Peptide model XXVIII: An exploratory ab initio and density functional study on the side-chain–backbone interaction in *N*-acetyl-L-cysteine-*N*-methylamide and *N*-formyl-L-cysteinamide in their γ_L-backbone conformations

M.A. Zamora, H.A. Baldoni, A.M. Rodriguez, R.D. Enriz, C.P. Sosa, A. Perczel, A. Kucsman, O. Farkas, E. Deretey, J.C. Vank, and I.G. Csizmadia

Abstract: A conformational and electronic study on the energetically preferred conformations (γ_L) of N- and C-protected L-cysteine (P-CONH-CH(CH_2SH)-CONH-Q, where P and Q may be H or Me) was carried out. After restraining the backbone (BB) conformation to its global minimum (γ_L or C_7^{eq}), all nine possible side-chain (SC) conformations were subjected to geometry optimization at the HF/3–21G and the B3LYP/6–31G(d,p) levels of theory. Seven of the nine side-chain conformers were located on the potential-energy surface. All conformers were subjected to an AIM (atoms in molecules) analysis. This study indicates that three of the seven optimized conformers exhibited either or both SC $\rightarrow$ BB- or BB $\rightarrow$ SC-type intramolecular hydrogen bonding. Five conformers, however, had distances between a proton and a heteroatom that suggested hydrogen bonding.

Key words: L-cysteine diamides, side-chain potential-energy surface, ab initio and DFT geometry optimization, AIM analysis, intramolecular hydrogen bonding.

Résumé : On a réalisé une étude conformationnelle et électronique des conformations énergétiquement privilégiée (γ_L) de la L-cystéine N- et C-protégée (P-CONH-CH(CH_2SH)-CONH-Q; P et Q peuvent être égaux à H ou à Me). Ainsi, en retreignant la conformation du squelette (BB) à son minimum global (γ_L ou C_7^{eq}), on a soumis les neuf conformations possibles pour la chaîne latérale (SC) à une optimisation de la géométrie aux niveaux HF/3–21G et B3LYP/6–31G(d,p) de la théorie. Sept des neuf conformères de la chaîne latérale se situent sur la surface d'énergie potentielle. Tous les conformères ont été soumis à une analyse par la méthode des atomes dans les molécules (AIM). Cette étude indique que trois des sept conformères optimisés présentent l'une ou l'autre ou les deux types de liaisons intramoléculaires SC $\rightarrow$ BB ou BB $\rightarrow$ SC. Toutefois, dans cinq des conformères les distances entre un proton et un hétéroatome suggèrent l'existence d'une liaison hydrogène.

Mots clés : diamides de la L-cystéine, surface d'énergie potentielle de la chaîne latérale, optimisation ab initio et DFT de la géométrie, méthode des atomes dans les molécules (AIM), liaison hydrogène intramoléculaire.

[Traduit par la Rédaction]

Introduction

Cysteine (Cys) is the sulfur analogue of serine (Ser). Both amino acid residues may conveniently be studied computationally in their N- and C-protected forms; Fig. 1 shows *N*-acetyl-L-Cys-*N*-methylamide (**I**) and *N*-formyl-L-Cys-amide (**II**). *N*-Formyl-L-Ser-amide (**III**) has already been studied extensively (1–5) as a continuation of a general investigation including Gly (6, 7), Ala (6–8), Val (9), and Phe (10–12).

In many ways the cysteine residue is similar to serine; one therefore expects the -CH_2-SH moiety of cysteine to be in-

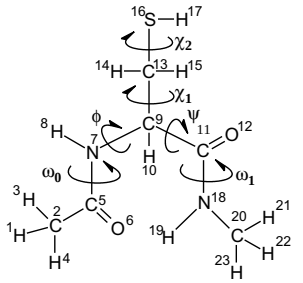
Received 20 March 2001. Published on the NRC Research Press Web site at http://canjchem.nrc.ca on 28 June 2002.

M.A. Zamora, H.A. Baldoni, A.M. Rodriguez, and R.D. Enriz. Department of Chemistry, National University of San Luis, Chacabuco 971, 5700 San Luis, Argentina.
C.P. Sosa. CRAY Inc. 1340 Mendota Heights Rd., Mendota Heights, MN 55120, U.S.A.
A. Perczel,[1] A. Kucsman, and O. Farkas.[2] Institute of Organic Chemistry, Eotvos University, 112 Budapest, PO Box 32, H-1117, Hungary.
E. Deretey, J.C. Vank, and I.G. Csizmadia.[3] Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada.

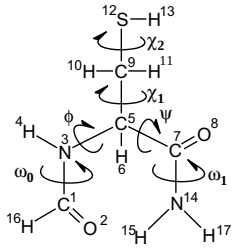
[1]Alternative address: Deparment of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.
[2]Alternative address: Department of Chemistry, Wayne State University, Detroit, MI 48202, U.S.A.
[3]Corresponding author (e-mail: icsizmad@alchemy.utoronto.ca).

 DOI: 10.1139/V02-076

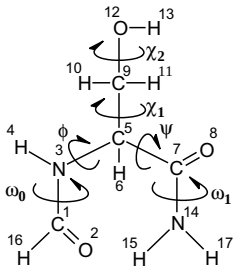
Fig. 1. γ_L conformations of MeCO-Cys-NHMe (**I**), HCO-Cys-NH_2 (**II**), and HCO-Ser-NH_2 (**III**). The numbering system of the atoms and the torsional angles ω_0, ω_1, ϕ, ψ, χ_1, and χ_2 are shown in this figure.



N-acetyl-L-Cys-*N*-methylamide
I



N-formyl-L-Cys-amide
II



N-formyl-L-Ser-amide
III

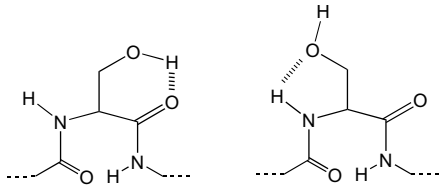
volved in intramolecular hydrogen bonding similar to that demonstrated in Fig. 2 for the -CH_2-OH group of serine.

Cysteine, a residue with a polar side chain, is often located on the surface of proteins, and thus is directly involved in various biochemical interactions. Cysteine can mediate structural changes, improve hydration of accessible area, or participate in intermolecular interactions. Thus, cysteine is frequently involved in a wide range of enzymatic reactions including those of cysteine proteases (13–15). In Scheme 1, a simplified mechanism of cysteine protease (HS-E-Im) is shown, where HS symbolizes the mercaptan side chain of a cysteine residue and Im denotes the imidazole moiety in the side chain of a histidine residue. The enzyme is not active at low pH due to protonation at Im, and it is also inactive at high pH due to deprotonation of the SH group. It is only active at neutral pH when internal proton transfer is possible from the SH group to Im.

A schematic mechanism for the hydrolysis of the peptide bond with the catalytic action of cysteine protease is shown in Scheme 2.

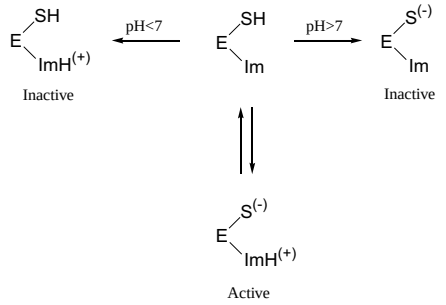
In addition to regular cysteine proteases, there are numerous other enzymes that contain free-sulfhydril groups. For example, there is a cysteine-containing enzyme called *transglutaminase*, which does not hydrolyze peptide bonds, but instead transfers an acyl group from one nitrogen to another. This field of study has recently been extensively reviewed (16). Vigorous theoretical investigations of this cysteine-containing pro-enzyme is currently being pursued by other research groups (17).

Fig. 2. A schematic representation of possible intramolecular hydrogen bonds for serine.



A **B**

Scheme 1.



The full conformational space of compound **I** was recently explored by ab initio RHF/3–21G computations (18). On the Ramachandran hypersurface of four independent variables $E = f(\phi, \psi, \chi_1, \chi_2)$ (Fig. 1), 47 conformers were located instead of the expected $3^4 = 81$ stable structures. The relative stabilities of the various conformers were analyzed in terms of side-chain–backbone interactions covering both hydrogen bonding and charge-transfer types. These results indicated that the γ_L-backbone conformation is the preferred form for this compound. In the present report, we have mainly focused our attention on the conformational intricacies of compound **I**, while restricting the backbone conformation to its global minimum (γ_L). A comparative study using different levels of theory of compounds **I–III** was also carried out.

Methods

Conformational analysis

IUPAC-IUB (19) recommended the use of 0° → +180° in the clockwise direction for rotation and 0° → –180° for counter-clockwise rotation (Scheme 3). For side-chain torsions, this implied the following range: $-180° \le \chi_1 \le 180°$ and $-180° \le \chi_2 \le 180°$. On the Ramachandran map, the $-180° \le \phi \le 180°$ and $-180° \le \psi \le 180°$ cut is indicated by the square drawn in broken lines (Fig. 3).

But for the graphical presentation of the side-chain conformational potential-energy surface (PES), we used the traditional cut ($0° \le \chi_1 \le 360°$ and $0° \le \chi_2 \le 360°$), similar to that previously suggested by Ramachandran and Sasisekharan (20).

Molecular computations

Initially, the nine side-chain conformations of cysteine in its γ_L-backbone conformation were determined using the topologically probable set of conformers as predicted by multidimensional conformational analysis (MDCA) (21–23).

Scheme 2.

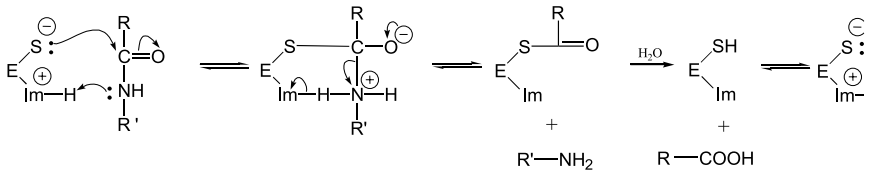
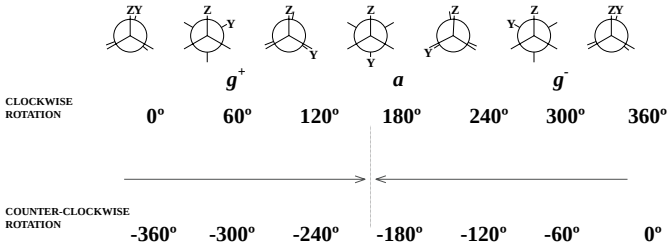


Table 1. Total energy values (Hartree) of the component molecules for isodesmic reactions.

Molecular system	HF/3–21G			B3LYP/6–31G(d,p)		
P-CONH-CH_2-CONH-**Q**	P = Q = H		P = Q = Me	P = Q = H		P = Q = Me
γ_L	–373.6487900		–451.2942433	–377.9165567		–456.5553698
β_L	–373.6477487		–451.2931883	–377.9158533		–456.5536252
CH_3-**R**	R = CH_2-SH	R = CH_2-OH	R = H	R = CH_2-SH	R = CH_2-OH	R = H
	–474.3472547	–153.2226808	–39.976877	–478.0228756	–155.0462074	–40.5238417

Scheme 3.



To investigate the side-chain–backbone conformational interaction effects, calculations were performed on compounds **I** and **II** using the Gaussian 94 (24) program at the RHF/3–21G (25) level of theory. This lower level is suitable for the purpose of comparing the present results with those obtained for serine (1, 2), even though the current trend is to rely on larger basis sets at a higher level of theory (8, 26). For this reason, geometry optimizations were also carried out for compounds **I** and **II** at the B3LYP/6–31G(d,p) level of theory.

Stabilization energies

The stabilization energies were calculated with respect to γ_L- as well as to $\beta_L(C_5)$-backbone conformation of N- and C-protected glycine (9, 27) using the following isodesmic (same number of the same type of bonds) reaction, where P and Q may be H (**II**) or CH_3 (**I**) and R = CH_2-SH:

$$\underset{\text{reference conformation } \gamma_L \text{ or } \beta_L}{\text{P-CONH-CH}_2\text{-CONH-Q} + \text{CH}_3\text{-R}} \rightarrow \underset{\text{conformation X}}{\text{P-CONH-CHR-CONH-Q}} + \text{CH}_3\text{-H}$$

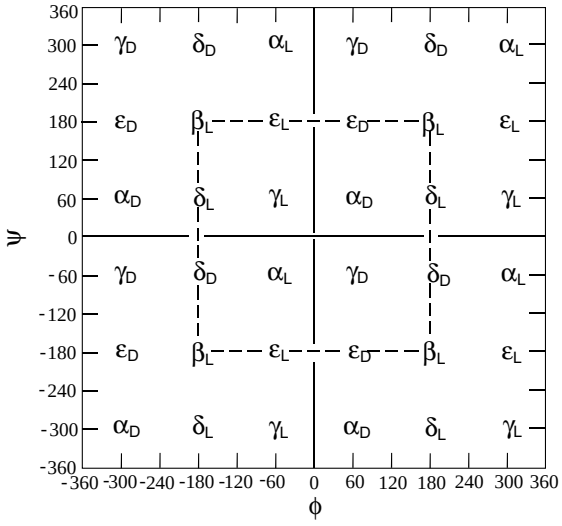
The stabilization energy may be calculated as follows:

$$\Delta E_{\text{stabil}} = \{E[\text{P-CONH-CHR-CONH-Q}]_X + E[\text{CH}_3\text{-H}]\} - \{E[\text{P-CONH-CH}_2\text{-CONH-Q}]_{\gamma_L \text{ or } \beta_L} + \text{CH}_3\text{-R}]\}$$

The energy values of the components are summarized in Table 1.

The results of the structure optimizations, namely, geometrical parameters, total energies, relative energies, and stabilization energies, are given in Tables 2–4. The total energies are given in Hartrees, and the relative energies and stabilization energies in kcal mol^{-1} (using the conversion factor: 1 Hartree = 627.5095 kcal mol^{-1}). The numbering system used for *N*-acetyl-L-cysteine-*N*-methyl amide (P = Q = Me) and *N*-formyl-L-cysteinamide (P = Q = H) is shown in Fig. 1.

Fig. 3. Topological representation of the Ramachandran map for an N- and C-protected amino acid P-CONH-CHR-CONH-Q (P and Q may be H or CH_3) showing two full cycles of rotation: $-360° \le \phi \le +360°$; $-360° \le \psi \le +360°$. The central box, denoted by broken line, represents the cut suggested by the IUPAC convention. The four quadrants denoted by solid lines are the conventional cuts. Most peptide residues exhibit nine unique conformations labeled as α_D (α_{LEFT}), ε_D, γ_D (C_7^{ax}), δ_L (β_2), β_L (C_5), δ_D (α'), γ_L (C_7^{eq}), ε_L, and α_L (α_{RIGHT}).



Topological analysis of electron density

The topology of the electron density was analyzed using AIM (atoms in molecules) (28). In this analysis, the gradient $[\nabla\rho(r)]$ and the Laplacian $[\nabla^2\rho(r)]$ play important roles. A critical point of the electron density along the line of two interacting atoms in a molecule is called the bond critical point (BCP), where $\nabla\rho(r) = 0$.

Table 2. Torsional angles and total energy values for compound **I** (CH_3CO-Cys-$NHCH_3$) optimized at HF/3–21G and B3LYP/6–31G(d,p) levels of theory.

Final geometry[a]	ϕ (5–7–9–11)	ψ (7–9–11–18)	χ_1 (7–9–13–16)	χ_2 (9–13–16–17)	ω_0 (2–5–7–9)	ω_1 (9–11–18–20)	Energy (Hartree)	ΔE_{rel} (kcal mol^{-1})	ΔE_{stabil} (γ_L) (kcal mol^{-1})	ΔE_{stabil} (β_L) (kcal mol^{-1})
HF/3–21G										
γ_L (g^+,g^+)	–85.211	64.199	52.657	67.777	–175.164	–179.297	–885.6785541	0.00	–8.74	–9.41
γ_L (g^+,a)	–86.016	63.387	54.989	–123.186	–177.042	–179.075	–885.6716533	4.33	–4.41	–5.07
γ_L (g^+,g^-)	Not found						—	—	—	
γ_L (a,g^+)	–87.218	74.967	–167.496	82.402	–177.482	–177.893	–885.6701841	5.25	–3.49	–4.15
γ_L (a,a)	Not found						—	—	—	
γ_L (a,g^-)	–86.474	69.587	–170.528	–74.251	–175.205	–178.714	–885.6744995	2.54	–6.20	–6.86
γ_L (g^-,g^+)	–87.265	68.555	–68.600	61.474	–176.407	–178.899	–885.6710358	4.72	–4.02	–4.69
γ_L (g^-,a)	–86.515	68.338	–55.821	–169.081	–172.746	–178.742	–885.6726444	3.71	–5.03	–5.70
γ_L (g^-,g^-)	–86.475	69.404	–54.352	–61.657	–172.469	–178.459	–885.6729162	3.54	–5.21	–5.87
B3LYP/6–31G(d,p)										
γ_L (g^+,g^+)	–82.61	66.19	52.84	64.39	–176.59	–178.01	–894.0647503	0.00	–6.49	–7.59
γ_L (g^+,a)	–82.61	66.33	55.53	–121.91	–178.88	–177.92	–894.0595386	3.27	–3.22	–4.32
γ_L (g^+,g^-)	Not found						—	—	—	
γ_L (a,g^+)	–83.87	79.56	–166.63	77.74	179.43	–174.77	–894.0575767	4.50	–1.99	–3.09
γ_L (a,a)	Not found						—	—	—	
γ_L (a,g^-)	–83.81	72.89	–170.93	–70.08	–177.51	–176.64	–894.0608365	2.46	–4.04	–5.13
γ_L (g^-,g^+)	–83.16	71.57	–68.84	54.12	–179.80	–177.38	–894.0583015	4.05	–2.45	–3.54
γ_L (g^-,a)	–84.19	70.68	–55.26	–176.57	–172.56	–176.93	–894.0587452	3.77	–2.72	–3.82
γ_L (g^-,g^-)	–84.47	73.02	–49.36	–58.06	–171.06	–176.27	–894.0597855	3.12	–3.38	–4.47

Note: The calculated relative energies (ΔE_{rel}) and stabilization energies (ΔE_{stabil}) are also shown.
[a]g^+, *gauche* (+60°); *a*, *anti* (180°); g^-, *gauche* (–300°).

Table 3. Torsional angles and total energy values for compound **II** (HCO-Cys-NH_2) optimized at HF/3–21G and B3LYP/6–31G(d,p) levels of theory.

Final geometry	ϕ (1–3–5–7)	ψ (3–5–7–14)	χ_1 (3–5–9–12)	χ_2 (5–9–12–13)	ω_0 (16–1–3–5)	ω_1 (5–7–14–17)	Energy (Hartree)	ΔE_{rel} (kcal mol^{-1})	ΔE_{stabil} (γ_L) (kcal mol^{-1})	ΔE_{stabil} (β_L) (kcal mol^{-1})
HF/3–21G										
γ_L (g^+,g^+)	–83.85	63.13	52.60	68.36	–174.99	–179.46	–808.0328587	0.00	–8.59	–9.24
γ_L (g^+,a)	–84.77	62.40	54.94	–123.34	–176.47	–179.23	–808.0260686	4.26	–4.33	–4.98
γ_L (g^+,g^-)	Not found						—	—	—	
γ_L (a,g^+)	–86.22	73.18	–167.32	83.69	–176.65	–178.01	–808.0244391	5.28	–3.31	–3.96
γ_L (a,a)	Not found						—	—	—	
γ_L (a,g^-)	–85.20	68.21	–169.98	–74.69	–174.83	–178.84	–808.0287635	2.56	–6.02	–6.67
γ_L (g^-,g^+)	–86.04	66.72	–69.12	63.08	–175.93	–179.17	–808.0253109	4.73	–3.85	–4.51
γ_L (g^-,a)	–84.73	66.71	–56.44	–169.94	–173.00	–178.96	–808.0268808	3.75	–4.84	–5.49
γ_L (g^-,g^-)	–84.47	67.84	–54.22	–61.86	–172.74	–178.61	–808.0271721	3.56	–5.02	–5.68
B3LYP/6–31G(d,p)										
γ_L (g^+,g^+)	–80.84	64.98	52.37	65.33	–177.20	–176.96	–815.4257909	0.00	–6.40	–6.84
γ_L (g^+,a)	–81.03	65.19	55.29	–121.44	–179.02	–176.65	–815.4206690	3.21	–3.17	–3.63
γ_L (g^+,g^-)	Not found						—	—	—	
γ_L (a,g^+)	–82.41	78.82	–166.65	78.58	–179.71	–171.45	–815.4186368	4.48	–1.91	–2.35
γ_L (a,a)	Not found						—	—	—	
γ_L (a,g^-)	–82.34	72.25	–170.54	–70.34	–177.42	–174.17	–815.4218597	2.46	–3.93	–4.38
γ_L (g^-,g^+)	–82.05	70.50	–69.32	55.31	–179.63	–175.20	–815.4192734	4.08	–2.31	–2.75
γ_L (g^-,a)	–82.03	70.48	–55.54	–178.87	–173.53	–174.19	–815.4198346	3.73	–2.66	–3.10
γ_L (g^-,g^-)	–82.11	73.25	–48.66	–58.68	–171.11	–172.97	–815.4210135	2.99	–3.40	–3.84

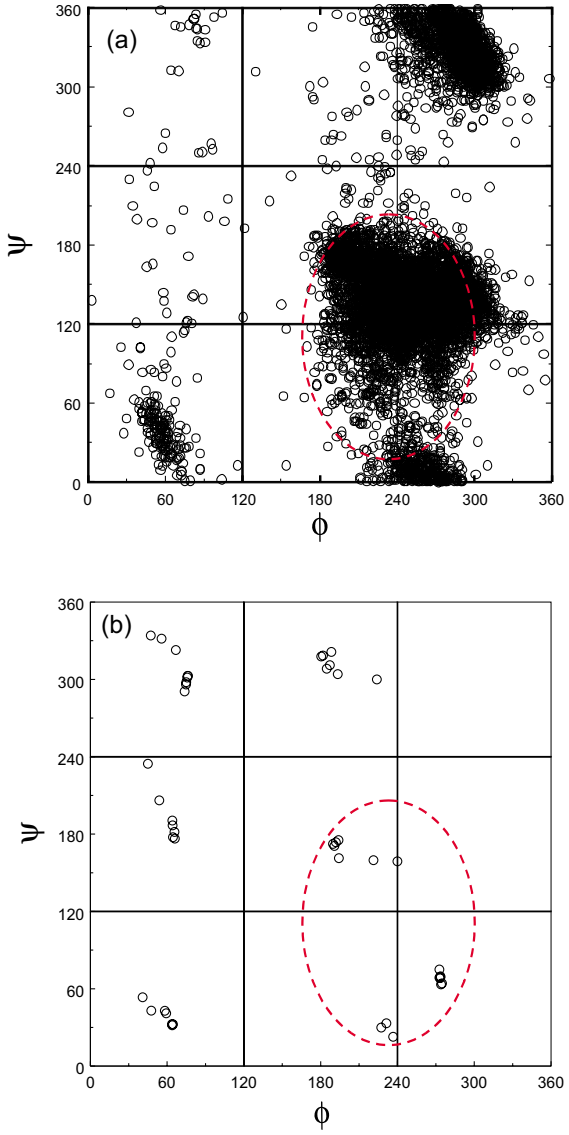
Note: The calculated relative energies (ΔE_{rel}) and stabilization energies (ΔE_{stabil}) are also shown.

Table 4. Torsional angles and total energy values for compound **III** (HCO-Ser-NH_2) optimized at HF/3–21G and B3LYP/6–31G(d,p) levels of theory.

Final geometry	ϕ (1–3–5–7)	ψ (3–5–7–14)	χ_1 (3–5–9–12)	χ_2 (5–9–12–13)	ω_0 (16–1–3–5)	ω_1 (5–7–14–17)	Energy (Hartree)	ΔE_{rel} (kcal mol^{-1})	ΔE_{stabil} (γ_L) (kcal mol^{-1})	ΔE_{stabil} (β_L) (kcal mol^{-1})
HF/3–21G										
γ_L (g^+,g^+)	–83.658	71.421	51.832	69.859	–176.197	–177.549	–486.9195667	0.00	–15.67	–16.32
γ_L (g^+,a)	Not found								—	—
γ_L (g^+,g^-)	Not found								—	—
γ_L (a,g^+)	–86.59	77.84	–170.03	74.00	–177.86	–176.82	–486.8996344	12.50	–3.16	–3.82
γ_L (a,a)	Not found								—	—
γ_L (a,g^-)	–83.397	62.518	179.732	–68.446	–173.778	179.794	–486.9119765	4.76	–10.91	–11.56
γ_L (g^-,g^+)	–85.262	97.114	–65.360	54.082	–177.938	–179.224	–486.9029219	10.44	–5.23	–5.88
γ_L (g^-,a)	–77.091	61.486	–43.944	–178.581	–168.802	179.636	–486.9075817	7.52	–8.15	–8.80
γ_L (g^-,g^-)	–77.250	62.991	–41.004	–75.808	–171.150	–179.957	–486.9076066	7.50	–8.17	–8.82
B3LYP/6–31G(d,p)										
γ_L (g^+,g^+)	–79.84	74.75	54.94	64.77	–178.17	–172.52	–492.4564571	0.00	–11.00	–11.44
γ_L (g^+,a)	Not found							—	—	
γ_L (g^+,g^-)	Not found							—	—	
γ_L (a,g^+)	–82.16	83.06	–167.97	69.62	179.37	–169.79	–492.4408888	9.76	–1.23	–1.68
γ_L (a,a)	Not found							—	—	
γ_L (a,g^-)	–80.61	60.41	–179.44	–64.18	–174.60	179.67	–492.4505714	3.69	–7.31	–7.75
γ_L (g^-,g^+)	–81.48	73.30	–65.61	53.10	177.63	–174.40	–492.4433657	8.21	–2.79	–3.23
γ_L (g^-,a)	–79.43	65.75	–47.87	174.60	–167.23	–175.97	–492.4457011	6.75	–4.25	–4.70
γ_L (g^-,g^-)	–79.22	67.70	–43.63	–69.67	–169.08	–174.94	–492.4463450	6.34	–4.66	–5.10

Note: The calculated relative energies (ΔE_{rel}) and stabilization energies (ΔE_{stabil}) are also shown.

Fig. 4. (*a*) Backbone conformers of all 10122 Cys residues taken from 974 non-homologous proteins (31). Using their backbone-dihedral parameters, all of the above Cys residues were plotted on a [ϕ, ψ] map. (*b*) Locations of calculated ab initio RHF/3–21G MeCO-Cys-NHMe backbone conformers (18) on a [ϕ, ψ] map.



The bond path is made up of a pair of gradient paths, originating at a BCP and terminating at a neighbouring nuclei. The necessary condition for two atoms to be bonded to each other is that their nuclei must be linked by a bond path. The bond path is regarded as "a universal indicator of bonded interaction" (29). The method is widely used for proving the existence of hydrogen bonds (30).

Results and discussion

The comparison of structural parameters from experimental databases (X-ray and NMR) with the ab initio (RHF/3–21G) results is an excellent validation of the computed results. Thus, the comparison of relative energies and the relative probabilities of conformers using a non-homologous database is a possibility for this cross-validation. Let us truncate the backbone of a protein into building units, e.g., amino acid diamides. We will assume that the probability of conformers in proteins depends only on its relative energy. This is a model where several stabilizing factors are neglected, such as inter-residue interactions, long-range effects, and hydration. By acknowledging the limitations of this approach, it is possible to correlate the relative energy of a conformer and the relative probability of the same backbone structure in an ensemble of proteins with known X-ray and NMR structures.

Using a recent (February 2002) X-ray- and NMR-determined protein data set of non-homologous proteins

Fig. 5. The possible nine molecular conformations of MeCO-Cys-NHMe. The two conformations that are crossed over are not minima, the remaining seven were located during HF/3–21G optimization.

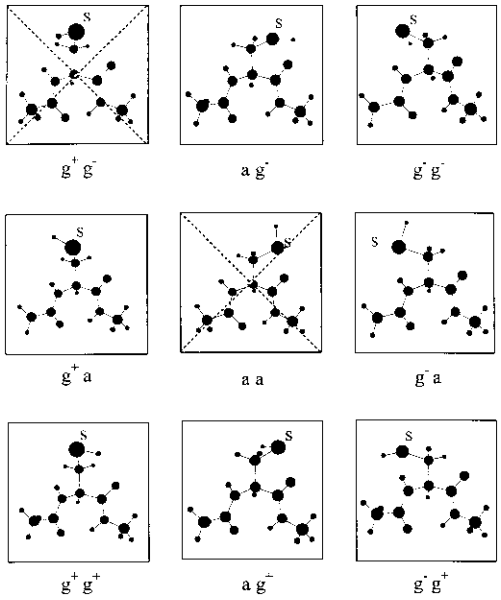


Fig. 6. Relative energies (kcal mol^{-1}) obtained for the different conformations of compounds **I–III**.

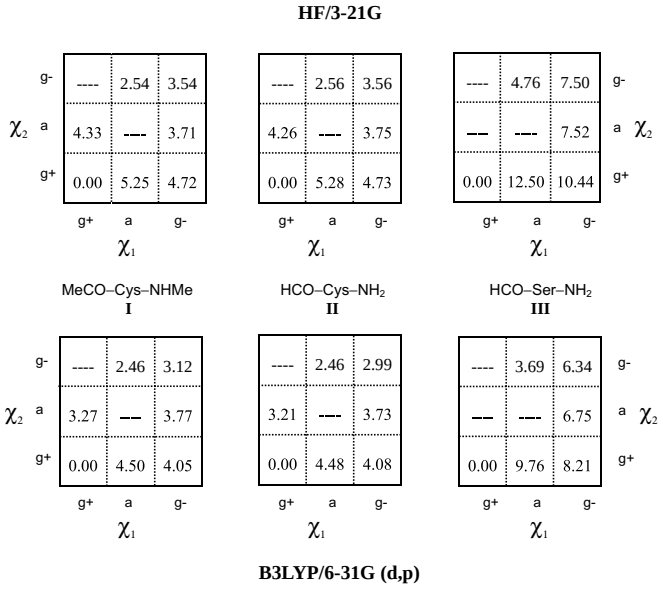


Fig. 7. Correlation of relative energies computed at DFT and HF/3–21G levels of theory for MeCO-Cys-NHMe (top), HCO-Cys-NH_2 (centre), and HCO-Ser-NH_2 (bottom).

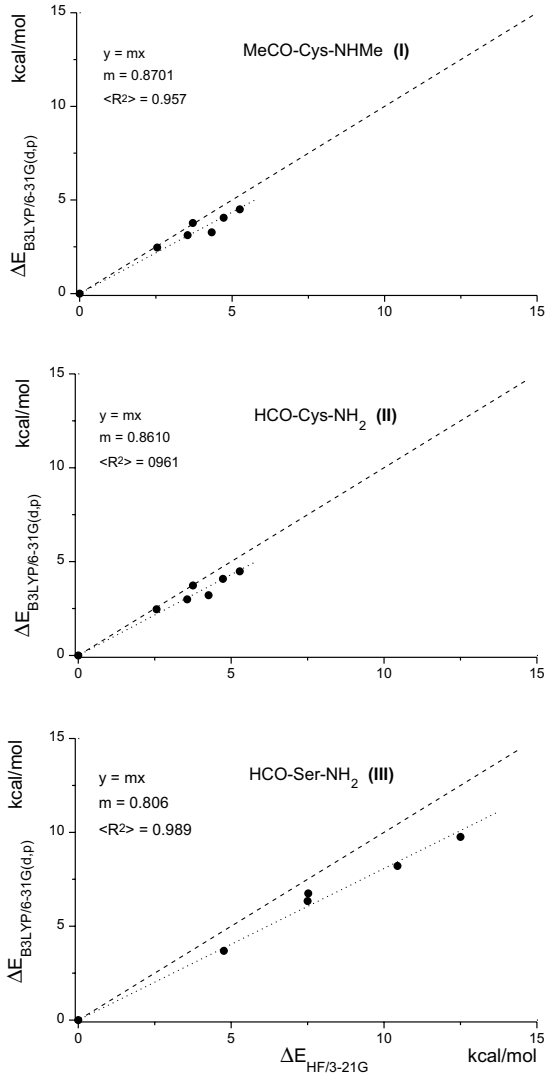
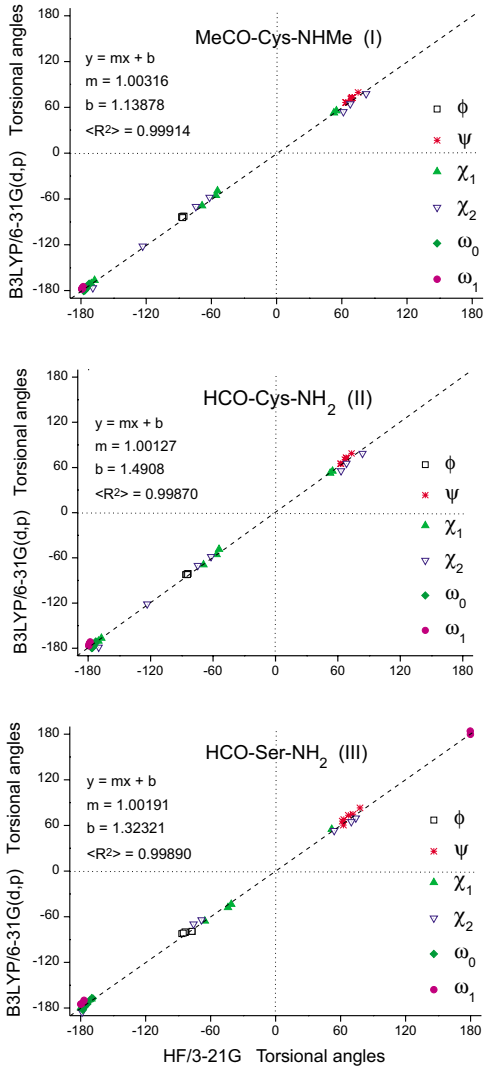


Fig. 8. Correlation of torsional angles computed at DFT and HF/3–21G levels of theory for MeCO-Cys-NHMe (top), HCO-Cys-NH_2 (centre), and HCO-Ser-NH_2 (bottom).



(31), a population-distribution map was generated (Fig. 4*a*). The backbone conformers of 10122 Cys residues, found in a total of 974 proteins, were plotted, showing ϕ vs. ψ values (Fig. 4*a*). The overall appearance of the Ramachandran surface is similar to other previously reported Ramachandran maps (Phe (12) and Ser (32)). Thus Cys shows an "average conformational character."

To perform a comparison between calculated and observed backbone conformers, an additional plot was made with the RHF/3–21G results (Fig. 4*b*). Comparing these data a promising overall similarity emerged. In the highly populated extended β-strand (β_L) and inverse γ-turn (γ_L) region, the location of the RHF/3–21G model structures and the largest probability of experimental- backbone conformers taken from the protein databases are similar (zone denoted with dotted lines in Fig. 4).

It should be noted that the seven γ_L conformers and the seven β_L conformers, together with the three δ_L conformers, are a subgroup formed by the most stable conformers of compound **I**. This low-energy region of the Ramachandran map (often quoted as the β region) is also the most dense area if X-ray- and NMR-determined main-chain probabilities are investigated. Such a correlation permit us to assume that if the diamide model is relevant to the description of main-chain folding of proteins, then the most stable conformers should have the lowest energy.

On the basis of the above results, in the present study we focused our attention on the conformational intricacies of the highly preferred backbone conformation of compounds **I** and **II**, this being the γ_L form.

Conformational study

The nine side-chain conformations, topologically predicted on the basis of MDCA, are depicted for compound **I** in Fig. 5. Two of the structures (g^+g^- and *aa*) are crossed out, since they were not found. On the basis of the $E = f(\chi_1, \chi_2)$ potential-energy surface (PES) generated for HCO-Ser-NH_2 (1), one may anticipate that these minima were eliminated due to a high-energy region or "mountain ridge" formed along the disrotatory diagonal axis. The other seven minima,

Fig. 9. Structures located on the side-chain PES of the γ_L-backbone conformations of N- and C-protected serine and cysteine. The stabilization energies (ΔE_{stabil}, kcal mol^{-1}) exerted by the side chain with respect to the glycine residue in its γ_L conformation are shown in brackets.

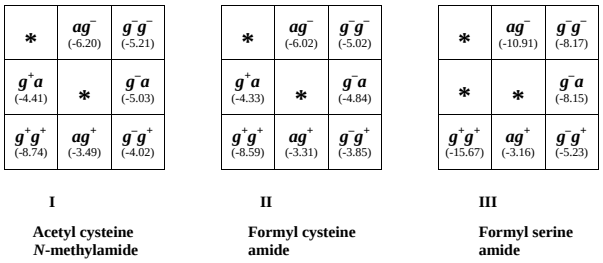
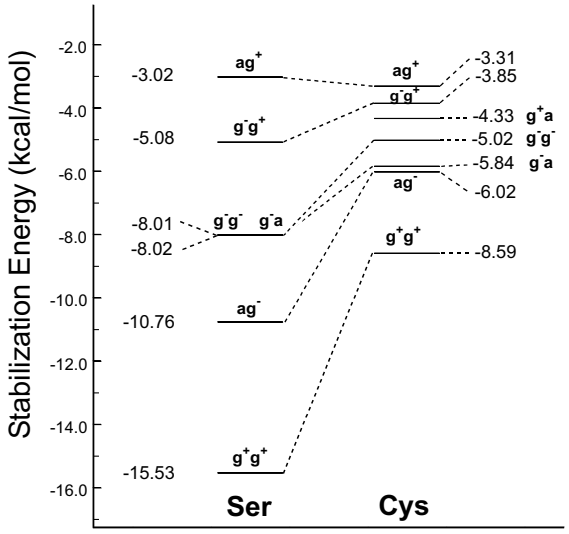


Fig. 10. Spectrum of conformational dependence of the side-chain stabilization energy on the γ_L-backbone conformation of protected Ser and Cys.



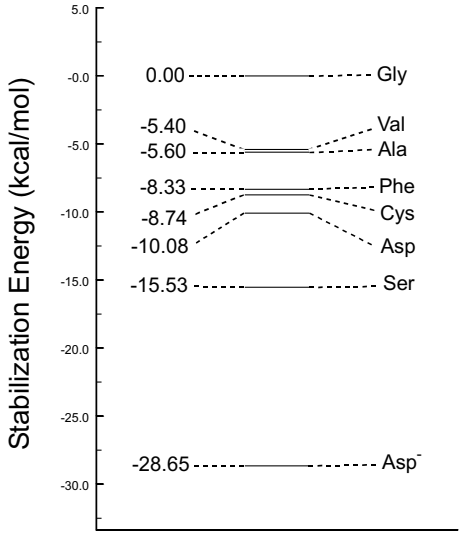
however, were successfully geometry optimized at the HF/3–21G level of theory (Table 2).

The structures of **I** and **II** were geometry optimized at the HF/3–21G level of theory and further optimized at the B3LYP/6–31G(d,p) level of theory (Tables 2 and 3). Their relative energy values are shown in a topological fashion in Fig. 6. For full comparison, HCO-Ser-CONH$_2$ (**III**) conformations were also evaluated at the B3LYP/6–31G(d,p) level of theory (Table 4).

The reliability of the HF/3–21G level of computations can be investigated here since we have results from HF/3–21G and B3LYP/6–31G(d,p) levels. The relative energies (ΔE_{rel}) of compounds **I**–**III**, computed at the two levels of theory, are compared in Fig. 7. Since the global minimum on the relative energy scale is, by definition, always zero, in order for the fitted line to pass over the origin, a $y = mx$ equation was fitted to the data points. While the slopes of the fitted lines are never unity, it is clear that the HF/3–21G results reproduce the trend quite well.

In addition to the relative stabilities, the accuracy of the key torsional angles (in the present case, ϕ, ψ, χ_1, χ_2, ω_0, and ω_1) is of great importance. The correlation of the above torsional angles computed at two levels of theory for compounds **I**–**III** are shown in Fig. 8. The least-square fit was of the type $y = mx + b$. For the cysteine and serine residues, m is not unity nor is b zero. The fitted lines, however, suggest a surprisingly good correlation.

Fig. 11. Relative side-chain stabilization energy values of various amino acids in the γ_L-backbone conformation.



Stabilization energies

Stabilization energies emerging from side-chain interaction with the backbone are traditionally calculated (7, 27) with respect to the global minimum, i.e, the γ_L conformation. It has been noted recently (33), however, that during the *cis*–*trans* isomerization process, the γ_L conformation may disappear for at least some of the amino acids. In other words, the γ_L conformer does not exist as a minimum-energy conformation on some of the *cis*-Ramachandran maps. For this reason, an alternative backbone conformation β_L was selected for calculating the stabilization energy (ΔE_{stabil}). The method of calculation, as outlined in the *Computational methods* section, is the same, but instead of the γ_L conformation of glycine, the β_L conformation is chosen as the reference conformation. Such a new standard may be important in the future, when an ab initio peptide database may include both *cis* and *trans* peptides. Nevertheless, the $\Delta E_{stabil}(\gamma_L)$ values are more practical if we wish to make comparisons with previously reported values.

Figure 9 compares $\Delta E_{stabil}(\gamma_L)$ values for compounds **I**–**III**. It is reassuring to see that the effect of the side-chain stabilization is not affected substantially if we change the protecting group from formyl to acetyl at the nitrogen end and from amide to methylamide at the carboxyl end.

The *N*-acetyl and *N*-methyl amide protecting groups of cysteine **I** are stabilized by the CH$_2$-SH side chain by ~0.16 kcal mol^{-1} more when compared with the same side chain stabilizing the *N*-formyl-cysteinamide (**II**) backbone. There are, however, marked differences when the oxygen is changed to sulfur in the side chain. Undoubtedly, such a change between serine and cysteine, as seen in Fig. 9, is also present in the relative energy values (ΔE_{rel}) presented in Fig. 6. The stabilization energy values are not only numerically different for Ser and Cys but they are also decreased

Fig. 12. Various types of intramolecular interactions in *N*-acetyl-L-cysteine-*N*-methylamide.

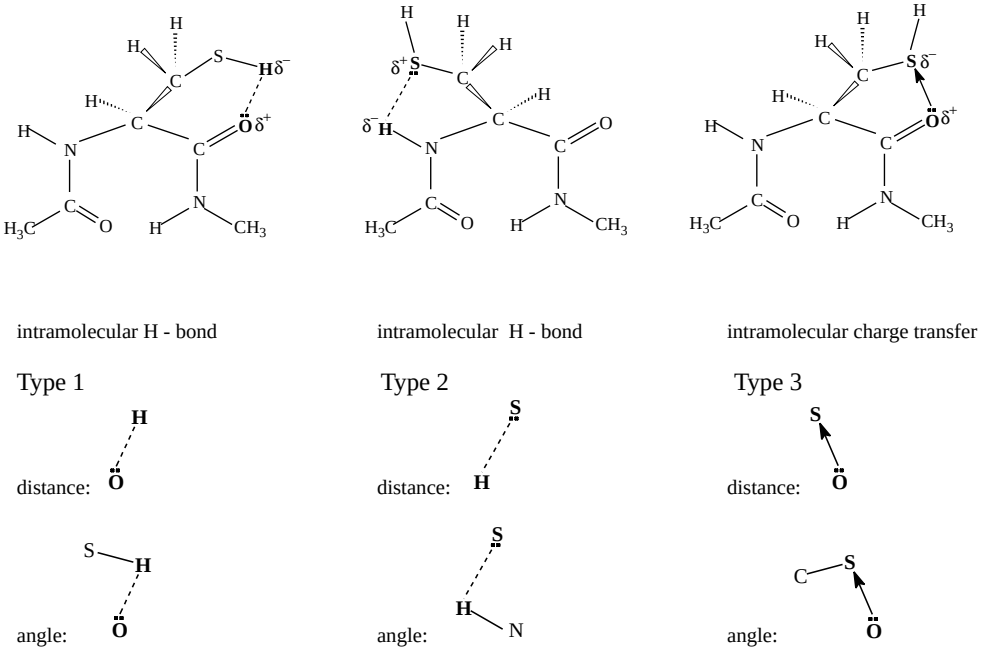


Table 5. Summary of intramolecular interactions in *N*-acetyl-L-cysteine-*N*-methylamide.

Final geometry	Energy (Hartree)	ΔE (kcal mol^{-1})	Type of interaction[a]	Distance (Å)[a,b]	Angle (°)[a]
γ_L (g^+,g^+)	–885.6785541	0.00	2	2.62	112.10
			1	2.30	119.68
γ_L (g^+,a)	–885.6716533	4.33	2	2.63725	111.035
			3	3.08	60.93
γ_L (a,g^-)	–885.6744995	2.54	1	2.43	115.77
γ_L (g^-,a)	–885.6726444	3.71	2	2.94	89.38
γ_L (g^-,g^-)	–885.6729162	3.54	2	2.96	92

[a]Definition is given in Fig. 11.
[b]Maximum threshold values are the sum of van der Waals radii (34, 35). Type 1 (O··H): 1.40 + 1.20 = 2.60 Å; Type 2 (S··H): 1.85 + 1.20 = 3.05 Å; Type 3 (S··O): 1.85 + 1.40 = 3.25 Å.

for Cys and increased for Ser, as seen in Fig. 10. When the global minima are compared, the side chain of Cys is stabilizing the backbone more than that of Phe but less than those of Asp and Ser. Such a "spectrum" of side-chain stabilization is presented in Fig. 11.

The size difference between O and S could play a major role in producing such remarkable energy differences. Additionally, since both make the same kinds of hydrogen bonds it is reasonable to assume that the Ser side chain is more favourably stabilizing than the Cys side chain because hydrogen bonds to and from oxygen atoms are stronger than sulfur atoms. This result is not surprising, since methanol is a liquid and methanethiol is a gas. This is in agreement with the ordering of the g^+g^+, g^+a, and ag^- conformations discussed in the next section.

Intramolecular interactions

The three types of intramolecular interactions, which may occur in the various conformers of cysteine, are shown in Fig. 12. The characteristic distances and angles, as well as the classification of interactions, are summarized in Table 5. Only the g^+a conformer has a heteroatomic distance that is short enough to suggest an O → S charge-transfer interaction.

In agreement with the data presented in Table 6, AIM analysis revealed that extensive side-chain–backbone intramolecular hydrogen-bonding-type interactions exists in three of the seven conformers. These were found to be g^+g^+, g^+a, and ag^-. Taking into account the distances and angles, however, we might consider that five of the seven γ_L conformations are stabilized by hydrogen bonding interactions (Table 5).

The g^+g^+ conformer has the greatest number of hydrogen bonds, so it is not surprising that it emerges as being the global minimum. In addition to the hydrogen bond, which is characteristic of the seven-member ring of the γ_L-backbone conformation, there are two additional hydrogen bonds corresponding to side-chain → backbone (C=O··H-S) and

Table 6. The topological properties of electron-density distribution of For-L-Cys-NH_2 at selected BCPs from B3LYP/6–31G(d,p)//HF/3–21G calculations.

Bond		g^+a	g^+g+	ag^-
O2··H15	ρ_b	0.0225	0.0245	0.0294
	∇^2_b	0.0659	0.0714	0.0679
	ε	0.0102	0.0114	0.0101
	H_b	–0.0041	–0.0047	–0.0037
H13··O8	ρ_b	—	0.0146	0.0117
	∇^2_b	—	0.0483	0.0417
	ε	—	0.2415	0.4547
	H_b	—	0.0065	0.0085
S12··H4	ρ_b	0.0137	0.0139	—
	∇^2_b	0.0488	0.0494	—
	ε	0.7443	0.7259	—
	H_b	0.0113	0.0114	—
Ring				
C5-C7-O8··H13-S12-C9	ρ_b	—	0.0114	0.0106
	∇^2_b	—	0.0484	0.0439
C5-N3-C1-O2··H15-N14-C7	ρ_b	0.0095	0.0097	0.0095
	∇^2_b	0.0502	0.0504	0.0484
C5-C9-S12··H4-N3	ρ_b	0.0134	0.0134	—
	∇^2_b	0.0623	0.0631	—

Note: Covalent bonds are denoted as A–B and hydrogen bonds specified as A··B.

Fig. 13. The contour map of the Laplacian of the electron density for the (g^+g^+) conformation of HCO-Cys-NH_2 calculated at the B3LYP/6–31+G(d,p)//RHF/3–21G level of theory. Bond paths are denoted by lines, bond critical points (BCPs) are denoted by black dots, ring critical points (RCPs) are denoted by open triangles, and the nuclei are denoted by crosses.

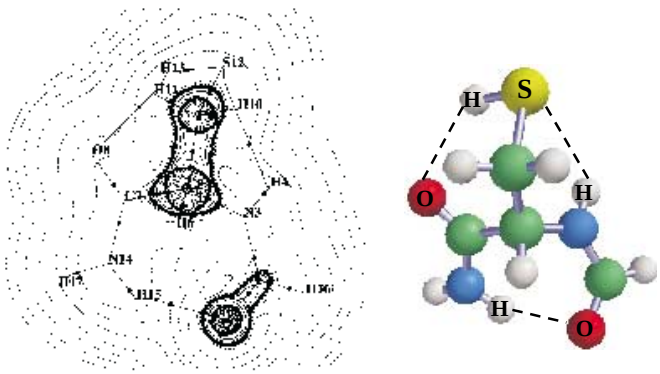
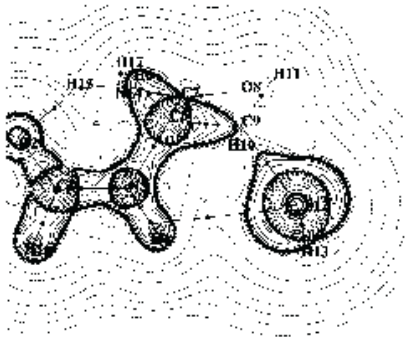


Fig. 14. The contour map of the Laplacian of the electron density for the (g^+a) conformation of HCO-Cys-NH_2 calculated at the B3LYP/6–31+G(d,p)//RHF/3–21G level of theory. Bond paths are denoted by lines, bond critical points (BCPs) are denoted by black dots, ring critical points (RCPs) are denoted by open triangles, and the nuclei are denoted by crosses.



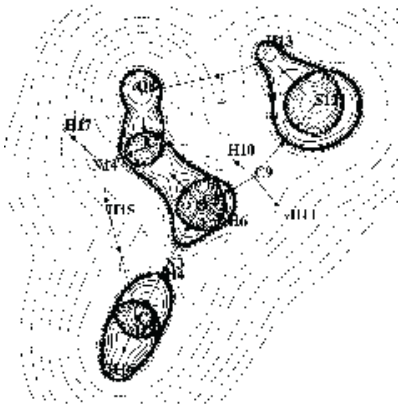
backbone→side-chain (N-H··S-H)-type hydrogen bonding (Fig. 13). The three ring-critical points, associated with the rings formed as the result of intramolecular hydrogen bonding are denoted by open triangles (Δ). The various BCPs are symbolized by solid dots (●).

Beyond the characteristic hydrogen bonding associated with the seven-member ring of the γ_L-backbone conformation, the g^+a conformer also exhibited an N-H··S-H-type hydrogen bonding. In Fig. 14, the ring-critical points are shown by open triangles (Δ) and the various BCPs are symbolized by solid dots (●).

Figure 15 shows the ag^- conformation, featuring a C=O··H-S hydrogen bond in addition to the hydrogen bond associated with the seven-member ring of the γ_L-backbone conformation.

In addition to these side-chain–backbone-type of hydrogen bonds, there exists the backbone—backbone C=O··H-N-type of hydrogen bond, typical in the seven-member ring of the γ_L conformer.

Fig. 15. The contour map of the Laplacian of the electron density for the (ag^-) conformation of HCO-Cys-NH_2 calculated at the B3LYP/6–31+G(d,p)//RHF/3–21G level of theory. Bond paths are denoted by lines, bond critical points (BCPs) are denoted by black dots, ring critical points (RCPs) are denoted by open triangles, and the nuclei are denoted by crosses.



Conclusions

In summary, we found seven different conformations for the side chain of compounds **I** and **II** when the backbone is in γ_L form. These results suggest that the side-chain stabilization is not substantially affected if we change the protecting group from formyl to acetyl at the nitrogen end and from amide to methylamide at the carboxyl end. In contrast, noticeable differences were found when comparing the conformational intricacies of compounds **II** and **III**: only six different conformations were obtained for compound **III** and the order of preference is also different. DFT calculations proved that polar side chains are able to interact with a peptide backbone, eliminating some otherwise legitimate minima through unfavourable backbone and side-chain torsional-angle combinations. The comparative study between compounds **II** and **III** illustrates this point very well. On the basis of our results, we might assume that the strength of hydrogen bonding involved is responsible for the different conformational behavior obtained for these compounds. On the other hand, it is interesting to note that both levels of theory (HF/3–21G and B3LYP/6–31G(d,p)) displayed closely related results indicating that HF/3–21G calculations are satisfactory to use in exploratory conformational studies.

Acknowledgments

We would like to thank Cray Inc. (Eagen, MN) for use of their resources and computer time. We would also like to thank the National Cancer Institute for the use of services at the Frederick Biomedical Supercomputing Center. This work was supported by grants from Universidad Nacional de San Luis (UNSL), and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) of Argentina. H.A.B. thanks CONICET for a postdoctoral fellowship. R.D.E. is a carreer researcher of CONICET. I.G.C. would like to thank Domus Hungarica Scientiarium et Atrium for supporting this international research collaboration.

References

1. O. Farkas, A. Perczel, J. Marcoccia, M. Hollosi, and I. Csizmadia. J. Mol. Struct. (Theochem) **331**, 27 (1995).
2. A. Perczel and I. Csizmadia. Int. Rev. Phys. Chem. **14**, 127 (1995).
3. A. Perczel, O. Farkas, and I. Csizmadia. J. Comp. Chem. **17**, 821 (1996).
4. A. Perczel, O. Farkas, and I. Csizmadia. J. Am. Chem. Soc. **118**, 7809 (1996).
5. A. Perczel, O. Farkas, I. Jakli, and I. Csizmadia. J. Mol. Struct. (Theochem) **455**, 315 (1998).
6. A. Perczel, J. Angyan, M. Kajtar, W. Viviani, J.-L. Rivail, J. Marcoccia, and I. Csizmadia. J. Am. Chem. Soc. **113**, 6256 (1991).
7. M. McAllister, A. Perczel, P. Csaszar, W. Viviani, J.-L. Rivail, and I. Csizmadia. J. Mol. Struct. (Theochem) **288**, 161 (1993).
8. G. Endredi, A. Perczel, O. Farkas, M. McAllister, G. Csonka, J. Ladik, and I. Csizmadia. J. Mol. Struct. (Theochem) **391**, 15 (1997).
9. W. Viviani, J.-L Rivail, A. Perczel, and I. Csizmadia. J. Am. Chem. Soc. **115**, 8321 (1993).
10. O. Farkas, M. McAllister, J. Ma, A. Perczel, M. Hollosi, and I. Csizmadia. J. Mol. Struct. (Theochem) **369**, 105 (1996).
11. A. Perczel, O. Farkas, and I. Csizmadia. Can. J. Chem. **75**, 1120 (1997).
12. I. Jakli, A. Perczel, O. Farkas, M. Hollosi, and I. Csizmadia. J. Mol. Struct. (Theochem) **455**, 303 (1998).
13. L. Polgar. FEBS Lett. **47**, 15 (1974).
14. A. Fersht. Enzyme structure and mechanism. 2nd ed. W.H. Freeman, New York. 1984. pp. 413–416.
15. A. Storer and R. Menard. Methods Enzymol. **244**, 486 (1994).
16. L. Muszbek, V. Yee, and Zs. Hevessy. Thromb. Res. **94**, 271 (1999).
17. I. Komároni, L. Kárpáthy, K. Morokuma, and L. Muszbek. 5th World congress of theoretically oriented chemists (WATOC). Book of Abstracts (p.p. 489). Royal Society of Chemistry, London. 1999.
18. M. Zamora, H. Baldoni, J. Bombasaro, M. Mak, A. Perczel, O. Farkas, and R. Enriz. J. Mol. Struct. (Theochem) **540**, 271 (2001).
19. IUPAC-IUB comission on biochemical nomenclature. Biochemistry, **9**, 3471 (1970).
20. I. Ramachandran and V. Sasisekharan. Adv. Protein Chem. **23**, 283 (1968).
21. I. Csizmadia. *In* The chemistry of the thiol group. *Edited by* S. Patai. John Wiley and Sons, New York. 1974.
22. M. Peterson and I. Csizmadia. J. Am. Chem. Soc. **100**, 6911 (1978).
23. M. Peterson and I. Csizmadia. *In* Progress of theoretical organic chemistry. Vol. 3. *Edited by* I.G. Czizmadia. Elsevier, Amsterdam. 1982.
24. Gaussian 94. Revision E.1, E.2, C.3, and D.2. M. Frisch, G. Trucks, H. Schlegel, P.W. Gill, B. Johnson, M. Robb, J. Cheeseman, T. Keith, G. Petersson, J. Montgomery, K. Raghavachari, M. Al-Laham, V. Zakrzewski, J. Ortiz, J. Foresman, J. Cioslowski, B. Stefanov, A. Nanayakkara, M. Challacombe, C. Peng, P. Ayala, W. Chen, M. Wong, J. Andres, E. Replogle, R. Gomperts, R. Martin, D. Fox, J. Binkley, D. Defrees, J. Baker, J. Stewart, M. Head-Gordon, C. Gonzalez, and J. Pople. Gaussian, Inc., Pittsburgh PA. 1995.
25. J. Binkley, J. Pople, and W. Hehre. J. Am. Chem. Soc. **102**, 939 (1980).

26. R. Frey, J. Coffin, S. Newton, M. Ramek, V Cheng, F. Momany, and L. Schafer. J. Am. Chem. Soc. **114**, 5369 (1992).
27. M. McAllister, G. Endredi, W. Viviani, A. Perczel, P. Csaszar, J. Ladik, J-L. Rivail, and I. Csizmadia. Can. J. Chem. **73**, 563 (1995).
28. R. Bader. Atoms in molecules: A quantum theory. Clarendon Press, Oxford. 1990.
29. R. Bader. J. Phys. Chem. A, **102**, 7314 (1998).
30. (*a*) D. Whiterfield and T. Tang. J. Am. Chem. Soc. **115**, 9648 (1993); (*b*) U. Koch and P. Popelier. J. Phys. Chem. **99**, 9747 (1995); (*c*) D. Whitefield, D. Lamba, T. Tangm, and I. Csizmadia. Carbohydrate Res. **286**, 17 (1996); (*d*) J. Platts, S. Howardm, and B. Bracke. J. Am. Chem. Soc. **118**, 2726 (1996); (*e*) D. Fang, P. Fabian, Z. Szekely, X. Fu, T. Tang, and I. Csizmadia. J. Mol. Struct. (Theochem) **427**, 243 (1998); (*f*) P. Popelier. J. Phys. Chem. A, **102**, 1873 (1998).
31. H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, P. Bourne. The Protein Data Bank. Nucleic Acids Res. **28**, 235 (2000). Last update: 26 Feb. 2002.
32. I. Jákli, A. Perczel, Ö. Farkas, P. Császár, C. Sosa, and I. Csizmadia J. Comp. Chem. **21**, 626 (2000).
33. (*a*) H. Baldoni, L. Torday, A. Rodríguez, G. Zamarbide, R. Enriz, C. Sosa, Ö. Farkas, I. Jákli, A. Perczel, and I. Csizmadia. 5th World congress of theoretically oriented chemists (WATOC). Book of Abstracts (p.p. 212). Royal Society of Chemistry, London, 1999; (*b*) H. Baldoni, G. Zamarbide, R. Enriz, E. Jáuregui, Ö. Farkas, A. Perczel, S. Salpietro, and I. Csizmadia. J. Mol. Struct. (Theochem) **500**, 97 (1998).
34. L. Pauling. The nature of the chemical bond. 3rd ed. Cornell University Press, Ithaca, New York. 1960.
35. J. Emsley. The elements. 3rd ed. Clarendon Press, Oxford. 1998.