2	176.8	63.3	−496.004545	2.26	1
A1	opt.	γ_D(a)	(g^-)	175.6	59.5	−38.5	−178.4	−168.8	−490.105971	5.65	0
D1	opt.			175.7	57.2	−31.0	−175.9	−172.1	−496.000442	4.84	0
A1	opt.	$\gamma_D(g^-)$	(g^+)	174.5	62.8	−39.1	−178.3	−35.8	−490.106569	5.27	0
D1	opt.			176.0	61.5	−39.0	−176.9	−36.8	−495.999374	5.51	0

[a] See text for labeling of the backbone conformers. See Table I for the methods used. Torsion angles in italics were fixed during geometry optimization. At the 3-21G RHF (method A1) and 6-311++G** B3LYP (method D1) levels the global minima are $\gamma_L(g^+)$ and δ_L(a), respectively.

[b] Side-chain conformers according to convention employed in the PDB.

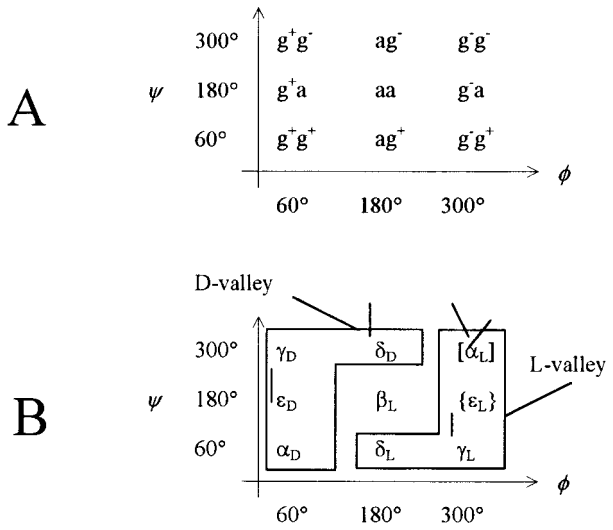
[c] Relative populations for computed *ab initio* relative energies are determined as $\exp(-\Delta E/RT)/\Sigma \exp(-\Delta E/RT)$, where $RT = NkT = 0.595371$ (kcal/mol) {$T = 300$ K, $k = 1.38 \times 10^{-23}$ [J/K], and the Avogadro's number (N) is 6.02×10^{23} (mol^{-1})}.

In For-L-Phe-NH_2, the three predicted orientations (g^+, a, and g^-) about χ_1 will also differ energetically (Scheme 1). On the other hand, for χ_2 only one orientation was found, that of the g^+ conformer, as reported earlier.[10]

Results and Discussion

STRUCTURES AND CONFORMATIONAL SHIFTS

Our peptide models, For-L-Val-NH_2 and For-L-Phe-NH_2, may have a maximum of nine characteristic backbone arrangements. For each backbone structure it is reasonable to assume three χ_1 orientations (g^+, a, and g^-) and a single χ_2 orientation (g^+), resulting in 27 idealized structures. From the 27 characteristic structures of For-L-Val-NH_2, 20 optimized to minima at the 3-21G RHF level.[7] In this study the entire structural data set was reoptimized for For-L-Val-NH_2 at the 6-311++G** B3LYP level with the result that 18 minima were obtained (Figs. 1 and 2). To make the conformational library of Val complete at both levels of theory, the remaining 7(9) structures of For-L-Val-NH_2 were



SCHEME 2. The ideal location of the nine conformers on the $E = E(\phi, \psi)$ surface, labeled according to the IUPAC-IUB guidance (A). Approximate location of optimized *ab initio* minima of For-*Xxx*-NHx (*Xxx* = Gly, Ala, Ser, Thr, Cys, Asp, Val, Phe, His), using the shorthand notation of the above nine minima (B). The α_L is in brackets because it may vanish. The ε_L stands for poly-proline-II type minima not yet assigned, during *ab initio* studies, to any P—CONH—CHR—NHCO—Q system. However, the existence of the α_L and ε_L conformations has been show, for example, in the dialanine diamide HCO-Ala-Ala-NH_2.

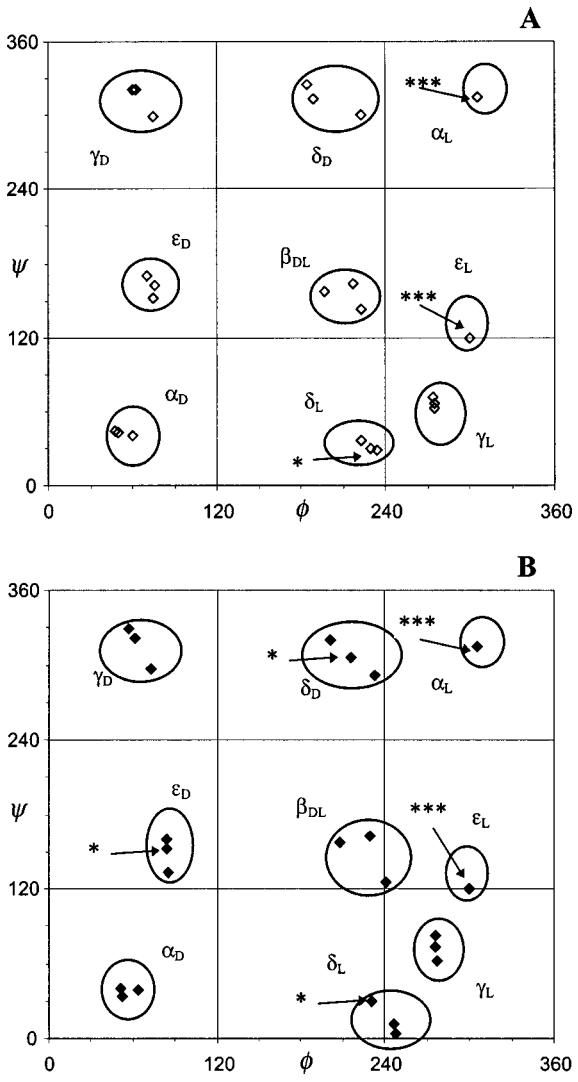


FIGURE 1. Distribution of the 27 conformers of HCO-L-Val-NH_2 on the Ramachandran plot determined at the RHF/3-21G (A) and B3LYP/6-311++G** (B) levels of theory. (A star stands fo a conformer whose structure was partially optimized using constrained ϕ and ψ values. See Table II for details.)

optimized using constrained $[\phi, \phi]$ torsional angles (see Table II). As usual,[4] the $\gamma_L(g^+)$ conformer, incorporating a strong intramolecular H-bond typical of inverse γ-turns, was found to be the most stable structure at the 3-21G RHF level. At the more dependable 6-311++G** B3LYP level, however, δ_L(a) becomes the most stable structure, even though δ_L(a), compared to $\gamma_L(g^+)$, has a high relative energy of 3.29 kcal mol^{-1} at the 3-21G RHF level. Interestingly, there is neither a H-bond nor any other obvious intramolecular stabilizing effect in δ_L(a).

Geometry optimizations, starting from idealized structures predicted by MDCA, result in slight structural changes for some conformers and in conformational migrations (where the initial and the final structures belong to different catchment regions) for others. The seven migrations observed for For-L-Val-NH_2, at the 3-21G RHF level, are as follows: $\alpha_L(g^-) \Rightarrow \delta_L(g^-)$, α_L(a) $\Rightarrow \delta_L$(a), $\alpha_L(g^+) \Rightarrow \gamma_L(g^+)$, $\delta_L(g^+) \Rightarrow \gamma_L(g^+)$, $\varepsilon_L(g^-) \Rightarrow \gamma_L(g^-)$, ε_L(a) $\Rightarrow \gamma_L$(a), and $\varepsilon_L(g^+) \Rightarrow \gamma_L(g^+)$. Two more migrations occur at the 6-311++G** B3LYP level: δ_D(a) $\Rightarrow \delta_L$(a) and ε_D(a) $\Rightarrow \gamma_D$(a). These migrations indicate flattening of the relevant portions of the PES. The disappearance of the two 3-21G RHF For-L-Val-NH_2 conformers is in line with the general trend[4, 22–24] that at higher levels of theory annihilation of minima (and saddle points) occurs frequently. Although no systematic mapping of the PES was performed, analysis of the nature of these migrations reveals the following important features about the PES of For-L-Val-NH_2: (a) neither α_L- nor ε_L-type backbone orientations correspond to minima at any side-chain orientation; (b) while all three polyproline-like structures (ε_L) "slide" to the γ_L catchment region, the α_L input conformers "end" either in the γ_L or in the δ_L

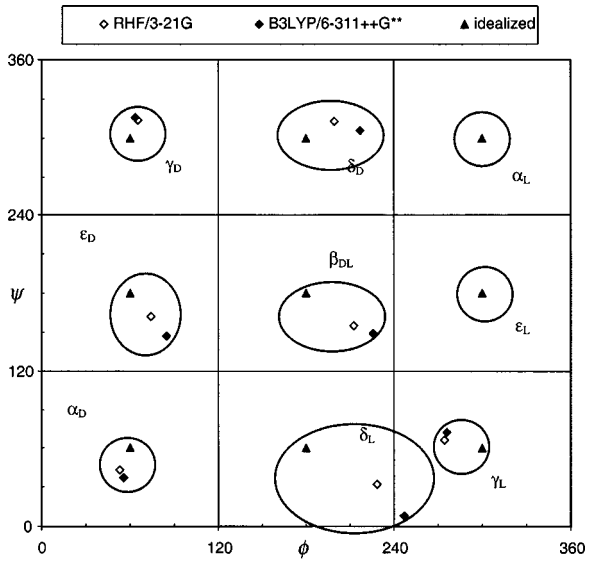


FIGURE 2. Location of the idealized and the averaged backbone conformers (3-21G RHF and 6-311++G** B3LYP) of For-L-Val-NH_2 on the Ramachandran surface (the ϕ and ψ values of conformers belonging to a common catchment region are averaged (e.g., $\gamma_L(g+)$, γ_L(a), and $\gamma_L(g^-) \Rightarrow \gamma_L$).

area of the Ramachandran surface (all these conformational changes can be rationalized by simple shifts in ψ); (c) side-chain orientations are always preserved during the migrations; and (d) migration of D-type structures occurs only during 6-311++G** B3LYP optimizations and the δ_D(a) $\Rightarrow$ δ_L(a) migration forms the unique example of a D to L structural shift (this conformational migration may be associated with the δ_L(a) conformer becoming the global minimum at the 6-311++G** B3LYP level). The lack of δ_D(a) and ε_D(a) migrations at the 3-21G RHF level is probably due to the fact that both initial conformers have a high relative energy at the 3-21G RHF level: $\Delta E[\delta_D(a)] = 10.9$ kcal mol^{-1} and $\Delta E[\varepsilon_D(a)] = 11.0$ kcal mol^{-1}.

In the next phase of our study the backbone conformational parameters of minima of the same catchment region but with different side-chain orientations were averaged. The results are presented in Table III and on Figure 2. As found earlier for alanine and serine diamide model systems,[4] the 3-21G RHF torsion angles are typically more similar to those of the idealized conformers than the more dependable 6-311++G** B3LYP optimized ones. The DFT structural shifts are small but significant, especially for the δ_L, δ_D, β_L, and ε_D conformers.

When comparing torsion angles of For-L-Val-NH_2, determined at the two levels of theory (18 common conformers), the average difference for the entire data set is only 4.9°, suggesting only small conformational changes upon reoptimization. Significant changes in torsions (>15°), are observed only for five conformers, as listed in Table IV. In two of the five cases, $\delta_L(g^-)$ and $\beta_L(g^+)$, both backbone torsional parameters change substantially, while for the other three structures, namely for δ_L(a), $\delta_D(g^-)$, and $\varepsilon_D(g^+)$, only one out of the three monitored torsional parameters shift significantly. There is no change exceeding 8° for any of the side-chain orientations. We can offer no obvious explanation for the five "large" conformational shifts. Unlike for model

TABLE III.
Average Backbone Conformational Parameters of Fully Optimized For-L-Val-NH_2.[a]

Conformer[b]	Number of Conf.[c]	6-311++G** B3LYP(3-21G RHF)	
		$\phi^{average}$	$\psi^{average}$
α_L	0 (0)	—	—
α_D	3 (3)	55.4 (52.5)	37.5 (42.9)
β_L	3 (3)	−133.9 (−147.7)	148.5 (154.8)
δ_L	2 (2)	−113.1 (−131.3)	8.2 (32.5)
δ_D	2 (3)	−143.1 (−160.9)	−54.0 (−47.0)
ε_L	0 (0)	—	—
ε_D	2 (3)	84.4 (73.7)	147.0 (161.9)
γ_L	3 (3)	−83.6 (−85.6)	72.9 (66.6)
γ_D	3 (3)	63.9 (65.5)	−44.4 (−46.4)

[a] $\phi^{average}$ and $\psi^{average}$ are in degrees. Values reported correspond to 6-311++G** B3LYP optimizations, values in parentheses were obtained at the 3-21G RHF level.
[b] Type of backbone conformers.
[c] In all nine backbone catchment regions only fully optimized conformers were averaged.

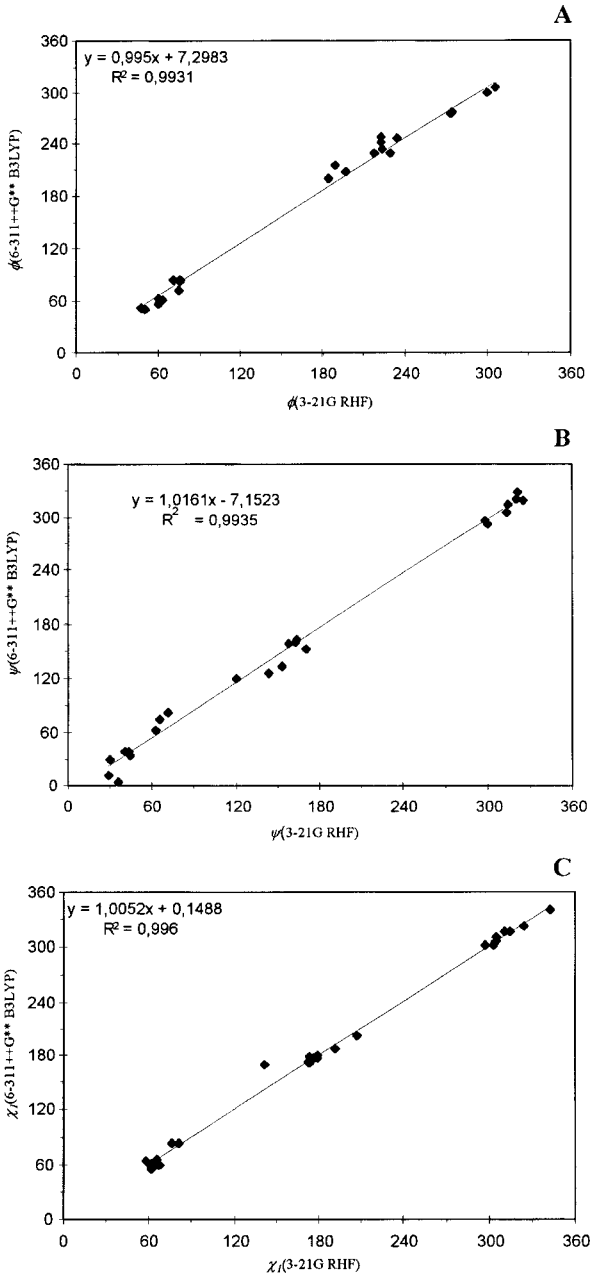


FIGURE 3. Pair-wise comparison of selected dihedral parameters (ϕ, ψ, and χ_1) of CHO-L-Val-NH_2 determined at 3-21G RHF and 6-311++G** B3LYP levels of theory.

systems of serine, where the conformers with higher relative RHF energies showed more tendency toward conformational shift upon reoptimization,[23] in the case of For-L-Val-NH_2 there is no such correlation. On the contrary, δ_L(a) is the structure that changes the most and becomes the global minimum at the 6-311++G** B3LYP level. Reoptimization at higher levels of theory often results in lower relative energies (cf. Table II and Fig. 3). However, a large conformational shift does not always imply a large change in relative energy, as can be seen for $\beta_L(g^+)$ in Table II. Furthermore, all three major dihedral angles (ϕ, ψ, and χ_1) used for describing molecular folding, determined at the 3-21G RHF and 6-311++G** B3LYP levels, correlate well. For example, the Pearson correlation coefficients,[25] R, between dihedral angles obtained at the 3-21G RHF and 6-311++G** B3LYP levels are very high: $R(\phi) = 0.993$, $R(\psi) = 0.994$, and $R(\chi_1) = 0.996$ (Fig. 3).

H-BONDS

The amino acids Val and Phe have apolar side chains; therefore, intramolecular H-bond formation is only possible between the carbonyl and the NH groups. The most important H-bond is formed between the oxygen of the formyl group, representing the carbonyl of residue $i-1$ in a peptide, and the proton of the NH_2 group of residue $i+1$ (Scheme 1). This strong H-bond is characteristic for γ-turns. [Two γ-turns are possible: the so-called inverse γ-turn (γ_L) and its conformational mirror image, the "normal" γ-turn, called here γ_D.] A second type of H-bond is formed between the proton attached to amide nitrogen and the carbonyl oxygen within the amino acid residue, resulting in a rather weak five-membered pseudoring system (Table V). These H-bond systems imply specific backbone torsions. In the γ_L conformation, $\phi \approx -75°$ and $\psi \approx +75°$. In γ_D, a conformational mirror image of γ_L, the same intramolecular H-bond pattern can be found, implying $\phi \approx +75°$ and $\psi \approx -75°$. The second type of H-bond results in an extended structure with $\phi \approx -150°$ and $\psi \approx +150°$, characteristic of β_L.

TABLE IV.
Significant ($\Delta\xi \geq 15°$) Conformational Differences between 3-21G RHF and 6-311++G B3LYP Optimized Conformers of For-L-Val-NH_2.**

Conf.[a]	$\lvert\Delta\phi\rvert$[b]	$\lvert\Delta\psi\rvert$[b]	$\lvert\Delta\chi_1\rvert$[b]	Average Shift[c]
$\beta_L(g^+)$	18.6	17.7	7.2	14.5
δ_L(a)	11.6	16.8	1.7	10.0
$\delta_L(g^-)$	24.8	31.7	6.2	20.9
$\delta_D(g^-)$	16.2	5.5	0.3	7.4
$\varepsilon_D(g^+)$	9.9	19.3	0.5	9.9

[a] Only fully optimized geometries are compared. See text for notation of backbone conformers. All values are in degrees.
[b] Differences between 3-21G RHF and 6-311++G** B3LYP conformational parameters.
[c] Averaged over three conformational variables (ϕ, ψ, and χ_1).

TABLE V.
H-Bond Parameters in For-L-Val-NH_2 at the 6-311++G B3LYP level.**

Conf.[a]	$r(O_1 \ldots N_2)$	$r(O_1 \ldots H—N_2)$	$\Theta(O_1 \ldots H—N_2)$	Conf.	$r(O_2 \ldots N_1)$	$r(O_2 \ldots H—N_1)$	$\Theta(O_2 \ldots H—N_1)$
$\gamma_L(g^+)$	3.04	2.21	137.39	$\beta_L(g^+)$	2.84	2.56	94.99
γ_L(a)	2.95	2.08	143.78	β_L(a)	2.68	2.24	104.13
$\gamma_L(g^-)$	2.98	2.12	141.57	$\beta_L(g^-)$	2.66	2.20	105.74
$\gamma_D(g^+)$	2.88	1.99	145.60				
γ_D(a)	2.81	1.88	151.24				
$\gamma_D(g^-)$	2.83	1.90	150.20				
γ average	2.92	2.03	144.96	β average	2.73	2.33	101.62

[a] For-L-Val-NH_2 conformers. Distances (r) in Å, angles (Θ) in degrees. The numbering of the atoms is according to Scheme 1/A.

Stabilization due to H-bonds is obvious: both γ_L and β_L structures have low relative energies. D-type backbone conformers are in general less favored for L-amino acid derivatives than any L-type conformer. Stabilization of γ_D compared to δ_L (at the 3-21G RHF level $\Delta E[\delta_L(g^+)] = 4.78$ kcal mol^{-1} > $\Delta E[\gamma_D(g^+)] = 2.21$ kcal mol^{-1}, at the 6-311++G** B3LYP level a similar relationship exists) is clearly due to H-bond formation.

BASIS SET DEPENDENCE OF RELATIVE ENERGIES

For the For-L-Val-NH_2 model, five sets of (relative) energies have been determined (see Table VI and Fig. 4A). All five data sets are referenced to the energy of the global minimum, $\gamma_L(g^+)$. Except for α_L and α_D, L-type conformers have significantly lower relative energies than their D-type counterparts. The α_L conformers themselves are slightly more stable than the appropriate α_D structures, except at the introductory 3-21G RHF level. Relative energies determined at the 3-21G RHF level show large variation as a function of the conformations. As found for other For-*Xxx*-NH_2 model systems, like *Xxx* = *Gly*,[4,18] Ala,[4,18] Ser,[4,23] and Phe,[9,10] the larger is the basis and/or the higher is the applied level of theory, the smaller is the spread of the relative energies. This trend is valid for valine (Fig. 4 and Table VI), although RHF/6-31+G*//RHF/3-21G relative energies can occasionally be lower than those computed at the highest, 6-311++G** B3LYP level.

As only limited energetic data were reported for For-L-Phe-NH_2 in the past,[9,10] in this article we present additional single-point energy calculations, which enhance our understanding about the energetics of this peptide model. All energies presented in Table VII (see also Fig. 4B) are referenced to the energy of $\gamma_L(g^+)$, the global minimum. The ΔE val-

TABLE VI.
***Ab Initio* Relative Energies (in kcal mol^{-1}) of For-L-Val-NH_2 at Five Different Levels of Theory.[a]**

Conformer	A1	B1	C1	D1	E1
$\alpha_L(g^+)$	8.0	5.5	5.9	5.4	5.0
α_L(a)	8.8	5.7	6.1	5.8	5.6
$\alpha_L(g^-)$	8.3	5.9	6.2	5.8	5.5
$\alpha_D(g^+)$	6.7	6.5	6.7	6.4	6.2
α_D(a)	7.9	6.6	6.8	6.4	6.1
$\alpha_D(g^-)$	8.7	8.7	8.3	7.9	7.9
$\beta_L(g^+)$	3.4	0.4	1.1	0.8	0.3
β_L(a)	1.8	−0.1	0.7	0.6	−0.1
$\beta_L(g^-)$	2.0	0.3	0.7	0.8	0.5
$\delta_L(g^+)$	4.8	3.9	4.0	4.3	4.5
δ_L(a)	3.3	1.3	1.8	−0.9	1.8
$\delta_L(g^-)$	4.2	3.5	3.6	1.0	3.7
$\delta_D(g^+)$	9.0	6.0	6.3	5.9	5.7
δ_D(a)	10.9	9.2	6.2	8.4	8.7
$\delta_D(g^-)$	7.9	6.6	6.5	6.3	6.0
$\varepsilon_L(g^+)$	6.3	1.3	2.6	2.3	1.1
ε_L(a)	7.0	3.4	4.0	3.6	2.8
$\varepsilon_L(g^-)$	7.2	3.6	4.2	3.7	2.9
$\varepsilon_D(g^+)$	10.8	9.1	9.1	8.9	8.9
ε_D(a)	11.0	10.2	10.0	10.4	10.5
$\varepsilon_D(g^-)$	12.4	10.6	10.4	10.4	10.5
$\gamma_L(g^+)$	0.0	0.0	0.0	0.0	0.0
γ_L(a)	0.4	0.7	0.3	0.6	1.1
$\gamma_L(g^-)$	0.7	1.9	1.2	1.2	1.8
$\gamma_D(g^+)$	2.2	1.6	1.2	1.4	2.2
γ_D(a)	5.6	5.1	3.7	4.0	5.3
$\gamma_D(g^-)$	5.3	6.3	4.6	4.6	6.3

[a] All relative energies are with respect to $E[\gamma_L(g^+)]$. $E[\gamma_L(g^+)]/E_h$ is {−490.114973, −492.862889, −496.004022, −496.006780, −493.010973} at the {A1 = RHF/3-21G, B1 = RHF/6-31+G*//RHF/3-21G, C1 = B3LYP/6-311++G**//RHF/3-21G, D1 = B3LYP/6-311++G**, E1 = RHF/TZ2P//B3LYP/6-31++G**} levels. See text for description of the conformers.

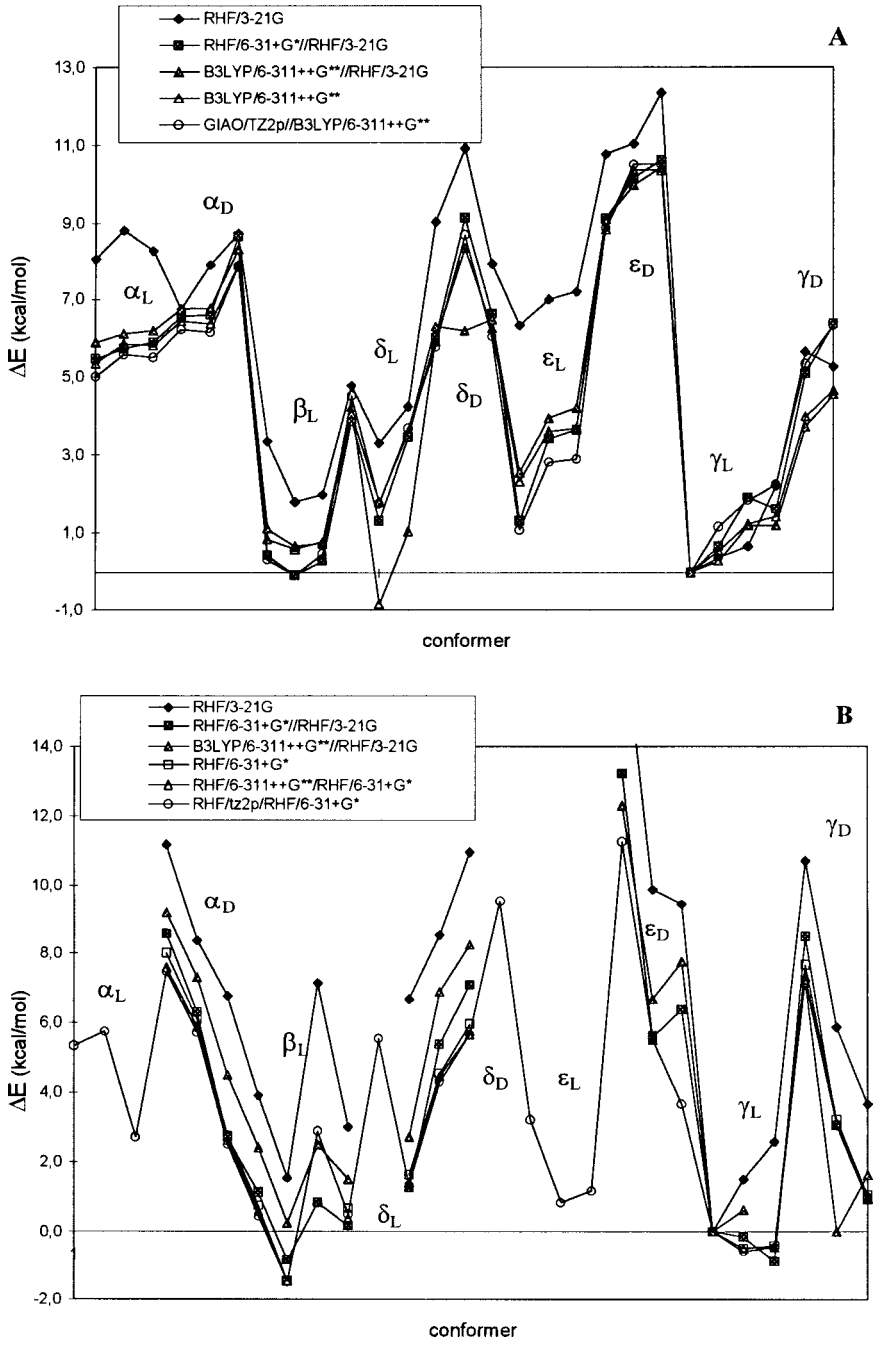


FIGURE 4. Comparison of relative energy differences of For-L-Val-NH_2 (A) and For-L-Phe-NH_2 (B) obtained at selected levels (see also Tables VI and VII).

ues of L-type conformers are typically lower than the respective values calculated for D-type conformers. This is in line with the experimental observation that in proteins D-type backbone conformers are much less frequent than their L-type counterparts.

The limitations inherent in the 3-21G RHF *ab initio* approach make some of the computed energies unreliable. This is a considerable hindrance, because characterization of conformational libraries of biomolecules at higher levels of theory is unrealistic. Therefore, it is of considerable importance to determine whether calibration of 3-21G RHF energies of For-L-Val-NH_2 against energies of higher level computations is feasible. As noted earlier, for the torsional angles ϕ, ψ, and χ_1 of For-L-Val-NH_2, the 3-21G RHF and the 6-311++G** B3LYP values show

TABLE VII.

***Ab Initio* Relative Energies (in kcal mol^{-1}) of For-L-Phe-NH$_2$ at Six Different Levels of Theory.[a]**

Conformer	A2	B2	C2	D2	E2	F1
$\alpha_L(g^+)$						5.34
α_L(a)						5.75
$\alpha_L(g^-)$						2.72
$\alpha_D(g^+)$	11.16	8.60	9.21	8.02	7.59	7.46
α_D(a)	8.39	6.32	7.31	6.07	5.86	5.71
$\alpha_D(g^-)$	6.75	2.71	4.48	2.76	2.61	2.48
$\beta_L(g^+)$	3.90	1.15	2.43	0.74	0.58	0.46
β_L(a)	1.54	−0.83	0.24	−1.44	−1.47	−1.46
$\beta_L(g^-)$	7.11	0.85	2.48			2.89
$\delta_L(g^+)$	2.98	0.17	1.48	0.68	0.40	0.48
δ_L(a)						5.55
$\delta_L(g^-)$	6.66	1.13	2.71	1.63	1.43	1.35
$\delta_D(g^+)$	8.54	5.39	6.87	4.55	4.47	4.28
δ_D(a)	10.96	7.09	8.25	5.94	5.65	5.65
$\delta_D(g^-)$						9.54
$\varepsilon_L(g^+)$						3.22
ε_L(a)						0.83
$\varepsilon_L(g^-)$						1.16
$\varepsilon_D(g^+)$	17.61	13.20	12.30			11.25
ε_D(a)	9.89	5.59	6.68	5.64	5.48	5.54
$\varepsilon_D(g^-)$	9.46	6.37	7.73			3.66
$\gamma_L(g^+)$	0.00	0.00	0.00	0.00	0.00	0.00
γ_L(a)	1.51	−0.18	0.62	−0.49	−0.52	−0.57
$\gamma_L(g^-)$	2.59	−0.87		−0.43	−0.47	−0.47
γ_D(a)	10.72	8.50	7.16	7.68	7.34	7.22
$\gamma_D(g^-)$	5.87	3.07		3.22	3.11	3.10
$\gamma_D(g^-)$	3.66	0.90	1.62	1.04	0.95	1.00

[a] All relative energies are with respect −648.462030 to $E[\gamma_L(g^+)]$. $E[\gamma_L(g^+)]/E_h$ is {−640.745238, −644.346283, −644.466682, −644.501279, −644.528642} at the {A2 = RHF/3-21G, B2 = RHF/6-31+G*//RHF/3-21G, C2 = B3LYP/6-311++G**//RHF/3-21G, D2 = RHF/6-31+G*, E2 = RHF/6-311++G**//RHF6-31+G*, F1 = RHF/TZ2P//RHF/6-31+G*} levels. See text for description of the conformers.

impressive correlation (Fig. 3), all R^2 values being higher than 0.993. Performing the same type of comparison for For-L-Phe-NH$_2$ (for structural data see Table I of ref. 9) between data determined at 3-21G and 6-31+G* RHF levels, the R^2 values are, once again, higher than 0.985. Comparing torsional values pair-wise, all conformational parameters responsible for the molecular fold correlate extremely well. This result further supports the important observation that increasing the applied level of theory changes the conformational preferences only marginally.

It is important to compare changes in relative energies as a function of the level of theory. Although relative energies do change when the *ab initio* method is altered, there is a clear linear correlation between relative energy values obtained at different levels [see ΔE(B3LYP/6-311++G**//RHF/3-21G) and ΔE(3-21G RHF), as well as ΔE(B3LYP/6-311++G**) and ΔE(B3LYP/6-311++G**//RHF/3-21G) trends in Fig. 5A and B]. The correlations are often significant: $R^2 = 0.882$ on Fig. 5A and $R^2 = 0.932$ on Fig. 5B. Note that the latter correlation can become as significant as $R^2 = 0.993$, if structures δ_L(a), $\delta_L(g^-)$, and δ_D(a) are removed from the data set. Note also that two [δ_L(a) and $\delta_L(g^-)$] out of the three structures responsible for deteriorating the linear correlation are among those conformers that shifted the most when structures were reoptimized at the higher *ab initio* level (see Table IV).

Another important problem of applied quantum chemistry is whether single-point energy calculations are good enough for flexible molecules like our peptide models or full geometry optimizations are required.[14, 16, 26] Comparison of our B3LYP/6-311++G**//RHF/3-21G and B3LYP/6-311++G** results for For-L-Val-NH$_2$ reveal the following: (a) the better correlation between ΔE(B3LYP/6-311++G**//RHF/3-21G) and ΔE(6-311++G** B3LYP) compared to correlation between ΔE(B3LYP/6-311++G**//RHF/3-21G) and ΔE(3-21G RHF) shows that the energetics are not very dependent on changes in the underlying reference geometries. (b) The single-point ΔE(B3LYP/6-311++G**//RHF/3-21G) energy calculations match well with the data set obtained from full 6-311++G** B3LYP geometry optimization. Furthermore, ΔE(RHF/6-31+G*//RHF/3-21G) and ΔE(RHF/TZ2P//B3LYP/6-311++G**) data sets also correlate impressively: $R^2 = 0.986$. In conclusion, the single-point calculations are almost as instructive as the 15 to 25 times lengthier geometry optimizations (Table V). This finding indicates that performing relatively inexpensive single-point calculations on the entire set of 3-21G RHF geometries could be used remarkably well to mimic higher level energy calculations, such as ΔE(RHF/TZ2P//B3LYP/6-311++G**). A similar analysis on For-L-Phe-NH$_2$, based on the available relative energies, results in a rather similar picture.

Recently, *ab initio* calculations in the presence of solvents are more common and readily available in QM packages. The Onsager model,[27] the polarized Continuum model (PCM) of Tomasi,[28] or the isodensity surface polarized continuum model (IPCM)[29] can be applied for amino acids and peptides. Despite their limitations, these approaches could help to mimic the properties of solvent en-

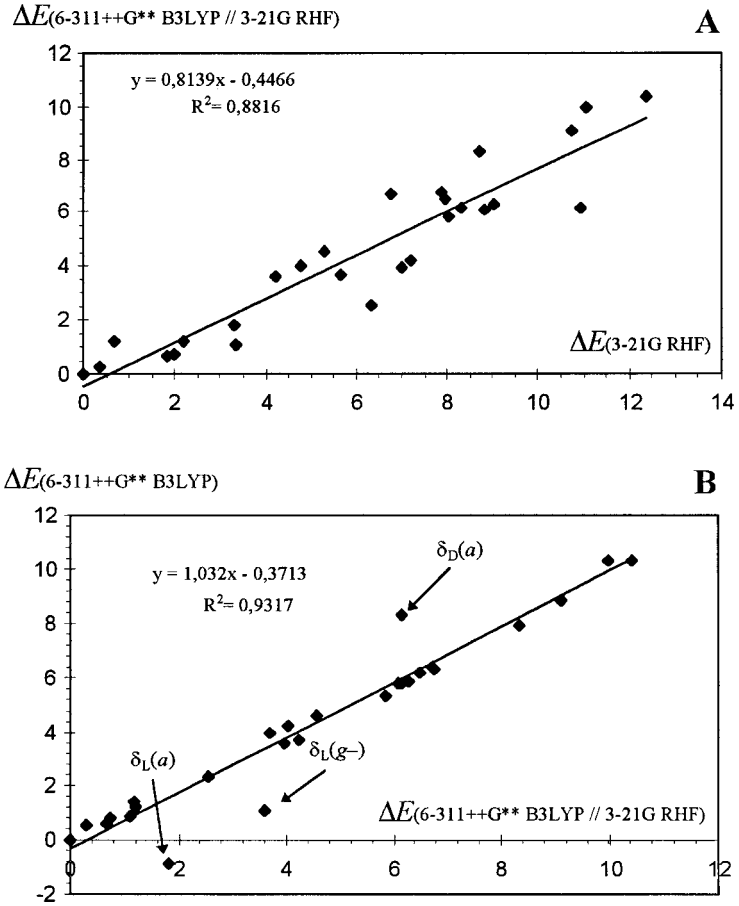


FIGURE 5. Cross-correlation of selected sets of relative energies of For-L-Val-NH_2 [$\Delta E_{RHF/3\text{-}21G}$ with $\Delta E_{B3LYP/6\text{-}311++G^{**}//RHF/e\text{-}21G}$ (A) and $\Delta E_{B3LYP/6\text{-}311++G^{**}}$ with $\Delta E_{B3LYP/6\text{-}311++G^{**}//RHF/3\text{-}21G}$ (B)].

vironment, providing a more realistic picture of "real-life" systems. However, as mentioned above, valine is a typical amino acid residue buried in the inside of proteins. Therefore, vacuum representing hydrophobic environment seems more adequate for this study. When relative energies and natural abundances are compared, neglect of solvent effects seems to be a good approximation (see the next paragraph).

RELATIVE ENERGIES VS. NATURAL ABUNDANCES

In this section we correlate relative *ab initio* energies of conformers of our model systems with relative "natural" abundances of backbone conformations extracted from a carefully filtered set of proteins of known X-ray structure. The energy of the basis set dependent global minimum (e.g., $\gamma_L(g^+)$ and δ_L(a) for For-L-Val-NH_2) was chosen as the reference value for determining relative *ab initio* energies of For-L-Val-NH_2 and For-L-Phe-NH_2. One of the basic assumptions of our analysis is that a low-energy structure will occur more frequently in proteins than a high-energy structure, i.e., "natural" abundance of an amino acid conformer, p_x, is determined by the relative energy of the molecular conformation. Furthermore, we assume that conformations of an amino acid residue (in this study we are interested in Val, Ile, Leu, Phe, Tyr, and Trp) in a protein are simply modeled by those of For-L-*Xxx*-NH_2, with *Xxx* = Val and Phe. This proposed model ignores several factors, most importantly interresidue interactions, long-range effects, and hydration. Finally, a Boltzmann-type exponential distribution is assumed for relative populations:

$$\left(\frac{p_x}{p_{\text{ref}}}\right) = e^{+\Delta E/m} \cdot e^{-b/m} \tag{1}$$

where $\Delta E = E_x - E_{\text{ref}}$. In logarithmic form

$$\Delta E = m \ln\left(\frac{p_x}{p_{\text{ref}}}\right) + b. \tag{2}$$

A protein data set, described elsewhere[20] and including over 10,000 valines, isoleucines, and leucines, over 6000 phenylalanines and tyrosines, and more than 2000 tryptophanes was used in this work. Each $\{\phi, \psi, \chi_1\}$ triad represents a characteristic conformer, as reported for For-L-Val-NH_2 in Table II. The subconformers of the target amino acid residues derived from proteins are expected to form clusters located around the centers of any of the 27 catchment regions. In this study these clusters are approximated by a series of volumes generated by parameter τ, where the volume is determined by $\{\phi+\tau, \psi+\tau, \chi_1+\tau\}$, and the $\{\phi, \psi, \chi_1\}$ triad refers to *ab initio* optimized torsional angles (see Tables VIII and IX and their footnotes for further details). With $\tau = 30°$, 45°, and 60°, increasing volumes are obtained, and consequently more and more residues (called Sum in Tables VII and IX) belong to a given catchment region. Naturally the volumes defined may overlap. The number of residues found in these overlapping regions are called Overlap in Tables VII and IX. It is instructive to see how unevenly the conformers are distributed among the 27 catchment regions. Finally, Pearson correlation coefficients, $R(\zeta)$, where $\zeta = \phi$, ψ, and χ_1, and standard errors are used as two measures of the linearity between "natural" abundance and relative energy (see Tables X and XI for relevant results).

Our systematic analysis involved *ab initio* energies from all conformational libraries (see Tables VI and VII for For-L-Val-NH_2 and For-L-Phe-NH_2, respectively) and the "natural" abundance of Val, Ile, Leu, Phe, Tyr, and Trp residues using conformational volumes of increasing size. Due to model limitations it is unrealistic to expect a perfect correlation with high R and low standard error. The results of Tables X and XI show that for Val the R value can be as high as -0.86 (Fig. 6). For Phe the best R values are significantly lower ($R = -0.77$) (Fig. 7), but this still shows correlation between *ab initio* calculated energies and "natural" abundance of these conformers in proteins. On Figure 6 relative energies of 18 fully optimized For-L-Val-NH_2 (ΔE[RHF/TZ2P//B3LYP/6-311++G**], Level E2) conformers are plotted against the "natural" abundance extracted for the same amino acid conformation from our protein database. When counting "natural" abundance values ($\ln[p_x/p_{\beta L(a)}]$), two sets of $\{\phi, \psi, \chi_1\}$ triads were used for For-L-Val-NH_2: one from 3-21G RHF and the other from 6-311++G**B3LYP optimized structures. The Pearson correlation coefficients are significant for both sets, $R = -0.86$ and $R = -0.81$, respectively (see Table X). Selected conformers of D-type (explicitly marked on Fig. 6) are primarily responsible for the lack of better linear correlations; for example, upon removal of both γ_D conformers from the analysis the R-values increase by some 10%. Systematic narrowing of the size of catchment volumes decreases drastically the number of residues found in a given region (cf. Tables IX and X); nevertheless, the R-values remain high: halving the radius ($60° \Rightarrow 30°$) results in a decrease from a total of 12,921 to 4606 valines (Table VII), while the Pearson correlation coefficient changes only to a small extent: R^{60}[Val(RHF)] $= -0.86 \Rightarrow R^{30}$[Val(RHF)] $= -0.79$. Therefore, enlargement of the volume does not improve significantly the standard error or the quality of the fit (Tables X and XI). This indicates that even for the smallest data set (with $\tau = 30°$) the distribution of the residues among the catchment regions is reasonably close to that detected for the largest volume ($\tau = 60°$). On the other hand, the standard error varies markedly with the change in the applied level of theory, indicating that either the location of the spheres or the relative energy differences change significantly as a function of the applied level of theory. Therefore, we probed fully optimized sets of energies [E(6-311++G** B3LYP), Level D2, or E(6-31+G* RHF), Level A2] and energies from single-point calculations [E(B3LYP/6-311++G**//RHF/3-21G), Level C, or E(RHF/6-31+G*//RHF/3-21G), Level B] against "natural" abundance. Monitoring the variation of the standard errors of the fitting as a function of the *ab initio* method, it looks as if the improved energy and not the improved geometry is responsible for the better fitting. However, the largest difference is induced if both α_L- and ε_L-type structures are incorporated into the analysis. Because none of these structures are minima for these model systems (these structures result from constrained geometry optimizations) the calculated *ab initio* energies are less reliable. Calculations at levels with index 2 (Table I) ignore these structures and indeed the resulting correlation is significantly better. At least two additional observations can be made by inspecting Tables X and XI. Although the *ab initio* energies of For-L-Val-NH_2 can be used to estimate the "natural" abundance of Ile or Leu conformers, these correlations are always less significant (compare, e.g., level E2 series in Table X) than those computed for valine. This suggests that the Ile and Leu residues, both having more side-chain rotamers than Val, can only partially be modeled by For-L-Val-NH_2. The structural difference between phenylalanine and tyrosine is minor, so a stronger

TABLE VIII.

Conformational Preferences of Valine (Val), Leucine (Leu), and Isoleucine (Ile) Residues in Selected Proteins with Known X-Ray Structure.[a]

	Val(RHF)			Val(DFT)			Ile(RHF)			Ile(DFT)			Leu(RHF)			Leu(DFT)		
CR[b]	30°	45°	60°	30°	45°	60°	30°	45°	60°	30°	45°	60°	30°	45°	60°	30°	45°	60°
$\alpha_L(g^+)$	174	250	294	144	240	289	91	195	295	80	193	288	13	29	38	11	29	37
α_L(a)	2568	2822	2984	2569	2822	2984	216	328	368	223	329	368	1644	1990	2202	1654	1998	2212
$\alpha_L(g^-)$	175	340	475	175	340	476	2362	2650	2793	2343	2650	2793	2271	3344	4224	2250	3327	4212
$\alpha_D(g^+)$	1	1	2	1	1	1	1	1	1	—	—	1	—	—	—	—	—	—
α_D(a)	6	13	14	9	11	14	—	2	4	1	2	4	4	9	18	2	9	18
$\alpha_D(g^-)$	5	6	7	6	6	7	3	5	9	3	4	8	41	74	89	40	75	88
$\beta_L(g^+)$	308	436	560	219	412	588	349	494	554	144	412	576	69	117	146	11	62	131
β_L(a)	285	1504	2936	574	2186	3532	126	319	379	163	350	408	73	403	923	116	621	1209
$\beta_L(g^-)$	187	659	943	425	834	1051	29	362	1333	106	771	1983	63	404	1027	201	719	1450
$\delta_L(g^+)$	11	36	73	11	37	74	33	136	243	21	122	236	4	15	24	2	15	22
δ_L(a)	4	22	81	13	48	236	3	11	26	6	18	75	—	8	44	3	17	112
$\delta_L(g^-)$	26	126	245	201	378	560	2	15	60	7	54	317	50	199	532	340	916	1780
$\delta_D(g^+)$	1	10	32	1	7	27	6	16	79	3	15	68	—	—	8	—	2	6
δ_D(a)	1	4	22	25	94	252	1	2	5	2	11	20	—	2	9	11	30	56
$\delta_D(g^-)$	—	4	35	—	25	114	1	2	18	3	22	89	—	1	22	2	16	138
$\varepsilon_L(g^+)$	50	97	169	50	97	169	19	60	140	18	60	141	8	23	32	7	23	32
ε_L(a)	723	1537	2736	717	1534	2727	39	95	162	38	96	162	373	758	1206	365	756	1202
$\varepsilon_L(g^-)$	37	143	302	38	146	306	586	1382	2515	586	1374	2513	421	1203	2206	388	1158	2187
$\varepsilon_D(g^+)$	—	—	1	—	—	—	—	—	1	—	—	—	—	—	1	—	—	—
ε_D(a)	—	5	8	3	6	8	—	—	—	—	—	1	—	1	5	—	2	5
$\varepsilon_D(g^-)$	—	—	—	—	—	—	—	—	3	—	3	—	—	1	4	—	1	3
$\gamma_L(g^+)$	5	11	51	5	22	100	3	10	58	2	7	46	4	9	18	3	9	20
γ_L(a)	29	166	843	25	144	760	1	4	20	1	3	18	18	74	411	15	60	357
$\gamma_L(g^-)$	4	22	98	4	19	78	48	230	1124	83	505	1895	97	261	716	106	324	845
$\gamma_D(g^+)$	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—
γ_D(a)	6	7	8	4	7	11	1	1	1	1	1	2	2	3	7	1	4	7
$\gamma_D(g^-)$	—	—	2	—	—	2	2	6	7	2	6	7	3	5	16	3	5	16
Sum	4606	8221	12,921	5219	9416	14,366	3922	6326	10,199	3836	7005	12,023	5158	8933	13,928	5531	10,178	16,145
Overlap	0	128	365	0	303	3671	0	156	1655	5	363	3396	0	152	1664	2	327	3706

[a] 650 nonhomologous proteins have been selected. Classification of the conformation of the three amino acid residues Val, Ile, and Leu into one of the 27 catchment regions (CR) is based on ϕ, ψ, and χ_1 torsional angles obtained for HCO-L-Val-NH_2 at the 3-21G RHF (RHF) or 6-311++G** B3LYP (DFT) levels and by allowing maximum deviations of 30°, 45°, and 60° from the computed torsional angles. The total number of amino acid residues (Val, Ile, and Leu) clustered according to the different schemes is given as Sum. The number of residues belonging to more than one catchment region is given as Overlap.

[b] See text for labeling of the backbone conformers. Designation of the side-chain conformations, g^+, a, and g^-, follows that used in the PDB.

TABLE IX.

Number of Phenylalanine (Phe), Tyrosine (Tyr), and Tryptophane (Trp) Residues in Selected Proteins of Known X-Ray Structure.[a]

	Phe(3-21G)[b]			Phe(6-31+G*)			Tyr(3-21G)			Tyr(6-31+G*)			Trp(3-21G)			Trp(6-31+G*)		
CR[b]	30°	45°	60°	30°	45°	60°	30°	45°	60°	30°	45°	60°	30°	45°	60°	30°	45°	60°
$\alpha_L(g^+)$	40	68	96	22	61	96	77	124	150	4l	113	147	77	122	147	58	114	149
α_L(a)	1145	1229	1256	1137	1228	1259	998	1075	1091	982	1075	1093	424	448	461	422	450	462
$\alpha_L(g^-)$	507	783	954	507	783	954	470	700	868	470	700	868	245	361	421	245	361	421
$\alpha_D(g^+)$	1	2	2	—	2	2	—	—	—	—	—	—	—	—	—	—	—	—
α_D(a)	6	10	13	5	9	14	1	3	7	1	3	7	1	2	2	1	2	2
$\alpha_D(g^-)$	78	97	106	75	98	106	73	96	104	72	97	105	19	24	26	19	24	26
$\beta_L(g^+)$	197	407	517	428	518	557	134	331	419	345	432	463	65	148	177	149	180	187
β_L(a)	6	35	114	99	223	385	13	38	137	118	259	392	7	17	51	40	86	146
$\beta_L(g^-)$	601	1048	1430	183	532	957	624	982	1330	217	592	907	170	303	454	56	149	275
$\delta_L(g^+)$	14	42	77	23	54	85	24	48	82	33	56	86	17	31	63	22	45	78
δ_L(a)	3	6	40	7	35	226	2	9	26	5	28	188	—	3	12	2	10	85
$\delta_L(g^-)$	239	459	692	552	1114	1334	234	419	603	497	1053	1250	43	85	157	219	423	495
$\delta_D(g^+)$	—	2	12	3	15	31	—	—	11	2	12	21	—	1	3	1	3	10
δ_D(a)	—	5	11	7	11	19	1	5	11	8	13	21	—	2	3	2	4	8
$\delta_D(g^-)$	—	2	21	1	28	92	—	1	25	3	35	111	—	—	5	—	8	30
$\varepsilon_L(g^+)$	3	9	32	—	8	25	5	23	52	2	14	44	2	13	23	2	11	19
ε_L(a)	212	353	494	208	358	496	232	374	481	219	372	481	81	134	186	81	133	184
$\varepsilon_L(g^-)$	159	505	1016	151	499	1014	122	404	801	118	400	801	89	231	373	86	228	373
$\varepsilon_D(g^+)$	—	1	1	—	—	2	—	1	2	—	—	1	—	—	—	—	—	—
ε_D(a)	—	—	1	—	1	1	—	1	1	—	1	1	—	—	1	—	1	1
$\varepsilon_D(g^-)$	—	—	2	4	4	6	2	3	8	4	6	6	—	—	—	1	1	1
$\gamma_L(g^+)$	—	3	12	—	5	24	3	7	19	2	12	42	—	2	24	—	8	40
γ_L(a)	28	91	287	81	255	573	23	74	252	67	240	530	8	35	104	33	97	207
$\gamma_L(g^-)$	53	188	536	84	295	751	35	120	413	59	208	548	17	37	119	22	66	212
$\gamma_D(g^+)$	—	1	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—
γ_D(a)	—	—	1	—	—	1	3	4	6	3	4	5	1	1	1	—	1	1
$\gamma_D(g^-)$	2	5	12	3	5	17	2	7	17	3	7	18	2	3	4	2	4	4
Sum	3294	5351	7736	3580	6141	9028	3078	4849	6916	3271	5732	8136	1268	2003	2817	1463	2409	3416
Overlap	0	189	1702	188	1001	2821	2	144	1420	166	897	2460	0	65	560	97	441	1098

[a] 650 nonhomologous proteins have been selected. Classification of the three amino acid residues Phe, Tyr, and Trp into one of the 27 catchment regions (CR) is based on ϕ, ψ, and χ_1 torsional angle triad obtained for HCO-L-Phe-NH_2 determmed at 3-21G RHF (3-21G) or 6-31+G* RHF (6-31+G*) levels and by allowing maximum deviations of $r = 30°$, 45°, and 60° from the computed torsional angles. The total number of amino acid residues (Phe, Tyr, and Trp) clustered according to the different schemes is given as Sum. The number of residues belonging to more than one catchment region is given as Overlap.

[b] See text for labeling of the backbone conformers. Designation of the side-chain conformations, g^+, a, and g^-, follows that used in the PDB.

TABLE X.

Fitting Parameters [Pearson Correlation Coefficients, *R*, and Standard Errors (in kcal mol^{-1}), σ] of *Ab Initio* Determined Relative Energies ($E - E_{ref}$) of Valine Diamide Conformers and Relative Probabilities ($\ln[p_x]/[p_{ref}]$) of Valine, Isoleucine, and Leucine Residues in Proteins.[a]

(A) R[b]	τ[c]	A1[d]	A2	B1	B2	C1	C2	D1	D2	E1	E2
Val//Val(RHF)[e]	30	−0.04	−0.59	−0.38	−0.76	−0.27	−0.68	−0.27	−0.68	−0.43	−0.79
Val//Val(RHF)	45	−0.27	−0.65	−0.57	−0.80	−0.46	−0.74	−0.49	−0.73	−0.61	−0.81
Val//Val(RHF)	60	−0.40	−0.67	−0.66	−0.79	−0.57	−0.76	−0.58	−0.78	−0.70	**−0.86**
Val//Val(DFT)	30	−0.07	−0.43	−0.43	−0.67	−0.31	−0.58	−0.34	−0.62	−0.46	−0.68
Val//Val(DFT)	45	−0.18	−0.51	−0.53	−0.74	−0.43	−0.66	−0.45	−0.69	−0.56	−0.76
Val//Val(DFT)	60	−0.29	−0.60	−0.62	−0.79	−0.53	−0.73	−0.53	−0.74	−0.64	**−0.81**
Ile//Val(RHF)	30	0.05	−0.42	−0.21	−0.53	−0.07	−0.46	−0.11	−0.41	−0.25	−0.51
Ile//Val(RHF)	45	−0.10	−0.53	−0.37	−0.65	−0.25	−0.59	−0.30	−0.59	−0.41	−0.64
Ile//Val(RHF)	60	−0.25	−0.56	−0.51	−0.68	−0.41	−0.64	−0.44	−0.67	−0.55	**−0.73**
Ile//Val(DFT)	30	0.09	−0.36	−0.19	−0.54	−0.05	−0.48	−0.09	−0.49	−0.23	−0.58
Ile//Val(DFT)	45	0.05	−0.29	−0.22	−0.46	−0.11	−0.43	−0.16	−0.50	−0.26	−0.55
Ile//Val(DFT)	60	−0.28	−0.53	−0.54	−0.67	−0.47	−0.64	−0.48	−0.71	−0.56	**−0.76**
Leu//Val(RHF)	30	0.33	−0.22	0.01	−0.48	0.15	−0.34	—	—	−0.04	−0.51
Leu//Val(RHF)	45	−0.31	−0.71	−0.53	−0.80	−0.41	−0.71	−0.46	−0.72	−0.57	−0.78
Leu//Val(RHF)	60	−0.38	−0.70	−0.59	−0.78	−0.52	−0.76	−0.53	−0.71	−0.63	**−0.77**
Leu//Val(DFT)	30	0.10	−0.29	−0.19	−0.49	−0.02	−0.37	−0.11	−0.36	−0.24	−0.45
Leu//Val(DFT)	45	−0.23	−0.55	−0.49	−0.70	−0.40	−0.65	−0.44	−0.63	−0.53	−0.68
Leu//Val(DFT)	60	−0.26	−0.53	−0.50	−0.66	−0.44	−0.65	−0.44	−0.66	−0.53	**−0.72**

(B) σ[f]	τ	A1[d]	A2	B1	B2	C1	C2	D1	D2	E1	E2
Val//Val(RHF)[e]	30	2.35	1.68	2.17	1.34	2.26	1.51	2.26	1.51	2.13	1.28
Val//Val(RHF)	45	2.21	1.67	1.88	1.31	2.04	1.49	2.01	1.50	1.82	1.29
Val//Val(RHF)	60	2.19	1.76	1.80	1.44	1.97	1.55	1.96	1.53	1.72	1.25
Val//Val(DFT)	30	2.28	1.97	2.07	1.62	2.17	1.79	2.15	1.72	2.04	1.60
Val//Val(DFT)	45	2.20	1.96	1.89	1.52	2.02	1.70	2.00	1.64	1.85	1.49
Val//Val(DFT)	60	2.14	1.91	1.76	1.46	1.90	1.63	1.90	1.59	1.72	1.41
Ile//Val(RHF)	30	2.45	1.86	2.40	1.74	2.45	1.82	2.44	1.89	2.38	1.79
Ile//Val(RHF)	45	2.47	1.96	2.31	1.77	2.41	1.87	2.37	1.89	2.27	1.79
Ile//Val(RHF)	60	2.52	2.12	2.24	1.88	2.37	1.97	2.34	1.96	2.18	1.80
Ile//Val(DFT)	30	2.34	1.86	2.31	1.69	2.35	1.75	2.34	1.81	2.29	1.69
Ile//Val(DFT)	45	2.32	2.18	2.27	2.03	2.31	2.07	2.30	2.06	2.25	1.99
Ile//Val(DFT)	60	2.51	2.23	2.20	1.96	2.31	2.01	2.29	1.91	2.16	1.79
Leu//Val(RHF)	30	2.09	1.53	2.21	1.38	2.19	1.48	—	—	2.21	1.35
Leu//Val(RHF)	45	2.49	1.62	2.21	1.40	2.38	1.62	2.32	1.56	2.14	1.40
Leu//Val(RHF)	60	2.23	1.60	1.94	1.38	2.06	1.45	2.04	1.50	1.88	1.34
Leu//Val(DFT)	30	2.45	1.95	2.42	1.78	2.46	1.90	2.45	1.91	2.39	1.84
Leu//Val(DFT)	45	2.39	1.87	2.14	1.61	2.25	1.71	2.21	1.72	2.09	1.62
Leu//Val(DFT)	60	2.23	1.87	2.00	1.65	2.07	1.67	2.07	1.65	1.96	1.52

[a] 650 nonhomologous proteins have been selected.

[b] Pearson correlation coefficient, $-1 \leq R \leq 1$.

[c] The radius (τ) of the hypersphere (30°, 45°, and 60°). The ϕ, ψ, and χ_1 values, calculated at 3-21G RHF or 6-311++G** B3LYP levels of theory, pinpoint the center of the sphere of radius τ determining the appropriate conformational cluster.

[d] The *ab initio* level of theory used for the energy calculation (for abbreviations see Table I).

[e] Type of amino acid residue (Val, Ile, and Leu) counted and conformationaly clustered in proteins by using the ϕ, ψ, and χ_1 conformational parameters of HCO-L-Val-NH_2 determined at RHF/3-21G (Val/RHF) and at B3LYP/6-311++G** (Val/DFT) levels of theory.

[f] Standard error, in kcal mol^{-1}.

TABLE XI.

Fitting Parameters of *Ab Initio* Relative Energies ($E - E_{ref}$) of Phenylalanine Diamide Conformers and Relative Probabilities ($\ln[p_x/p_{ref}]$) of Phenylalanine, Tyrosine, and Tryptophane Residues in Proteins.[a]

(A) Perason Corr. Coefficient[b]	Radius Size[c]	A[d]	B	C	D	E	F1	F2
Phe//Phe(3-21G)[e]	30	—	−0.46	—	−0.44	−0.44	−0.45	−0.18
Phe//Phe(3-21G)	45	−0.48	−0.70	−0.59	−0.64	−0.64	−0.65	−0.54
Phe//Phe(3-21G)	60	−0.59	−0.74	−0.71	−0.73	−0.74	**−0.75**	−0.55
Phe//Phe(6-31+G*)	30	—	−0.66	—	−0.59	−0.60	−0.61	−0.50
Phe//Phe(6-31+G*)	45	−0.46	−0.68	−0.57	−0.66	−0.67	−0.68	−0.42
Phe//Phe(6-31+G*)	60	−0.60	−0.74	−0.72	−0.76	−0.76	**−0.77**	−0.53
Tyr//Phe(3-21G)	30	−0.18	−0.56	−0.42	−0.50	−0.51	−0.53	−0.17
Tyr//Phe(3-21G)	45	−0.48	−0.68	−0.62	−0.62	−0.64	−0.65	−0.51
Tyr//Phe(3-21G)	60	−0.56	−0.73	−0.69	−0.73	−0.74	**−0.75**	−0.51
Tyr//Phe(6-31+G*)	30	−0.24	−0.57	−0.45	−0.54	−0.56	−0.57	−0.36
Tyr//Phe(6-31+G*)	45	−0.43	−0.68	−0.59	−0.68	−0.69	−0.71	−0.37
Tyr//Phe(6-31+G*)	60	−0.64	−0.78	−0.75	−0.74	−0.75	**−0.76**	−0.51
Trp//Phe(3-21G)	30	—	−0.48	—	−0.49	−0.51	−0.53	0.09
Trp//Phe(3-21G)	45	−0.22	−0.58	−0.45	−0.58	−0.60	−0.61	−0.19
Trp//Phe(3-21G)	60	−0.51	−0.75	−0.70	−0.78	−0.79	**−0.80**	−0.43
Trp//Phe(6-31+G*)	30	—	−0.76	—	−0.69	−0.70	−0.71	−0.36
Trp//Phe(6-31+G*)	45	−0.50	−0.74	−0.68	−0.71	−0.72	−0.73	−0.39
Trp//Phe(6-31+G*)	60	−0.56	−0.76	−0.73	−0.75	−0.76	**−0.77**	−0.36

(B) σ[f]	Radius Size[c]	A[d]	B	C	D	E	F1	F2
Phe//Phe(3-21G)[e]	30	—	1.93	—	1.79	1.78	1.77	2.20
Phe//Phe(3-21G)	45	2.11	1.73	1.96	1.68	1.67	1.66	2.08
Phe//Phe(3-21G)	60	2.10	1.73	1.76	1.64	1.63	1.60	2.11
Phe//Phe(6-31+G*)	*30*	—	1.49	—	1.60	1.59	1.57	1.85
Phe//Phe(6-31+G*)	45	2.09	1.72	1.93	1.75	1.73	1.70	2.02
Phe//Phe(6-31+G*)	60	2.09	1.75	1.73	1.69	1.67	1.64	2.11
Tyr//Phe(3-21G)	30	2.16	1.82	2.12	1.72	1.70	1.69	2.29
Tyr//Phe(3-21G)	45	1.99	1.66	1.85	1.57	1.55	1.53	2.05
Tyr//Phe(3-21G)	60	1.83	1.51	1.61	1.39	1.36	1.34	1.84
Tyr//Phe(6-31+G*)	30	2.12	1.80	2.03	1.86	1.84	1.82	2.09
Tyr//Phe(6-31+G*)	45	2.03	1.64	1.84	1.62	1.60	1.57	1.96
Tyr//Phe(6-31+G*)	60	1.86	1.51	1.60	1.48	1.46	1.42	1.92
Trp//Phe(3-21G)	30	—	1.55	—	1.36	1.35	1.33	1.97
Trp//Phe(3-21G)	45	1.96	1.64	1.80	1.49	1.47	1.46	2.09
Trp//Phe(3-21G)	60	1.85	1.41	1.50	1.26	1.23	1.21	1.90
Trp//Phe(6-31+G*)	30	—	1.33	—	1.46	1.44	1.42	1.97
Trp//Phe(6-31+G*)	45	1.87	1.46	1.57	1.47	1.45	1.42	1.94
Trp//Phe(6-31+G*)	60	1.92	1.50	1.54	1.47	1.45	1.42	2.00

[a] 650 nonhomologous proteins were selected all of high resolution structure.

[b] Pearson correlation coefficient ($-1 \leq R \leq 1$).

[c] The radius (r) of the hypersphere (30°, 45°, and 60°). The ϕ, ψ, and χ_1 values, calculated at RHF/3-21G or RHF/6-31+G* levels of theory. Pinpoint the center of the sphere of radius r determining the appropriate conformational cluster.

[d] The used *ab initio* level of theory for energies calculation (for abbreviations see Methods).

[e] Type of amino acid residue (Phe, Tyr, and Trp) counted and conformationaly clustered in proteins by using ϕ, ψ, and χ_1 conformational parameters of HCO-L-Phe-NH_2 determined at RHF/3-21G (Phe/3-21G) and at RHF/6-31+G* (Phe/6-31+G*) levels of theory.

[f] Standard error, in kcal mol^{-1}.

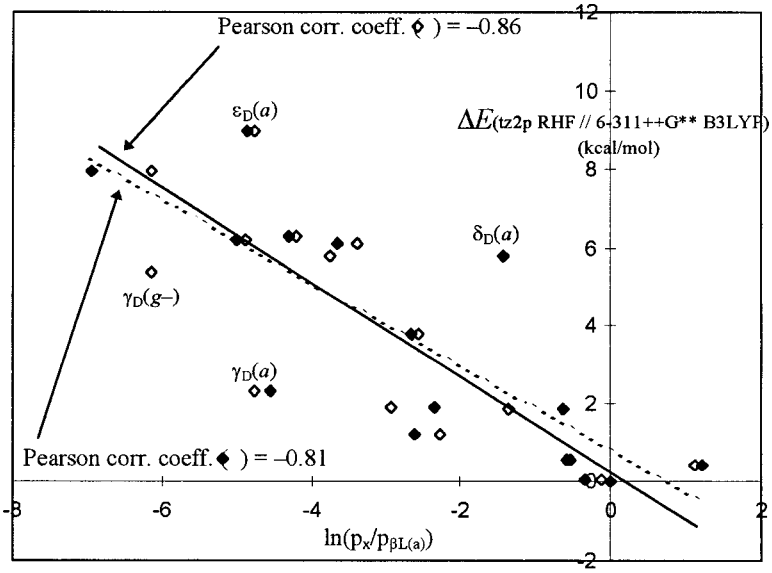


FIGURE 6. Comparison of *ab initio* calculated relative energies and "natural" probabilities of Val residues. Relative energies are for 18 fully optimized For-L-Val-NH_2 [ΔE(RHF/TZ2p/B3LYP/6-311++G**)] (level E2) conformers. "Natural" probabilities are for valines extracted from proteins with known X-ray structures from PDB (Table VIII and X) (27 different 6-311++G** B3LYP and 3-21G RHF conformers were used as template structures for clustering the conformers of valines found in our protein database. Using RHF-optimzed structures a total of 12,921 and for DFT determined references a total of 14,366 valines were conformationaly assigned). To determine natural probabilities ($\ln[p_x/p_{\beta L(a)}]$) of conformer *x* the relevant ϕ, ψ, and χ_1 values were taken from RHF/3-21G (filled diamond) and from B3LYP/6-311++G** (open diamond) fully optimized structures (for energetic values see Table VI).

resemblance is expected. In fact, all three types of aromatic rings are rather similar, having only two side-chain torsional angles (χ_1 and χ_2). Therefore, relative energies determined for For-L-Phe-NH_2 could be used equally well to predict "natural" abundance of Tyr and Trp conformers (see level E2 or F1 series in Table XI).

Conclusions

Multidimensional conformational analysis (MDCA) predicts[2] 3^n distinct structures for any amino acid composed of n relevant torsional angles, such as ϕ, ψ, χ_1, and χ_2.[21] Therefore, after restricting χ_2 to a single value, 27 conformers are expected for our model systems For-L-Val-NH_2 (Val) and For-L-Phe-NH_2 (Phe). Nevertheless, at the 3-21G RHF level the PESs of Val and Phe exhibit only 20 and 19 minima, respectively. Reoptimizing the structures of these conformers at higher levels of theory results in some additional conformational migrations: at the 6-311++G** B3LYP level 18 minima are found for Val, while at the 6-31+G* RHF level 16 minima are found for Phe. Relative energies of these minima were analyzed in conjunction with their conformational properties. Strong correlation was found between any two sets of *ab initio* relative energies.

For peptides and protein fragments, due to fortunate cancellation of errors, 3-21G RHF calculations provide realistic structural information. Even when some of the 3-21G RHF structures cease to be minima at higher levels of theory, the conformational data could be used to obtain remarkably high quality single point energies.

The calculated torsional angles define centers for catchment regions of different relative energy, which can be employed to understand "natural" abundances of conformers of protein building units of known X-ray structure. Hydrophobic amino acid residues, like Val (Ile, Leu) and Phe (Tyr, Trp) form the core of proteins, where only negligible amount of water is present. Consequently, *ab initio* calculations referring to the gas phase are expected to be useful for predicting conformational and energetic (probability) properties of these peptide building units. Analyzing thousands of conformers of Val and Phe residues, a significant correlation was found between their "natural" abundances and the relative energies of the molecular conformations of our Val and Phe model compounds. Although

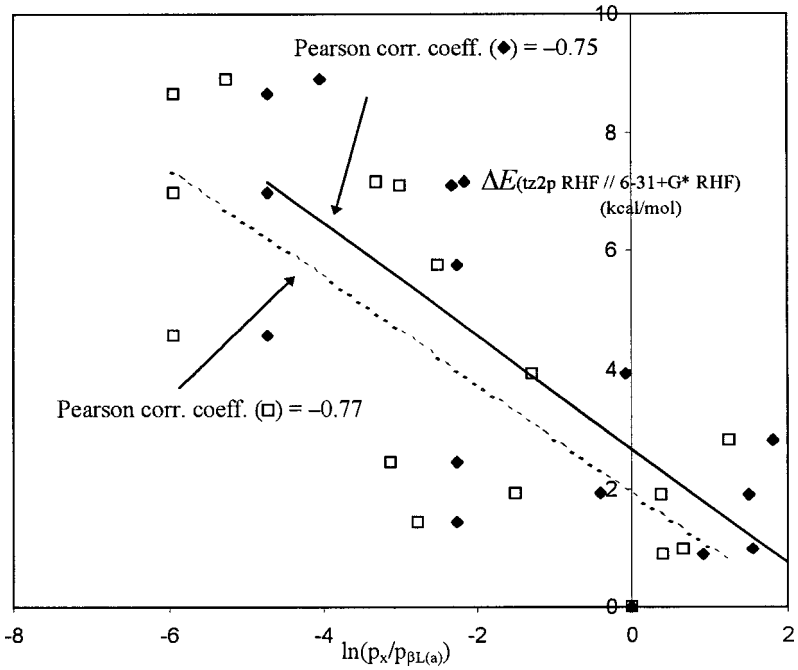


FIGURE 7. Comparison of *ab initio* calculated relative energies and "natural" probabilities of Phe residues. Relative energies are for 16 For-L-Phe-NH_2 (ΔE[RHF/TZ2p//RHF/6-31+G*]) conformers. Natural probabilities are for phenylalanines extracted from protein with known X-ray structures from the PDB (Tables IX and XI) (27 6-31+G* RHF or 3-21G RHF conformers were used as template structures for clustering the conformers of phenylalanines found in our protein database. Using optimized structures from lower level *ab initio* calculations, a total of 7736, while for a higher level of computations a total of 9028 phenylalanines were conformationaly assigned). To determine "natural" probabilities ($\ln[p_x/p_{\beta L(a)}]$) of conformer *x* the relevant ϕ, ψ, and χ_1 values were taken from RHF/3-21G (diamond) and from RHF/6-31+G* (open diamond) fully optimized structures (for energetic values, see Table VII).

the correlation is sensitive to the type and level of *ab initio* calculation performed, in the most favorable cases the Pearson correlation coefficient can be as high as −0.86. This leads to the important conclusion that it is the relative stability of the subconformations of the building units of the core of proteins that determines their structural fold. In other words, although additional factors (like the molecular environment) are certainly of importance, conformational preferences, at least within the core of proteins (where most apolar residues are), are primarily determined by local stability, modeled successfully by our simple diamide systems. Although promising, this idea should be further tested on other amino acid residues, on other model systems, and on larger peptide fragments.

Acknowledgments

The research described was supported by grants from the Hungarian Scientific Research Fund (OTKA T017604, T024044, and T032486).

References

1. Ramakrishnan, C.; Ramachandran, G. N. Biophys J 1965, 5, 909.
2. Perczel, A.; Ángyán, J. G.; Kajtár, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Czsizmadia, I. G. J Am Chem Soc 1991, 113, 6256.
3. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J Mol Biol 1997, 112, 535.
4. Császár, A. G.; Perczel, A. Prog Biophys Mol Biol 1999, 71, 243.
5. Perczel, A.; Csizmadia, I. G. In The Amide Linkage, Structural Significance in Chemistry, Biochemistry and Materials Sciences; Greenberg, A.; Breneman, C. M.; Liebman, J. F., Eds.; Wiley–Interscience: New York, 2000, pp. 409.
6. McAllister, M. A.; Endrédi, G.; Viviani, W.; Perczel, A.; Császár, P.; Ladik, J.; Rivail, J.-L.; Csizmadia, I. G. Can J Chem 1995, 73, 1563.
7. Viviani, W.; Rivail, J.-L.; Perczel, A.; Csizmadia, I. G. J Am Chem Soc 1993, 115, 8321.
8. Perczel, A.; Farkas, Ö.; Csizmadia, I. G. J Am Chem Soc 1996, 118, 7809.

9. Perczel, A.; Farkas, Ö.; Császár, A. G.; Csizmadia, I. G. Can J Chem 1997, 75, 1120.
10. Jákli, I.; Perczel, A.; Farkas, Ö.; Hollósi, M.; Csizmadia, I. G. J Mol Struct (Theochem) 1998, 445, 303.
11. (a) Roothaan, C. C. J. Rev Mod Phys 1951, 23, 69; (b) Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. Ab Initio Molecular Orbital Theory; Wiley–Interscience: New York, 1986.
12. (a) Parr, R. G.; Yang, W. Density-Functional Theory of Atoms and Molecules; Oxford Science Publication: Oxford, 1989; (b) Becke, A. D. J Chem Phys 1993, 98, 5648; (c) Lee, C.; Yang, W.; Parr, R. G. Phys Rev B 1988, 37, 785.
13. Møller, C.; Plesset, M. S. Phys Rev 1934, 46, 618.
14. Császár, A. G. J Am Chem Soc 1992, 114, 9568.
15. Barone, V.; Adamo, C.; Lelj, F. J Chem Phys 1995, 102, 364.
16. Császár, A. G. J Phys Chem 1996, 100, 3541.
17. Frisch, M. J.; et al. Gaussian94, Revision B.2; Gaussian Inc.: Pittsburgh, PA, 1995.
18. Endrédi, G.; Perczel, A.; Farkas, Ö.; McAllister, M. A.; Csonka, G. I.; Ladik, J.; Csizmadia, I. G. J Mol Struct (Theochem) 1997, 391, 15.
19. Schäfer, A.; Huber, C.; Ahlrichs, R. J Chem Phys 1994, 100, 5829.
20. (a) Hobohm, U.; Charf, M.; Schneider, R.; Sander, C. Protein Sci 1992, 1, 409; (b) Hobohm, U.; Sander, C. Protein Sci 1994, 3, 522.
21. IUPAC-IUB Commission on Biochemical Nomenclature, Biochemistry 1970, 9, 3471.
22. Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C., III; Pople, J. A. J Am Chem Soc 1991, 113, 5989.
23. Jákli, I.; Perczel, A.; Farkas, Ö.; Sosa, C.; Csizmadia, I. G. J Comput Chem 2000, 21, 656.
24. Perczel, A.; Farkas, Ö.; Csizmadia, I. G., to appear.
25. Wonnacott, T. M.; Wonnacott, R. J. Introductory Statistics; Wiley: New York, 1990, 5th edn.
26. Császár, A. G. J Mol Struct 1995, 346, 141.
27. Onsager, L. 1936, 58, 1486.
28. Miertus, S.; Scrocco, E.; Tomasic, J. Chem Phys 1981, 55, 117.
29. Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. J Chem Phys 1996, 100, 16098.