Peptide Models XXIII. Conformational Model for Polar Side-Chain Containing Amino Acid Residues: A Comprehensive Analysis of RHF, DFT, and MP2 Properties of HCO-L-SER-NH_2

IMRE JÁKLI,[1] ANDRÁS PERCZEL,[1] ÖDÖN FARKAS,[1] ATTILA G. CSÁSZÁR,[2] CARLOS SOSA,[3] IMRE G. CSIZMADIA[1,4]

[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*
[3]*Silicon Graphics Computer System, 655 E. Lone Oak Drive, Eagan, Minnestoa 55123*
[4]*Department of Chemistry, University of Toronto, Toronto, Ontario, Canada, M5S 1A1*

Received 29 June 1999; accepted 14 January 2000

ABSTRACT: Geometric and energetic properties of a diamide of serine, HCO-NH-L-CH(CH_2OH)CO-NH_2, are investigated by standard methods of computational quantum chemistry. Similarly to other amino acid residues, conformational properties of HCO-L-Ser-NH_2 can be derived from the analysis of its $E = E(\phi, \psi; \chi_1, \chi_2)$ hypersurface. Reoptimization of 44 RHF/3-21G conformers at the RHF/6-311++G^{**} level resulted in 36 minima. For all conformers, geometrical properties, including variation of H-bond parameters and structural shifts in the torsional space, are thoroughly investigated. Results from further single-point energy calculations at the RHF, DFT, and MP2 levels, performed on the entire conformational data set, form a database of 224 energy values, perhaps the largest set calculated so far for any single amino acid diamide.
A comprehensive analysis of this database reveals significant correlation among energies obtained at six levels of *ab initio* theory. Regression parameters provide an opportunity for extrapolation in order to predict the energy of a conformer at a high level by doing explicit *ab initio* computations only for a few selected conformers. The computed conformational and relative energy data are compared with structural and occurrence results derived from a nonhomologous

Correspondence to: A. Perczel; e-mail: perczel@para.chem.elte.hu

Contract/grant sponsor: Hungarian Scientific Research Foundation; contract/grant numbers: OTKA T032486, T030841, and T017192
Contract/grant sponsor: NSERC of Canada

© 2000 John Wiley & Sons, Inc.

protein database incorporating 1135 proteins. © 2000 John Wiley & Sons, Inc.
J Comput Chem 21: 626–655, 2000

Keywords: serine; geometry; energy prediction; *ab initio*; statistics

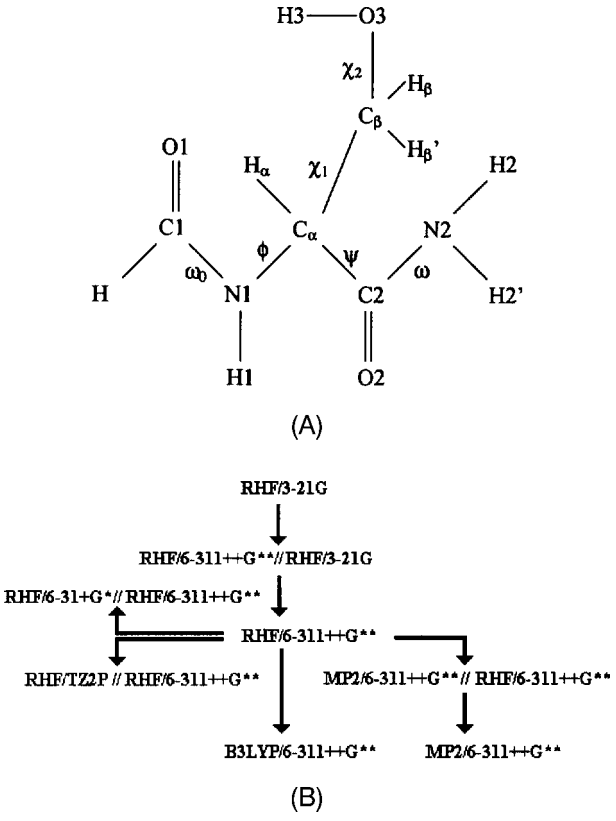
Introduction

Point mutation and other gene-technology tools have greatly widened the frontier of molecular biology. During structure/function studies of proteins, replacement of any amino acid in a given sequence has become a routine task. Such changes may affect not only the local environment of the modified unit but the conformation of the entire macromolecule. Neither the dynamics of the related motions nor the extent of the structural shifts are well understood. These and other problems, associated with molecular flexibility, can be investigated straightforwardly by methods of computational quantum chemistry.[1–8]

It is generally accepted that the 3D structures of proteins can be divided into self-consistent units. This idea of self-organizing subsets is strongly supported by X-ray diffraction and NMR studies that analyzed the conformers of proteins (Brookhaven Protein Databank[9, 10]). Although the stability and the conformational properties of these domains, modules, motives, and secondary structural elements are of tremendous interest,[11–19] formational consequences of a point mutation are exceptionally difficult to predict. One possible strategy to gain insight into structural variations of peptides and proteins is to scan the geometrical and physico-chemical properties of the building units computationally.[1] At the present time, investigation of the potential energy hypersurface (PES) of a peptide (or protein) can be accomplished only at empirical and semiempirical levels of theory. Indeed, many of the molecular mechanics (MM: MM+,[20] AMBER,[21] CHARMM,[22] and OPLS[23, 24]) and semiempirical molecular orbital methods (e.g., MINDO/3,[25] MNDO,[26] AM1,[27] and PM3[28]) have been employed extensively for systems too large for *ab initio* studies. A hindrance of this approach is that the results obtained from empirical (force field) and semiempirical computations often differ from those derived from accurate *ab initio* computations.[29] For example, the different force fields often yield substantially different relative energies and energy order for the conformers of the same macromolecule. Furthermore, even the number of conformers can change significantly from method to method. A complementary strategy is to perform nonempirical quantum chemical calculations on molecular models of moderate size. The unrealistic dependence of the computational result on the method applied is less common at these theoretical levels. Therefore, *ab initio* results are more and more favored for small peptide models, such as PCO—$(Xxx)_n$—NHQ, with Xxx = Gly, Ala, Val, Phe, Ser, . . ., P, Q = H, CH_3, . . ., and $n = 1, 2, 3, \ldots$. Although *ab initio* computations are still expensive, there is an ever-growing number of successful applications.[1, 30–44] Exploratory investigations, in which a systematic topological search was the principal goal, were typically performed using moderate basis sets (e.g., 3-21G, 4-21G) at the Hartree–Fock (HF) level. Nevertheless, computationally more demanding studies, similar to the present one, have started to appear.[45, 46]

Serine is directly involved in a variety of biochemical interactions. Together with threonine, serine is the anchor residue and the binding point of O-glycosylated oligosaccharide antennas attached to proteins. As a residue with a polar side chain, Ser is often located on the surface of proteins and involved in intermolecular electrostatic interactions. It can mediate structural changes, improve hydration of the accessible surface area, get involved in intermolecular adhesion and enzymatic interactions, like that characteristic of serine proteases. The role of Ser in these interactions can be investigated computationally. For example, the catalytic mechanism of aspartylglucosaminidase was recently simulated by *ab initio* computations, showing similarities to the well-known mechanism of serine proteases.[47, 48] The catalytic triad of a serine protease itself was also studied by a quantum mechanical approach.[49] To have a better understanding of serine-containing macromolecules, detailed structural studies are required, wherein conformational properties of simple serine-containing peptide models are investigated.

In the late 1980s, pioneering *ab initio* studies have been carried out on selected conformers of HCO-L-Ser-NH_2 by Scarsdale et al.,[32] Siam et al.[50] These calculations were mostly restricted to the +*syn-clinal* ($g+, g+$) side-chain orientation of the molecule. This simplification was based on the premise that the



SCHEME 1.

variation of the side chain may not affect the topology of the $[\phi, \psi]$ surface. The working hypothesis of these early *ab initio* calculations was confirmed neither by NMR and X-ray[52, 53] techniques nor by recent computational investigations.[51] The coupling between side-chain and backbone orientations initiated additional, more elaborate studies.[39–41, 54, 55] The analysis presented below provides new data on the conformational properties of HCO-L-Ser-NH_2 (Scheme 1A), and is based on information for 36(44) conformers optimized at different *ab initio* levels.

Computational Details and Nomenclature

AB INITIO COMPUTATIONS

A thorough basis set study has been carried out previously on N-formyl-alaninamide[45] at the restricted Hartree–Fock (RHF) and the MP2 levels of theory. A total of 11 basis sets were used; the smallest and largest basis sets were 3-21G and 6-311++G**, respectively. While the total energy gradually lowered with increasing basis set size, the relative stabilities, as measured by the conformational energy differences (ΔE), converged in a nonmonotonic fashion. If anything, the convergence was oscillatory. Interestingly enough the ΔE(RHF/3-21G) values were remarkably close to the converged, or nearly converged, final ΔE(MP2/6-311++G**) results the same pair of basis sets are used, and now reported here, for the case of N-formyl-L-serinamide.

The conformers of HCO-L-Ser-NH_2 were optimized at the RHF level of theory[56] using the standard 6-311++G** basis set. Further geometry optimizations were performed on selected conformers at the B3LYP/6-311++G** and MP2/6-311++G** levels. Single point energies were obtained for all conformers at the RHF/6-31+G*//RHF/6-311++G**, RHF/6-311++G**//RHF/3-21G, RHF/TZ2P[57]//RHF/6-311++G**, and MP2/6-311++G**//RHF/6-311++G** levels (see Scheme 1B). Computations were performed with the program package Gaussian94.[58] Conformational parameters and relative energy values are summarized in Tables IA–IV. The energies listed are relative to $E_{\gamma L(g+,g+)}$, the global minimum of HCO-L-Ser-NH_2.

DATABASE

Data for 9511 Ser residues were collected from 1135 nonhomologous proteins.[59, 60] All entries correspond to high-resolution X-ray structures taken from the 1998 issue of the Protein Databank (PDB).[9, 10] All serine residues were placed on a $[\phi, \psi]$-type Ramachandran map (see Fig. 1).

NOMENCLATURE FOR BACKBONE ORIENTATIONS

The backbone structure of peptides and proteins is often described by three torsional variables per amino acid residue: ϕ, ψ, and ω (Scheme 1A). A further simplified potential energy surface, $E = E(\phi, \psi)$, called the Ramachandran map, is employed to visualize the energy cost associated with changes in the relative orientation of two geminally located amide bonds. An important property of the Ramachandran map is that it can be divided into so-called conformational (catchment) regions. These regions are employed as conformational descriptors for backbone orientations of peptides, and are labeled differently by different research groups. According to Karplus, for example, the Ramachandran surface has twelve distinct regions, labeled as α_L, α_R, β_S, β_P, γ, γ', δ_L, δ_R, ε, ε', ε'', and ζ. This division of the map is similar to the one proposed previously by Efimov.[61] Prior to these descriptions,

TABLE IA.
Torsional Angles of Optimized *Ab Initio* Conformers of HCO-L-Ser-NH_2 at the RHF/6-311++G and RHF/3-21G Levels of Theory.**

Conf.[a,b]	ω_0	ϕ	ψ	ω	χ_1	χ_2
$\beta_L(g-,a) \Rightarrow \gamma_L(g-,a)$	— (−179.46)	— (−137.86)	— (160.53)	— (177.19)	— (−64.62)	— (167.81)
$\beta_L(g-,g+)$	177.65 (177.01)	−174.03 (−179.13)	165.53 (172.54)	179.3 (179.41)	−94.39 (−89.02)	57.55 (55.32)
$\beta_L(g+,g-)$	171.47 (175.01)	−153.30 (−166.61)	173.98 (174.83)	173.21 (176.66)	67.37 (67.22)	−66.15 (−60.11)
$\beta_L(g+,a)$	176.21 (179.41)	−156.76 (−170.14)	173.34 (175.08)	172.55 (177.67)	70.06 (68.26)	−162.77 (−171.93)
$\beta_L(a,a)$	172.85 (174.98)	−155.95 (−171.36)	174.89 (−173.38)	174.02 (−179.27)	−170.4 (−172.95)	162.56 (157.23)
$\beta_L(a,g+)$	172.8 (175.23)	−157.29 (−170.31)	−179.1 (−171.5)	174.69 (179.82)	−165.68 (−171.8)	89.48 (90.14)
$\delta_L(g-,g-)$	−166.39 (−172.21)	−134.06 (−151.98)	17.07 (35.73)	171.86 (176.91)	−51.72 (−46.76)	−69.5 (−54.07)
$\delta_L(g-,a) \Rightarrow \alpha_L(g-,a)$	— (−166.62)	— (−130.24)	— (29.98)	— (176.41)	— (−52.75)	— (−166.21)
$\delta_L(a,g-) \Rightarrow \gamma_L(a,g-)$	— (−173.51)	— (−127.86)	— (33.18)	— (176.41)	— (−172.42)	— (−60.39)
$\delta_L(g+,a)$	−170.57 (−172.71)	−114.38 (−118.21)	11.79 (20)	174.12 (177.4)	53.42 (51.5)	161.49 (160.01)
$\gamma_L(g-,g-)$	−171.19 (−171.15)	−86.96 (−77.26)	73.12 (62.95)	−174.17 (−179.96)	−54.95 (−41)	−71.96 (−75.8)
$\gamma_L(g-,a)$	−169.29 (−168.8)	−86.99 (−77.05)	71.14 (61.42)	−174.96 (179.61)	−57.28 (−43.95)	−175.94 (−178.53)
$\gamma_L(g-,g+)$	176.42 (−177.95)	−86.66 (−85.29)	82.85 (67.1)	−172.81 (−179.22)	−67.55 (−65.37)	61.64 (54.07)
$\gamma_L(a,g-)$	−177.22 (−173.78)	−86.45 (−83.42)	69.16 (62.5)	−178.11 (179.79)	−178.33 (179.75)	−71.9 (−68.42)
$\gamma_L(a,g+) \Rightarrow \beta_L(a,g+)$	— (−177.86)	— (−86.59)	— (77.83)	— (−176.83)	— (−170.03)	— (74.02)
$\gamma_L(g+,g+)$	−176.82 (−176.2)	−85.43 (−83.66)	72.24 (71.42)	−175.09 (−177.55)	54.44 (51.84)	69.82 (69.86)
$\alpha_L(a,a)$	−170.61 (−174.55)	−73.91 (−62.39)	−37.99 (−42.81)	171.58 (−179.72)	−172.87 (179.89)	−167.36 (−168.71)
$\alpha_L(g-,a)$	−164.89 (−169.36)	−80.77 (−70.53)	−14.24 (−24.94)	172.39 (−179.77)	−52.65 (−47.53)	−172.3 (−174.52)
$\alpha_L(g-,g-)$	−166.59 (−171.05)	−82.60 (−71.98)	−14.9 (−23.68)	173.44 (−179.17)	−53.63 (−45.62)	−76.4 (−77.51)
$\varepsilon_D(g-,g-) \Rightarrow \gamma_D(g-,g-)$	— (−173.8)	— (68.93)	— (178.24)	— (179.55)	— (−58.12)	— (−67.32)
$\varepsilon_D(g-,g+)$	−168.61 (−165.26)	56.02 (64.49)	−136.97 (177.92)	174 (179.68)	−60.3 (−63.53)	65.38 (67.4)
$\varepsilon_D(a,a)$	−164.54 (−166.85)	71.28 (68.42)	−170.18 (−172.85)	−176.31 (−179.03)	−157.75 (−162.47)	178.95 (−179.31)
$\varepsilon_D(a,g+)$	−164.61 (−167.12)	69.72 (66.89)	−167.07 (−168.95)	−179.73 (179.75)	−161.14 (−166.37)	81 (82.38)
$\varepsilon_D(g+,g-)$	−170.13 (−175.46)	89.91 (99.84)	−141.95 (−116.85)	177.77 (179.85)	84.61 (76.29)	−66.63 (−68.69)
$\varepsilon_D(g+,a)$	−175.18 (178.31)	41.83 (43.02)	−115.63 (−105.53)	171.23 (176.85)	64.47 (61.25)	−171.34 (−171.58)

TABLE IA.
(Continued)

Conf.[a,b]	ω_0	ϕ	ψ	ω	χ_1	χ_2
$\gamma_D(g-,g-)$	170.61 (170.77)	75.33 (74.66)	−50.48 (−55.23)	−177.54 (−178.23)	−60.61 (−57.62)	−79.95 (−78.2)
$\gamma_D(g-,a)$	170.18 (170.32)	77.05 (75.42)	−55.49 (−56.12)	179.53 (−178.81)	−59.79 (−58.64)	172.99 (176.01)
$\gamma_D(g-,g+)$	177.18 (177.67)	76.48 (72.21)	−55.35 (−57.54)	−178.59 (−178.3)	−57.39 (−61.11)	90.55 (79.1)
$\gamma_D(a,g-) \Rightarrow \alpha_D(a,g-)$	— (172.41)	— (67.46)	— (−31.2)	— (−176.13)	— (−163.77)	— (−39.85)
$\gamma_D(a,a)$	−179.32 (178.25)	75.33 (73.99)	−74.56 (−64.98)	172.47 (−179.46)	−168.08 (−177.46)	−160.45 (−157.97)
$\gamma_D(a,g+)$	179.59 (176.57)	74.23 (71.3)	−61.43 (−52.17)	−178.11 (−175.86)	−179.23 (170.84)	59.58 (48.99)
$\gamma_D(g+,g-)$	−178.51 (−179.61)	80.21 (77.96)	−37.71 (−45.24)	−169.55 (−176.74)	83.51 (81.92)	−64.89 (−62.21)
$\gamma_D(g+,a)$	172.4 (168.47)	58.45 (51.9)	−27.74 (−28.66)	−173.04 (−177.43)	70.5 (65.81)	173.12 (173.28)
$\gamma_D(g+,g+)$	170.05 (171.24)	62.66 (62.93)	−18.59 (−40.34)	−176.01 (−177.71)	46.5 (41.7)	53.81 (48.73)
$\alpha_D(g-,a)$	165.47 (170.43)	66.39 (60.25)	33.17 (37.6)	−173.43 (−179.9)	−55.62 (−57.97)	179.43 (179.01)
$\alpha_D(g+,g+)$	165.89 (167.53)	48.28 (46.39)	50.81 (53.58)	−170.33 (−177.4)	58.42 (56.24)	65.08 (62.1)
$\alpha_D(a,g-)$	170.4 (174.86)	64.30 (62.32)	34.72 (34.14)	−174.34 (179.56)	−162.32 (−168.06)	−67.43 (−65.39)
$\alpha_D(a,g+)$	172.39 (177.17)	66.54 (60.1)	39.84 (43.84)	−172.81 (−179.76)	−148.85 (−156.93)	78.64 (79.05)
$\delta_D(g-,g-)$	173.68 (173.92)	−148.44 (−157.92)	−50.65 (−51.78)	−171.96 (−178.56)	−50.59 (−47.49)	−37.88 (−32.19)
$\delta_D(a,g+)$	175.77 (174.52)	−147.48 (−173.29)	−72.03 (−49.35)	−173.3 (−175.69)	173.59 (163.93)	73.42 (68.23)
$\delta_D(g+,g-)$	173.12 (173.74)	−153.88 (−159.3)	−62.82 (−67.57)	−172.52 (−179.08)	55.27 (52.28)	−86.1 (−76.04)
$\delta_D(g+,a)$	−179.98 (179.87)	−154.02 (−163.74)	−58.32 (−63.28)	−173.26 (179.71)	59.7 (55.09)	−168.36 (−169.82)
$\delta_D(g-,g+) \Rightarrow \gamma_D(g-,g+)$	— (174.77)	— (146.26)	— (−33.93)	— (−175.62)	— (−75.23)	— (74.37)
$\delta_D(a,a) \Rightarrow \beta_L(a,a)$	— (174.31)	— (−172.41)	— (−55.07)	— (−177.17)	— (168.63)	— (165.94)

[a] The backbone conformers are labeled according to the set of abbreviation introduced in the past[31]: $\alpha_L, \alpha_D, \beta_L, \gamma_L, \gamma_D, \delta_L, \delta_D, \varepsilon_L$, and ε_D.
[b] The values in parenthesis are calculated at RHF/3-21G level of theory.

TABLE IB.

Structural and Energetic Parameters of Selected *Ab Initio* Conformers of HCO-L-Ser-NH_2 at the B3LYP/6-311++G (A) and MP2/6-311++G** (B) Levels of Theory.**

Method	Conf.	ω_0	ϕ	ψ	ω	χ_1	χ_2	N—Cα—C	E (Hartree)	ΔE (kcal mol^{-1})
A	$\beta_L(g-,g+)$	176.98	−176.73	169.85	−179.22	−88.79	54.49	106.34	−492.6003925	3.20
B		179.84	−177.66	160.76	168.54	−94.07	54.98	106.73	−491.3068282	4.56
A	$\beta_L(a,a)$	172.80	−159.05	−178.58	176.04	−170.77	164.35	107.18	−492.6007172	3.00
B		171.9	−155.34	175.87	170.27	−171.93	159.13	107.49	−491.308436	3.55
A	$\beta_L(a,g+)$	173.13	−159.04	−176.33	175.07	−168.23	86.67	107.64	−492.6016711	2.40
B		171.68	−156.42	−176.69	170.84	−166.37	82.13	107.88	−491.3095185	2.88
A	$\gamma_L(a,g-)$	−175.75	−81.59	63.08	−179.97	−178.29	−66.63	112.16	−492.6006617	3.03
B		−175.62	−84.02	69.84	−173.55	179.73	−67.41	110.9	−491.3086594	3.41
A	$\gamma_L(g+,g+)$	−177.46	−81.39	72.20	−174.57	55.67	66.87	110.94	−492.6054963	0.00
B		−177.62	−82.43	75.91	−169.66	54.13	66.03	110.19	−491.3141004	0.00
A	$\varepsilon_D(g+,g-)$	−173.38	98.32	−121.63	177.21	77.99	−67.07	108.01	−492.6002788	3.27
B		−167.97	79.93	−143.21	175.95	86.08	−60.81	109.23	−491.3074246	4.19
A	$\varepsilon_D(g+,a)$	172.41	53.99	−29.50	−174.50	67.52	174.22	119.22	−492.5874125	11.35
B		173	58.43	−29.09	−170.94	70.6	174.92	118.41	−491.2962503	11.20
A	$\alpha_D(a,g+)$									
B		171.1	62	43.65	−169.48	−155.27	69.06	111.43	−491.292331	13.66

TABLE IIA.

Statistical Parameters of the Pair-Wise Comparison[a] of Structural Parameters Determined at RHF/3-21G, RHF/6-311++G, B3LYP/6-311++G**, and MP2/6-311++G** Levels of Theory for HCO-L-Ser-NH_2 Molecules.**

	ω_0	ϕ	ψ	ω	χ_1	χ_2	$\tau_{(N—C\alpha—C)} - 110^\circ$
(a) RHF/3-21G ⇔ RHF/6-311++G**[a]							
R^2 [b]	0.884	0.991	0.988	0.368	0.998	0.998	0.872
Pearson cor. coef.[c]	0.940	0.995	0.994	0.606	0.999	0.999	0.934
standard error (σ)[d]	2.649	8.819	12.533	1.502	5.081	4.507	1.259
b[e]	38.108	−3.832	2.802	147.757	−7.131	−6.057	−0.870
m[e]	0.788	1.007	0.983	0.181	1.029	1.031	1.100
(b) RHF/3-21G ⇔ B3LYP/6-311++G**							
R^2 [b]	0.863	0.996	0.999	0.564	1.000	0.999	0.986
Pearson cor. coef.[c]	0.929	0.998	1.000	0.751	1.000	1.000	0.993
standard error (σ)[d]	2.613	5.593	2.599	0.959	1.196	3.355	0.664
b[e]	0.265	−1.711	0.339	138.457	−3.162	2.359	−126.546
m[e]	1.000	0.986	1.012	0.234	1.007	0.982	1.141
(c) RHF/3-21G ⇔ MP2/6-311++G**							
R^2 [b]	0.637	0.986	0.990	0.358	0.998	0.998	0.979
Pearson cor. coef.[c]	0.798	0.993	0.995	0.598	0.999	0.999	0.990
standard error (σ)[d]	3.906	11.354	10.600	1.070	3.490	4.116	0.748
b[e]	66.847	3.305	−2.285	166.961	−8.513	7.760	−142.940
m[e]	0.626	0.965	1.049	0.076	1.039	0.956	1.291
(d) RHF/6-311++G** ⇔ B3LYP/6-311++G**							
R^2 [b]	0.961	0.997	0.992	0.924	0.999	1.000	0.980
Pearson cor. coef.[c]	0.981	0.998	0.996	0.961	0.999	1.000	0.990
standard error (σ)[d]	1.446	5.247	8.595	1.492	3.294	2.121	0.621
b[e]	−17.245	1.908	2.096	−20.757	5.089	4.077	−96.022
m[e]	1.099	0.986	0.967	1.115	0.971	0.976	0.874
(e) RHF/6-311++G** ⇔ MP2/6-311++G**							
R^2 [b]	0.975	0.999	0.999	0.821	0.999	0.999	0.991
Pearson cor. coef.[c]	0.988	0.999	1.000	0.906	1.000	1.000	0.995
standard error (σ)[d]	1.124	3.343	2.642	2.392	2.387	2.873	0.387
b[e]	25.460	7.125	−3.150	93.332	−0.755	8.950	−111.249
m[e]	0.856	0.966	1.017	0.486	1.012	0.952	1.014
(f) B3LYP/6-311++G** ⇔ MP2/6-311++G**							
R^2 [b]	0.933	0.992	0.991	0.626	0.998	0.999	0.974
Pearson cor. coef.[c]	0.966	0.996	0.995	0.791	0.999	0.999	0.987
standard error (σ)[d]	1.696	8.195	9.726	2.844	3.771	3.612	0.790
b[e]	42.333	6.462	−3.079	117.100	−5.388	3.550	−13.456
m[e]	0.759	0.974	1.039	0.352	1.033	0.981	1.123

[a] All available values reported in Tables IA, IB, and IV were used for all four levels of theory.
[b] The square of Pearson correlation coefficient.
[c] Pearson correlation coefficient [cf. eq. (4)].
[d] Standard error [cf. eq. (3)].
[e] The parameters of the fitted line $\{y = mx + b\}$.

TABLE IIB.

***Ab Initio* Conformational Shifts between the Appropriate Dihedral Parameters Calculated for Structures Obtained at RHF/3-21G and RHF/6-311++G** Levels of Theory for HCO-L-Ser-NH_2.**

Conf.[a]	$\Delta\omega_0$	$\Delta\phi$	$\Delta\psi$	$\Delta\omega$	$\Delta\chi_1$	$\Delta\chi_2$
$\beta_L(g-,a) \Rightarrow \gamma_L(g-,a)$	—	—	—	—	—	—
$\beta_L(g-,g+)$[b]	−0.6	−5.1	7.0	0.1	5.4	−2.2
$\beta_L(g+,g-)$	3.5	−13.3	0.9	3.4	−0.1	6.0
$\beta_L(g+,a)$	3.2	−13.4	1.7	5.1	−1.8	−9.2
$\beta_L(a,a)$	2.1	−15.4	11.7	6.7	−2.5	−5.3
$\beta_L(a,g+)$	2.4	−13.0	7.6	5.1	−6.1	0.7
$\delta_L(g-,g-)$	−5.8	−17.9	18.7	5.0	5.0	15.4
$\delta_L(g-,a) \Rightarrow \alpha_L(g-,a)$	—	—	—	—	—	—
$\delta_L(a,g-) \Rightarrow \gamma_L(a,g-)$	—	—	—	—	—	—
$\delta_L(g+,a)$	−2.1	−3.8	8.2	3.3	−1.9	−1.5
$\gamma_L(g-,g-)$	0.0	9.7	−10.2	−5.8	14.0	−3.8
$\gamma_L(g-,a)$	0.5	9.9	−9.7	−5.4	13.3	−2.6
$\gamma_L(g-,g+)$	5.6	1.4	−15.8	−6.4	2.2	−7.6
$\gamma_L(a,g-)$	3.4	3.0	−6.7	−2.1	−1.9	3.5
$\gamma_L(a,g+) \Rightarrow \beta_L(a,g+)$	—	—	—	—	—	—
$\gamma_L(g+,g+)$	0.6	1.8	−0.8	−2.5	−2.6	0.0
$\alpha_L(a,a)$	−3.9	11.5	−4.8	8.7	−7.2	−1.3
$\alpha_L(g-,a)$	−4.5	10.2	−10.7	7.8	5.1	−2.2
$\alpha_L(g-,g-)$	−4.5	10.6	−8.8	7.4	8.0	−1.1
$\varepsilon_D(g-,g-) \Rightarrow \gamma_D(g-,g-)$	—	—	—	—	—	—
$\varepsilon_D(g-,g+)$	3.4	8.5	−45.1	5.7	−3.2	2.0
$\varepsilon_D(a,a)$	−2.3	−2.9	−2.7	−2.7	−4.7	1.7
$\varepsilon_D(a,g+)$	−2.5	−2.8	−1.9	−0.5	−5.2	1.4
$\varepsilon_D(g+,g-)$	−5.3	9.9	25.1	2.1	−8.3	−2.1
$\varepsilon_D(g+,a)$	−6.5	1.2	10.1	5.6	−3.2	−0.2
$\gamma_D(g-,g-)$	0.2	−0.7	−4.8	−0.7	3.0	1.8
$\gamma_D(g-,a)$	0.1	−1.6	−0.6	1.7	1.2	3.0
$\gamma_D(g-,g+)$	0.5	−4.3	−2.2	0.3	−3.7	−11.5
$\gamma_D(a,g-) \Rightarrow \alpha_D(a,g-)$	—	—	—	—	—	—
$\gamma_D(a,a)$	−2.4	−1.3	9.6	8.1	−9.4	2.5
$\gamma_D(a,g+)$	−3.2	−2.9	9.3	2.3	−9.9	−10.6
$\gamma_D(g+,g-)$	−1.1	−2.3	−7.5	−7.2	−1.6	2.7
$\gamma_D(g+,a)$	−3.9	−6.6	−0.9	−4.4	−4.7	0.2
$\gamma_D(g+,g+)$	1.2	0.3	−21.8	−1.7	−4.8	−5.1
$\alpha_D(g-,a)$	5.0	−6.1	4.4	−6.5	−2.4	−0.4
$\alpha_D(g+,g+)$	1.6	−1.9	2.8	−7.1	−2.2	−3.0
$\alpha_D(a,g-)$	4.5	−2.0	−0.6	−6.1	−5.7	2.0
$\alpha_D(a,g+)$	4.8	−6.4	4.0	−6.9	−8.1	0.4
$\delta_D(g-,g-)$	0.2	−9.5	−1.1	−6.6	3.1	5.7
$\delta_D(a,g+)$	−1.3	−25.8	22.7	−2.4	−9.7	−5.2
$\delta_D(g+,g-)$	0.6	−5.4	−4.7	−6.6	−3.0	10.1
$\delta_D(g+,a)$	−0.1	−9.7	5.0	−7.0	−4.6	−1.5
$\delta_D(g-,g+) \Rightarrow \gamma_D(g-,g+)$	—	—	—	—	—	—
$\delta_D(a,a) \Rightarrow \beta_L(a,a)$	—	—	—	—	—	—

[a] The backbone conformers are labeled according to the set of abbreviation introduced in the past (ref. 31): $\alpha_L, \alpha_D, \beta_L, \gamma_L, \gamma_D, \delta_L, \delta_D, \varepsilon_L$, and ε_D.

[b] Differences are calculated from the subtraction of RHF/6-311++G** conformational values from the appropriate RHF/3-21G values.

TABLE IIC.
The Most Important Conformational Shifts Observed between RHF/3-21G and RHF/6-311++G *Ab Initio* Conformers of HCO-L-Ser-NH_2.**

Conf.[a]	$\Delta\omega_0$[b]	$\Delta\phi$	$\Delta\psi$	$\Delta\omega$	$\Delta\chi_1$	$\Delta\chi_2$	Average Shift[c]	ΔE(RHF/3-21G)[d]	ΔE(RHF/6-311++G**)
$\delta_L(g-,g-)$	−5.8	**−17.9**	**18.7**	5.0	5.0	**15.4**	11.3	11.0	6.4
$\varepsilon_D(g-,g+)$	3.4	8.5	**−45.1**	5.7	−3.2	2.0	11.3	16.3	9.5
$\varepsilon_D(g+,g-)$	−5.3	9.9	**25.1**	2.1	−8.3	−2.1	8.8	4.9	4.9
$\gamma_D(g+,g+)$	1.2	0.3	**−21.8**	−1.7	−4.8	−5.1	5.8	14.0	10.9
$\delta_D(a,g+)$	−1.3	**−25.8**	**22.7**	−2.4	−9.7	−5.2	11.2	15.7	10.1

[a] The backbone conformers are labeled according to the set of abbreviation introduced previously (ref. 31): $\alpha_L, \alpha_D, \beta_L, \gamma_L, \gamma_D, \delta_L, \delta_D, \varepsilon_L$, and ε_D.
[b] Differences are calculated from the subtraction of RHF/6-311++G** conformational parameters from the appropriate RHF/3-21G values.
[c] Average shift over the six ($\omega_0, \phi, \psi, \omega, \chi_1$, and χ_2) conformational variables (in degrees).
[d] Relative energy values (in kcal mol^{-1}), compared to $E^{\gamma L(g+,g+)}$.

Zimmerman and coworkers[62] suggested to distinguish 16 regions, labeled as A, B, C, D, E, F, G, H, A*, B*, C*, D*, E*, F*, G*, and H*. Richardson and Richardson,[63] Rooman and coworkers,[64] and Thornton et al.[65] also proposed alternative ways to cluster the different backbone conformers. The above approaches are all based on an analysis of X-ray data of proteins. Based only on the topology of two geminally located amide bonds encompassing a chiral center, we proposed our own labeling scheme. We speculated that a maximum of nine basically distinct backbone structures can appear on the $E = E(\phi,\psi)$ PES. Following IUPAC-IUB recommendations, the *gauche*+ (*g*+), the *anti* (a), and the *gauche*− (*g*−) descriptors can be used for their notation. Ideal locations of the nine expected minima can be traced on Scheme 2A.

We introduced the following short-hand notation for the 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 $\varepsilon_D \equiv (g+,a)$[31] (Scheme 2B). This nomenclature is consistent, unambiguous, and it is based entirely on the conformational properties of peptide units. It also incorporates the nomenclature traditionally used for peptides [e.g., α-structures := $(\alpha_L)_n$, β-structures := $(\beta_L)_n$, inverse γ-turns := γ_L]. This description reflects the "chirality" of the Ramachandran surface and the relative energetics of the appropriate pairs of conformer: L-amino acid residues favor conformations from the L-valley, while D-amino acids favor conformers within the D-valley. This convention provides additional features for fine tuning the notation of more specific conformers: e.g., $[(\alpha_L)_n]^{3_{10}}$ or $[(\alpha_L)_n]^{\pi}$ could stand for 3_{10}- or π-helices, obviously members of the larger α-helix family. The only slight problem arises when one has to specify right-handed and left-handed helical conformations with this notation:

$$\alpha_{\text{Right-handed}} = (\alpha\text{-helix})_{\text{right}} \Rightarrow (\alpha_L)_n \Rightarrow (\alpha_P)$$
$$\alpha_{\text{Left-handed}} = (\alpha\text{-helix})_{\text{left}} \Rightarrow (\alpha_D)_n \Rightarrow (\alpha_M)$$

It is to be mentioned that IUPAC-IUB recommends, in general, to distinguish the two forms of helical structures by using the P and M rather than the R and L subscripts.

NOMENCLATURE FOR SIDE-CHAIN ORIENTATIONS

In serine, which contains two vicinal H^β (H^β_A and H^β_B) protons and a hydroxyl group, distinct χ_1 rotamers are expected. The β carbon atom is a prochiral center next to the chiral C^α. Thus, for HCO-L-Ser-NH_2 the three predicted orientations (*gauche*+, *anti*, and *gauche*−) about χ_1 will differ energetically. Similarly to χ_1, the χ_2 rotation (—C^β—O^γ—) is expected to result in three energetically distinct rotamers. This results in a total of nine different side-chain conformers for each stable backbone structure of this peptide model.

Structure and Stability

Even for a simple peptide model such as HCO-L-Ser-NH_2, characterization of the conformational ensemble by higher order Møller–Plesset (MP), coupled cluster (CC) or other high quality methods of computational quantum chemistry is unrealistic.

TABLE IIIA.

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

Conf. Number	Conf.[b] Code	ΔE(RHF/ 3-21G)[c]	ΔE(RHF/6-31+G*// RHF/6-311++G**)[d]	ΔE(RHF/ 6-311++G**// RHF/3-21G)[e]	ΔE(RHF/ 6-311++G**)[f]	ΔE(RHF/TZ2P// RHF/6-311++G**)[g]	ΔE(MP2/ 6-311++G**// RHF/6-311++G**)[h]
1	$\beta_L(g-,a)$	15.39		6.92			
2	$\beta_L(g-,g+)$	5.56	4.13	3.15	3.86	3.78	4.43
3	$\beta_L(g+,g-)$	9.09	5.10	4.02	4.51	4.59	4.96
4	$\beta_L(g+,a)$	11.18	6.61	5.30	5.78	6.08	6.63
5	$\beta_L(a,a)$	3.79	2.84	2.31	2.38	2.73	3.31
6	$\beta_L(a,g+)$	3.28	2.38	1.98	2.10	2.20	2.74
7	$\delta_L(g-,g-)$	10.97	6.81	6.06	6.36	6.53	7.12
8	$\delta_L(g-,a)$	13.28		6.00			
9	$\delta_L(a,g-)$	8.32		4.15			
10	$\delta_L(g+,a)$	8.64	4.84	3.81	4.33	4.94	5.11
11	$\gamma_L(g-,g-)$	7.51	5.33	4.98	4.89	5.02	5.33
12	$\gamma_L(g-,a)$	7.52	5.04	4.55	4.53	4.80	5.17
13	$\gamma_L(g-,g+)$	10.44	6.31	4.93	5.71	5.84	6.11
14	$\gamma_L(a,g-)$	4.76	3.02	2.19	2.78	2.80	3.32
15	$\gamma_L(a,g+)$	12.51		6.77			
16	$\gamma_L(g+,g+)$	0.00	0.00	0.00	0.00	0.00	0.00
17	$\alpha_L(a,a)$	20.61	14.10	13.06	13.06	13.08	12.92
18	$\alpha_L(g-,a)$	13.13	7.01	6.09	6.41	6.53	6.72
19	$\alpha_L(g-,g-)$	12.52	6.76	5.95	6.26	6.22	6.54
20	$\varepsilon_D(g-,g-)$	20.51		13.35			
21	$\varepsilon_D(g-,g+)$	16.33	10.21	9.63	9.52	9.58	8.97
22	$\varepsilon_D(a,a)$	9.35	7.51	6.77	6.97	7.31	7.29
23	$\varepsilon_D(a,g+)$	10.07	8.28	7.61	7.82	7.99	7.85
24	$\varepsilon_D(g+,g-)$	4.87	4.89	5.08	4.90	4.83	4.24
25	$\varepsilon_D(g+,a)$	18.47	12.99	12.28	12.13	12.31	10.92

TABLE IIIA.
(Continued)

Conf. Number	Conf.[b] Code	ΔE(RHF/ 3-21G)[c]	ΔE(RHF/6-31+G*// RHF/6-311++G**)[d]	ΔE(RHF/ 6-311++G**// RHF/3-21G)[e]	ΔE(RHF/ 6-311++G**)[f]	ΔE(RHF/TZ2P// RHF/6-311++G**)[g]	ΔE(MP2/ 6-311++G**// RHF/6-311++G**)[h]
26	$\gamma_D(g-,g-)$	12.88	8.46	7.12	7.85	8.09	7.66
27	$\gamma_D(g-,a)$	12.02	7.57	6.16	6.88	7.39	6.83
28	$\gamma_D(g-,g+)$	12.52	7.83	6.51	7.30	7.61	7.12
29	$\gamma_D(a,g-)$	10.59		7.68			
30	$\gamma_D(a,a)$	12.68	9.31	8.41	8.64	8.95	8.68
31	$\gamma_D(a,g+)$	11.99	9.35	8.41	8.78	8.95	8.67
32	$\gamma_D(g+,g-)$	9.39	6.98	6.56	6.75	6.65	6.22
33	$\gamma_D(g+,a)$	17.13	12.83	11.98	12.05	12.14	11.18
34	$\gamma_D(g+,g+)$	14.03	11.30	10.35	10.92	10.76	10.21
35	$\alpha_D(g-,a)$	12.89	7.76	6.45	7.09	7.48	7.17
36	$\alpha_D(g+,g+)$	12.14	9.17	8.79	8.57	8.40	7.80
37	$\alpha_D(a,g-)$	9.07	6.71	6.04	6.33	6.32	6.08
38	$\alpha_D(a,g+)$	20.49	15.18	13.89	14.31	14.19	13.79
39	$\delta_D(g-,g-)$	15.65	11.55	10.66	10.90	10.73	10.70
40	$\delta_D(a,g+)$	15.71	10.73	10.11	10.06	10.27	9.52
41	$\delta_D(g+,g-)$	10.51	7.72	7.55	7.35	7.57	6.87
42	$\delta_D(g+,a)$	11.19	7.77	7.37	7.20	7.82	7.27
43	$\delta_D(g-,g+)$	12.29		8.37			
44	$\delta_D(a,a)$	17.23		11.04			
	Average	11.56	7.62	7.05	7.09	7.34	7.10
	Average of all x_L	9.40	5.35	4.85	4.86	5.40	5.36
	Average of all x_D	13.20	9.24	8.73	8.68	8.83	8.34

[a] Relative energies (in kcal mol^{-1}) with respect to $E^{\gamma L(g+,g+)}$.
[b] The backbone conformers are labeled according to the set of abbreviation introduced in the past (ref. 31): $\alpha_L, \alpha_D, \beta_L, \gamma_L, \gamma_D, \delta_L, \delta_D, \varepsilon_L$, and ε_D.
[c] $E^{\gamma L(g+,g+)} = -486.9195667$ hartree at RHF/3-21G level of theory.
[d] $E^{\gamma L(g+,g+)} = -489.6582967$ hartree at RHF/6-31+G*//RHF/6-311++G** level of theory.
[e] $E^{\gamma L(g+,g+)} = -489.786993$ hartree at RHF/6-311++G**//RHF/3-21G level of theory.
[f] $E^{\gamma L(g+,g+)} = -489.7936854$ hartree at RHF/6-311++G** level of theory.
[g] $E^{\gamma L(g+,g+)} = -489.811645$ hartree at RHF/TZ2P//RHF/6-311++G** level of theory.
[h] $E^{\gamma L(g+,g+)} = -491.3081862$ hartree at MP2/6-311++G**//RHF/6-311++G** level of theory.

TABLE IIIB.

Statistical Parameters of the Pair-Wise Comparisons of Relative Energies[a] Calculated at Different Levels of Theory for HCO-L-Ser-NH_2.

	ΔE(RHF/ 3-21G)	ΔE(RHF/ 6-31+G*// RHF/6-31++G**)	ΔE(RHF/ 6-311++G**// RHF/3-21G)	ΔE(RHF/ 6-311++G**)	ΔE(RHF/TZ2P// RHF/6-311++G**)	ΔE(MP2/ 6-311++G**// RHF/6-311++G**)	ΔE(MP2/ 6-311++G**)	ΔE(B3LYP/ 6-311++G**)
(A) Pearson Corr. Coef.[b]								
ΔE(RHF/3-21G)	**1.000**	0.960	0.928	0.951	0.956	0.961	0.996	0.996
ΔE(RHF/6-31+G*// RHF/6-311++G**)	0.960	**1.000**	0.993	0.999	0.998	0.993	0.995	0.989
ΔE(RHF/6-311++G**// RHF/3-21G)	0.928	0.993	**1.000**	0.995	0.993	0.981	0.986	0.975
ΔE(RHF/6-311++G**)	0.951	0.999	0.995	**1.000**	0.998	0.990	0.992	0.982
ΔE(RHF/TZ2P// RHF/6-311++G**)	0.956	0.998	0.993	0.998	**1.000**	0.991	0.994	0.986
ΔE(MP2/6-311++G**// RHF/6-311++G**)	0.961	0.993	0.981	0.990	0.991	**1.000**	0.9998	0.990
ΔE(MP2/6-311++G**)	0.996	0.995	0.986	0.992	0.994	0.9998	**1.000**	0.990
ΔE(B3LYP/6-311++G**)	0.996	0.989	0.975	0.982	0.986	0.990	0.990	**1.000**
(B) Standard Error[c]								
ΔE(RHF/3-21G)	—	1.320	1.742	1.465	2.092	1.317	0.705	0.494
ΔE(RHF/6-31+G*// RHF/6-311++G**)	0.945	—	0.391	0.158	0.187	0.394	0.571	0.668
ΔE(RHF/6-311++G**// RHF/3-21G)	1.200	0.377	—	0.315	1.417	0.642	0.913	0.954
ΔE(RHF/6-311++G**)	0.998	0.150	0.310	—	0.206	0.452	0.678	0.809
ΔE(RHF/TZ2P// RHF/6-311++G**)	1.683	0.176	1.655	0.204	—	0.422	0.613	0.696
ΔE(MP2/6-311++G**// RHF/6-311++G**)	0.799	0.334	0.564	0.403	0.381	—	0.111	0.521
ΔE(MP2/6-311++G**)	0.446	0.486	0.824	0.609	0.556	0.109	—	0.534
ΔE(B3LYP/6-311++G**)	0.325	0.581	0.865	0.743	0.639	0.539	0.555	

TABLE IIIB.
(Continued)

	ΔE(RHF/ 3-21G)	ΔE(RHF/ 6-31+G*// RHF/6-31++G**)	ΔE(RHF/ 6-311++G**// RHF/3-21G)	ΔE(RHF/ 6-311++G**)	ΔE(RHF/TZ2P// RHF/6-311++G**)	ΔE(MP2/ 6-311++G**// RHF/6-311++G**)	ΔE(MP2/ 6-311++G**)	ΔE(B3LYP/ 6-311++G**)
(C) The Value of **b** from $y = mx + b$								
ΔE(RHF/3-21G)	**0.000**	0.846	2.056	1.162	2.449	−0.165	−1.070	−0.060
ΔE(RHF/6-31+G*// RHF/6-311++G**)	0.010	**0.000**	0.587	0.170	−0.060	−0.696	−0.691	0.041
ΔE(RHF/6-311++G**// RHF/3-21G)	−0.329	−0.473	**0.000**	−0.332	0.755	−1.091	−0.861	−0.221
ΔE(RHF/6-311++G**)	−0.073	−0.146	0.391	**0.000**	−0.208	−0.791	−0.706	−0.003
ΔE(RHF/TZ2P// RHF/6-311++G**)	−0.111	0.079	0.803	0.234	**0.000**	−0.571	−0.621	0.039
ΔE(MP2/6-311++G**// RHF/6-311++G**)	0.645	0.682	1.212	0.838	0.634	**0.000**	−0.132	0.586
ΔE(MP2/6-311++G**)	0.718	0.638	0.917	0.711	0.628	0.133	**0.000**	0.687
ΔE(B3LYP/6-311++G**)	0.065	0.049	0.382	0.140	0.067	−0.529	−0.630	**0.000**
(D) The Value of **m** from $y = mx + b$								
ΔE(RHF/3-21G)	**1.000**	1.341	1.347	1.397	1.111	1.583	1.575	1.516
ΔE(RHF/6-31+G*// RHF/6-311++G**)	0.688	**1.000**	1.029	1.051	1.062	1.172	1.169	1.135
ΔE(RHF/6-311++G**// RHF/3-21G)	0.639	0.959	**1.000**	1.011	0.768	1.117	1.092	1.076
ΔE(RHF/6-311++G**)	0.647	0.950	0.980	**1.000**	1.009	1.111	1.105	1.070
ΔE(RHF/TZ2P// RHF/6-311++G**)	0.719	0.939	1.049	0.987	**1.000**	1.100	1.097	1.075
ΔE(MP2/6-311++G**// RHF/6-311++G**)	0.583	0.842	0.861	0.883	0.893	**1.000**	1.014	0.956
ΔE(MP2/6-311++G**)	0.630	0.847	0.890	0.891	0.900	0.986	**1.000**	0.952
ΔE(B3LYP/6-311++G**)	0.655	0.861	0.883	0.901	0.905	1.025	1.029	**1.000**

[a] Relative energy data sets computed at eight different levels of theory are reported in Tables IB and IIIA.
[b] Pearson correlation coefficient [cf. eq. (4)].
[c] Standard error [cf. eq. (3)].

TABLE IV.
The N—C$^{\alpha}$—C′ Values of HCO-L-Ser-NH$_2$ as a Function of Its Conformations Calculated at RHF/3-21G and RHF/6-311++G Levels of Theory.**

Conf.[a]	RHF/6-311++G** (RHF/3-21G)[d] N—C$^{\alpha}$—C′[b]	Shift Values[c]
$\beta_L(g-,a)$	— (108.3)	— (−1.7)
$\beta_L(g-,g+)$[b]	106.96 (105.45)	−3.04 (−4.55)
$\beta_L(g+,g-)$	107.58 (105.94)	−2.42 (−4.06)
$\beta_L(g+,a)$	107.91 (106.3)	−2.09 (−3.7)
$\beta_L(a,a)$	107.68 (106.06)	−2.32 (−3.94)
$\beta_L(a,g+)$	107.98 (106.4)	−2.02 (−3.6)
$\delta_L(g-,g-)$	113.53 (111.34)	3.53 (1.34)
$\delta_L(g-,a)$	— (112.62)	— (2.62)
$\delta_L(a,g-)$	— (111.32)	— (1.54)
$\delta_L(g+,a)$	114.39 (112.94)	4.39 (2.94)
$\gamma_L(g-,g-)$	111.06 (112.97)	1.06 (2.97)
$\gamma_L(g-,a)$	111.23 (113.21)	1.23 (3.21)
$\gamma_L(g-,g+)$	109.74 (110.25)	−0.26 (0.25)
$\gamma_L(a,g-)$	111.04 (111.02)	1.04 (1.02)
$\gamma_L(a,g+)$	— (108.48)	— (−1.52)
$\gamma_L(g+,g+)$	110.85 (108.97)	0.85 (−1.03)
$\alpha_L(a,a)$	113.15 (111.82)	3.15 (1.82)
$\alpha_L(g-,a)$	115.21 (114.06)	5.21 (4.06)
$\alpha_L(g-,g-)$	115.29 (114.47)	5.29 (4.47)
$\varepsilon_D(g-,g-)$	— (109.9)	— (−0.1)
$\varepsilon_D(g-,g+)$	110.03 (109.75)	0.03 (−0.25)
$\varepsilon_D(a,a)$	110.54 (109.52)	0.54 (−0.48)
$\varepsilon_D(a,g+)$	110.6 (109.56)	0.6 (−0.44)
$\varepsilon_D(g+,g-)$	109.27 (106.62)	−0.73 (−3.38)
$\varepsilon_D(g+,a)$	111.26 (111.81)	1.26 (1.81)
$\gamma_D(g-,g-)$	115.46 (114.79)	5.46 (4.79)
$\gamma_D(g-,a)$	114.56 (114.38)	4.56 (4.38)
$\gamma_D(g-,g+)$	114.01 (114.01)	4.01 (4.01)
$\gamma_D(a,g-)$	— (117.78)	— (7.78)
$\gamma_D(a,a)$	111.59 (112.8)	1.59 (2.8)
$\gamma_D(a,g+)$	113.89 (115.94)	3.89 (5.94)
$\gamma_D(g+,g-)$	114.72 (113.75)	4.72 (3.75)
$\gamma_D(g+,a)$	118.62 (120.07)	8.62 (10.07)
$\gamma_D(g+,g+)$	118.28 (118.14)	8.28 (8.14)
$\alpha_D(g-,a)$	113.89 (112.26)	3.89 (2.26)
$\alpha_D(g+,g+)$	112.78 (110.62)	2.78 (0.62)
$\alpha_D(a,g-)$	113.38 (111.83)	3.38 (1.83)
$\alpha_D(a,g+)$	112.53 (110.66)	2.53 (0.66)
$\delta_D(g-,g-)$	110.53 (109.16)	0.53 (−0.84)
$\delta_D(a,g+)$	108.65 (109.6)	−1.35 (−0.4)
$\delta_D(g+,g-)$	108.98 (107.62)	−1.02 (−2.38)
$\delta_D(g+,a)$	109.47 (107.87)	−0.53 (−2.13)
$\delta_D(g-,g+)$	— (109.44)	— (−0.56)
$\delta_D(a,a)$	— (108.38)	— (−1.62)

[a] The backbone conformers are labeled according to the set of abbreviation introduced previously (ref. 31): $\alpha_L, \alpha_D, \beta_L, \gamma_L, \gamma_D, \delta_L, \delta_D, \varepsilon_L$, and ε_D.
[b] N—C$^{\alpha}$—C′ values in degrees.
[c] The shift-values are $\tau^{N—C\alpha—C'} - 110^{\circ}$.
[d] Values in parenthesis were obtained at RHF/3-21G levels of theory.

On the other hand, for models of this size RHF, density functional theory (DFT), and MP2 methods are readily applicable. Therefore, these methods are used extensively in this article to determine geometries and energies for all minima on the PES of HCO-L-Ser-NH$_2$.

CONFORMATIONAL PARAMETERS

According to multidimensional conformational analysis (MDCA), a maximum of $3^2 * 3^2 = 81$ conformers are expected on the $E = E(\phi, \psi, \chi_1, \chi_2)$ PES of HCO-L-Ser-NH$_2$: nine different side-chain orientations for each of the nine backbone structures. From the 81 expected conformers 44 structures have been found at the RHF/3-21G level.[40] The other 37 input structures migrated to one of the 44 RHF/3-21G minima. (The term "migration" is used if, following an optimization, the initial and the final structures belong to different catchment regions.) When these 44 RHF/3-21G structures were reoptimized at the RHF/6-311++G** level, eight further migrations were observed:

$$\beta_L(g-,a) \Rightarrow \gamma_L(g-,a),$$
$$\delta_L(g-,a) \Rightarrow \alpha_L(g-,a),$$
$$\delta_L(a,g-) \Rightarrow \gamma_L(a,g-),$$
$$\gamma_L(a,g+) \Rightarrow \beta_L(a,g+),$$
$$\varepsilon_D(g-,g-) \Rightarrow \gamma_D(g-,g-),$$
$$\gamma_D(a,g-) \Rightarrow \alpha_D(a,g-),$$
$$\delta_D(g-,g+) \Rightarrow \gamma_D(g-,g+), \quad \text{and}$$
$$\delta_D(a,a) \Rightarrow \beta_L(a,a)$$

Consequently, at the RHF/6-311++G** level the number of minima found is 36. In the eight migrations, while the ϕ, ψ values were affected, the changes in χ_1 and χ_2 were only of second order. In other words, for all migrations the backbone of the final conformer belongs to a different conformational region than that of the initial conformer,

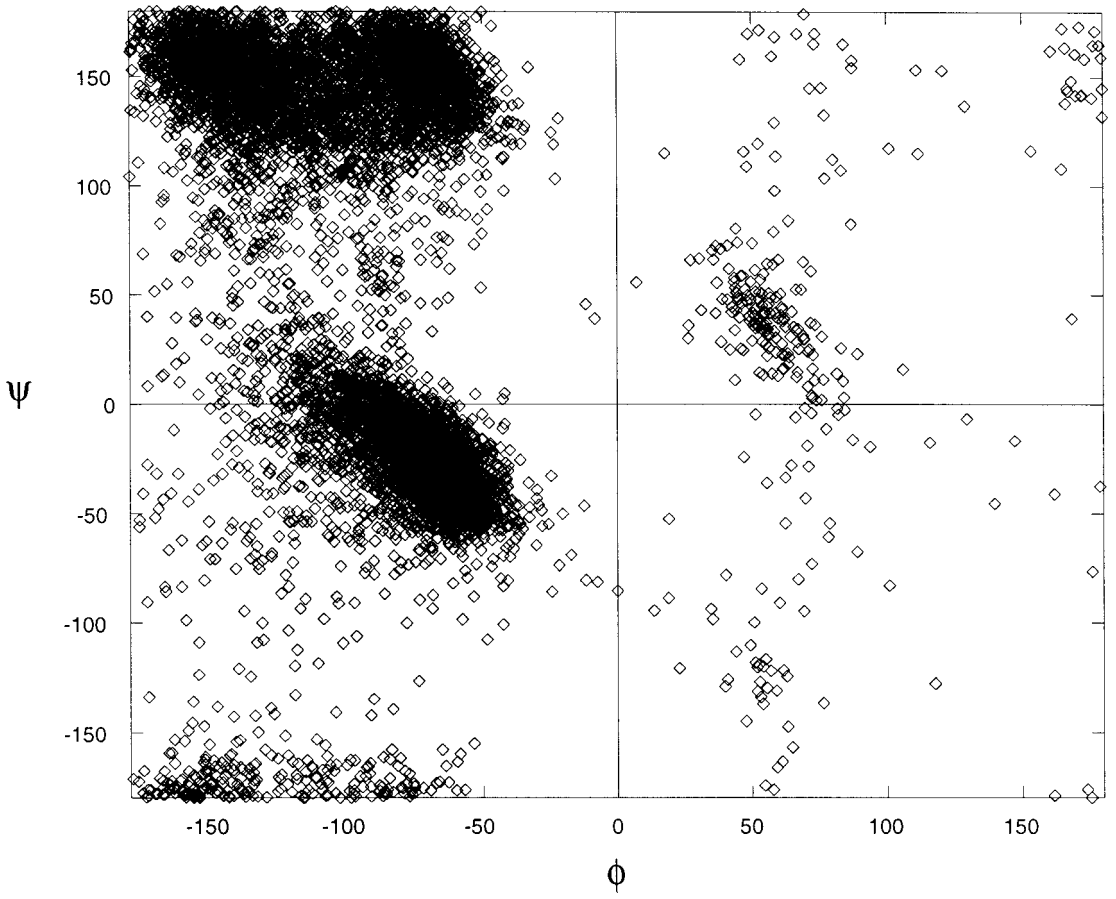


FIGURE 1. Distribution of the ϕ, ψ values of 9511 serine residues observed in proteins.

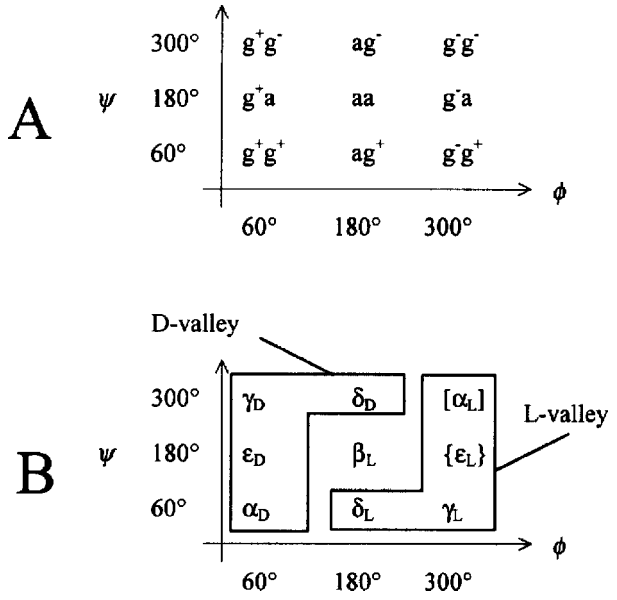
while orientation of the side chain remains the same. (For visualization of the eight migrations on a Ramachandran-type surface, see Fig. 2A.) The four migrations involving L-type initial conformers resulted in backbone conformers still belonging to the "L-valley." Similarly, three out of the four D-type initial conformers that migrated produced conformers belonging to the D-valley. Only the $\delta_D(a,a) \rightarrow \beta_L(a,a)$ migration (Fig. 2A) resulted in a final conformer that belongs to the other main valley. Therefore, this is perhaps the largest structural shift observed here.

It is of interest to trace the above eight migrations on the relevant full $E = E(\chi_1, \chi_2)$ side-chain PES, only available at the RHF/3-21G level of theory (Fig. 2B). The use of the $E = E^{\mathrm{RHF/3\text{-}21G}}(\chi_1, \chi_2)$ data set is so far the only possibility to provide interpretation why certain minima migrate and others do not when reoptimized at a different level. If the initial structure is located in a shallow catchment region (Fig. 2B), it is easy to rationalize the occurring backbone shift. This occurs for the

$$\begin{aligned}
\delta_D(a,a) &\Rightarrow \beta_L(a,a),\\
\gamma_D(a,g-) &\Rightarrow \alpha_D(a,g-),\\
\gamma_L(a,g+) &\Rightarrow \beta_L(a,g+),\\
\delta_D(g-,g+) &\Rightarrow \gamma_D(g-,g+),\\
\delta_L(g-,a) &\Rightarrow \alpha_L(g-,a), \quad \text{and}\\
\delta_L(a,g-) &\Rightarrow \gamma_L(a,g-)
\end{aligned}$$

migrations. For example, it is obvious that the RHF/3-21G $\gamma_L(a,g-)$ structure belongs to a much deeper catchment region than $\delta_L(a,g-)$. On the other hand, it is much harder to find the driving force of migrations like those of $\beta_L(g-,a) \rightarrow \gamma_L(g-,a)$ and $\varepsilon_D(g-,g-) \rightarrow \gamma_D(g-,g-)$. Perhaps the available RHF/3-21G surfaces are not detailed and accurate enough to establish an unambiguous description?

It is of interest to compare the conformational shifts at the RHF/3-21G, RHF/6-311++G**, B3LYP/6-311++G**, and MP2/6-311++G** levels in a pair-wise fashion. All four dihedral para-

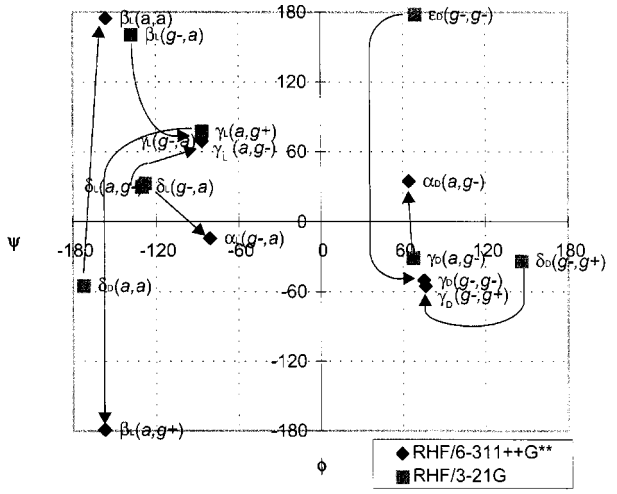


SCHEME 2. (A) The nine legitimate conformers on the Ramachandran surface labeled according to the IUPAC-IUB guidance. (B) Optimized *ab initio* minima of HCO-Xxx-NH_2 (Xxx = Gly, Ala, Val, Phe) using a short-hand notation for the above nine minima. The α_L is in bracket, because this minimum may vanish for HCO-Xxx-NH_2 models, assigned so far only for Xxx = Ser and Thr. The braces around ε_L signify that this minimum has been not yet assigned in any P—CONH—CHR—NHCO—Q system. Nevertheless, existence of the ε_L conformation has been shown in dialanine diamide: HCO-Ala-Ala-NH_2.

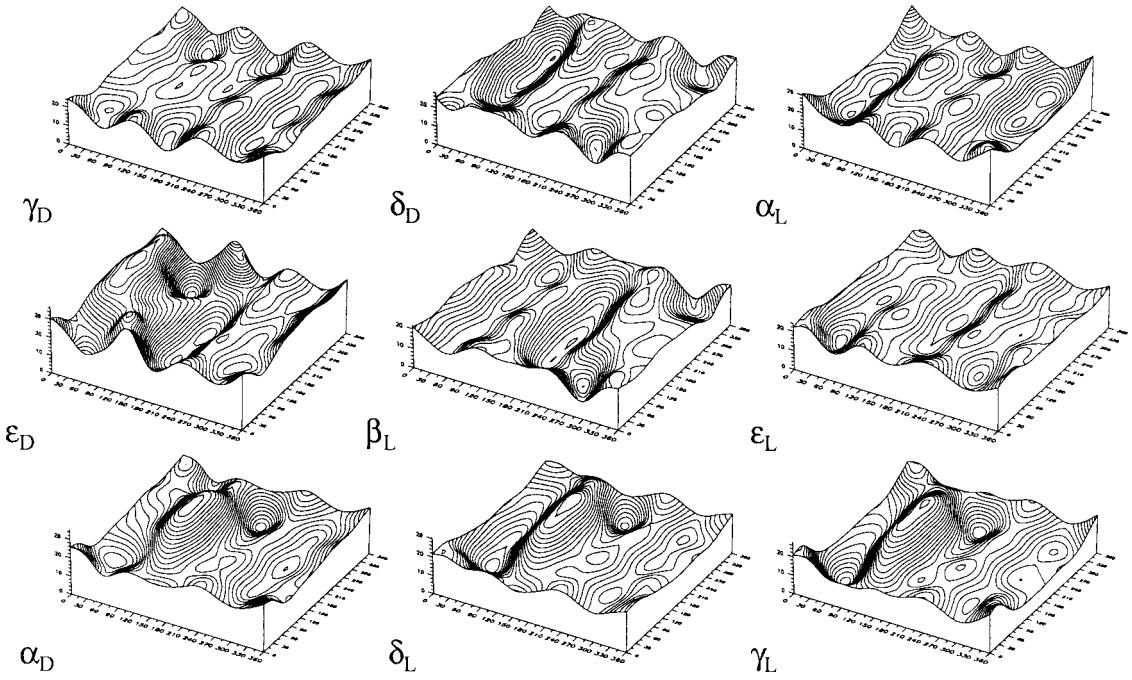
meters, ϕ, ψ, χ_1, and χ_2, used for the description of the molecular folding (Table IIA, Fig. 2C), correlate well among the different theoretical levels. For example, the Pearson correlation coefficient [r; eq. (4), *vide infra*] indicates convincing linearities (e.g., $r_{\phi}^{\text{RHF/3-21G}\Leftrightarrow\text{RHF/6-311++G}^{**}} = 0.995$, $r_{\psi}^{\text{RHF/3-21G}\Leftrightarrow\text{RHF/6-311++G}^{**}} = 0.994$, $r_{\chi 1}^{\text{RHF/3-21G}\Leftrightarrow\text{RHF/6-311++G}^{**}} = 0.999$, $r_{\chi 2}^{\text{RHF/3-21G}\Leftrightarrow\text{RHF/6-311++G}^{**}} = 0.999$) (Table IIA, form a) between dihedral angles obtained at the RHF/3-21G and RHF/6-311++G** levels. The statistical analysis of all the other five crosscorrelations (Table IIA, forms b to f) show a rather similar picture, although in the case of DFT and MP2 only a small number of structures are available. The correlation between ω values is typically much lower (e.g., $R_{\omega 1}^{\text{RHF/3-21G}\Leftrightarrow\text{RHF/6-311++G}^{**}} = 0.606$), fluctuating between $r_{\omega 1}^{\text{RHF/3-21G}\Leftrightarrow\text{MP2/6-311++G}^{**}} = 0.60$ and $r_{\omega 1}^{\text{RHF/6-311++G}^{**}\Leftrightarrow\text{B3LYP/6-311++G}^{**}} = 0.96$. Besides the dihedral parameters, selected bond angle values were also the subject of crosscomparison. For example, the values of the N—C^{α}—C′ bond angle, calculated at four different theoretical levels, correlate with high significance ($r \geq 0.93$) (Table IIA). These findings suggest that the geometrical properties of the present peptide model change in a predictable way between different theoretical levels. This is especially true for the ϕ, ψ, χ_1, and χ_2 parameters.

The almost perfect correlation of data obtained at different levels of theory makes the analysis of the shift values of special interest. Because only the RHF/3-21G and the RHF/6-311++G** data sets are considered to be complete,[45] the analysis below incorporates these data only. Differences ($\Delta\omega_0, \Delta\phi, \Delta\psi, \Delta\omega, \Delta\chi_1$, and $\Delta\chi_2$) were averaged over the entire ensemble (216 dihedral parameters). The obtained average shift, 5.3°, suggests that no significant conformational change occurs upon reoptimization. Notable changes were observed only in 5 of the 36 conformers. Most importantly, showing higher than average sensitivity to variation of the size of the basis are those conformers that are characterized by high relative energy (ΔE) (Table IIC). The conformer $\delta_L(g-,g-)$ is a good example, because it belongs to those L-type structures that have high $\Delta E^{\text{RHF/6-311++G}^{**}}$ (6.36 kcal mol^{-1}, Table IIIA). Similarly, the $\varepsilon_D(g-,g+)$, $\gamma_D(g+,g+)$, and $\delta_D(a,g+)$ conformers have ≈10 kcal mol^{-1} higher energy than the global minimum, and they exhibit a large dihedral angle variation at the RHF level when the basis is changed from the compact 3-21G to 6-311++G**. Among these five structures only $\varepsilon_D(g+,g-)$ shows an unexpected behavior: although a remarkable backbone shift is associated with the reoptimization on a larger basis set ($\Delta\psi = 25.1°$), this conformer has one of the lowest energy among the D-type conformers ($\Delta E^{\text{RHF/6-311++G}^{**}} = 4.90$ kcal mol^{-1}).

In general, the RHF/6-311++G** optimizations resulted in conformational parameters that are closer to values derived from an X-ray analysis of proteins (cf. Table VA) than those obtained with the compact 3-21G basis. This is the case when β_L-, γ_L-, and δ_L-type backbone orientations are analyzed and compared. However, for the building unit of the α-helix (i.e., the α_L conformation) the opposite is true: the ϕ, ψ; values determined at the higher level of theory show less similarity to averages observed in proteins than values calculated at the RHF/3-21G level. In proteins, there are two forms of the helical structure to be considered, the 3_{10}-helix and the normal α-helix. The different packing of these two secondary structural elements is monitored by the periodic ϕ, ψ; value pairs: $\phi \approx -60°$, $\psi \approx -30°$



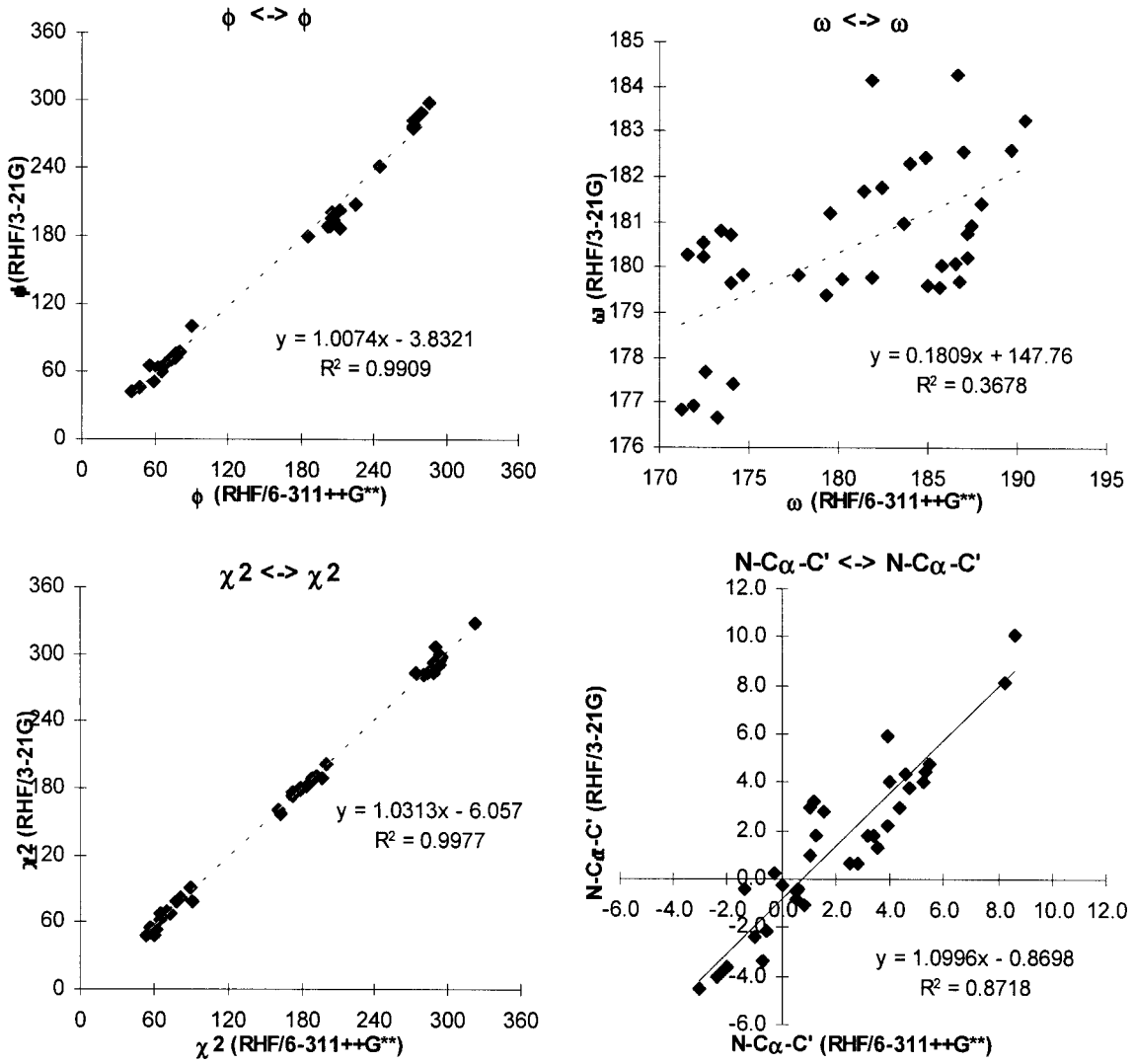
(A)



(B)

FIGURE 2. (A) The $\beta_L(g-,a) \Rightarrow \gamma_L(g-,a)$, $\delta_L(g-,a) \Rightarrow \alpha_L(g-,a)$, $\delta_L(a,g-) \Rightarrow \gamma_L(a,g-)$, $\gamma_L(a,g+) \Rightarrow \beta_L(a,g+)$, $\varepsilon_D(g-,g-) \Rightarrow \gamma_D(g-,g-)$, $\gamma_D(a,g-) \Rightarrow \alpha_D(a,g-)$, $\delta_D(g-,g+) \Rightarrow \gamma_D(g-,g+)$, and $\delta_D(a,a) \Rightarrow \beta_L(a,a)$ migrations on the Ramachandran surfaces, observed when RHF/3-21G structures are reoptimized at RHF/6-311++G**. (B) RHF/3-21G side-chain maps associated with the nine different backbone orientations. (C) The pair-wise comparison of selected dihedral parameters of HCO-L-Ser-NH_2 determined at the RHF/3-21G and RHF/6-311++G** levels of theory.

is typical for a 3_{10}-, while $\phi \approx -54°$, $\psi \approx -45°$ is representative for an α-helix. At both levels of theory, in the conformational ensemble of HCO-L-Ser-NH_2 three helical building units were determined: $\alpha_L(a,a)$, $\alpha_L(g-,a)$, and $\alpha_L(g-,g-)$. At the RHF/3-21G level the average backbone parameters ($\phi \approx -68.3°$, $\psi \approx -30.5°$) of these α_L-conformers are close to the expected values of a 3_{10}-helix: $\Delta\phi \approx -8.3°$, $\Delta\psi \approx -0.5°$. The averages $\phi \approx -79.1$ and $\psi \approx -22.4$ calculated at the RHF/6-311++G** level



(C)

FIGURE 2. (Continued)

show a larger deviation ($\Delta\phi \approx -19.1°$, $\Delta\psi \approx -7.6°$) from the expected data, suggesting a building unit that could provide an even sharper secondary structure than that of a 3_{10}-helix. In conformational terms the α_L- and δ_L-structure families are closer to each other at the higher level of theory. According to RHF/6-311++G** computations, both the α_L and δ_L sets are located close to the "bridge region," which interconnects the broad β-structure area and the narrower α-helical region of the Ramachandran surface.

In the case of HCO-L-Ser-NH_2, we determined the average shift value due to basis set enlargement at the RHF level. On the whole the conformational properties are rather similar at the two levels of theory (RHF/6-311++G** and RHF/3-21G) (Table IIB). To be able to scale these changes, a comparison was made between the HCO-L-Ser-NH_2 and HCO-L-Ala-NH_2 dipeptide models at the above two levels of theory (see Tables VA and VB). The ideal values ($[\phi, \psi]^{ideal}$) reported in Table VB are those predicted by MDCA. Both model compounds at both levels of theory exhibit rather similar backbone conformers, giving a very similar center for the conformational regions.[31] It looks as if the conformational changes are more characteristic to the level of theory applied than to the nature of the amino acid residue. This conclusion is based on the fact that the side-chain conformations were averaged in calculating $\phi^{average}$ and $\psi^{average}$. Nevertheless, it is possible that

TABLE VA.
Average Backbone Conformational Parameters of HCO-L-Ser-NH_2 at RHF/6-311++G Levels of Theory.**

Conf.[a]	Averaged[c]	$\phi^{average}$	$\psi^{average}$	Average $(\tau^{N—C\alpha—C'} - 110°)$	$(\tau^{N—C\alpha—C'} - 110°)$ from *ab initio* Ac-L-Ala-NHMe[d]	Aver. $(\tau^{N—C\alpha—C'} - 110°)$ by X-ray and (NMR)[d]
β_L	5 (6)	−159.5 (−165.9)[b]	173.7 (176.4)	−2.4 (−3.6)	−4	−1 (0)
δ_L	2 (4)	−124.2 (−132.1)	14.4 (29.7)	4.0 (2.1)	3	2 (1)
γ_L	5 (6)	−86.5 (−82.2)	73.7 (67.2)	0.8 (0.8)	−2	−2 (−)
α_L	3 (3)	−79.1 (−68.3)	−22.4 (−30.5)	4.5 (3.5)	4	3 (3)
ε_D	5 (6)	65.8 (68.6)	−146.4 (−154.7)	0.4 (−0.5)	0	0 (0)
γ_D	8 (9)	72.5 (69.8)	−47.7 (−47.9)	5.1 (5.8)	—	— (—)
α_D	4 (4)	61.4 (57.3)	39.6 (42.3)	3.2 (1.3)	2	3 (2)
δ_D	4 (6)	−151.0 (−173.4)	−61.0 (−53.5)	−0.6 (−1.3)	1	2 (−1)

[a] Backbone conformers. The $\phi^{average}$ and $\psi^{average}$ are in degrees calculated at RHF/6-311++G** level of theory.
[b] Values in parenthesis were obtained at RHF/3-21G levels of theory.
[c] Number of RHF/6-311++G** side-chain conformers within the same backbone catchmant region. Values in parenthesis are the same calculated at RHF/3-21G levels of theory.
[d] Shift-values (from Reference [Karplus etc.] Figure 5A, D, and H) located in the vicinity pinpointed by the present RHF/6-311++G** calculations. [The *ab initio* RHF/4-21G values used by Karplus originates from L. Schäfer and M. Cao.[67]]

such a conclusion is not valid for all amino acids because we are comparing two amino acids with not very different side chains (i.e., —R = —CH_3 and —CH_2OH). However, phenylananine (—R = —CH_2Ph) also behaves similarly.[43, 44]

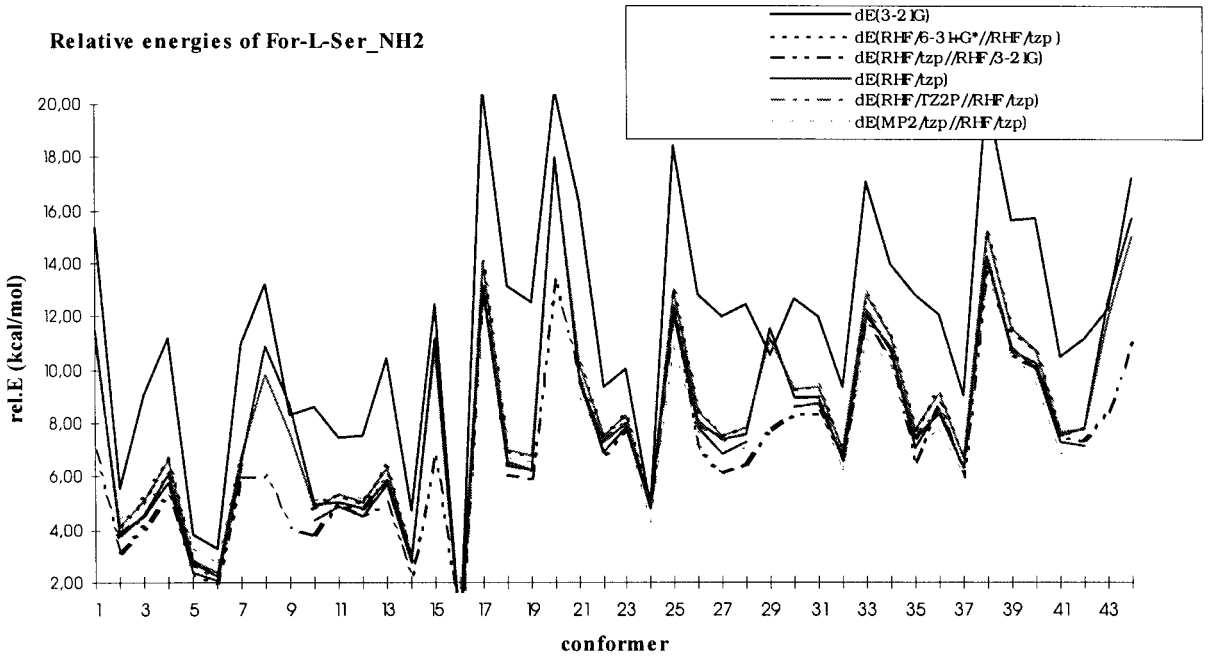
N—C^{α}—C′ BOND ANGLE PROPERTIES

Ab initio structural results provide the basis for a search for the correlation of selected bond lengths and bond angles with backbone orientations. This possibility for structural comparisons was first investigated for peptide models by Schäfer and coworkers.[66] They proposed to scan, for example, the value of the N—C^{α}—C′ angle ($\alpha^{N—C\alpha—C'}$) over the entire conformational space. The value of $\alpha^{N—C\alpha—C'}$ varies markedly as function of the ϕ, ψ dihedral space. Schäfer and Cao[67] demonstrated for N-acetyl-N′-methylalaninamide that the value of the above bond angle is predictable upon knowledge of the ϕ, ψ values. This prediction was checked later against data taken from high resolution protein structures. Karplus[68] performed a data search on 70 diverse proteins and found a limited bond angle variation close to ±5° from the standard 110°. The conclusion is that the change in $\alpha^{N—C\alpha—C'}$ is largely

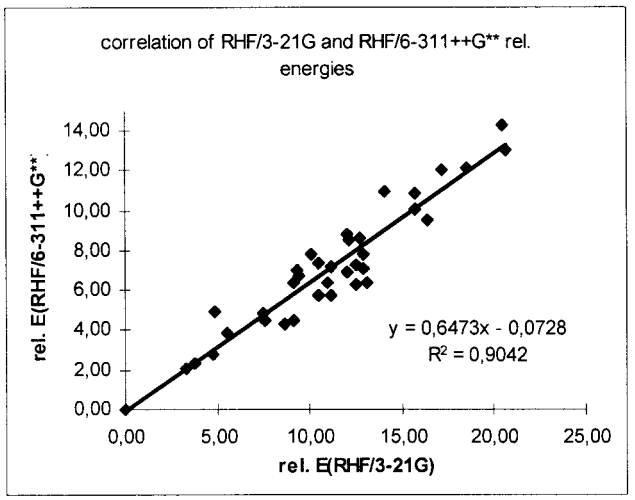
TABLE VB.
Selected Conformational Parameters of HCO-L-Ala-NH_2, Calculated at Three Different Levels of Theory.

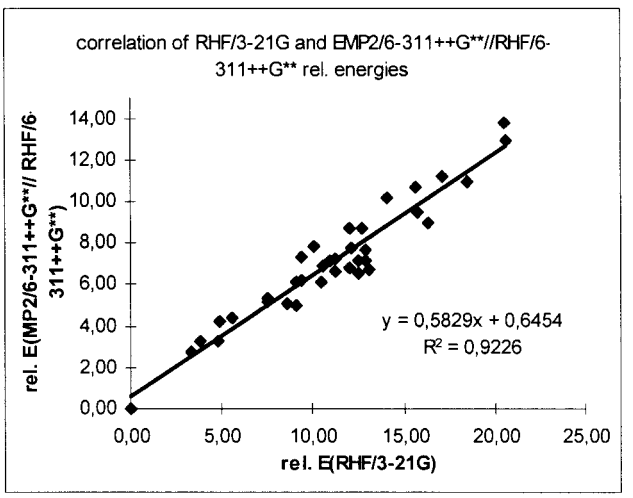
	RHF/6-311++G**[c]		MP2/6-311++G**[c]		Ideal	
Conf.[a]	ϕ	ψ	ϕ	ψ	ϕ	ψ
β_L	−155.1 (−168.3)[b]	161.0 (170.6)	−157.1	163.2	−180.0	180.0
δ_L	−112.8 (−128.1)	13.2 (29.8)	—	—	−180.0	60.0
γ_L	−86.2 (−84.5)	78.8 (67.3)	−82.8	80.6	−60.0	60.0
α_L	—	—	—	—	−60.0	−60.0
ε_D	— (67.2)	— (−171.9)	—	—	60.0	−180.0
γ_D	75.3 (74.0)	−55.4 (−57.4)	74.4	−49.1	60.0	−60.0
α_D	69.0 (63.8)	26.9 (32.7)	63.1	35.5	60.0	60.0
δ_D	−165.2 (−178.6)	−42.1 (−44.1)	−166.0	−39.9	−180.0	−60.0

[a] Backbone conformer type (ϕ and ψ are in degree).
[b] Values in parenthesis were obtained at RHF/3-21G levels of theory.
[c] Calculated parameters are from Endredi et al.[45]



A





B C

FIGURE 3. (A) Comparison of relative energy differences obtained at the RHF/3-21G, RHF/6-311++G**, and RHF/6-311++G**//RHF/3-21G levels of theory (see also Table IIIA). (B) and (C) Correlation of $\Delta E_{\text{RHF/3-21G}}$ with $\Delta E_{\text{RHF/6-311++G**}}$ and $\Delta E_{\text{RHF/3-21G}}$ with $\Delta E_{\text{MP2/6-311++G**//RHF/6-311++G**}}$.

independent of the type of amino acid residue, and varies with the backbone orientation in a concerted way.

$\alpha^{\text{N—C}\alpha\text{—C}'}$ shifts are reported in Table IV and Figure 4 for our serine diamide model, using $\alpha^{\text{N—C}\alpha\text{—C}'} = 110°$ as the reference point. The pairwise comparison of ($\alpha^{\text{N—C}\alpha\text{—C}'} - 110°$) values obtained at RHF/3-21G and RHF/6-311++G** levels shows a linear relationship (Table IIA and Fig. 2C) and strong correlation ($r = 0.934$). The $\alpha^{\text{N—C}\alpha\text{—C}'}$

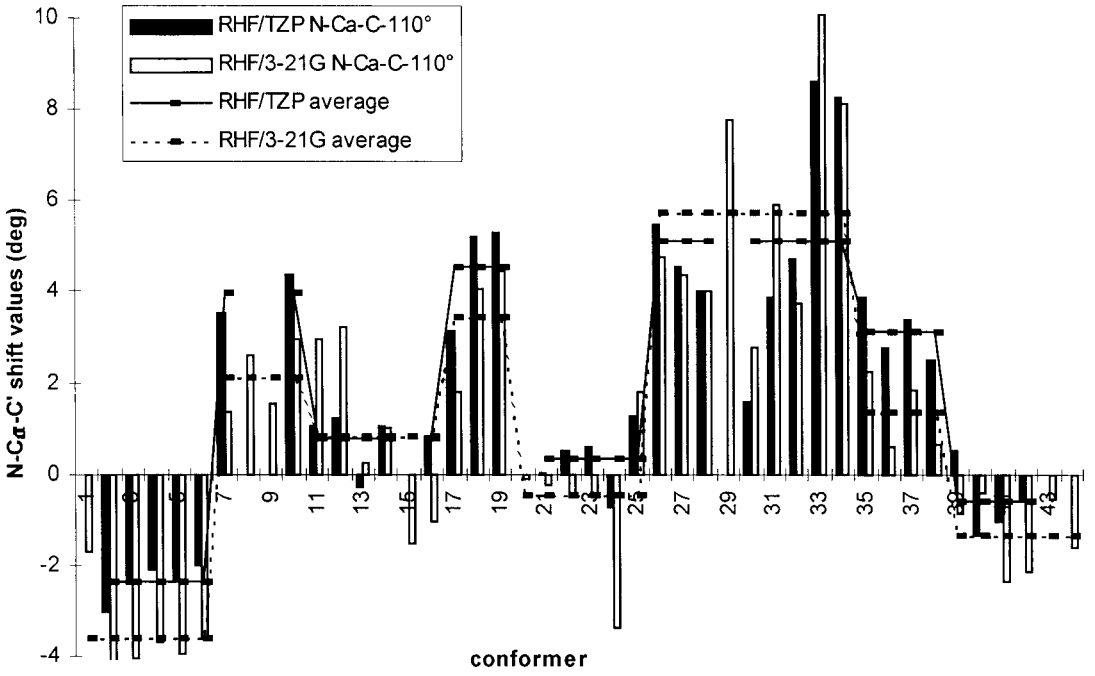


FIGURE 4. The N—C$^{\alpha}$—C′ shifts and their averages over the same backbone conformers of HCO-L-Ser-NH_2 as observed in structures calculated at the RHF/3-21G and RHF/6-311++G** levels (see also Table IV).

shifts of the different side-chain orientations were averaged for all backbone conformational types (Fig. 4). For all conformers but γ_D and γ_L we found that the average shift values, relevant for each conformational cluster, are up-shifted when the RHF/6-311++G** data are compared to RHF/3-21G values. The unexpected behavior of γ-turns (γ_D and γ_L) could originate from the strong intramolecular hydrogen bond, affecting directly the value of $\alpha^{N—C\alpha—C'}$. For all eight types of backbone conformations (Table VA) we compared individually the experimental (X-ray and NMR[68]) and theoretical (RHF/6-311++G**) average shift values ($\alpha^{N—C\alpha—C'} - 110°$). The comprehensive analysis was extended to both Ala and Ser models. Comparing the *ab initio* $\alpha^{N—C\alpha—C'} - 110°$ values of HCO-L-Ala-NHMe (RHF/4-21G data from Schäfer and Cao[67]) and of HCO-L-Ser-NH_2 (RHF/3-21G data), a high similarity was noticed for the β_L, δ_L, α_L, ε_D, and α_D regions of the Ramachandran surface (Table VA). The difference is around 2–3° when the same shift values, associated with γ_L and δ_D clusters, are compared. Beside the pair-wise comparison of the *ab initio* calculated shift values of HCO-L-Ala-NHMe and HCO-L-Ser-NH_2, the calculated data were checked against experimental results. The X-ray data compare well with *ab initio* results in the case of our serine model, especially when the δ_L-, α_L-, ε_D-, and α_D-structures are investigated (Table VA). The similarity is still acceptable for the extended (or β_L), γ_L and δ_D regions ($-3.6° \approx -1°$, $0.8° \approx -2°$, etc.) (Table VA). In the case of the β_L and α_D conformers, the enlargement of the basis from 3-21G to 6-311++G** improves the correlation between calculated and experimentally determined values. In contrast, for δ_L- and α_L-type backbone orientations enlargement of the basis has an opposite effect. To sum up, the correlation between experimentally determined and *ab initio* calculated conformational dependence of $\alpha^{N—C\alpha—C'}$ shift values is remarkable.

HYDROGEN-BONDING PROPERTIES

The H-bond analysis of HCO-L-Ser-NH_2, having a flexible and polar side chain, can reveal important structural features. This peptide model has the potential to form several types of intramolecular H-bonds (backbone–backbone [bb/bb] and backbone–side chain [bb/sc]), also operative in proteins. Both the amide group, forming the core of the backbone, and the hydroxymethyl group of the side chain can be a proton donor and/or a proton acceptor. Writing the symbol of the acceptor first followed by that of the donor group, three different types of interaction are possible: bb/bb, bb/sc, and sc/bb.

By analyzing the H-bond pattern and stability in HCO-L-Ser-NH_2, it is possible to compare interac-

TABLE VI.
Backbone–Backbone-Type Hydrogen-Bond Parameters of HCO-L-Ser-NH_2 at the RHF/6-311++G Level of Theory.**

Conf.[a]	$d_{O1\ldots N2}$[b]	$d_{O1\ldots H—N2}$	$\Theta_{O1\ldots H—N2}$
$\gamma_L(g-,g-)$	3.01 (2.79)[c]	2.19 (1.90)	139.4 (145.3)
$\gamma_L(g-,a)$	2.99 (2.77)	2.15 (1.88)	140.3 (146.1)
$\gamma_L(g-,g+)$	3.19 (2.93)	2.43 (2.09)	132.2 (140.4)
$\gamma_L(a,g-)$	3.06 (2.87)	2.25 (2.00)	137.6 (143.4)
$\gamma_L(g+,g+)$	3.04 (2.88)	2.23 (2.04)	138.2 (140.1)
$\gamma_D(g-,g-)$	2.91 (2.79)	2.04 (1.89)	144.8 (148.5)
$\gamma_D(g-,a)$	2.92 (2.80)	2.05 (1.90)	144.7 (148.5)
$\gamma_D(g-,g+)$	2.98 (2.82)	2.12 (1.93)	143.0 (146.8)
$\gamma_D(a,a)$	3.01 (2.85)	2.19 (1.98)	139.1 (143.7)
$\gamma_D(a,g+)$	2.97 (2.83)	2.14 (1.94)	139.4 (146.7)
$\gamma_D(g+,g-)$	3.18 (2.99)	2.35 (2.13)	140.1 (143.5)
$\gamma_D(g+,a)$	2.85 (2.72)	1.96 (1.78)	148.4 (154.1)
$\gamma_D(g+,g+)$	2.86 (2.73)	2.00 (1.82)	144.4 (150.1)
γ average	**3.00 (2.83)**	**2.16 (1.94)**	**140.9 (145.9)**
	$d_{O2\ldots N1}$	$d_{O2\ldots H—N1}$	$\Theta_{O2\ldots H—N1}$
$\beta_L(g-,g+)$	2.63 (2.58)	2.17 (2.05)	106.4 (110.7)
$\beta_L(g+,g-)$	2.62 (2.58)	2.15 (2.07)	107.1 (109.7)
$\beta_L(g+,a)$	2.63 (2.60)	2.20 (2.10)	104.2 (108.4)
$\beta_L(a,a)$	2.62 (2.58)	2.15 (2.06)	106.8 (110.5)
$\beta_L(a,g+)$	2.62 (2.59)	2.16 (2.08)	106.4 (109.7)
β average	**2.62 (2.59)**	**2.17 (2.07)**	**106.3 (109.8)**

[a] For L-Ser-NH_2 conformers. Distances (d) in Å and angles (Θ) are in degrees.
[b] The numbering of the atoms are according to that of Scheme 1A.
[c] Values in parenthesis are those obtained at RHF/3-21G levels of theory.

tions in a common molecular frame as function of the applied level of theory. While the RHF/3-21G calculation slightly overestimates the polar (e.g., H-bond) interactions, the RHF/6-311++G** computations provide a more realistic description. This holds when comparing the H-bond parameters of HCO-L-Ser-NH_2 in its two forms of γ-turns as a function of the basis set size (see Table VI). Both γ_D and γ_L backbone conformers incorporate a strong seven-membered bb/bb-type H-bond. By using the larger basis set in all 13 γ-turns (eight γ_D and five γ_L) both the heavy atom ($d_{O1\ldots N2}$) and the H-bond ($d_{O1\ldots H—N2}$) distances are lengthened: on average, this increase is around 0.2 Å (Table VI). Furthermore, the H-bond angle ($\Theta_{O1\ldots H—N2}$) is more bent by about 5°, if the 6-311++G** basis set is used instead of the 3-21G. In β_L conformers, contrary to γ-turn–type structures, where the interaction is strong between backbone atoms, a rather weak H-bond is present between the NH and the CO groups. Due to this weak interaction, only minor H-bond parameter changes follow the increase of the basis set size: in average $\Delta d_{O2\ldots N1}$ is 0.03 Å and $\Delta d_{O2\ldots H—N1}$ is 0.1 Å (Table VI). The H-bond parameters were found to vary from zero up to 0.22 Å. Such lengthening of H-bond distances is expected to be accompanied by a significant change in the relative energies (Tables IIIA and IIIB).

Beside the bb/bb-type H-bonds, serine diamide can form two additional types of intramolecular H-bonds: bb/sc and sc/bb. In total, four different interactions are possible between backbone and side-chain atoms, analyzed here in detail at both levels of theory (Table VII). The —OH group of the polar side chain can be a proton donor (first half of Table VIII, abbreviated as sc/bb) as well as a proton acceptor (second half of Table VII, labeled as bb/sc). The example of $\delta_L(g-,g-)$ shows that one should be cautious when attempting to convert changes in H-bond characteristics into relative energies. This RHF/3-21G structure contains a medium-strong sc/bb-type hydrogen bond ($d_{O1\ldots O3}$ = 2.93 Å,

TABLE VII.
Side-Chain/Backbone (sc/bb) and Backbone/Side-Chain (bb/sc) Type Hydrogen-Bond Parameters of HCO-L-Ser-NH_2 at RHF/6-311++G Levels of Theory.**

H-Bond Type	Conf.[a]	$d_{O1\ldots O3}$[b]	$d_{O1\ldots H—O3}$	$\Theta_{O1\ldots H—O3}$
sc/bb	$\beta_L(g-,g+)$	2.87 (2.71)[c]	2.05 (1.83)	144.0 (149.1)
sc/bb	$\delta_D(g-,g-)$	2.95 (2.74)	2.27 (1.94)	129.0 (137.2)
sc/bb	$\gamma_D(g+,g+)$	2.95 (2.78)	2.50 (2.18)	109.4 (118.8)
sc/bb	$\gamma_D(g+,g-)$	2.82 (2.70)	2.01 (1.85)	142.2 (144.9)
sc/bb	$\varepsilon_D(g+,g-)$	2.75 (2.61)	1.93 (1.75)	142.6 (145.7)
	average	**2.87 (2.71)**	**2.15 (1.91)**	**133.4 (139.1)**
	Conf.	$d_{O2\ldots O3}$	$d_{O2\ldots H—O3}$	$\Theta_{O2\ldots H—O3}$
sc/bb	$\gamma_L(a,g-)$	2.86 (2.72)	2.25 (2.03)	121.7 (126.5)
sc/bb	$\gamma_L(g+,g+)$	2.85 (2.75)	2.22 (2.07)	123.6 (126.1)
sc/bb	$\alpha_D(g+,g+)$	3.05 (2.93)	2.48 (2.32)	118.3 (120.9)
sc/bb	$\alpha_D(a,g-)$	2.81 (2.70)	2.16 (1.98)	125.2 (129.1)
	average	**2.89 (2.78)**	**2.28 (2.10)**	**122.2 (125.7)**
	Conf.	$d_{O3\ldots N1}$	$d_{O3\ldots H—N1}$	$\Theta_{O3\ldots H—N1}$
bb/sc	$\delta_L(g+,a)$	2.73 (2.63)	2.35 (2.19)	101.2 (104.7)
bb/sc	$\gamma_L(g-,g-)$	2.82 (2.66)	2.63 (2.20)	90.7 (106.4)
bb/sc	$\gamma_L(g-,a)$	2.77 (2.60)	2.56 (2.12)	91.1 (107.0)
bb/sc	$\alpha_L(g-,a)$	2.72 (2.63)	2.34 (2.14)	102.0 (108.3)
bb/sc	$\alpha_L(g-,g-)$	2.81 (2.70)	2.45 (2.21)	100.5 (108.4)
	average	**2.77 (2.64)**	**2.47 (2.17)**	**97.1 (107.0)**
	Conf.	$d_{O3\ldots N2}$	$d_{O3\ldots H—N2}$	$\Theta_{O3\ldots H—N2}$
bb/sc	$\beta_L(a,a)$	2.85 (2.69)	2.16 (1.91)	125.3 (132.5)
bb/sc	$\beta_L(a,g+)$	2.89 (2.72)	2.17 (1.92)	128.7 (134.7)
bb/sc	$\varepsilon_D(a,a)$	2.86 (2.73)	2.13 (1.94)	129.1 (134.1)
bb/sc	$\varepsilon_D(a,g+)$	2.87 (2.73)	2.12 (1.92)	130.8 (136.2)
bb/sc	$\varepsilon_D(g+,g-)$	2.85 (2.69)	2.11 (1.85)	130.0 (138.0)
bb/sc	$\varepsilon_D(g+,a)$	2.90 (2.75)	2.18 (1.97)	127.6 (132.5)
bb/sc	$\delta_D(g+,g-)$	2.92 (2.74)	2.18 (1.92)	130.0 (136.9)
bb/sc	$\delta_D(g+,a)$	2.88 (2.70)	2.17 (1.91)	127.2 (133.5)
	average	**2.88 (2.72)**	**2.15 (1.92)**	**128.6 (134.8)**

[a] For L-Ser-NH_2 conformers. Distances (d) in Å and angles (Θ) are in degrees.
[b] The numbering of the atoms are according to that of Scheme 1A.
[c] Values in parenthesis are those obtained at RHF/3-21G levels of theory.

$d_{O1\ldots H—O3}$ = 2.36 Å, and $\Theta_{O1\ldots H—O3}$ = 117.1°). Reoptimization of the $\delta_L(g-,g-)$ fold using the larger basis set resulted in a conformational shift (Table IIC). Although the recalculated structure is still in the $\delta_L(g-,g-)$ catchment region, the sc/bb-type H-bond is no longer present ($d_{O1\ldots O3}$ = 3.52 Å, $d_{O1\ldots H—O3}$ = 3.27 Å, and $\Theta_{O1\ldots H—O3}$ = 97.5°). Therefore, due to the lack of the H-bond stabilizing term, one would expect an increase in relative energy. In general, a normal H-bond is expected to have a contribution of some 2–3 kcal mol^{-1}. Instead of the predicted increase, a decrease of 4.6 kcal mol^{-1} was computed upon replacement of the 3-21G basis by 6-311++G** (Table IIC). The latter value is approximately the amount due to the energy lowering effect of the enlarged basis set (estimated as ≈4.5 kcal mol^{-1} and discussed in details later). It looks as if the "loss" of an H-bond would not have impact on the total energy of the molecule. Beside the $\delta_L(g-,g-)$ conformation, where reoptimization

TABLE VIII.

The Fitting Parameters between Relative Energies of Serine Diamide Conformers Compared to $\gamma_L(g+,g+)$ and the Relative Probabilities of Similar Backbone Structures in a Set of Proteins with Known X-ray Structure.

Size[a]	Method[b]	m[c]	b[d]	Stand. Error[e]	Sum[f]	Overlap[g]
60	RHF/6-311++G**	−0.84	6.67	2.80	10622	4395
60	RHF/6-311++G**//RHF/3-21G	−0.46	6.69	3.09	10374	4364
60	RHF/3-21G	−0.59	11.09	4.53	10374	4364
45	RHF/6-311++G**	−0.66	7.04	2.96	7214	2640
45	RHF/6-311++G**//RHF/3-21G	−0.42	6.90	3.15	7074	2640
45	RHF/3-21G	−0.61	11.49	4.46	7074	2640
30	RHF/6-311++G**	−0.63	7.05	3.00	3637	1314
30	RHF/6-311++G**//RHF/3-21G	−0.37	6.90	3.21	4027	1375
30	RHF/6-21G	−0.28	11.48	4.59	4027	1375

[a] The radius of the hypersphere pinpointed by the ϕ, ψ, χ_1, and χ_2 conformational values calculated with the given method.
[b] The used *ab initio* level of theory to determine the ϕ, ψ, χ_1, and χ_2 conformational parameters and the relative energies.
[c] The slope of the fitted line (see also Fig. 5).
[d] The $\Delta E(0)$ value or vertical intercept.
[e] Standard error (in kcal mol^{-1} units) calculated according to eq. (3).
[f] The sum of series residues incorporated in the analysis.
[g] The number of overlapping serine residues.

has resulted in the elimination of the H-bond, two additional conformers [$\gamma_D(g+,g+)$ and $\alpha_D(g+,g+)$] were found where the H-bonds are weakened in structures determined with the larger basis set. The remaining sc/bb-type H-bonds (Table VII), found in the $\beta_L(g-,g+)$, $\delta_D(g-,g-)$, $\gamma_D(g+,g-)$, $\varepsilon_D(g+,g-)$, $\gamma_L(a,g-)$, $\gamma_L(g+,g+)$, and $\alpha_D(a,g-)$ conformers, look less affected by the inclusion of diffuse and polarization functions in the basis. On average, the H-bond distance and H-bond angle modifying effect of the basis-set size is in the range of 0.2 Å and 5°, respectively.

The last case to be mentioned is that of the bb/sc-type H-bond (second half of Table VII), where a backbone amide proton is donated to the oxygen atom (O3) of the side chain. In this category a significant difference is observed, depending on whether the first or the second amide group is involved in the interaction. If the NH of the first amide group (N1) is involved in the interaction, as for $\delta_L(g+,a)$, $\gamma_L(g-,g-)$, $\gamma_L(g-,a)$, $\alpha_L(g-,a)$, and $\alpha_L(g-,g-)$, the H-bond looks weakened in structures determined at the 6-311++G** basis set. For example, $\Delta d_{O3\ldots H—N1} = 0.3$ Å, resulting in a rather large average value ($d_{O3\ldots H—N1} = 2.47$ Å) for the five 6-311++G** structures (Table VII). On the other hand, if the NH of the second (N2) amide group is donated to the side-chain oxygen atom (O3), the H-bond is strong in both sets of structures (e.g., $d^{6\text{-}311++G^{**}}_{O3\ldots H—N2} = 2.15$ Å, $d^{3\text{-}21G}_{O3\ldots H—N2} = 1.92$ Å). In conclusion, H-bond parameters are lengthened at the RHF/6-311++G** level of theory compared to RHF/3-21G data, but the way how the individual H-bonds affect the relative energy of the conformer is not easy to establish.

ENERGETICS

Energies of HCO-L-Ser-NH_2 conformers were first determined at the RHF/3-21G level, followed by single-point calculations at the RHF/6-311++G**//RHF/3-21G level. Afterwards, all these structures were reoptimized at the RHF/6-311++G** level and three additional sets of single-point calculations (RHF/6-31+G*//RHF/6-311++G**, RHF/TZ2P//RHF/6-311++G**, and MP2/6-311++G**//RHF/6-311++G**) were performed on all available conformers (Scheme 1B). Finally, B3LYP/6-311++G** and MP2/6-311++G** optimizations were made on selected conformers. All sets of energies ($\Delta E_{RHF/3\text{-}21G}$, $\Delta E_{RHF/6-311++G^{**}//RHF/3\text{-}21G}$, $\Delta E_{RHF/6\text{-}311++G^{**}}$, $\Delta E_{MP2/6\text{-}311++G^{**}//RHF/6\text{-}311++G^{**}}$, $\Delta E_{RHF/6\text{-}31+G^{*}//RHF/6\text{-}311++G^{**}}$, $\Delta E_{RHF/TZ2P//RHF/6\text{-}311++G^{**}}$, and $\Delta E_{B3LYP/6\text{-}311++G^{**}}$, and $\Delta E_{MP2/6\text{-}311++G^{**}}$) were referenced to $E_{\gamma L(g-,g-)}$ and analyzed (Table IB, Table IIIA, and Fig. 3A). Comparing the ΔE values, those of L-type conformers (1 to 19) are typically lower than values calculated for D-type conformers (20 to 44) (Table IIIA). In the case of the RHF/6-311++G** data, for example, all L-type backbone structures (x_L)

but $\alpha_L(a,a)$ have lower relative energy than any D-type backbone orientations (x_D) but $\varepsilon_D(g+,g-)$. Typically, for calculations better than RHF/3-21G the average of all x_L is lower than 5.5 kcal mol^{-1}, while the average of the D-type backbone orientations is larger than 8 kcal mol^{-1}. This significant relative energy difference between structures from the L-and the D-valley is very similar for all RHF calculations and the MP2 single-point calculations as well. This, in turn, supports the observation that D-type backbone conformers are much less frequent than L-type ones in globular proteins.

BASIS SET DEPENDENCE OF RELATIVE STABILITIES

Due to the considerable limitations built into the RHF/3-21G *ab initio* model, energetic and conformational results obtained with it cannot always be trusted. Nevertheless, characterization of 40 or much more distinct conformers of large biomolecules by more expensive methods of *ab initio* theory is unrealistic. Therefore, it is of importance to attempt to calibrate our lowest level results against more accurately computed values to gain further insight into the expected accuracy of the RHF/3-21G *ab initio* model.

For structures, we wish to compare ϕ, ψ, χ_1, and χ_2 values obtained by geometry optimization at the RHF/6-311++G** level with those obtained at the RHF/3-21G level. Let Z stand for any of the torsional angles ϕ, ψ, χ_1, and χ_2. We found a linear relationship between $Z^{\text{RHF/6-311++G**}}$ and $Z^{\text{RHF/3-21G}}$. The statistical analysis of the pair-wise correlations (Table IIA), described earlier, concluded that all four major conformational parameters correlate extremely well: the Pearson correlation coefficients are typically 0.994 or better. This agrees with our earlier finding that at the HF level the conformational parameters do not change substantially when reoptimized using a larger basis set. Furthermore, as Figure 3 attests, there is a linear relationship between $\Delta E_{\text{RHF/6-311++G**}}$ and $\Delta E_{\text{RHF/6-311++G**//RHF/3-21G}}$ values. Therefore, the clear decrease in relative energy, described above, is due to the enlarged basis set and not to the improved geometry.

Taking advantage of the unusually large set of energies calculated for the present peptide model, questions of general interest can be addressed. For example, for the relative energies we attempted to condense the effect of basis set enlargement into a single value. By correlating any two ΔE sets, straight lines of the following form

$$\Delta E^{\text{Improved method}} = m\Delta E^{\text{RHF/3-21G}} + b \qquad (1)$$

have been obtained. The fitted parameters are summarized in Table IIIB, and two of these correlations are shown in Figure 3B and C. All energy trends were crosscorrelated (Table IIIB), Pearson correlation coefficients (r), standard errors (σ), and fitting parameters (m and b) were determined. Values in the $8 * 8$ correlation matrix are organized in a way that from left to right and from up to down the total energy of the energy series increases (cf. Table IIIA footnotes). In simple terms, the results from smaller basis sets are to the left, while higher quality results are more to the right. The diagonal elements refer to autocorrelation and the off-diagonal elements to crosscorrelation. The pair-wise comparison of all eight relative energy sets (Table IIIB) revealed that they all do crosscorrelate at an unexpectedly high level. Sorting the off-diagonal elements of the *Pearson* matrix, the best r value is close to 0.9998 and even the poorest correlation is as good as 0.928. (The Pearson correlation matrix is the only symmetrical one among the matrices containing the statistical results.) For example, the correlation, $r = 0.995$, found between $\Delta E_{\text{RHF/6-311++G**//RHF/3-21G}}$ and $\Delta E_{\text{RHF/6-311++G**}}$ data sets can be considered as typical, with a low standard error ($\sigma = 0.32$ kcal mol^{-1}). The high significance implies that with a small error margin the energy of any set of conformers of a given theoretical level can be predicted by simply using the linear relationship between two selected levels of theory. The last two subtables (C and D) of Table IIIB compiles the regression parameters, providing an opportunity to predict *ab initio* total energies. From any energy set any other could be predicted, but in practice only the estimation of the results of expensive methods from inexpensive ones is of interest (e.g., $\Delta E_{\text{MP2/6-311++G**}} = 0.63\Delta E_{\text{RHF/3-21G}} + 0.718$, the expected standard error is 0.45 kcal mol^{-1}). The quality of the prediction is typically better between methods at higher levels, and in general, the magnitude of the standard error decreases as the level of applied theory increases.

To make predictions, the *a priori* knowledge of the fitting parameters, determined from the fitting of any two sets of energies, is necessary. Fortunately, due to the nearly perfect correlation observed, it is possible to reduce dramatically the number of the conformers used for the determination of the fitting parameters. We found that it is enough to use less than one-quarter of the total energy points

and already rather accurate fitting parameters can be obtained. Instead of taking the subset of conformers in a random way, estimation of the fitting parameters can be improved if structures are selected both from the small- and from the high-energy side of the set. Undoubtedly, the more points are calculated explicitly the better the result is, but in practice, after the use of carefully selected five to eight points the improvement is insignificant. (This can be demonstrated straightforward by a "Jackknife" or a "random subsampling" method.) To sum up, the result of a few but expensive energy calculations can be used in conjunction with any complete energy set obtained at a low theoretical level to determine the fitting parameters, which could then be used to predict the energies of all other conformers. To test the power and efficiency of such an energy estimation, B3LYP/6-311++G** and MP2/6-311++G** energies of serine diamide were determined (the *ab initio* data reported in Table IB form a good enough basis to perform such an extra- or interpolation of the energies).

Another important problem of *ab initio* computations is whether single-point energy calculations are good enough, or full geometry optimization is required to determine the energy order of the full conformational set.[69–71] Comparing the values of ΔE(RHF/6-311++G**//RHF/3-21G) with those of ΔE(RHF/6-311++G**) and the results of MP2 single point calculations [ΔE(MP2/6-311++G**//RHF/6-311++G**)] with values determined by MP2 geometry optimizations [ΔE(MP2/6-311++G**)], the answer is clear that single-point calculations are just as instructive as the 15 to 25 times more expensive geometry optimizations (Table IIIB). (Note that $r^{\Delta E(\mathrm{RHF/6\text{-}311++G^{**}//RHF/3\text{-}21G})\Leftrightarrow\Delta E(\mathrm{RHF/6\text{-}311++G^{**}})} = 0.995$ and $r^{\Delta E(\mathrm{MP2/6\text{-}311++G^{**}//RHF/6\text{-}311++G^{**}})\Leftrightarrow\Delta E(\mathrm{MP2/6\text{-}311++G^{**}})} = 0.9998$.)

It is also important to establish how accurate the most favored B3LYP density functional theory (DFT) method is compared to MP2, because for larger molecules the CPU requirements of the two energy computations can be drastically different, favoring density functional techniques. The Pearson correlation coefficient between ΔE(MP2/6-311++G**) and ΔE(B3LYP/6-311++G**) energies is close to 0.99 (Table IIIB), indicating that B3LYP results in energy data of MP2 quality.

These finding are, of course, more theoretical than practical at this point. However, they outline the possibilities that, at least for peptides and protein fragments, inexpensive RHF/3-21G or RHF/6-311++G** energy (and perhaps property) results can be scaled, and thus approximate results from much higher levels of the *ab initio* armamentarium.

Correlation between Natural Occurrence of SER Conformers and Their Computational Stability

Comparison of structural parameters taken from experimental databases (X-ray and/or NMR) with relevant *ab initio* results reveals interesting features. In the following, we are assuming that the probability p_x of a diamide structure (conformer x) in proteins depends only on its relative energy. In this simple model several factors are neglected, such as interresidue interactions, long-range effects, hydration, etc. Notwithstanding the limitations of this approach, it makes possible the correlation of relative energies and the relative probability of the same backbone conformation in an ensemble of proteins with known X-ray (or NMR) structure. Choosing $\gamma_L(g+,g+)$ as the reference conformer for the energy scale, the relative population is assumed to be related to relative stabilities in a Boltzman-type exponential distribution:

$$\left(\frac{p_x}{p_{\gamma_L}}\right) = e^{+(\Delta E/m)} \cdot e^{(-m/b)}, \tag{2}$$

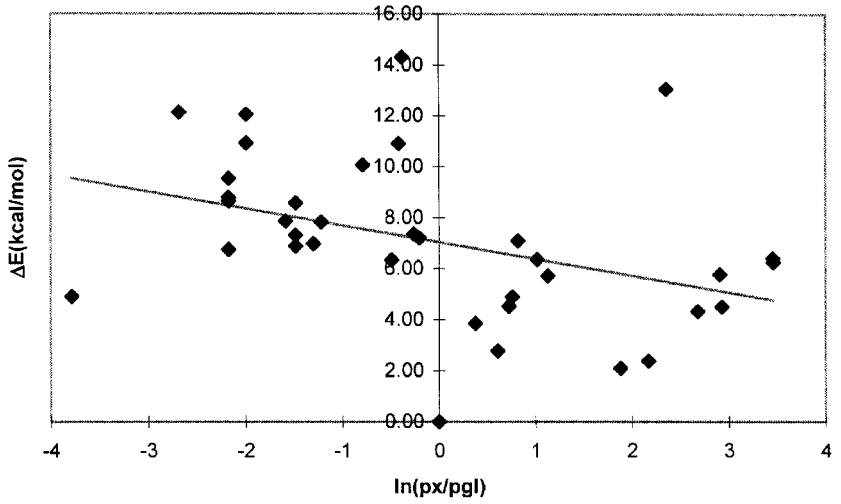
where $\Delta E = E_x - E_{\gamma L}$. In the logarithmic form of this equation,

$$\Delta E = m \cdot \ln\left(\frac{p_x}{p_{\gamma_L}}\right) + b, \tag{3}$$

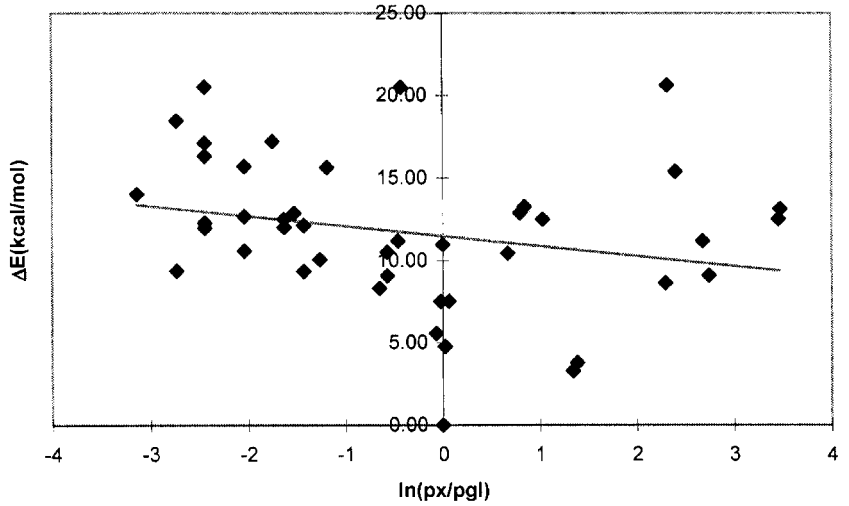
the correlation between relative energies and relative probabilities should be linear.

Equations (2) and (3) are expected to be valid only in the gas phase. If relative populations ($p_x/p_{\gamma L}$) could be measured experimentally in the gas phase (e.g., by FTIR), and if accurate ΔE values could be computed by *ab initio* methods, then eqs. (2) and (3) are expected to be valid. However, it should be remembered that here we are correlating solid state occurrences of various serine residue conformations in globular proteins with computed ΔE values of an isolated serine model peptide, HCO-L-Ser-NH_2.

A protein data set, described elsewhere,[59, 60] was used in this work, where as many as 9511 Ser residues were found in 1135 proteins. Each set of conformational parameters (ϕ, ψ, χ_1, and χ_2) from RHF/6-311++G** and RHF/3-21G calculations fix the center of a 4D hypersphere. The experimentally



A



B

FIGURE 5. Correlation of the relative energy of serine diamide conformers (relative to its $\gamma_L(g+,g+)$ conformation) and the relative probabilities of similar backbone structures in an ensemble of proteins with known X-ray structure. (A) 36 RHF/6-311++G** ϕ, ψ, χ_1, and χ_2 conformational values were used to center the 36 hyperspheres (with a radius of 45°; see text) incorporating a sum of 7214 serines, among which 2853 overlapped. The standard error of the fit is 2.96. (B) 44 RHF/3-21G ϕ, ψ, χ_1, and χ_2 conformational values were used to center the 44 hyperspheres (with a radius of 45°; see text) incorporating a sum of 7074 serines, among which 2640 overlapped. The standard error of the fit is 4.46.

determined structures form a cluster about these conformational centers. The cluster falls within a circle of a given radius predetermined by a vectorial sum of $\Delta\phi, \Delta\psi, \Delta\chi_1$, and $\Delta\chi_2$. Although it is more proper to express the values of angular deviations in radii, if they are to be regarded as distances, nevertheless, here, for the sake of convenience, degrees are used. Setting the size of the radius as 60°, 45°, and 30°, different number of serine residues (called *sum* in Table VIII) was assigned within the selected conformational area. These hyperspheres can overlap. The number of serine residues found

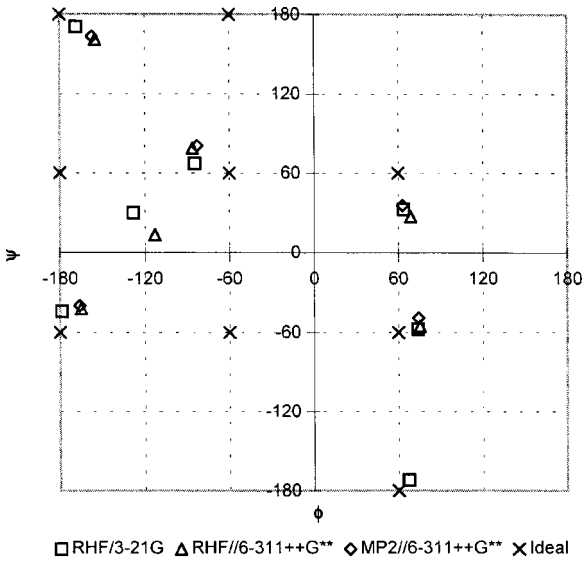
in these overlapping regions (called *overlap* in Table VIII) was counted. The standard error (4) and the Pearson-correlation coefficient (r) (5) are calculated as:

$$S_{y-x} = \left(\left[\frac{1}{n(n-2)} \right] \left[n \sum y^2 \left(\sum y \right)^2 - \frac{\left[n \sum xy - (\sum x)(\sum y) \right]^2}{n \sum x^2 - (\sum x)^2} \right] \right)^{1/2} \quad (4)$$

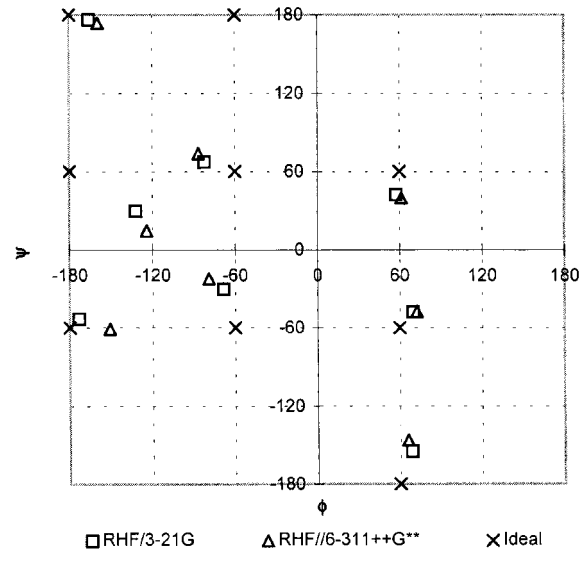
and

$$r = \frac{\left[n(\sum xy) - (\sum x) - (\sum y) \right]}{\sqrt{\left(n \sum x^2 - (\sum x)^2 \right) \left(n \sum y^2 - (\sum y)^2 \right)}} \quad (5)$$

They measure the accuracy and the correlation of the fit, respectively. S_{y-x} is in energy units (i.e., kcal mol^{-1}), because y symbolizes ΔE and x symbolizes $\ln(p_\chi / p_{\gamma L})$. (Due to the limitations of the diamide approximation, it would not be realistic to expect an exceptionally good fit (Fig. 5), namely a standard error close to zero.) It is interesting to note that the enlargement of the allowed region (the increase of the volume of the hypersphere) does not improve significantly the standard error of the fit (Table VIII). This indicates that even for the smallest data set (obtained with a radius of 30°) the distribution of the 3827 ± 200 serine residues among the conformational clusters is close to that calculated for more than 9500 residues. On the other hand, the standard error varies markedly with the change of the applied levels of theory (RHF/3-21G, RHF/6-311++G**, and RHF/6-311++G**//RHF/3-21G). This suggests that either the precise locations of the hyperspheres or the relative energy differences change significantly with the applied method. The observation that the 6-311++G** energies ($E^{\mathrm{RHF/6\text{-}311++G^{**}}}$) and the 6-311++G** single point energies ($E^{\mathrm{RHF/6\text{-}311++G^{**}//RHF/3\text{-}21G}}$) are close (Tables IIIA, IIIB and Fig. 3) is of considerable interest. Upon monitoring the variation of the standard errors of the line fitting as a function of the *ab initio* method (Table VIII), it looks as if the improved energy and not the improved geometry is responsible for the better fitting. In summary: (a) RHF/3-21G geometries for peptides, due to a fortunate cancellation of errors, provide structural data that are of remarkable quality; and (b) single-point energy calculations provide a quick way to get valuable energetic information. While promising, the computational strategy based on these observations should be further tested on other amino acid residues and on other models.



A



B

FIGURE 6. (A) The RHF/3-21G, RHF/6-311++G**, MP2/6-311++G**, and ideal locations of the different backbone conformers of HCO-L-Ala-NH_2 on the Ramachandran surface. (B) The ideal and the average backbone locations of HCO-L-Ser-NH_2 calculated at the RHF/3-21G and RHF/6-311++G** levels of theory.

Finally, one is tempted to analyze the small number of available MP2 information (e.g., calculated for Ala) and perform some extrapolations for Ser (see Fig. 6A and B) as outlined above. In terms of

ϕ, ψ locations, the introduction of electron correlation in the case of the alanine diamide did not result in major conformational changes when compared to RHF/6-311++G** backbone values (Table VB): only the ψ value of α_D is shifted to a noticeable extent. On the other hand, the δ_L structure migrated to the nearby global minimum γ_L. For most of the backbone conformers the RHF/6-311++G** backbone torsional values are closer to those calculated at MP2/6-311++G** than values computed at RHF/3-21G. Once again, we have to stress the general conclusion that, due to fortuitous cancellation of errors, the RHF/3-21G geometries are better than expected. Turning now to total energies, electron correlation could, in principle, modify the energy order of the available conformers. This is indeed the case if rather special conformational orientations are investigated. However, in general, as demonstrated here the linearity between ΔE values holds.

Conclusions

For HCO-L-Ser-NH_2 multidimensional conformational analysis (MDCA) predicts 81 distinct structures. Previously, 44 of these structures proved to be minima at the RHF/3-21G level. In this study, these structures were reoptimized at the RHF/6-311++G** level. Eight conformational migrations were observed, leading to 36 distinct conformers. Among these, only five structures differ significantly from their RHF/3-21G parents. Conformational properties of the relevant structures, as well as relative energy differences obtained at eight different levels of theory were analyzed. Beside these "major changes," only smaller conformational shifts were observed for all the other structures. We were able to demonstrate that strong correlation exists between structural and energetic properties determined at different levels of theory. As found earlier, the polar side chain may have a significant impact on selected backbone structures. In consequence, three α_L main-chain folds were located, although the α_L structure vanishes in the case of HCO-Xxx-NH_2 (where Xxx = Gly, Ala, Val, Phe). On the other hand, no ε_L-type backbone conformers were found. These model conformers were used as template structures to describe the location of 9511 Ser residues taken from 1135 nonhomologous proteins. In conclusion, RHF/3-21G geometries for peptides, due to a fortunate cancellation of errors, provide structural data that are of remarkable quality. In general, single-point energy calculations provide a quicker way to get valuable energetic information, then with full geometry optimization, and preserving ΔE trends. While promising, these observations should be further tested on other amino acid residues and on other more extended peptide models.

Acknowledgments

Special thanks to Silicon Computer Ltd, Hungary, and to Chemicro Ltd., Hungary, for hardware support.

References

1. Császár, A. G.; Perezel, A. Prog Biophys Mol Biol 1999, 71, 243.
2. (a) Howard, J. C.; Ali, A.; Scheraga, H. A.; Momany, F. A. Macromolecules 1975, 8, 607; (b) Némethy, G.; Miller, M. H.; Scheraga, H. A. Macromolecules 1980, 13, 914.
3. Paine, G. H.; Scheraga, H. A. Bipolymers 1986, 25, 1547.
4. Bruccoleri, R. E.; Karplus, M. Biopolymers 1987, 26, 137.
5. Ripoll, D. R.; Scheraga, H. A. Biopolymers 1988, 27, 1283.
6. Schulz, G. E. Annu Rev Biophys Chem 1988, 17, 1.
7. Wolfe, S.; Bruder, S.; Weaver, D. F.; Yang, K. Can J Chem 1988, 66, 2703.
8. (a) Lambert, M. H.; Scheraga, H. A. J Comp Chem 1989, 10, 798; (b) Lambert, M. H.; Scheraga, H. A. J Comp Chem 1989, 10, 770; (c) Lambert, M. H.; Scheraga, H. A. J Comp Chem 1989, 10, 817.
9. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D., Jr.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J Mol Biol 1977, 112, 535.
10. Abola, E. E.; Bernstain, F. G.; Bryant, S. H.; Koetzle, L. F.; Weng, J. In Data Commission of the International Union of Crystallography; Allen, F. H.; Bergerhoff, G.; Sievers, R., Eds.; Cambridge: Bonn, 1987, p. 107.
11. Mirskz, A. E.; Pauling, L. Proc Natl Acad Sci USA 1936, 22, 439.
12. Liquori, A. M. Q Rev Biophys 1969, 2, 65.
13. Lattman, E. E.; Rose, G. D. Proc Natl Acad Sci USA 1993, 90, 439.
14. Aubry, A.; Marraud, M.; Protas, J.; Neel, J. C R Acad Sci Paris 1974, 287C, 163.
15. Aubry, A.; Ghermani, N.; Marraud, M. Int J Pept Protein Res 1984, 23, 113.
16. Raj, P. A.; Soni, S. D.; Ramasubbu, N.; Bhandary, K. K.; Levine, M. J. Biopolymers 1990, 30, 73.
17. Stroup, A. N.; Cole, L. B.; Dhingra, M. M.; Gierach, L. M. Int J Pept Protein Res 1990, 36, 531.
18. Gething, M.-J.; Sambrook, J. Nature 1992, 335, 33.
19. Perczel, A.; Foxman, B. M.; Fasman, G. D. Proc Natl Acad Sci USA 1992, 89, 8210.
20. The HyperMM+ force field is derived from the public domain code developed by: Allinger, N. L. J Am Chem Soc 1977, 99, 8127.
21. Wiener, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profetajr, S.; Wiener, P. J Am Chem Soc 1984, 106, 765.

22. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
23. Jorgensen, W. L.; Tirado–Rivers, J. J Am Chem Soc 1988, 10, 1657.
24. Pranate, J.; Wierschke, S.; Jorgensen, W. L. J Am Chem Soc 1991, 113, 2810.
25. Bingham, R. C.; Dewar, M. J. S.; Lo, D. H. J Am Chem Soc 1975, 97, 1302.
26. Dewar, M. J. S.; Thiel, W. J Am Chem Soc 1977, 99, 4899.
27. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Steward, J. J. P. J Am Chem Soc 1986, 107, 3902.
28. Stewart, J. J. P. J Comp Chem 1989, 10, 221.
29. Rodriguez, A. M.; Baldoni, H. A.; Suvire, F.; Nieto–Vasquez, R. G.; Zamarbide, R. D.; Farkas, Ö.; Perczel, A.; Csizmadia, I. G. J Mol Struct (Theochem), to appear.
30. McAllister, M. A.; Perczel, A.; Császár, P.; Viviani, W.; Rivail, J. L.; Csizmadia, I. G. J Mol Struct (Theochem) 1989, 290, 161.
31. Perczel, A.; Ángyán, J. G.; Kajtár, M.; Vivini, W.; Rivail, J. L.; Marcoccia, J. F.; Csizmadia, I. G. J Am Chem Soc 1991, 113, 6256.
32. Scarsdale, J. N.; Van Alsenoy, C.; Klimkowski, V. J.; Schäfer, L.; Monany, F. A. J Am Chem Soc 1983, 105, 3438.
33. Schäfer, L.; Klimkowski, V. J.; Momany, F. A.; Chuman, H.; van Alsenoy, C. Biopolymers 1984, 23, 2335.
34. Wiener, S. J.; Singh, U. C.; O'Donnel, T. J.; Kollman, P. J Am Chem Soc 1984, 106, 6243.
35. Head–Gordon, T.; Head–Gordon, M.; Frisch, M. J.; Brooks, C. L., III; Pople, J. A. Int J Quantum Chem Biol 1989, 16, 311.
36. Böhn, H. J.; Brode, S. J Am Chem Soc 1991, 113, 7129.
37. Head–Gordon T.; Head–Gordon, M.; Frisch, M. J.; Brooks, C. L., III; Pople, J. A. J Am Chem Soc 1991, 113, 5989.
38. Viviani, W.; Rivial, J. L.; Perczel, A.; Csizmadia, I. G. J Am Chem Soc 1993, 115, 8321.
39. Farkas, Ö.; Perczel, A.; Marcoccia, J. F.; Hollósi, M.; Csizmadia, I. G. J Mol Struct (Theochem) 1995, 331, 27.
40. Perczel, A.; Farkas, Ö.; Csizmadia, I. G. J Comput Chem 1996, 17, 821.
41. Perczel, A.; Farkas, Ö.; Csizmadia, I. G. J Am Chem Soc 1996, 118, 7809.
42. Perczel, A.; Farkas, Ö.; Marcoccia, J. F.; Csizmadia, I. G. Int J Quantum Chem 1997, 61, 797.
43. Farkas, Ö.; McAllister, M. A.; Ma, J. H.; Perczel, A.; Hollósi, M.; Csizmadia, I. G. J Mol Struct (Theochem) 1996, 396, 105.
44. (a) Perczel, A.; Farkas, Ö.; Császár, A. G.; Csizmadia, I. G. Can J Chem 1997, 75, 1120; (b) Jákli, I.; Perczel, A.; Farkas, Ö.; Hollósi, M.; Csizmadia, I. G. J Mol Struct (Theochem) 1988, 444, 303.
45. Endrédi, G.; Perczel, A.; Farkas, Ö.; McAllister, M. A.; Csonka, G. I.; Ladik, J.; Csizmadia, I. G. J Mol Struct (Theochem) 1997, 391, 15.
46. Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schäfer, L. J Am Chem Soc 1992, 114, 5369.
47. Perakyla, M.; Kollman, P. A. J Am Chem Soc 1997, 199, 1189.
48. Beveridge, A. J. Protein Sci 1996, 5, 1355.
49. Perakyla, M.; Rouvinen, J. Chem Eur J 1996, 2, 1548.
50. Siam, K.; Klimkowski, V. J.; van Alsenoy, C.; Ewbank, J. D.; Schäfer, L. J Mol Struct (Theochem) 1987, 152, 261.
51. Perczel, A.; Daudel, R.; Ángyá, J. G.; Csizmadia, I. G. Can J Chem 1990, 68, 1882.
52. Aubry, A.; Ghermani, N.; Marraud, M. Int J Pept Proetin Res 1984, 23, 113.
53. Karle, I. L. J Am Chem Soc 1979, 181, 4.
54. Jákli, I.; Perczel, A.; Farkas, Ö.; Hollósi, M.; Csizmadia, I. G. J Mol Struct (Theochem) (1998), to appear.
55. Jákli, I.; Perczel, A.; Farkas, Ö.; Csizmadia, I. G. J Comput Chem 1998, to appear.
56. Pulay, P.; Forgarasi, G.; Pang, F.; Boggs, J. E. J Am Chem Soc 1979, 101, 2550.
57. Schäfer, A.; Huber, C.; Ahlrichs, R. J Chem Phys 1992, 100, 5829.
58. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head–Grodon, M.; Gonzalez, C.; Pople, J. A. Gaussian94; Gassian, Inc.: Pittsburgh, PA, 1995.
59. Hobohm, U.; Scharf, M.; Schneider, R.; Sander, C. Protein Sci 1992, 1, 409.
60. Hobohm, U.; Sander, C. Protein Sci 1994, 3, 522.
61. Efimov, A. V. Prog Biophys Mol Biol 1993, 60, 201.
62. Zimmerman, S. S.; Pottle, M. A.; Nemethy, G.; Scheraga, H. A. Macromolecules 1977, 10, 1.
63. Richardson, J. S.; Richardson, D. C. Prediction of Potein Structure and the Principles of Protein Cnformation; Fasman, G. D., Ed.; Plenum Press: New York, 1989, p. 1.
64. Rooman, M. J.; Kocher, J. P. A.; Widak, S. J. Biochemistry 1992, 31, 10226.
65. Thornton, J.; Jones, D. T.; MacArthur, M. W.; Orengo, C. M.; Swindells, M. B. Philos Trans R Soc Lond 1995, B348, 71.
66. Schäfer, L.; van Alsenoy, C.; Scarsdale, J. N. J Chem Phys 1982, 76, 1439.
67. Schäfer, L.; Cao, M. J Mol Struct (Theochem) 1995, 333, 201.
68. Karplus, P. A. Protein Sci 1996, 5, 1406.
69. Császár, A. G. J Am Chem Soc 1992, 114, 9569.
70. Császár A. G. J Mol Struct 1995, 346, 141.
71. Császár, A. G. J Phys Chem 1996, 100, 3541.